MASTER

Embodying dynamics to transport modeling
linking ALBATROSS and MaDAM

Kwak, M.A.

Award date:
2011

Embodying Dynamics to Transport Modeling : Linking ALBATROSS and MaDAM

by

Min Ah Kwak

A Thesis submitted to
the group Design and Decision Support System (DDSS)
in the Faculty of Architecture, Building, and Planning
in partial fulfillment of the requirements for the degree of

Master of Science

at the
EINDHOVEN UNIVERSITY OF TECHNOLOGY

August, 2011

Exam committee members: T.A. Arentze,
A.W.J. Borgers,
H.J.P. Timmermans

Abstract

Traffic is inherently dynamic. It refers to not only time-varying characteristic but also a complex interplay of multiple facets. As a novel approach to add the dynamic nature to transport modeling, it is attempted to make an integrated transport model which bridges the activity-based approach and dynamic traffic assignment. Whereas the conceptual linkage has been made between Albatross and MaDAM, the operational linked system is implemented on ALBATROSS and OmniTRANS. What concerns critically for linking the two models by either concepts or techniques is compatibility. Dealing with compatibility requires several adjustments of the models and data. First, Albatross is extended so as to handle a local study area in its parameter setting of different population fraction sizes for study area and outer area in the synthesis module and its generation of a study-area based O-D trip tables from total flows of an entire synthetic Dutch population. Second, data from Albatross is converted to Omnitrans required formats. Different zoning systems are handled by zone mapping files and Albatross O-D trip tables are imported to the database of Omnitrans. Based on the adaptations, the linked model system is operated and tested. Third, it is examined whether sample size has an impact on prediction of traffic flows. The test of equality of variance suggested that more random variation of predicted results exists for smaller sample sizes. Lastly, the linked model system is evaluated in terms of its face-validity of predictions of impacts of a scenario of the Netherlands for the year 2020. The system is sensitive to changes in model assumptions defined as a scenario and the results are internally consistent, implying that traffic assignment result is in line with Albatross prediction.

Key words: Dynamics, Activity-based approach, Dynamic traffic assignment, Compatibility

Author: Min Ah Kwak (Student id number: 0641665)

Thesis Supervisors:

Theo Arentze (t.a.arentze@tue.nl)
Eindhoven University of Technology

Aloys Borgers (a.w.j.borgers@tue.nl)
Eindhoven University of Technology

Harry Timmermans (h.j.p.timmermans@tue.nl)
Eindhoven University of Technology

BSB

Acknowledgements

The moment to think of what to write for the acknowledgement has finally come. It makes me look back on the past years. A thousand emotions and thoughts crowded on my mind. I could not hide my joy upon the finishing of the thesis writing. And at the same time, I'm sorry for not being very sincere on this work and taking care of family also. In the midst of work, I often asked to myself whether the thesis would have finished some day. There have been many big developments in one's life, as I became one's spouse and one's mother during the past three years. Completing master's thesis is much more invaluable accomplishment especially for me than for any other regular students since I have had roles as a mother, wife, and student while I was working on the thesis. Every experience was the first to me. It was one of the hard times in my life with which I have struggled, as I felt like I was precariously balancing my three roles, much like a juggler trying to keep three balls afloat. The thesis tunnel is sometimes thought to be dark and blocked with obstacles, making me frustrated. Eventually, I was successful to come out of the tunnel and I would like to return thanks to them who support me.

I would like to express heartfelt appreciation to my thesis supervisors, Prof. Theo Arentze, Prof. Aloys Borgers, and Prof. Harry Timmermans, without whom this work would have not materialized. As for Theo, his constant encouragement and guidance have been invaluable in completing the thesis. Wherever I stayed either in Eindhoven or in my home country after Amy was born and wherever he stayed either in Eindhoven or in Switzerland for his sabbatical leave, he was always sincere to my work. I appreciate his comments and advices on my writings. What drives me to keep continue work without giving it up was his thoughtful understanding of my circumstances, rather than rushing me to finish. Aloys assured me that I can make it whenever I felt difficulty working with a company coming from a different research area that I've never learned and work culture totally different from academic institutes. Particularly, I'm grateful that he actively arranged many administrative things that I must do if I were in Eindhoven. As for Harry, I still remember the day we first met in front of Joe's office in Korea. I thank for giving me an opportunity to study in urban planning group of DDSS at the TU/e. I'm very honored that he is a member of committee of the graduation project.

This thesis would not have been possible unless Omnitrans International supports the project. First of all, I appreciate Erik de Romph to give me an opportunity to conduct the graduation project at Omnitrans International in collaboration with Eindhoven University of Technology. I also would like to show my gratitude to Michiel Bliemer of Goudappel Coffeng. He always sincerely responded to my questions by e-mail. His support of academic knowledge helps my understanding of the topic. As this was the first, biggest and long-term project that I've ever participated, Peter Kant's advices and tips on the progress of the project were very useful. I would particularly to acknowledge the contribution of Jeroen van Oorspronk, Edwin Mein and Feike Brandt for technical support of Omnitrans application.

Finally to my family and friends, I really appreciate for their patience and supports. Especially to my husband Minseok and my daughter Amy, I was so sorry for not being a good wife and a mother and I'm grateful for their patience and encouragements. Although it was really hard to do all my duties at best and those worries discouraged me to go forward, what I am now may not be present unless there was extreme devotion of my parents and parents-in-law. Special thanks to my best friend, Andrea Wright, she was always there with a sympathetic ear whenever I was frustrated in managing my responsibilities in a good way. As a final comment, I would say this master's thesis not my individual work but a joint effort of people around me.

Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 RESEARCH BACKGROUND AND OBJECTIVES

Transport modeling is used for forecasting the future travel demand caused by demographic changes or assessing the likely impacts of new policies or investment in transportation infrastructure. It has been carried out by the four step model (FSM) as the pivotal model of the conventional trip-based approach (Hensher and Button, 2002). Researchers have explored alternative approaches to supplement the limitations of the traditional FSM which is characterized by a trip-based as well as a time-independent and this leads to a paradigm shift from trip based approach to activity based approach in transport modeling. It has been pointed out that existing trip-based models is behaviorally less sensitive so as to reflect the complexity of travel behaviors of individuals and social trends. The models do not fully reflect reality and predict travel patterns accurately.

Two research fields have been actively being developed solving the shortcomings of the FSM. First, what attempted to capture interrelationships between trip-related attributes for more accurate prediction is activity-based modeling (ABM) in transport. Unlikely to trip-based approach, trips are understood in a broader context of a daily activity pattern in the sense that a travel from one location to another takes place to participate activities. Trip information can be implicitly derived from activity pattern of individuals. Understanding travel behavior in connection with activity participation enables more accurate and consistent prediction than relying only on trip records. Moreover, simultaneously but independently, dynamic traffic assignment (DTA) models have been developed, replacing static traffic assignment (STA) methods. It predicts how traffic flows change over time by propagating the traffics onto the given network. Numerous formulations and solutions approaches have been introduced ranging from mathematical programming, to variational inequality, optimal control, and simulation-based (Peeta and Ziliaskopoulos, 2001). The rationale behind both these changes in modeling approaches is that travel behavior or traffic movement is not static but dynamic. In other words, traffic changes over time within a day and results from a complex interplay between many factors. Dynamic aspects mean that not only it is time-varying but also interrelations and interdependencies between travel-related facets exist (Ettema, 1996).

These two streams of research have achieved great advancements in embodying dynamics to transport modeling. Activity-based approach regards travel as a derived demand from a wish to participate in activities that satisfies one's need (Ben-Akiva and Bowman, 1998). The intention of an ABM is to predict in an integrated fashion which activities are conducted where, for how long, when, where and the transport mode used when travel is involved (Timmermans et al, 2002). On the other hand, simulation-based DTA techniques allow one to predict how traffic flows evolve over time throughout a day (Peeta and Ziliaskopoulos, 2001). Unfortunately, much of the research efforts in activity-based modeling and DTA techniques have been undertaken independently. It would yield inconsistent results and fail to fully benefit from the potentials of the two fields if a model adopts only one of these approaches for modeling. In other words, using an ABM with a STA model does not consider temporal dynamics of traffic movements captured in ABM. Likewise, employing trip-based models, which predict travel demands only for coarse time periods, for example peak and off-peak hours, does not make the most use of the ultimate aim of DTA techniques (Lin et al, 2008). Therefore, a comprehensive model system which links ABM and DTA models needs to be formulated to realize true dynamic transport modeling which is behaviorally realistic and sensitive.

Developing a comprehensive model system which links ABM and DTA models is proposed in order to embody dynamics to transport modeling. Even though there have not been many studies combining both models in a unified framework, two studies, which attempted to establish such a link, have been reported. Lam and Yin (2001) presented a conceptual activity-based and time-dependent traffic assignment model. A time-dependent activity utility profile concept is introduced to formulate activity choice as a multinomial logit model. For route choice modeling, dynamic user equilibrium condition as a variational inequality approach of DTA methods is used. However, the research did not attempt to make an operational system but merely formulated a conceptual model. More similarly to our approach, there is an ongoing project which attempts to connect the two approaches. Lin et al (2008) proposed an integrated model connecting CEMDAP, a micro-simulation ABM, and VISTA, a simulation-based DTA method. This indicates that integration of the two types of models is receiving increasing attention and that no operational unified systems exist to date. Hence, the main objective of this research is to link two different approaches methodologically and computationally beyond the conceptual foundations.

1.2 RESEARCH QUESTIONS

The central aim of this research is to integrate ABM and DTA in a single framework. The main research question dealt in the thesis is twofold: i) how can an ABM (in specific ALBATROSS) and DTA (in specific, MaDAM) be linked? and ii) how well does the linked model system perform? The linkage between ABM and DTA on a conceptual as well as operational level involves development of an interface which enables a two-way interaction of the models. Thus, compatibility of the models needs to be achieved. The question of testing the linked model system involves evaluating both technical performance of the system and validity (accuracy and sensitivity) of predictions of traffic movements.

Specified research questions to be answered in separate chapters are:

i) How to handle study area travel demand in Albatross national model?

Spatial scope that the two models consider is different. That is, Albatross is a national model which predicting travels of entire Dutch population. On the other hand, Omnitrans traffic assignment is typically conducted at smaller study area due to computational limitation. The adaptation is made on Albatross side. It has not changed that Albatross predicts travel of the entire population. However, the algorithm which finds entry or exit zone along the study area cordon is developed and thus external trips in relation to study area are indirectly predicted.

ii) What are data requirement for each modeling component of the linked system?

Adjusting spatial scope of Albatross result, the two models are possible to connect. Compatibility of the models is the most critical requirement to build linkage between independent models. Not only data file formats (e.g., file extensions or data variables) but also zoning system needs to correspond to each other. For example, the output of Albatross, say O-D trip tables, must be compatible with the input format of MaDAM. That is, the data needs to be converted into the readable format in Omnitrans. Moreover, zoning system of Albatross and MaDAM also has to match each other. Otherwise, result of traffic assignment wrongly represents traffic patterns.

iii) Is predictability of the model affected by sample fraction size?

Albatross predicts activity-travel patterns of individuals. Like many micro-simulation models, generation of population data for simulation, population synthesis, is one of the fundamental steps in Albatross. Synthesis typically involves not entire but fraction of population due to computational feasibility. Computation time matters with sample fraction size. For instance, synthesis of 10% of population generally takes 8 hours on standard PC. In practice, synthesis of a fraction of population has not been problematic to predict long-term strategic travel pattern. However, representativeness of the results at Omnitrans would be influenced by fraction size in that Omnitrans describes traffic pattern at finer disaggregated level. It is more microscopic in the sense that traffic condition on transport network is available every hour within the day. Impact of the level of population fraction on prediction of travel behavior can be examined by analyzing simulation output with a set of different fraction sizes.

iv) How does the linked model system perform?

Validation of the model is an essential component of model building. The linkage is made between the two independent models. Thus, the validation of the linked model system mainly concerns whether the linkage is established. Scenario analysis is used here to evaluate the system. Adoption of the scenario not only displays the system performance but also demonstrates possible application of the linked model system in practice.

1.3 LAYOUT OF THE THESIS

The thesis is organized as follows. Chapter 2 serves as literature review providing conceptual grounds and progress of fundamental approaches and theories in transport modeling. The four step process, activity-based modeling approach, and traffic assignment are described. Chapter 3 introduces the two models and the relevant application software. The model Albatross works in ALBATROSS application. And traffic assignment model, MaDAM, runs in Omnitrans software. From chapter 4, the thesis gets into discussion of practical issues of the linked model system. Chapter 4 deals with study area delineation problem. Path finding algorithm developed for predicting external trip flow always accompanying study area is discussed in detail. Chapter 5 introduces modeling scheme of the linked model. In particular, it focuses on data requirements and conversion. Chapter 6 presents the analysis whether the resulting traffic pattern varies in relation to different sample fraction size of population. Chapter 7 illustrates the sensitivity analysis for validation of the linked model system. Chapter 8 concludes the thesis with summary and future research suggestion.

2 TRANSPORT MODELING APPROACHES

2.1 INTRODUCTION

New approaches and corresponding models for transport modeling have been developed as to supplement and solve problems of the previous models. In 1960's, dramatically increasing auto ownership and car use caused large amount of investment to transportation infrastructure such as road. The policies are typically planned and executed for a longer period, say twenty or thirty years. To assess the likely impacts of these long-term infrastructure supply strategies, transport models that predict travel demand on the long run was called for, resulting in the development of aggregated models such as spatial interaction and entropy models. These models predict traffic flows between zones and the outcomes were used to measure future needs for road capacity. It was assumed that travel was the result of four subsequent decisions which can be modeled separately, implying that each decision is an isolated phenomenon. The aggregated four-step models regards that travelers first decide whether or not to travel (trip generation), then choose where to go (trip distribution), what transport mode to use (mode choice), and which route to take (route choice).

In 1970's these aggregated models were criticized for the lack of explanation of travel behavior on human decision making. Thus, individual behavior rather than zonal traffic flow became the focus of modeling. Furthermore, the characteristic of transportation policies changed. There had been a shift in the focus of policies from long-term investment strategies to short-term market-oriented policies. This facilitates a need for a model that could predict behavioral responses to new strategies. Consequently, disaggregated travel demand models had been developed in micro-economics and psychology which analyze choice behavior. The dominant model of these disciplines is discrete choice models based upon theories of utility maximizing and individual choice behavior. It enabled better prediction of response to market-oriented policies.

However, disaggregated models focusing on trips also received criticism. Even though these models incorporate principles of human behavior into traffic modeling, the conceptual foundation is limited in the sense that they consider travel as a demand in its own right and neglect the question why people make trips. Moreover, these disaggregated trip-based models often fail to recognize the existence of linkages among trips, assuming that trips in trip chains made by an individual are treated as separate and independent entities in the analysis. For instance, one trip may be delayed or even cancelled due to the postponement of the previous trips. And the transport mode for one trip always influences that of the next trip, particularly for car use. Therefore, modeling of transportation has proved to be inaccurate due to this misspecification: an inappropriate representation of travel behavior relationships (Jones, *et al.*, 1990). In addition to these assumptions of the traditional trip-based models, social, technological, and behavior changes induce increasingly complex travel and activity patterns. For example, flexible working hours in the help of advancements in telecommunication causes more complex behavior responses than simple changes in mode choice of just one trip. Hence, alternative modeling approaches which capture the relationships between travel and non-travel aspects are required. Furthermore, the constraints concept must be applied to modeling to reflect the time dimension (Arentze and Timmermans, 2000). This paradigm shift has appeared as activity-based modeling approach in transportation

As explained above, transport models and modeling methodologies have been developed following the paradigm shift. This chapter reviews transport modeling approaches and relevant models from classic to state-of-the-art in detail. First of all, classifications of transport models are discussed. The section touches on advantages and disadvantages of existing approaches for understanding trip behaviors in terms of different units of analysis: trip, tour, and activity. Following section presents the seminal and famous transport model, the four step model. State-of-the-art transport modeling approaches relevant to the research are activity-based modeling and dynamic traffic assignment. Section 4 and 5 introduce the concepts and various models of those modeling approaches. This chapter terminates with summary of the chapter.

2.2 CLASSIFICATIONS OF TRANSPORT MODELS

Before presenting the four step model, classifications of transport models needs to be discussed first. Several classifications are possible depending on which criteria are considered for classification. Two classifications, which are important in this research in the sense that it relates to level of detail for forecasting transport, are introduced here. According to what unit of modeling or analysis is taken, models can be divided into aggregated/disaggregated models or trip-/tour-/activity-based models.

2.2.1 Aggregated and Disaggregated Models

Division of transport models into aggregated and disaggregated model is one of the most fundamental classifications. Aggregated models were used almost without exception in transportation studies up to the late 1970s. This type of models uses zonal characteristics at an aggregated level without taking account of variations in purpose, person type, etc. (Hensher and Button, 2000). For example, population characteristics of each zone instead of the choice behavior of single travelers are used for prediction of the number of trips generated at each zone. These aggregated zonal models have been criticized for lack of any explanation in terms of human decision making, and this leads to a problem of inaccuracy and insensitivity of conventional models in particular to policy-related and societal changes. On that account to this, another type of models which is based on individual observation instead of aggregated zonal data, named disaggregated models has developed. This type of models, which became increasingly popular during the 1980s, offers substantial advantages over the traditional methods in that they include a reasonable amount of travel variation, and thus a higher level of accuracy in prediction (Ortuzar and Willumsen, 2001). ALBATROSS used in this research belongs to disaggregated model because modeling of travel behaviors depends on activity schedules of every individual.

2.2.2 Trip-, Tour-, and Activity-Based Models

Another classification of models is possible in terms of viewpoints on trips and unit of modeling. Travel behavior can be analyzed in multiple levels. It can be understood as a single trip that is independent from others, trip chain that multiple trips are connected, or activity participation. According to what unit of modeling of travel behavior is, a model is categorized into trip-based, tour-based, or activity-based approach. The simplest and oldest subclass divides daily schedule into trips. Some more recent models combine trips explicitly into tours. The last subclass combines the tours in a daily schedule (Ben-Akiva and Bowman, 1998). Activity-based model is the latest among the three approaches. However, it does not automatically mean that the approach is the best which have the least shortcomings and performs better than others. Appropriateness of a model relies on purposes and unit of analysis.

2.2.2.1 Trip-Based Models

In the beginning stage of transportation research, most models that adopted a disaggregate approach assume an individual trip as a unit of modeling, implying that no linkages and interdependencies between trips exist. The key assumption of the approach is that travel demand is derived as its own right. In the traditional matrix-based models (aggregated approach), understanding of travel behavior depends on the cells in the matrices. Furthermore, they represent time as simply a cost of making trips. Trip making is limited by constraints such as opening hours of facilities and car allocation within the household. Even though these constraints largely affect trip patterns, they are not incorporated in modeling. For example, traffic congestion on the way to home after work would cause late arrival at home. In turn, this might result in cancellation of shopping on that day because the shop closes before he arrives there. Since single trips are modeled separately from each other, this type of impact cannot be captured by the trip-based models. The assumptions of trip-based models induce a problem that internal consistency of the travel pattern is not guaranteed.

2.2.2.2 Tour-Based Models

In the tour-based model the trips are explicitly connected in tours. A more recent approach, tour-based approach, was first designed in the late 1970's and 80's in the Netherlands (Daly, van Zwam and van der Valk, 1983). Tour-based models combine trips into tours based on the fact that all travel can be viewed in terms of round-trip journeys based at the home. A tour is assumed to have primary activity and destination that is the major motivation for the journey (Bowman and Ben-Akiva, 2001). The modeling of tours improves representation of travel behavior in precision. And it is enabled by explicitly incorporating temporal-spatial constraints among activity stops in tours. However, this approach lacks a connection among multiple tours taken in the same day, thereby failing to capture the effects of inter-tour temporal-spatial constraints.

2.2.2.3 Activity-Based Models

The last subclass explicitly links the tours and explicitly models the time dimension. Spatial separation is the essence of travel demand (Hensher and Button, 2000). The objects that satisfy human's needs are spatially separated and this motivates movement from one location to another. As this intuitive explanation about trip motivation, activity-based models view travel as a derived demand for activity participation rather than an independent demand. The models provide much more insights in individuals' travel behaviors than traditional trip- or tour-based models. This individual-based disaggregate model simulate activity patterns of individuals over the day. Fundamental modeling unit is the day and individual or even the household. From the daily schedule, travel information is implicitly derived and the model makes use of all the information included in the schedule. An interval between activities is travel. Given information about activity location and start/end time, we can derive travel time and traffic route in accordance with transport mode. Moreover, household interaction and institutional system are captured in the model because the daily schedule is resulted from direct consideration of these constraints. And, activity scheduling of 24 hours time period allows modeling travel within the context of overall daily time-use (both durations and time-of-day).

2.3 THE FOUR STEP MODEL

The traditional aggregated trip based approach has been employed in transport modeling for at least the past thirty years. The most extensively used model in this tradition is the so-called classic four step model (FSM). As the name tells us, the modeling process usually proceeds in a sequence of four sub-models. The four stages relates to: i) trip generation, ii) trip distribution, iii) modal split, and iv) route assignment. Each sub-model addresses an intuitively reasonable question. Trip generation forecasts the number of trips that will be made from particular locations. Trip distribution determines where the trips will be destined to. Modal split is concerned with how the trips will be divided among various modes of travel and finally trip assignment predicts the routes that the travelers will take. The model has not been developed at once as a single model, but rather, years of experimentation and development have resulted in a general structure illustrated in Figure 2.1. The FSM takes into account both demand and supply side of transportation system in a single framework. Since the model represents an important position in the history of transport modeling, it is meaningful and helpful to discuss it here as an overarching framework. The model description is mainly based on the book 'Modeling Transport' by J. de Ortúzar and L.G. Willumsen and the course material provided from Delft University of Technology (Bovy, *et al.*, 2006).

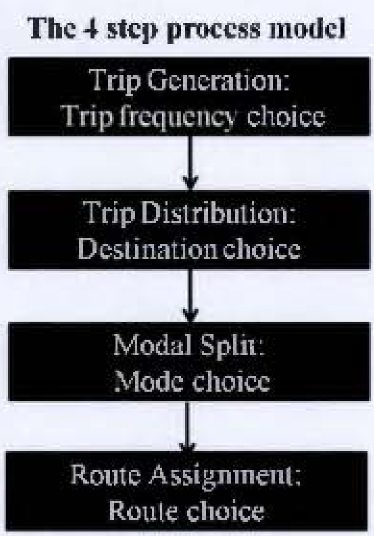


Figure 2.1 Modeling scheme of the four step model

The objective of the trip generation model is to predict the total number of trips generated from and attracted to each zone of the study area. In other words, the number of out-going trips is determined for each origin. Similarly for each destination, the number of incoming trips is determined. From another point of view, trip generation can be viewed as modeling of trip frequency, whether and how frequent an individual makes trips or not on a day. Trips can be distinguished between home-based and non-home-based trips. By definition in the book of Ortúzar and Willumsen, home-based trips are trips starting or ending at the home. Non-home-based trips have neither trip end at home. Trip ends are modeled as productions or attractions. The home-end of a trip is always the production because it is the household and its activity demands that give rise to, or produce, all trips. The beginning of a non-home-based trip is also production. Trip attraction is the non-home end of any home-based trip or the destination of any non-home based trip. Therefore, in trip generation stage, trip production and attraction are modeled. As trips are produced at inhabitant and activity location, socioeconomic factors such as income, car ownership, or household structure, are typically used for prediction of trip productions. In contrast to trip production, where household travel surveys can provide a rich database, it is difficult to obtain adequate information to estimate trip attraction. For this reason, it has proved far more tractable to develop models for the productions than for the attractions. Thus, the normal convention is that the productions are taken as well-defined but the attractions are merely an indication of the relative attractiveness of different zones (Bates, 2000). The variables that have been found to best explain this, however, are based on characteristics of the land use, such as employment level and roofed space available for commercial and other services. Trip generation can be computed by several modeling approaches: regression analysis, cross-classification, and discrete choice models. Regression methods assume a statistical relationship between the number of trips generated and socio-economic or demographic variables. In particular, the simplest and the most popular regression technique, the linear regression analysis establishes linearity between the variables. In other words, each independent variable exerts a linear influence on the dependent variable. Second, cross-classification or the so-called category analysis model separates a population into relatively homogeneous groups based on certain socio-economic characteristics and associates a trip rate to each group to calculate the total number of trips. Lastly, discrete choice models such as binary logit can be used to predict trip generation. Binary logit can represent the choice whether an individual will make one or more trips. Applying binary choice models consecutively in the nested structure, it predicts the probability that an individual will make no, one or more round tours for specific trip purposes on an average day.

The following step involves distribution of trips onto space from a 'production' (origin) to 'attraction' zones (destinations) as a function of the travel quality between zones. In other words, it links productions to attractions. For each origin, it is determined which fraction of outgoing trips go to which destination. As a result of this stage, we obtain a so-called origin-destination (O-D) trip matrix or table which specifies the number of trips that go from each origin to each destination. Traffic flow quantities can be conveniently represented by an array and the representation is the O-D table illustrated in Figure 2.2. Although multidimensional O-D table which has at least three dimensions is possible if disaggregation dimension such as mode or time of day exists, the simplest O-D table has a two dimensional array of cells where rows and columns represent origins and destinations dimension respectively. Whereas the place of origin is represented by rows of the table, columns of the table correspond to the place of destination. Thus, the margins of the O-D table refer to trip production and attraction computed by the trip generation model. Given the margin of the O-D table, what is predicted in this stage is T_{ij}, the number of trips between origin *i* and destination *j*. Either gravity model or growth factor methods can be used for calculation of T_{ij}. Gravity model is motivated by Newton's law of gravity that gravitational pull between two objects decreases as a function of the distance between the objects. The model explicitly relates flows between zones to inter-zonal impedance to travel. Impedance measures include distance, time, or cost of travel between zones. The concept is translated into the general form, written as follows:

$$T_{ij} = \alpha O_i D_j f(c_{ij})$$

where, T_{ij} is the number of trips between origin i and destination j

O_i is the number of production of zone i

D_j is the number of attraction of zone j

$f(c_{ij})$ is a generalized function of the travel costs between origin i and destination j

α is a proportionality factor

In general, it is assumed that the number of trips between *i* and *j* are proportional to the production and attraction of each zone and proportional to some decreasing function of the cost of traveling between them. However, trip production and attraction are not always known. Whether they are known or unknown imposes constraints, and hence, various models can be derived. But the general equation structure of the models is not significantly different from each

Zones		Destinations			$\sum_j T_{ij}$
		1	...	j	
Origins	1				O_1
	...				O_2
	i			T_{ij}	O_i
	$\sum_i T_{ij}$	D_1	D_2	D_j	

Trip Production

Trip Attraction

Figure 2.2 Design of the O-D trip table

other. Only actual quantities of trip production (O_i) or attraction (D_j) are replaced with relative measure of production or attraction. Growth factor method, on the other hand, is scaling an existing matrix by applying multiplicative factors to each cell of a base year matrix to predict future traffic flow. These trips can be made by different means of transportation and mode choice is modeled in the following stage of trip distribution. Predicted trips in the O-D matrix are allocated to modes. It is a choice problem of what transport mode to use for traveling from origin i to destination j. The best known widely used choice model is the multinomial logit (MNL) models. By MNL, the probability of using a specific mode is calculated. And the probability is multiplied by the number of trips between an origin and destination. It results in O-D table disaggregated by transport modes. Since the MNL structure is not always suitable when some of the modes are inherently similar to each other (i.e., similarity between train and bus in terms of seat availability), the nested logit models may also be used, precisely to reflect this aspect. The alternatives sharing common attributes are grouped into a nested structure. Within a nested structure, the first choice may be between the car and public transport, while, conditional on choosing public transport, there may be a further choice between bus and train. The principle of the MNL is discussed in section 'utility-maximizing models' of activity-based modeling approaches.

While the modeling steps presented until now concerns modeling travel demand between zones with specific transport modes, the following step, route assignment, describes supply modeling of network infrastructure by assigning the trips by each mode to the corresponding transportation network. By distributing travel demand onto the network, one could know how the network system performs or is used. This assignment step is discussed in the latter part of this chapter as a separate section because substantial amount of advancement in modeling techniques has been taken place as an independent field of research and it is of the research interest.

2.4 ACTIVITY-BASED MODELING

Activity-based modeling approach has received its attention since the early 1980's in order to overcome some of the drawbacks of disaggregated trip-based demand models. In particular, increasing concern for environmental issues has stimulated the legislation such as the Clean Air Act Amendments (CAAA) in 1991. This encompasses a major change in transportation planning and proposed policies concerning ridesharing, telecommuting, or congestion pricing. This explicitly encouraged the development of activity-based models as the most appropriate tool for predicting the complex impacts of these kinds of policies.

Stated simply, the motivation for activity based approach is that travel decisions are activity based (Ben-Akiva and Bowman, 1998). The approach provides better theoretical grounds for description and explanation of travel behavior. It assumes that travel demand is a derived demand in the sense that travelers arrange their travel to perform their activities. Another important assumption is acknowledgement of dependency of travel and non-travel aspects. The modelers realized that travel patterns are the result and the manifestation of the implementation of activity programs over time and space. In other words, activity patterns emerge as the result of complex interplay between physical environment, the institutional context, the transportation system, and individual's need to participate in activities, within a particular economic, political, social and cultural context (Timmermans, *et al*, 2002). Thus, activity-based approach has attempted to understand travel behavior in the broader context of activity participation. Travel patterns are revealed in the broader context of structure of activities, of the individual or household, with a framework emphasizing the importance of time and space constraints. Thus, underlying factors of travel behavior regarded important. The analysis examines why, where and when activities are taken place, and how activity engagement is related to the spatial and institutional organization.

As it considers travel behavior as a derivative of activities, individuals' activity scheduling behavior is emphasized in the approach. Activity program or daily schedule of an individual, which is basis for analysis in this approach, does not describe single-facet of one trip, but rather addresses complex decisions concerning multiple dimensions of various trips and activities. The reason behind this is that travel is not simple but complex behavior as multiple decision dimensions are interrelated for making trips and these travel decision are constrained by temporal, spatial, institutional context, such as for instance, opening hours, available transport mode and destination. With the daily schedule of an individual, it can be known what activities, at what destinations, at what times and for how long are conducted. In turn, the organization of the schedule implicitly explains trip information such as travel time and transport mode. Hence, models which describe activity scheduling of the execution of activity patterns are potentially useful tools for predicting travel behavior. This approach enables better understanding of travel behavior with the capability to predict how individuals respond to changes in their travel environment and how the responses are temporally correlated (Lam and Yin, 2001).

Several different modeling approaches have been developed in an effort to address the concepts explained above. First, constraints-based models based on the theory of space-time geography systematically identify the set of possible activity patterns, given various constraints. As the theory emphasizes the importance of the effects of space and time on activity engagement, this type of models evaluates the feasibilities to implement specific activity programs in constrained spatial-temporal setting. In addition, discrete choice theory describing the choice of activity patterns led to the development of utility-maximizing models. Finally, computational models from artificial intelligence and cognitive science are capable of describing activity scheduling in terms of human reasoning process. These modeling approaches are presented in the following.

The main objective of this section is to introduce the four modeling approaches of activity-based modeling, presenting concepts and models exist up-to-date. Nevertheless, it is necessary to look into theories in geography and urban planning from which the various modeling approaches originate. Therefore this section is organized as follows. First, theoretical underpinnings in geography and urban planning will be discussed. Then, the four modeling approaches of activity-based modeling will be explained concretely.

2.4.1 Theoretical Underpinnings in Geography and Urban Planning

Since the mid-eighties, there has been a rapid growth of interest in activity-based analysis in transportation research. As a reaction to the traditional approaches of aggregated models, there has been a search for paradigms with strong assumptions of the approaches relaxed. And this is called behavioral research, which focuses on the individual as the central unit of modeling. However, many concepts of activity-based approach are not new. They were originally founded in other disciplines such as geography and urban planning, although the interest gradually faded in the disciplines in recent years. Chapin and Hägerstrand initialized a behavioral research examining activity patterns of individuals and their associated travel. Both theories are complementary of each other, focusing on motivations for activity participation and space-time constraints, respectively.

2.4.1.1 Chapin's Theory of Activity Pattern in Land Use Planning

Chapin emphasized the importance of activity patterns in land use planning. He contends that activity patterns determine the demand for certain facilities at particular times and spaces. He identified four driving factors which manages individual's decision about activity engagement. They are propensity, opportunity, appropriate circumstances, and environmental context. 'Propensity' refers to needs and personal characteristics. 'Opportunity' accounted for spatial factors, referring to availability of services and facilities for activities. Temporal aspects are explained as 'appropriate circumstances'. However, he did not fully develop this line of thought in his empirical work, mostly concentrating on the interrelationship between activity pattern and socio-psychological propensity factor, which concerns personal characteristics and desires. His theory lacks sophistication of the necessary concept that activity pattern is a result of mixture of both spatial and temporal dimensions (Arentze and Timmermans, 2000). Rather than relating the dimensions in an integrated fashion with other factors, they are considered as separate concepts.

2.4.1.2 Hägerstrand's Space-time Geography

An approach which incorporates both space and time in a coherent framework to understand activity pattern is the Hägerstrand's space-time geography. It is also called as *time geography*. The fundamental principle underlying time geography is that all activities have both spatial and temporal dimensions that cannot be meaningfully separated (Miller, 2004). Moreover, the basic assumption of time geography is that time and space are scarce goods and that these space-time constraints largely influence daily activity patterns. Hence, the most important aspect of his theoretical reasoning is the constraints that restrict individual's choice in specific space-time environment. Constraints that limit the ability of individuals to travel and participate in activities include: i) the person's capabilities for trading time for space in movement (physical constraints); ii) the need to couple with others at particular locations for given durations, thus restricting the ability to participate in activities at other locations (coupling constraints); and iii) the ability of public or private authorities to limit physical presence from some locations in space and time (authority constraints).

Two central time geographic concepts are the space-time path and prism (see Figure 2.3 and 2.4). The space-time path represents trajectories of the individual's physical movement in space with respect to time. It describes activities in terms of duration and location. And the sequence of activities is represented as a path. A person must trade time for space through movement to participate in activities. And this tradeoff can be improved or worsen by the role of transportation. The slope of the curve illustrates the relationship. The steeper the slope is, the less efficiency in trading time for space. Moreover, the path highlights the constraining effects of an individual's need to be at different locations at different times. Extended concept of the space-time path is the space-time prism. It delimits the possible locations for the space-time path. Two anchoring location, vertical line, is where the fixed activities take place. Between the fixed activities, the

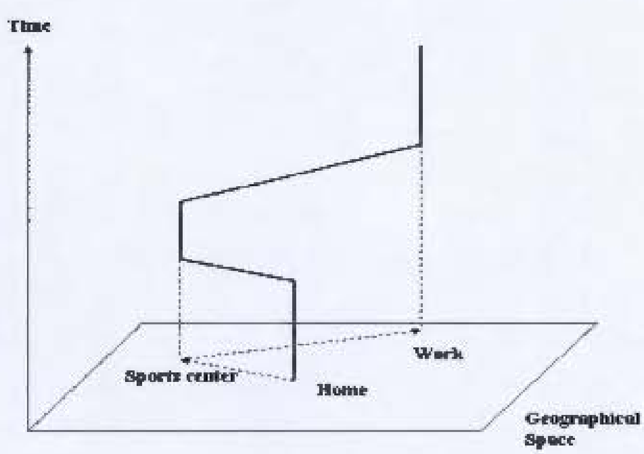


Figure 2.3 A space-time path

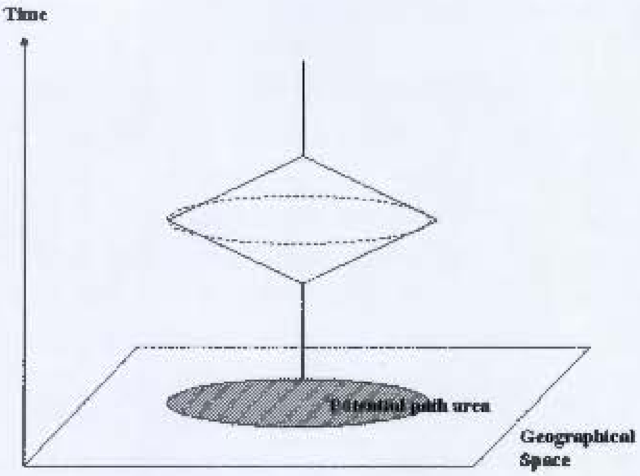


Figure 2.4 A space-time prism

possible space time path, which consists of flexible activities, can be displayed by a prism. The interior of the prism is the potential path space where the person could occupy during this travel episode.

Their theories are complimentary to each other because different focus was taken to explain activity-travel behavior. While the primary focus in Chapin's theory concerned characteristics of activity patterns with relation to propensity, motivational factors, Hägerstrand's work put emphasis on understanding which activity patterns can be realized in particular spatial-temporal settings. What attempted to integrate the theories was an approach of Cullen and Godson (1975). Their work bridges propensity and opportunity factors of Chapin, constraints concept of Hägerstrand and human decision-making process.

2.4.2 Activity-Based Modeling Approaches

Recently, activity-based approach has received its attention in transportation research. The basics are not new at all as pioneering works of Chapin and Hägerstrand shows. However, since the early 1980s, it was hard to find publications on activity analysis in urban planning and geography (Timmermans, *et al.*, 2002). Rather, scholars in different disciplines have consistent interest in the study of activity patterns. The concepts have advanced considerably, resulting in various theories and models. They are classified into constraints-based models, utility-maximizing models, computational models and micro-simulation models. Hägerstrand's time geography progressed towards constraints-based models. Utility maximizing models applying discrete choice theory to activity pattern choice are rooted in micro-economics. Computational process models explaining activity scheduling behavior in terms of human psychological reasoning process are engendered by artificial intelligence and cognitive science.

2.4.2.1 Constraints-Based Models

Based on the Hägerstrand's work, several models, namely constraints-based models have been developed. These models systematically examine whether the activity patterns can be realized within a particular space and time environment. The approach identifies feasible activity schedules as a function of various kinds of constraints, though it was not so much predicting activity-travel patterns. Various models exist from Lenntorp's PESASP (1978) to Mastic (Dijst, 1995), the mechanism of them are quite similar. To evaluate the feasibility of activity schedules under specified space-time setting, all possible activity sequences need to be generated first. A combinatorial algorithm is typically used here. Then, a set of activity programs is tested whether it can be happened in light of predefined constraints. The output is a list of feasible activity patterns. For parsimonious modeling, CARLA (Jones, *et al.*, 1983) generates activity arrangements in an ordered way to avoid wasting time on testing infeasible patterns. Also, heuristic rules reduce the number of alternatives examined. These feasible activity schedules can be used as measure of flexibility of time-space environment. This type of models is mostly used to assess political impacts of changes in the time-space environment on activity-travel behaviors. For example, BSP (Huigen, 1986) was used to see how school closure affects the travel times of the students. Nevertheless, the approach has a limitation that it ignores possible adaptation or adjustment behavior to changes in constraints. An individual may encounter unforeseen or unexpected events during the execution of the activities. Under this situation, adjustment behavior such as rescheduling of the activity plan can take place because people tend to seek alternative ways and still do have substantial choices. Moreover, an individual's preferences for specific activity patterns are not known by the models. Detailed explanations of the models can be found in the literatures of Ettema (1996), Arentze and Timmermans (2000), and Timmermans, *et al* (2002).

2.4.2.2 Utility-Maximizing Models

While Hägerstrand's time geography have developed to constraints-based models, Chapin's empirical work on the interrelationship between socio-psychological factors and activity patterns have led to choice models. Choice approach describes a problem of choosing an option out of a choice set consisting of multiple alternatives, for instance mode choice among car, bus, and train. This approach which is applied to modeling of daily activity pattern and scheduling behavior, viewed observed behavior as a representation of preferences. And individuals choose an alternative maximizing their utility. These utility-maximizing models such as discrete choice and logit models are dominant in transportation since 1980's. Theoretical grounds of discrete choice theory are provided by micro-economics and psychology. In particular, micro-economic consumer theory provides main principles underlying discrete choice models. However, classic consumption theory cannot be applied to description of discrete choice problem as it is. This is because discrete choice is not a continuous consumption of goods, but rather is a choice out of a number of mutually exclusive alternatives. Thus, consumption bundles with the amount of goods can be transformed into a set of choice alternatives in bundle of its characterizing attributes. Utilities are determined for each alternative separately as a function of the attributes and the total utility of an alternative is a weighted sum of part-worth utilities of attributes. And people choose an alternative that maximizes utility.

Difference in behavioral assumptions led to various types of models (Ettema, 1996). Models are differentiated by whether choice behavior is considered deterministic or stochastic (deterministic vs. non-deterministic choice models), that is, whether a random component, to account for various sources of error, is introduced (strict vs. random utility models), and which distribution is assumed for the error terms (probit vs. logit models).

The type of discrete choice model which is extensively used in transportation is the multinomial logit (MNL) model. The MNL is non-deterministic (random) logit model. Individual choice behavior is regarded probabilistic in nature, expressing the chance that an alternative is chosen as a probability. Random utility theory accounts for the stochastic characteristic, considering utility as a random variable. This implies inconsistency and intransitivity of choices and these come from unobserved factors not included in the model, taste variations, or measurement error. To capture these factors, random component describing various sources of error is added to deterministic part of the model. The error term is assumed to follow some statistical distribution and the assumption on statistical distribution differentiate between probit and logit model. Under the assumption of independently and identically Gumbel distributed error terms, the choice probabilities are represented by the well-known MNL. However, a serious limitation of the MNL is independence of irrelevant alternatives (IIA) property following from the assumption that the error components of separate alternatives are independently and identically distributed (IID). IIA property means that ratio of choice probability between two alternatives is not affected by adding a third alternative to the choice set. When two or more alternatives 'intuitively' have something in common that is not captured in the deterministic part of utility, their random errors are correlated and the IID assumption is not hold. For example, in case 'train' alternative is introduced to choice between 'car' and 'bus' alternatives, commonality of public transit between bus and train such as seat availability unequally influence choice probability of 'car' alternative and 'bus' alternative. Thus, incorrectness of IID assumption incorrectly exhibits IIA property. Intuition suggests relaxation of IIA property and the solution in the case mentioned is the nested logit model which describes multi-dimensional choices where a natural hierarchy exists in the decision process. Within the nested structure, lower dimensions of the hierarchy are conditioned by the outcomes of the higher dimensions. And the utility of a higher dimension alternative depends on the expected utility arising from the conditional dimension alternatives.

Activity-based choice modeling is characterized by its complexity in that activity scheduling relates to multiple decision dimensions such as destination, activity type or mode, and the decision-making takes place at different points in time. Models can be classified in terms of the choice dimensions modeled, the extent of interdependency between decision dimensions, and the decision-making process. The classification below is based upon works of Ettema (1996), Arentze and Timmermans (2000), and Timmermans, *et al.* (2002).

First, joint choice models regarded activity patterns as a result of a single choice rather than combination of partial choices on specific decision dimensions. Thus, choice of an activity pattern implicitly determines decision dimensions of travel. Adler and Ben-Akiva's model (1979) describes a single choice of a complete activity pattern. As another model in this approach, STARCHILD (Recker, *et al.*, 1986a, 1986b) incorporates a sub-model which diminishes the choice set. Rather than considering all feasible activity patterns, they are reduced into a smaller set of distinct independent activity patterns by removing inferior patterns based on a set of decision objectives. Moreover, the model includes household interaction such as car allocation and joint activity with other household members. These two models are operationalized as a MNL and in an application performed well in terms of goodness-of-fit and parameter significance. However, they are limited due to rigorous assumptions of independency between choice alternatives and the decisions are made at one point in time. Obviously, activity-scheduling is a multi-dimensional choice process, and alternatives undoubtedly share common dimensions.

The second group of models complements the weakness of the previous model by breaking down the activity scheduling process into a number of partial decisions, represented in a hierarchical nested decision structure. This nested logit models embodies the interrelationships between decisions made on different dimensions. Representative models of nested logit models are the daily activity schedule model proposed by Ben-Akiva and Bowman (1996) and Ettema, *et al.*'s HCG model (1997). The daily activity schedule model is an extension of tour-based models. An activity schedule consists of a set of tours and decisions for each tour are structured as different hierarchical levels. The first decision involves choice of a daily activity pattern. It concerns primary activity type, structure of the primary tour, the number and purpose of the secondary tour. The second choice relates to timing of the primary tour. Then, destination and mode choice of the primary tour is modeled. In case there is a secondary tour, the same decision procedure as that of the primary tour repeats for the secondary tour. The structure is illustrated in Figure 2.5. Choice at a certain level is affected by the expected maximum utility derived from the lower level alternatives. Ettema, *et al.*'s HCG model is successively run by two modules. The first module describes long-term travel and activity decisions largely influenced by exogenous variables. Then the second module runs daily activity and trip scheduling based on the outcomes of the previous module. The model adopts the form of nested structure. The number and type of out-of-home activities are decided, and then choice of the trip pattern is made. Having the specific trip pattern, destination and mode choice are modeled. In addition to these, there are less complicated nested logit models developed by Kawakami and Isobe (1989) and Wen and Koppleman (1999). PETRA (Fosgerau, 1998) which illustrates all tours undertaken in a day also belongs to this group. The nested logit models have problems in modeling the timing and

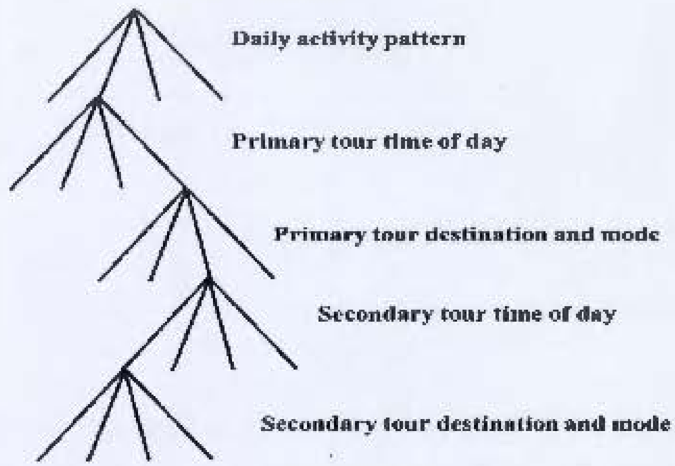


Figure 2.5 Structure of nested logit model (Ben-Akiva and Bowman, 1997)

duration of activities. For discrete choice, continuous time dimensions must be discretized into coarse time of day ranges. Thus, the models are not so much sensitive to behavior responses in terms of temporal dimensions.

Third, conjoint choice models are derived from data collected through experimental designs, though many models rely on revealed observed behavior in general. The advantage of conjoint choice model is that researchers have control over covariance structure of the independent variables. The exemplary model is COBRA (Wang and Timmermans, 2000). The model exhibits the potential of a new methodology of modeling activity patterns. However, conjoint design is not enough to establish a comprehensive model and it requires special data collection efforts.

The last group of utility-maximizing models depicts activity-scheduling as a sequential choice process. Activity patterns are outcome of consecutive choices of separate activities or trips. An important characteristic of some of this type of models is that decisions are made after completion of the previous ones, implying decision-making is processed during the execution of the activity schedule. VISEM (Fellendorf, *et al*., 1995) and PCATS and PCATS-RUM (Kitamura and Fujii, 1998) simulate activity-travel behavior. VISEM simulates the choices of an activity chain by drawing from empirical probability functions. Activity sequence is fixed and the decisions are made for a separate activity. Although the authors call PCATS a computational process model, it is regarded as utility-maximizing model because it uses utility function at least for the discrete choices in the process (Arentze and Timmermans, 2000). For example, the utility-maximizing, nested logit model of activity type choice is used in PCATS. It is a model of activity engagement and travel that incorporates the concept of Hägerstrand's space-time prism and generated activities and trips within given prisms. In addition to prism constraints, availability of transport modes and potential activity locations are constraining factors. Thus compared to nested-logit models, a different utility structure and a sequential scheduling process are assumed for the model system. Another model system, CEMDAP at the University of Texas at Austin, is another operational activity-based model. The system comprehensively models the daily activity-travel pattern of individuals. Provided various input variables of land-use, socio-demographic, activity system, and transportation level-of-service, the simulator implements a predefined econometric modeling system that represents choice behavior. Econometric models such as regression, hazard duration, or MNL are implemented in CEMDAP (Bhat, et al., 2004). Each decision variable is associated with an instance of one of modules, say mode choice with MNL module. Applying specific module to certain decision dimension, it predicts the corresponding choice. As a result, the output of the system is the predicted activity-travel patterns for all individuals in the simulation sample.

2.4.2.3 Computational Process Models

Utility-maximizing models are criticized in that individuals are not necessarily led to the optimal choice in reality. Rather than exhaustively evaluating solutions, people rely on heuristic process that is context-dependent. And the attempt to describe human decision-making in light of human reasoning process in the brain was initialized by researches in cognitive psychology and artificial intelligence (AI). In the disciplines, problem solving tasks are understood as a part of mental process. And the computer is regarded the most appropriate tool for displaying intelligent behavior as this discipline started from the analogy between the human brain and the computer in terms of data processing and storage. Thus, the objective is to build operational models which can simulate decision-making processes. As applicable to activity-based modeling, activity scheduling behavior can be modeled with this approach. Namely, computational process models are a computer program implementation of a production system model that specifies how a task is solved (Gärling, et al., 1994).

There have been two paradigms, resulting in different types of models (Ettema, 1996). Symbolic search space paradigm assumes that reasoning process is based on a symbolic representation of the real world. The symbolic representation constitutes the search space and this is a domain where the solution for a specific problem is driven. The search space consists of several states, moving from initial states that the search starts, through intermediate states, to goal states matching acceptable solutions for the problem (Ettema, 1996). And it is guided by heuristic rules. This algorithmic procedure in problem-solving has been introduced by Simon (1990). He contends that limitation of computational capability of human brain and complexity of everyday decisions supports the reason why humans rely on approximate methods to reduce the amount of information to be processed. Therefore, the solution acquired as a result of heuristic search may not be optimal but sufficient. This paradigm is operationalized by computational process models or production systems. Computational process models focus on the process of decision-making and capture schedule constraints explicitly. Heuristic is formalized as IF (Condition)…, THEN (Action) rules. Production system is composed of if-then rules, short term memory of current states, and a controlling mechanism of specific rules to apply. The overall process is one of matching rules against the current states while carrying out suited action in an iterative manner until the terminate state is reached. Thus, activity patterns can be represented as a sequence of condition-action pairs. Alternatively, the connectionist paradigm describes intelligent processes by activation of certain nodes in the networks connected by weighted links. This representation is similar to the physical structure of the brain as human reasoning process is exchange of electric activities among neurons that are connected to each other. The concept is operationalized in the form of neural networks consisting of nodes and links. A node itself is not meaningful at all, but patterns of activated nodes imply a solution to a specific task. As it is based on learning processes, the network must be trained prior to application to determine link weights and firing rules. The attractiveness of the rule-based model is that activity scheduling behavior can be better represented with more flexibility than econometric models which use systems of equations to capture relationships among attributes and predict the probability of decision outcomes (Bhat, et al., 2004). Moreover, rescheduling is described by some models unlike the utility-maximizing models. Nevertheless, the disadvantages of the models are difficulty of calibration of the models and generalization of modeling process.

The first model in this tradition is SCHEDULER (Gärling, *et al.*, 1989). In a conceptual framework, the model focuses on the process of how individuals organize activities. It follows reactive heuristic process which means preliminary schedule is adapted in response to the outcome of previous decisions. Activities are chosen from long-term calendar of the household. Then, activities are sequenced in consideration of temporal constraints and total travel distance incorporating 'nearest-neighbor' heuristics. What follows is mental execution of sequence of activities that evaluates its feasibility under certain circumstances. It results in more detailed information. Production system is adopted to operate the model. Hence, the decisions are based on a set of heuristic functions. Golledge, *et al.* (1994), applied this model to predict the activity patterns of commuters in response to the introduction of telecommuting.

AMOS (Kitamura, *et al.*, 1993), a partial model of SAMS, focuses on adaptation behavior in response to a changed environment. Ettema (1996) and Arentze and Timmermans (2000) discuss the model in their publication. Adaptation of the schedules is guided by learning processes in which individuals gain knowledge about the varying environment. Hence, the model can be used as an evaluation tool on how individuals adapt to the new policies. Given a baseline activity-schedule, the adaptation behavior is encouraged by recognition of changes in travel environment. Having recognized a need for behavioral changes, the search begins by identifying possible response options, and then individuals try out all the alternatives until a satisfactory pattern is established. In other words, the overall procedure involves the selection of a basic response which narrows the domain of search. This is followed by the search for one feasible adjustment and the decision to accept the adjustment or continue the search. These are operationalized by the specific module of the system in sequence; baseline activity-travel pattern synthesizer, response option generator, activity-travel adjuster, and evaluation routine. The outcome of the activity-travel adjuster, the adjusted activity pattern is measured by the evaluation routine. AMOS has a potential to be very informative in that it predicts switches to specific policy changes from a baseline schedule, implying a policy-specific switching model. However, custom development is required for each policy and it does not forecast long-run effect of policy changes. More aspects about the weakness of the model are documented in publication of Bowman and Ben-Akiva (1996), and Bowman (1998).

Another model, SMASH (Ettema, *et al.*, 1994), is meaningful in the way the activity schedule is constructed. Sequential process is assumed and the schedule is constructed and adopted from an empty schedule through an iterative process of four basic actions defined as addition, deletion, substitution of activities in activity agenda and activity schedule, and stop of the process. A decision is made each time whether or not to accept the current schedule and stop the building process. In each step, production system conducts a generic non-exhaustive search enumerating all possibilities. And the choice between schedule adjustment and schedule acceptance is implemented as a nested logit model. Thus, SMASH incorporates both production system and utility-maximization. A key feature of the model is description of schedule construction process. Though, it has shortcomings that individuals must step through the entire schedule building process. Moreover, non-exhaustive search heuristics may be inadequate because human decision making involves a certain search domain, thus systematically restricting alternatives.

The latest computational models are ALBATROSS (Arentze and Timmermans, 2000) that at the same time is the topic of this research, and TASHA (Miller and Roorda, 2003). It is only recently that operational models of the household activity scheduling process have begun to merge. ALBATROSS is a rule-based model where prediction of activity patterns is based on choice heuristics represented by decision trees. The model system is characterized by its comprehensiveness in that Albatross ideally includes all relevant variables of decision variables and explanatory variables. Moreover, Albatross is flexible as it describes many different types of activity patterns and different response types also. That is to say, the system consists of a set of agents by which specific functions are performed. The agents handle the available data sets required for scheduling decisions, the induction of decision rules from empirical data, the simulation of activity patterns, the analysis of activity patterns supporting the display of basic properties, the reporting of model performance, the calculation of indicators, and the evaluation of scenarios. More detailed explanation of the model system is presented as a separate chapter. Toronto Area Scheduling model for Household Agents, TASHA in short, operationalized the conceptual model of household decision making presented by Miller (2002). The conceptual model, a basis for implementation of TASHA, is well described in the publication of Miller (2002). TASHA generates activity and travel patterns for a twenty-four hour typical weekday for all persons in a household by individually simulating the behavior of every household in a representative sample of households. This model uses conventional trip diary data from survey, implying model transferability to other region where such trip data exist. And the concept of the project, defined by Axhausen as a coordinated set of activities tied together by a common goal or outcome, is one of the main features of the model, but has not yet been implemented. Rule-based method is used to organize activities into projects and to make schedules for interacting household members.

2.4.2.4 Micro-Simulation Models

Micro-simulation models integrate household activities, land use distributions, regional demographics and transportation networks in time-dependent fashion. Unlike the previous models developed by application of theoretical basis to empirical data, micro-simulation models are rather data-driven. The best-known micro-simulation models are TRANSIMS (Transportation Analysis and SIMulation System), developed at the Los Alamos National Laboratory in the United States, and its successor MATSim (Multi-Agent Transport Simulation) developed jointly in Zurich and Berlin (Balmer, *et al.*, 2008)

The objective of TRANSIMS is to develop a set of mutually supporting realistic simulations, models, databases that employ advanced computational and analytical techniques to create an integrated regional transportation system. It is a microscopic simulation model which simulates all entities of the system, for instance, travelers, vehicles, intersections, etc., as individual objects. The conceptual advantage of a micro-simulation is that it can be made arbitrarily realistic in principle. TRANSIMS consists of a set of modules. Demographic data is disaggregated to obtain individual households and household members. It creates synthetic population by combining available data using proportional fitting method. For each individual with socio-demographic characteristics, a set of activities and activity locations for a day is generated. Modeling of choices on mode and route which connects activities at different location is followed. Then, the traffic micro-simulation executes individual's activity schedules simultaneously onto the transportation network. The model has received much attention for this transportation micro simulation. It imitates the movement and interactions of travelers throughout transportation system of metropolitan area. More detailed description of the modules refers to a technical report of TRANSIMS (Smith, *et al.*, 1995).

The next generation of micro-simulation models is MATSim. The description of the system here is mainly referred to dissertation of Balmer (2007). It is an agent-based simulation model where each traveler is modeled as an individual agent in the simulation. Multi-Agent Transport Simulation Toolkit (MATSim-T, 2006) is presented as a modular platform for transport planning software. An algorithmic core of the toolkit is MATSim-EA. It tries to iteratively optimize each individual's demand. The basic concept for MATSim-EA originates from evolutionary algorithm presented by Bäck (1996). The idea behind this optimization process is first to create a population of objects, say activity plans in this context. The objects are assessed against a fitness function, which evaluates how successful the object was. Then, the object that fulfills the function the best, the highest score, survive while the others die. Applying the concept of evolutionary algorithm to activity-scheduling behavior, MATSim-EA processes three pieces of tasks; i) each agent independently generates a plan for a day, ii) all agent's plans are simultaneously executed in the simulation of the physical system, MATSim-EXEC, iii) for each executed plan, then the performance is calculated by a utility function. MATSim-EXEC does actual microscopic traffic simulation including interaction between the agents of the system with each other. The stochastic, queue based, agent traffic simulation is used here to simulate the trip agents during the defined time period. Given the scores of the plans, agents can change plans again. And the next execution of the revised plans takes place. The system iterates the above process until a stable state is reached. The strong feature of the system is that each algorithm or module has its specific purpose, splitting up the demand modeling process into pieces. In other words, it is no longer forced to follow the four step process in its predefined way. One can iteratively run or combine any algorithms of the system (Balmer, 2007). For the details of the model system, see several papers and the website of MATSim (http://www.matsim.org/). A list of papers and reports to MATSim is provided in the publication of Balmer, *et al.* (2008).

2.5 TRAFFIC FLOW PREDICTION MODELS

Traffic assignment techniques are used to predict traffic flows on networks, given a certain travel demand. In addition to traffic assignment models, there is another class of traffic flow prediction models, namely (pure) simulation models (De Romph 1994, Bliemer 2001). Or alternatively termed, they can be divided into equilibrium and non-equilibrium models. However, neither of these two models have a generally accepted definition. Simulation models "only" propagate the traffic flow along the routes through time, without taking route choice into account. These models also can be called either dynamic traffic propagation models or dynamic network loading models. Traffic propagation models instead of simulation models will be used throughout the thesis to avoid any confusion of terminology between simulation models and simulation-based assignment models. On the contrary, assignment models are not necessarily time based models that do consider route-choice. That is, time aspects can be assumed to be either constant or variant. This section presents detailed description of traffic propagation models and assignment models. Basic concepts and a wide variety of different approaches will be provided.

2.5.1 Simulation Models

In general, simulation is defined as dynamic representation of some part of the real world achieved by building a computer model and moving it through time (Drew 1968). Computer aided simulations are frequently employed in traffic and transportation system analysis in order to describe inherent dynamic nature of traffic. Traffic behavior is described over extended periods of time by conducting numerical experiments on a computer. They are basically dynamic time propagation models because the state of the model at time $t+1$ is calculated based on the state at time t (De Romph, 1994). Traffic propagation models can be classified in several ways. Since real-time computational tractability matters in simulation, the basic classification is in terms of the choice of granularity; macroscopic, mesoscopic, and microscopic models.

Macroscopic models are mathematical models that formulate the relationships among traffic flow characteristics at flow level, which does not discern individual vehicles but they are treated as a fluid-like continuum. These models represent traffic stream in terms of aggregated measures such as density and flow rate. The underlying assumption is originated from hydrodynamic theory that traffic streams as a whole are comparable to fluid streams. The most widely used dynamic macroscopic model is the hydrodynamic model, namely LWR model, developed by Lighthill and Whitham (1955) and Richards (1956); see Daganzo (1997) for a review. Mesoscopic models handle packet of vehicles. On the other hand, microscopic models are vehicle-oriented models where the movement of individual vehicles is tracked, capturing the interaction between individual drivers. Microscopic properties such as the state of each vehicle (i.e., the position and the speed of a single vehicle) are represented by the models. And this class of models usually has stochastic characteristics while macroscopic models are deterministic. As it allows analysis of individual vehicles, microscopic models are used to investigate local behavior of traffic. Although route choice behavior is not incorporated in the models, they are not useless because the models can be used as a dynamic network loading component in an assignment model (Bliemer, 2001). But, this type of models is not well suited for network problems and application is rather limited because they are designed for small networks. Description of traffic propagation models will not be provided because it is out of the scope of this thesis.

2.5.2 Assignment Models

Traffic assignment usually is the last step in the conventional FSM as the process of allocating the trips in one or more trip matrices to their routes in the network. The outcome of route assignment is link flows and link travel time. The resulting traffic pattern, determined by following certain rules or principles of route choice of travelers, provides information about the performance or the use of the network. Furthermore, the results can be used for planning and design purposes by getting insight the state of the network in response to changes.

The interaction between transport supply and travel demand produces a flow pattern on the network. It is a traffic assignment model that incorporates and manifests the interaction of the two elements. Traffic assignment models determine an optimal trade-off between supply and demand. According to what assumption is considered, several models have been formulated. And the classification of those models is possible on different levels. However, one of the most fundamental theoretical assumptions used is how the time aspects are treated within the model. In other words, whether the time, travel demand and cost is regarded as a constant or a variant determines a distinction between static and dynamic traffic assignment models. As for input to assignment model, travel demand and transport network is required. Travel demand is given as O-D matrix and transport network consists of nodes, links, and link characteristics such as maximum speed and capacity. The general scheme of assignment framework typically is composed of the two sub problems; route choice and network loading. The route choice model produces route flows from the trips of O-D table. Having determined the route flows by the route choice model, the next step is to load the flows onto the network. The flows are transferred to the network loading model to simulate the flows over the network. Network loading is also referred to as traffic flow propagation that merely propagates the route flows though the link. Therefore, route choice and network loading is combined in traffic assignment models. The objective of this section is to review the traffic assignment models, with emphasis on dynamic assignment models, which are superior to static models given the inherent dynamic nature of traffic. Nevertheless this does not mean that static assignment technique is inferior because static models can provide basic insights to problems.

2.5.2.1 Static Traffic Assignment Models

A well established and widely accepted class of traffic assignment models is static traffic assignment (STA) models. The review is mainly referred to Ortúzar and Willumsen (2001) and Bovy, *et al.* (2006). STA assumes that traffic conditions such as travel demand and link cost functions are constant over time during the time span of interest. In other words, link flows and link travel times are time independent. A given travel demand is allocated on the transportation network and spatial distribution of the traffic volume are obtained as a result of the assignment. The resulting traffic volume represents average conditions of the network for the time period.

It is commonly observed that different drivers often use different routes when traveling between the same two points. Or, sometimes even the same driver chooses different routes under the same situation. This can be explained by two factors: congestion and stochastic effects. Stochastic elements refer to variability of perceptions or level of knowledge on the link attributes. Congestion effects make some ideal routes less attractive as delays increase with greater usage, implying whether or not capacity of the link is restrained. These two elements can be dimensions for distinction of STA models. Different types of methods identified according to them are depicted in Table 2.1.

Table 2.1 Types of static traffic assignment model

		Stochastic elements	
		No	Yes
Congestion effects	No	All-or-nothing	Pure stochastic
	Yes	Wardrop's user equilibrium	Stochastic user equilibrium

The simplest assignment method is all-or-nothing (AON) assignment. It is assumed that link capacity is not restrained, which means that the link travel time is fixed and does not vary depending on the congestion of the link. In addition, all drivers consider the same attributes for route choice and perceive them in the same way. As a consequence, all drivers are assigned to the singly most attractive route even if there is another path with the same travel time. This method is unrealistic in that only a single route is utilized by travelers. In contrast to all-or-nothing assignment, stochastic methods emphasize behavior differences of drivers. This method allocates the trips between a certain O-D pair over several routes, assuming that route choice behavior varies from person to person. Because travelers choose their perceived shortest path, distribution over multiple routes between the same origin and destination is possible.

In reality, every link in the network has capacity limits and this can be expressed by relational function between the level of flow and travel costs of links. Thus, more realistic assignment needs to take capacity restraints into account. Wardrop's user equilibrium method considers congestion of the link. Several routes between O-D pairs are still possible with this method in that level of service attributes of links depend on link loads, even though every driver behaves in the same way. Travelers will search for efficient routes to travel between two points and it is expected that they reach some stable set of choices. The stable status, equilibrium of the network, is best described by Wardrop. It is defined that no individual driver can reduce his path costs by switching route under equilibrium condition and travel time of all used route alternatives must be equal. The assignment method which considers both behavioral difference of drivers and congestion of the network is stochastic user-equilibrium assignment (SUE). Equilibrium condition of this assignment is defined that no travelers can improve his perceived travel time by unilaterally changing route. Route choice behavior depends on travel time experienced and this is affected by link flow. The only difference with Wardrop's equilibrium is that SUE model uses perceived travel cost which varies from person to person. SUE performs better than other model in that it reflects both properties in modeling.

STA models are widely accepted models having both advantages and disadvantages. Although a number of shortcomings of STA have been reported and they mostly concern time independent assumption of the model, STA is an extremely valuable approach to traffic analysis as it allows quick estimation of the use of traffic networks and an initial appreciation of the situation. Initial estimates are used to perform more detailed analysis or more demanding dynamic assignment. The reason why the results of static assignment are used for dynamic assignment is that calculation of the initial turning movement proportions at junctions is performed with STA. Despite the widespread use, shortcomings of them have become more apparent as congestion becomes important in transportation research. The major limitation is inability to fully capture the true dynamics of trip departure and real-time routing behavior. They cannot satisfactorily be captured by a static modeling approach. For example, departure time change caused by congestion cannot be modeled unless a time-varying element is considered. Modeling not immediate but average conditions of traffic has the risk of underestimating the congestion effect. Therefore, the need for better models, dynamic assignment models in this context, became apparent and will be discussed in the following section.

2.5.2.2 Dynamic Traffic Assignment Models

Dynamic traffic assignment (DTA) techniques have evolved substantially since the introduction of the first analytical dynamic assignment model by Merchant and Nemhauser (1978a, 1978b), addressing many of the limitations of STA methods. They depart from the standard STA models to deal with time-varying flows (Peeta and Ziliaskopoulos, 2001). By considering the time-varying nature of traffic flows, dynamic models can produce practically empirically more realistic and useful estimates of state variables such as speeds, queue lengths, delays and congestion effects so as to better assess the environmental and functional impacts of transportation planning measures. Basically, the dynamics are introduced by adopting discrete time periods or continuous time functions in the models. However, the existence of time variance by itself does not imply that the model is dynamic, unless events in one period can influence events in the next (De Romph, 1994). Hence, the true dynamic is a comprehensive concept, which also includes dynamics caused by interactions between travel facets. For example, departure time choice needs to be modeled in dynamic way as well because travelers change departure time in response to traffic situation out there.

While STA models do route choice and network loading in time-independent way, DTA models solve the sub problems dynamically. The DTA models typically consist of dynamic route choice and dynamic network loading (DNL). And the DTA problem can be solved by combining the route choice problem and DNL problem. The outcomes of assignment are dynamic link and route flows, dynamic link travel times, and dynamic route travel costs. Consequently, they can be transferred back into route choice again such that adaptation behavior of drivers can take place in response to changes in transportation environment. Likewise, DTA itself is an iterative procedure, converging to traffic equilibrium. The DTA scheme is depicted in Figure 2.6.

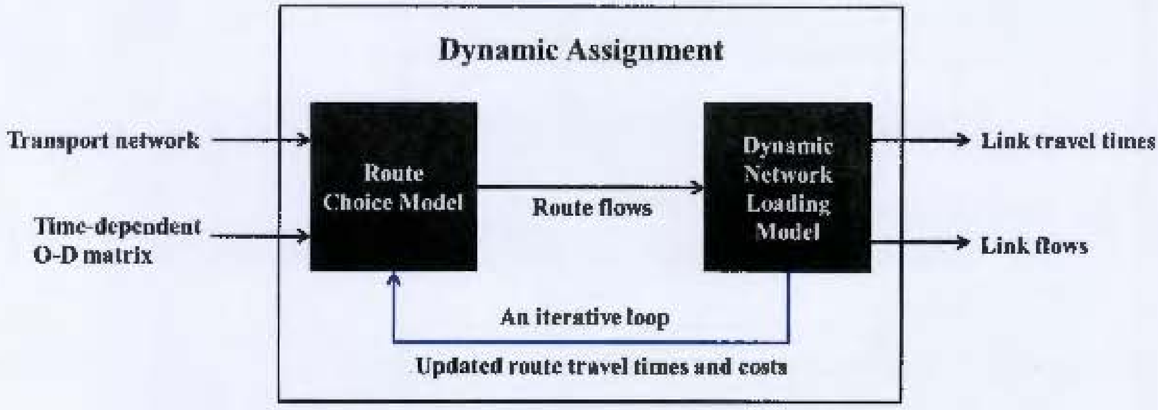


Figure 2.6 Dynamic assignment schematic frameworks

Although a wide variety of different approaches to DTA problem has been developed, none of the models provides a universal solution for general networks (Peeta and Ziliaskopoulus, 2001). This results from inherently ill-behaved system properties of general DTA problem. That is to say, it needs to adequately represent traffic realism and human behavior. Hence, most researchers or practitioners are often faced with a trade-off between traffic realism and mathematical tractability when solving the assignment problem. Division into two groups of approaches in DTA is possible in terms of these two conditions: analytical (mathematical) and heuristic approach (De Romph, 1994). The main interest of analytical approach is to provide a theoretically sound mathematical formulation of the problem. It seeks system optimum or user equilibrium. Thus, convergence of algorithms and unique solution matters in this approach. The approach has some advantages over heuristic models in terms of its ability of determining solution properties, which implies uniqueness of a solution model. And it is theoretically very powerful. However, the amount of realism is rather limited because assignment problem is solved with simplified assumptions, precluding capture of all the details of the problem. Furthermore, most of them can only be applied to a small problem. As problem size grows larger, it is computationally cumbersome or not practical due to their analytical nature. On the contrary, heuristic approach pragmatically models traffic, reflecting reality as closely as possible, without bothering too much about a consistent mathematical structure, that is, convergence of algorithms on uniqueness of solutions. It captures dynamics and driver behavior, preventing the guarantee of the standard mathematical properties. This approach seeks for effectiveness, robustness, and deployment of efficiency. It has a higher degree of flexibility, which ensures traffic realism. However, the structure of the models is sacrificed to the flexibility of the models. Another classification in terms of methodology, which is more specific, is proposed by Peeta and Ziliaskopoulos (2001). Analytical approach is further classified into mathematical programming, optimal control, variational inequality models, whereas simulation-based models are part of the heuristic approach. For comprehensive overview of the four methodologies, see Peeta and Ziliaskopoulos (2001). While analytical models emphasize derivation of theoretical insights, simulation-based models concentrate on enabling practical deployment for realistic network.

Among the four DTA methodologies above mentioned, simulation-based dynamic traffic assignment models will be further explained here because MaDAM, the macroscopic traffic simulation model used in this research, does solve assignment problem in a quite similar way, without adding iterative loop in the model. The advent of large and powerful computers has prompted simulation techniques to gain significance. Simulation-based DTA, which gained greater acceptability in the context of real-world deployment (Ben-Akiva, *et al.*, 1998), employs DTA techniques in a simulation environment for real-time application by systematically combining dynamic network assignment models and traffic simulation models. It adopts car following simulation techniques of traffic propagation models. This class of models is used primarily for traffic operational studies to capture the evolution of traffic flows in a network, which result from the decisions of individual travelers making route choice decisions. It uses a traffic simulator to replicate the complex traffic flow dynamics. The advantages of simulation-based DTA methods include: i) it captures the time-dependent interactions between the demand for the network and the supply of the network, ii) it represents the travel choices in great detail, and iii) it predicts the locations and impacts of traffic congestion. However, the models have some drawbacks that concern their inability to derive the associated mathematical properties. Moreover, computational burden can be operationally restrictive. Thus, trade-off between solution accuracy and computational efficiency could be a matter. The key issue of simulation-based DTA models is that theoretical insights cannot be analytically derived as the complex traffic interactions are modeled using simulation. Several models are implemented and available as software products such as CONTRAM (Leonard, *et al.*, 1978), DynaMIT (Ben-Akiva, *et al.*, 1998), and DYNASMART (Mahmassani, *et al.*, 1998).

2.6 CONCLUSION

This chapter has reviewed transport modeling approaches and associated models from classic to state-of-the art. Basic concepts of approaches and key characteristics of several models were described. To summarize, it can be explained by the flow of history in transport modeling. Many researchers and practitioners have searched for better models which replicate real-world as closely as possible and therefore provide the analyst with more accurate outcomes and forecasts. Those efforts have appeared as development of new approaches making up for limitations of transport models of earlier generation, rather than making something out of nothing. The modeling approaches shifted from initially trip-based models at aggregated level to activity-based models at disaggregated level. The perspective change on traffic movement has played a pivotal role for the paradigm shift. While the earlier models assumes travel demand in its own right without consideration of individual travel behavior and temporal aspects, the latest models view travel is a derived demand to satisfy human needs, appeared as outcomes of consecutive choices between travel facets. The most critical point is that modelers have begun to recognize traffic is essentially dynamic process. To complement the shortcomings of traditional four step process model, two research fields have been actively being developed: Activity-based approach captures interrelationships between trip-related attributes by investigating individual's daily activity pattern. And thus, activity scheduling behaviors are modeled with this approach and it enabled more accurate and consistent prediction than depending only on trip records. Replacing static traffic assignment, there has been substantial amount of advancements in dynamic traffic propagation models, though it was independently developed from activity-based models. Incorporating time-varying aspects to capture dynamic nature of traffic, DTA models and dynamic traffic propagation models predict how traffic flows change over time by propagating the dynamic travel demand onto the given transport network. In the following chapter, the two instances of activity-based models and dynamic traffic propagation models, ALBATROSS and MaDAM will be discussed in detail.

3 THE MODELS: ALBATROSS and MaDAM

3.1 INTRODUCTION

This chapter introduces two models and its operational model system in the field of activity-based travel demand modeling and dynamic traffic simulation modeling. Commissioned by the Dutch Ministry of Transportation and Public works, urban planning group at Eindhoven University of Technology has developed ALBATROSS. Another is MaDAM developed by Goudappel Coffeng BV and Omnitrans International. The attempt to building the linked model system in this research is facilitated by two application software, ALBATROSS and OmniTRANS. Albatross is a micro-simulation system which predicts activity-patterns of individuals of every household based on decision rules incorporating various constraints (Arentze and Timmermans, 2000), and hence, travel demand is predicted. OmniTRANS is an integrated multi-modal transportation planning package. It delivers a set of tools used to address an extensive range of transport modeling problems. In particular, MaDAM simulates the propagation of traffic throughout the network, taking into account queuing and junction effects, and automatically detecting network flow conditions (OmniTRANS Manual), and thus traffic flow on the network is predicted.

This chapter is structured as follows. Section 2 presents conceptual model and its operational model system of ALBATROSS. Conceptual description of Albatross first identifies the modeling components, and then combines the separate components into an integrated model, consequently constructing the modeling framework. The Albatross system is described in terms of functionality of modules. Section 3 introduces OmniTRANS software package, also with functionality. Then, dynamic traffic propagation model, MaDAM, is conceptually described without introduction of mathematical details.

3.2 ALBATROSS

Straightforwardly speaking, activity scheduling and organizing the related travel are characterized as complexity. Manifested traffic pattern is the result of complex decision making process to participate activities at different locations considering spatial-temporal settings and institutional context. Decisions regarding activity scheduling and its implementation relate to multiple dimensions to be decided and they are interrelated to each other. Moreover, the circumstances also are complex as physical environment changes from moment to moment, transportation situation is uncertain, and multi-day variations exist. This complexity may dramatically influence the way individuals organize their daily life, including travel. Unlike the utility-maximizing activity-based models, ALBATROSS assumes that individuals do not systematically compare all possible activity schedules leading to optimal choice, but rather follow learning process guided by circumstance-specific heuristics. The insight is the starting point of development of ALBATROSS, abbreviation for A Learning Based Transportation Oriented Simulation System.

3.2.1 Conceptual Framework

Modeling problem in the context of activity-based travel demand modeling involves activity scheduling. Prior to modeling activity scheduling behavior, components for modeling need to be identified first. This section introduces the key concepts by definition. The concepts are translated into components of ALBATROSS. And, linking the components in an integrated fashion formulates the ALBATROSS conceptual framework. Subsequently, activity scheduling mechanism will be discussed, as the core of the model is modeling activity-scheduling behavior.

3.2.1.1 Components of Activity-Based Travel Demand Modeling

As mentioned in the previous chapter, activity-based travel demand modeling predicts which activities are conducted where, when, for how long, with whom, which transport mode is used, and ideally the implied route choice decisions, taking spatial, temporal, institutional settings into account. These decision dimensions are modeling components of the ALBATROSS model. In the following, concepts and definitions of the components will be explained. The description is based on the book of Albatross version 1 (Arentze and Timmermans, 2000).

A day and a household is a unit of prediction in the model. ALBATROSS predicts activity schedules of every individual of a household except children for a given day. Any given day can be divided into a series of *episodes*. Hence, summing up the activity duration of all episodes of the day is 24 hours. Each episode represents another activity or trip. An episode is characterized by a start time, an end time, and duration. A question of which activities to perform is largely affected by decisions made at the level of household. *Activity calendar*, which is a set of activities that individual need to complete within a particular time horizon, say a month or a year, is household-specific as it is based upon long-term household decisions and household characteristics. Whereas time frame of activity calendar can be a week, a month, or a year, what consists of a list of activities that are planned for a particular day is *activity program*. Activity program is drawn from household activity calendar and the activities of a program are daily activities. *Activity schedule*, which is more common to us, is an ordered sequence and timing of the daily activities. However, not all the activities in the schedule can be executed as they were planned due to unplanned events. Realized activity schedule is called *activity pattern* and it can be explicitly defined in terms of activity location, transport mode, timing, duration, and travel party.

Activities can be classified in two ways in terms of the spatial location where an activity is engaged and obligation to conduct activities. The first classification divides into *in-home* and *out-of-home* activities. They literally mean that activities executed at home or other than home. Another is categorized into *mandatory* and *discretionary* activities. While mandatory activities are the activities that individuals must perform with some degree of obligation, discretionary activities are those in which an individual chooses to engage or not to engage, which means the individual is not obliged to perform.

Activities are taken place at certain places in *physical environment* which is tangible. Even though there are more activity locations beyond an individual's cognition, imperfect information of space limits activity locations that individuals conceive in physical environment. The only those locations that are familiar to an individual is called *cognitive environment*. Thus, a destination is chosen out of location alternatives of cognitive environment. Moreover, activity location is characterized by a set of attributes. Activity locations are specified with certain *land-use* narrowing down the possible destinations for conduct of a particular activity.

Spatial separation of activity locations causes travel between locations to complete activity schedule. What facilitates the necessary link between different locations is transportation system. Even though the transportation system provides opportunity to overcome distances, transportation system itself also acts as a constraint because travelers always consider how long it does take to reach a destination in terms of temporal dimensions such as opening hours of a shop, implying that travel speed matters for using the system. People organize travel in different ways in order to increase the overall utility of activity participation. In other words, to minimize travel cost, people make *trip chains*. Trip chain is defined as a series of trips that originates and terminates at home and includes two or more stops between two home-based trips. Transport mode choice is always involved for planning out-of-home activities. Travelers consider *available transport modes* at the time of usage. In particular, availability of an individual car, for example, constrains the mode choice of a member of a household where there is only one car in a household.

As a member of society, there are some rules and regulations with which individuals must comply, which are referred to as *institutional context*. Examples of institutional context include opening hours of a shop and the minimum age of eighteen for getting a driver's license or buying liquors. Especially, opening hours or available time slots of services dictates the earliest possible starting time and latest possible end time for activity participation. Therefore, for scheduling of each activity, an individual considers the temporal aspects in terms of institutional context.

To sum up, activity scheduling involves sub-decisions on activity type, destination, time of day, travel party, and transport mode. The related choices are modeled as a function of activity calendar, an individual's cognitive environment, availability of transport modes, land-use of activity locations, and institutional context (Arentze and Timmermans, 2000).

3.2.1.2 Activity Scheduling Decision-Making Process

Having identified which variables are modeled in the Albatross model, this section describes assumptions about the decision-making process in activity scheduling and execution of schedules. That is to say, how an individual arrives at activity schedules of a particular day from long-term activity calendar at the level of household, and additionally, how an individual chooses between feasible activity patterns and rescheduling during the conduct of the activity schedule will be discussed here. Rather than viewing choice as an optimal solution resulting from systematic comparisons of all possible patterns as utility-maximizing models, the Albatross hypothesizes that activity scheduling and execution evolve through learning process that decisions are guided by heuristics that are formulated, updated, or dissipated by experiences.

Decision-making regarding scheduling and execution of activities involves not only short-term decisions as daily activity schedule, but also long-term as activity calendar on a monthly or yearly basis (Ben-Akiva and Bowman 1996, Ettema 1996, and Arentze and Timmermans 2000). Even though activities in the daily activity schedule seem to be resulted from short-term decisions specific for the day, those activities are considerably influenced by long-term decisions at household-level. Long-term decisions of a household such as work location or residence choice largely influences composition of activity calendar of a household. Besides, decisions on marital status or number of children affect the activities that a household takes part in. Reflecting the relationship between socio-demographic variables and lifestyle of a household, different households have different activity calendars. This long-term household activity calendar determines largest part of activity types that individuals take part in daily life. The activities of the household activity calendar are allocated to household members, resulting in activity program for every individual within the household. Then, individuals organize those activities in some specific sequencing and timing. The activity schedule is converted into activity patterns as activity scheduling occurs. Activity scheduling is a series of decision making process which consists of sub-decisions for each episode in the activity schedule. The sub-decisions for each episode are choice about destination, activity start times, duration, if travel is involved, transport mode, and travel party.

It should be noted, however, that individual's choice is rather limited because of various constraints. The ALBATROSS identified six types of constraints, extended from those of Hägerstrand's (Arentze and Timmermans, 2000). Constraints are incorporated in the model in such way that it uses constraints to see what the possibilities, in other words, what feasible decisions are, for each time a decision is made. Primarily, this concerns spatial-temporal constraints: earliest possible start time and latest possible end time of an activity reflecting institutional context and what locations are within reach and what transport mode is available and could be used. This corresponds to what geographers have called space-time prisms described in the previous chapter.

Lastly, it is postulated that decision-making process relevant for activity scheduling and rescheduling is based on learning mechanism. In other words, individuals apply particular heuristics for certain situation to solve the problem, activity scheduling in this context. This approach is different from econometric models in which optimal solution is pursued by exhaustively examining all possible options. Rather, choice might be sub-optimal in heuristic approach as the individual has not tried every conceivable options resulted from adoption of certain rules that is situation-specific. Experiences make heuristics to be newly formulated, be adapted in response to changes, be reinforced by repetitive behaviors with positive results, or be dissipated by negative experiences. Initially, individuals may try every possible activity patterns, given that they have very little knowledge about the environment. Whereas positive experience reinforce the choice of the activity pattern in the future leading to habitual behavior, negative experience attenuate the use of the pattern, eventually resulting in disappearance of it. As their experience repeats, individuals gradually have preferences for certain activity patterns. Consequently, individuals develop scripts, defined by Arentze and Timmermans (2000) as ready-made heuristics under specific conditions. During the execution of the activity pattern, individuals often encounter unforeseen events that adaption needs to be made. It should be noted that under this changed environment, individuals do not consider all options, and rather they use more generic rules that have proven to be adequate and satisfactory in the past. Based on the understanding of concepts underlying the Albatross model, what is following is description of operational system of the conceptual model.

3.2.2 The ALBATROSS System

The conceptual framework is operationalized as ALBATROSS system. The system is a rule-based model where a schedule results from applying choice heuristics that individuals develop. Like the other computational models describing scheduling, Albatross assumes a sequential decision making process that sub-decisions are made in a specific sequence. Choices of activity type, destination, time of day, travel party, and transport model are explicitly predicted by the model.

The description of the model system is based on the current version of ALBATROSS, Version 4, also used in this research. Albatross-4 is estimated on the national travel survey dataset (MON). The travel survey substitutes activity diary data because transformation of the survey into activity diary data format was possible. As the data collected for the entire of the Netherlands involving 45,000 person-days, Albatross-4 is fully operational on a national scale.

Albatross is rather a broader model system in its operation, beyond the fundamental module that generates schedules. Before starting a prediction run for scheduling, Albatross first generates a synthetic population for which schedules are to be predicted. The next step involves the main part of the system to predict the activity schedules of the individuals in each household. During the prediction run, three system components interact; they are the scheduling engine which controls scheduling process, the decision unit from which the choice outcomes are delivered, and the inference system that derives required information about conditions using analytical rule-based models. All relevant input and output of the three system cores are stored at database of the system. Regarding a post-processing step of the results obtained from prediction runs, Albatross has additional agents that further process predicted activity schedules in several ways to derive useful information.

As micro-simulation models involve construction of a data set representing the characteristics of the individuals of a household as the first step, a population needs to be synthesized in Albatross prior to predict the schedules of the individuals. Like many other micro-simulation models, a method to population synthesis in Albatross is based on the conventional approach originally developed by Beckman, *et al.* (1996), Iterative Proportional Fitting (IPF). The ultimate use of IPF is to estimate a multi-way table. A representative sample of the population and demographic census data are necessary to construct the multi-way table. Margins of the table are given by demographic data and initial proportions in cells of the table are derived from the sample data. It is a process of finding cell proportions as the initial sample data is gradually adjusted through repeated calculations to fit the given margins while correlation structure in the multi-way table is maintained. As a preprocessing step to apply IPF, demographic data is transformed into household data using relation matrices because Albatross needs household level data as well and demographic census data is not available at household level (Arentze, et al., 2008). The methodology of population synthesis will be further discussed as a separate chapter in the following.

Central to the model system is the scheduling engine, which simulates the scheduling of activities of individuals across the day. Intuitively, scheduling and execution phase entails rescheduling behavior so as to improve the preliminary schedule or adjust the revised schedule due to constraints and unexpected events. Rescheduling operator, however, has not been implemented in the current operational model, and thus, only preliminary schedule which is the first outcome of scheduling process is predicted in Albatross. Individuals solve scheduling problems following some specified sequence as the scheme of Figure 3.1 shows. Generated

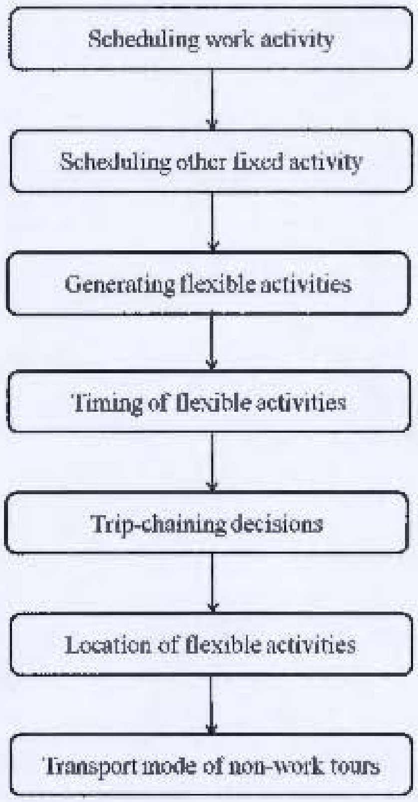


Figure 3.1 The scheduling process of Albatross

activity schedules, resulting from the sequential decision making process, describe multiple facets of activity and travel; for a given day which activities are conducted, when it starts, for how long, where, and, if traveling is involved, the travel mode and chaining of trips. Within the predefined sequence of decision dimensions, Albatross uses a priority-based scheduling process. For example, mandatory activities are scheduled prior to discretionary activities. Moreover, location choice of flexible activities precedes those of mode choices.

The distinctive feature of Albatross is its constraints-based search to solve the scheduling problem. In other words, various constraints are satisfied for every decision making step through the entire scheduling process. Each time a decision is made it uses constraints to see what feasible decisions are. Albatross delineates the choice set or choice range from all scheduling constraints given by available data. Primarily, this concerns spatial-temporal constraints: earliest possible start time and latest possible end time of an activity and what locations are within reach and what transport mode is available and could be used.

The fundamental level of modeling is a particular household and a particular day. But this doesn't imply that Albatross models household-level decision making. Rather, household interactions are incorporated in scheduling of individuals of a household in several ways. Car allocation problem between household members is taken into account in mode choice of individual members where there is only one car and more than one driving license. Although scheduling of each household member is implemented simultaneously, the current schedule of other members has a direct impact on decision making of another person, alternating choice options. Lastly, the presence of children, if any, is taken into account as a condition for scheduling decisions.

The Decision Unit contains choice options and decision rules. Although the choice alternatives are predefined by the external data, the structure of decision rules are empirically derived from activity-travel diary and the induction is executed by learning mechanism. A decision tree (DT) formalism is adopted to represent choice heuristics. DT consists of a list of condition variables and action variables. Condition variables relate to attributes of household, physical environment, or transportation system. Action variables refer to the available choice alternatives. The primary reason why the system takes DT representation rather than unstructured rules lies in its property that exhaustiveness, exclusiveness, and consistency among choice options are guaranteed. As a method of learning mechanism, CHAID-related algorithm is used in the system. Inductive learning algorithms from AI can also be used, but CHAID-based algorithm is more powerful for modeling probabilistic rules (i.e., rules that predict choice probabilities rather than a deterministic choice). The possibilities of the two DT induction methods are explored and this is described in concrete in the book ALBATROSS version 1 (Arentze and Timmermans, 2000). The process model uses a total of 26 decision trees to derive decisions in the scheduling process. Complementary to the decision unit is the inference system that consists of analytical and rule-based models for deriving needed information about constraints from available data about the study area and the household considered. These models represent basic knowledge about scheduling constraints by calculating temporal constraints or defining dynamic location choice sets, for example.

The database is where a set of data files and generated activity schedules in run-time are stored. The input data concerns information of study area; household attribute, land-use pattern such as opening hours and possible location for specific activities, and transportation system. What is important in transportation system is a set of travel time matrices under free-flow, shortest path as well as real-traffic conditions. On the other hand, output data includes schedule information; observed schedule and predicted schedule.

The scenario agent allows users to define various scenarios to simulate the effects of policy measures or population composition change. The scenario builder provides the interface to manipulate system parameters such as composition of the sample fraction of population or change attribute settings such as transportation environment and land-use. Simulation of traffic flows on the network and adjusting travel time data dependent on network capacity may improve accuracy of prediction. Nevertheless in the present model Albatross-4, traffic propagation does not take place and thus does not include feedback of travel-time realizations to the scheduler. No iterations occur in a prediction run.

Having the predicted schedule for each person of the synthetic population, the set of schedules can be analyzed in several ways by reporter agent of Albatross. It supports the display of basic properties of complete activity patterns, tours, trips, together with derived statistics. In particular, trip matrices possibly disaggregated on some third dimension such as mode can be generated. Relevant to assessment of environmental and transportation policies, the system calculates as set of indicators. Moreover, similarity between observed and generated activity schedules is measured by goodness-of-fit analysis. To account for interdependencies between decision dimensions as well as sequential order of elements, multidimensional sequential alignment method (MDSAM) (Joe, *et al.*, 1997), is used. In addition, impact assessment can be done by comparison of activity pattern under base-line and scenario condition. Rather than the observed set, the activity pattern under zero-change condition is used because a bias is unavoidable, thus the found differences cannot be attributed to the scenario conditions alone. Albatross does not run multiple times to test the statistical significance of the found differences, but one time. In other words, the output is randomly divided in a number of parts and then the statistics are calculated. This is similar to n times run on $1/n$ of population. Given the predicted schedules under both conditions, they are compared in terms of the frequency tables, indicators, and trip matrices generated between the null scenario and the scenario condition.

3.3 OmniTRANS and MaDAM

In this research, OmniTRANS software package is used for modeling dynamic traffic assignment. Particularly, MaDAM, which is a traffic simulation heart of OmniTRANS, is run as the subsequent step after prediction of travel demand by Albatross. This section touches on a brief description of the software package and conceptual details of the MaDAM without providing mathematical description of the model.

3.3.1 OmniTRANS

OmniTRANS is a transport planning software package. It is a comprehensive tool that not only supports addressing of various transport modeling problems but also satisfies different users such as planners, consultants, and researchers with different purposes and needs. The application of OmniTRANS is extensive as it allows a user to manage data and projects with user-friendly system interface, manipulate modeling process in a user-defined way, apply diverse renowned transport modeling techniques to problems, and analyze the model outcomes in several ways.

OmniTRANS provides a rich interface for managing and manipulating the system. Project and data management and network editing are available through interfaces. The data and models relevant for analysis are managed as a project. The project templates allow different users to work in collaboration by establishing uniform standards and styles. And potentially large and complex datasets are easy to control in OmniTRANS. Transport modeling data generally relates to multiple dimensions such as transport mode, time, and etc. This multiplicity makes data very complex and difficult to manage. In Omnitrans all data and process results are constructed and stored in an origin-destination 'Matrix Hypercube' in terms of six dimensions: purpose, mode, time period, user, result, and iteration (PMTURI). Although conventional origin-destination matrices typically are two-dimensional, Omnitrans provides more expanded concept of trip matrices into PMTURI structure. While input data can be structured as PMT(U)-combination, output skims can be stored in PMTURI-combination where result and iteration are added. Therefore, a trip matrix is always referenced by means of a PMTU-combination. For flexibility, users are allowed to define the structure of the dimensions as one intends. Thus, efficient and consistent storage and management of data are possible.

What enables unlimited potential for modeling in OmniTRANS is its scripting language which is used for editing and organizing modeling jobs and managing the execution of the jobs. OmniTRANS job language (OJL) is based on the Ruby object oriented programming. The Ruby is attractive in that it is open source language and no significant programming knowledge or skills are required to use it. OJL adopts the entire native Ruby language. But, Omnitrans-specific classes are added to Ruby standard classes to deal with transport modeling problems. Users can interact with the system by explicitly defining job scripts so that the task specified by the users can take place. Therefore, OJL is powerful in developing innovative transport models of any scope and style.

A number of aspects of travel can be modeled by the help of modeling techniques available in OmniTRANS. Basically, it models both travel demand and traffic assignment, yet it is addressed in many ways. It is a multi-modal and multi-temporal system. On the one hand, travel behavior can be analyzed on an aggregated or disaggregated level. For example, traffic can be described for multi-users classes. On the other hand, modeling can be done in a static or dynamic way in terms of time aspects.

Additionally, OmniTRANS provides useful tools to make use of the modeling outcomes. The system facilitates the graphic representation, the visualization, the comparison of the results, and the report generation, and thus, the results can be presented in a form that is easily legible and understandable so that the users can draw any meaningful implication from it. Regarding the graphic representation and visualization, modeling data and results can be represented in a graphic format such as chart and graph. Furthermore, animated display in relation to time periods is possible by synchronizing animated displays for multiple time slices. And for the visual representation, users can manipulate labels and annotations. Given the charts or graphs, scenarios and dimensions can be compared within OmniTRANS. Finally, the data and the results together with the graphic representations can be presented in nicely formatted reports.

3.3.2 MaDAM

Although MaDAM (Macroscopic Dynamic Traffic Assignment Model) itself is not a dynamic assignment model, it is run in the DTA framework as a small part of the DTA system, specified for dynamic network loading. Within the DTA framework, the MaDAM, the macroscopic traffic propagation model, simulates route flows onto the given network. As the name of the model suggests, it is the macroscopic model taking principles from fluid mechanics. Attractiveness of the macroscopic simulation model is that it is faster and deterministic, and easier to use than microscopic simulation models or analytical macroscopic models. Microscopic simulation models are slower to configure the network and are not applicable to large networks due to the fact that everything is analyzed on the level of individual vehicles. On the other hand, this macroscopic simulation model does not put too much emphasis on theoretical and mathematical structure of problem as analytical macroscopic models do, but this doesn't mean that the simulation models have little mathematics (De Romph, 1994). The following text explains how MaDAM works within the DTA framework and distinctive features of the model in detail. The explanations refer to several research papers published by Omnitrans International, the Omnitrans software package manual, and the website (http://www.omnitrans-international.com).

Here, MaDAM will be described in terms of the DTA framework because the ultimate use of the model is to solve the traffic assignment problem. As explained in the previous chapter, dynamic assignment consists of two sub problems; route choice and dynamic network loading. Likewise, route choice and MaDAM are combined for DTA in Omnitrans (see Figure 3.2). A set of routes is necessary to simulate traffic. This implies that routes must be prepared based on the O-D trip pairs. This route generation is implemented as an intrinsic part of DTA, and thus, route choice is modeled as a pre-processing step in the software to running of MaDAM for dynamic flow propagation. Many methods are available for route generation (Raadsen, *et al.*, 2009b), but the MaDAM uses turn-fractions based method, as follows. The process of route generation is to estimate the division of flow over the network. This concerns the calculation of splitting rates for turning movements at exit links of nodes. Each exit links has turning proportions which is defined as the portion of traffic volume entering a node and leaving the node through one of the

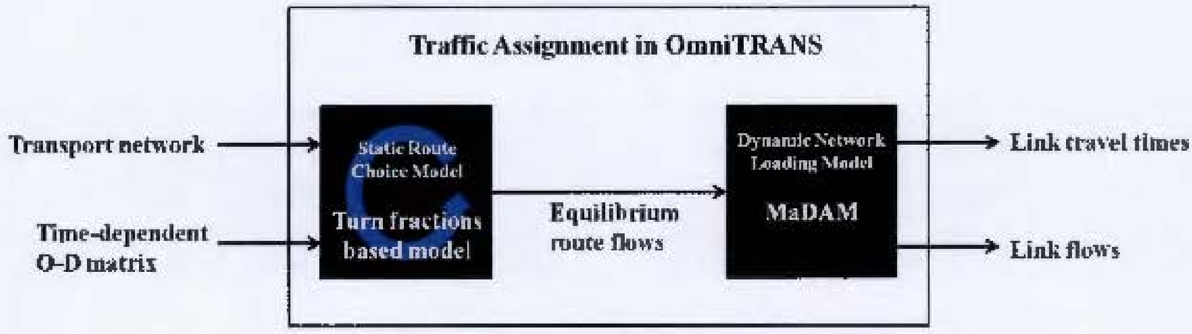


Figure 3.2 Traffic assignment process in Omnitrans

exit links. Turn-fractions method is computationally friendly in that routes are not traced during DNL and splitting rates are defined at the node level per exit links, not per route. Splitting rates need to be determined on beforehand and static assignment is used to define the initially turning fractions in Omnitrans. Static route choice model is implemented in an iterative framework where static equilibrium split proportions are converged over iterations. Therefore, a set of turn fractions are static and have no time element. However, true dynamics can be captured only when the route choice varies over time in relations to traffic conditions and this is implemented in StreamLine framework which is the latest DTA system in Omnitrans. To make time-dependent route choice possible, a technique of layering is used. Vehicles departing at the beginning of a simulation period are only handled using the initial turn fractions resulted from static assignment. Then, newly departed vehicles traffic is propagated using additional layer of fractions that is based on the updated route choice behavior. That is to say, the flows departing at different moments in time employs different set of turn proportions.

Having defined the turning proportions at each exit link, the next step is to load the trips departing from each origin onto the network at the start of simulation period. And this dynamic network loading is implemented by MaDAM. The MaDAM propagation model is a second order model originally based on Papageorgiou's METANET which is developed from the so called LWR model through a first order cell transmission model (Daganzo 1994). Cell based models are widely used as macroscopic DTA model and such models divide links into equal length segments or cells in order to propagate traffic. Each cell holds information on its relevant variables such as density and speed. Vehicles move from one cell to another. Similarly in MaDAM, Omnitrans automatically divides links into smaller segments to capture more dynamics and more accuracy of prediction. Through dynamic network loading, the trips are propagated along each link through predetermined routes and when they reach nodes, they are split between the candidate exit-links using turn proportions. By the end of the modeling period, all trips have reached their destinations. It is noted, however, that at any given point in the network, we do not know where the trips come from or where they are going to because trips are moving through time and are aggregated over all O-D pairs. As a consequence of dynamic network loading, a series of loads are obtained. Those indicators of traffic loads include density, flow, or speed of each link. Link loads are computed for smaller segments into which the system automatically has been divided into. And for each segment and each time step, the traffic situation is computed conditioning upon the amount of traffic on the road and situation upstream or downstream. In the system, computation of link loads can be done for every second. However, storing the results on this level of resolution will not be practical. For this reason, aggregation to some extent is needed so that the results are manageable or usable for drawing some implications from them. By default, the system stores the values for 15 minute periods, but time step size can be controlled by users by setting.

Lastly, additional but optional features of MaDAM are junction modeling, cordon assignment, and animated representation of output. Junction is defined as any node that has three or more connected links. Inclusion of delays caused by the presence of junctions can improve the assignment results significantly, especially in urban area since delays are an important factor affecting travel costs. The module is designed to determine the turning delays for motorized traffic. It calculates the average delay per vehicle for each turning movement on the basis of the junction layout and turning flows. As a link is divided into segments for computation of link loads, a link at junctions are segmented into approach lanes for junction modeling. It also adopts fluid dynamics principles but specified for traffic behaviors at junctions. For each approach lane, delays are calculated and then converted into speed. Another feature involves cordon assignment. By defining cordon, traffic propagation is not always simulated for the entire network, but can be implemented for the area of interest. Cordon matrix is not used, but it is manipulated by cordon selection tool of Omnitrans. Furthermore, the results can be visualized in animated style as a video display. The video shows changes of network condition during the simulation period. Variations of traffic flow over time or other attributes such as speed and density are represented.

3.4 CONCLUSION

This chapter described the two models relevant to the research: Albatross and MaDAM. Whereas Albatross predicts travel demand by generating the activity schedules of individuals, MaDAM simulates the predicted travel demand onto the transportation network by propagating the route flows. Linkage of the two models would result in an alternative for the the four-step modeling process which still is used predominantly in practice. Albatross' system heart is the scheduling engine which predicts the activity schedules. Since it is a micro-simulation model so that simulation of agents is necessary, the population synthesizer estimates individual-household attributes using IPF method as a preceding step. Having generated a (synthetic) population, the scheduler runs the prediction of the activity schedules of every individual of households. The scheduling process follows a predefined order of decision-making steps relevant to activity participation: activity type, time of day, trip chain, activity location, and transport mode, etc. The scheduling process is additionally supported by the decision unit and the inference system, making Albatross a rule-based system. The decision unit contains the tree-like structured decision rules and choice alternatives of which the individuals make use when decision-making. Albatross determines available choice options or ranges taking into consideration the various constraints and this is facilitated by the inference system consisting of analytical models. The resulting activity schedules can be further processed to generate a report with summary statistics, system performance measures, etc. In particular, what is the most relevant to the research, trip matrices with a third dimension such as time of day or mode can be derived from the activity schedules which matches the trip tables that are input to traffic assignment implemented by MaDAM in Omnitrans. MaDAM itself is a dynamic network loading model that merely simulates the traffic streams through specific routes in the network. This indicates that routes for propagating the traffic need to be generated and this is done in a step preceding the execution of MaDAM in Omnitrans. The route choice is modeled to derive turn-fractions. Turn proportions at every exit link are defined by running a static assignment that is iterated until equilibrium is reached. Thus, it results in static equilibrium route sets and, which, in turn, are used as input to MaDAM. Dynamic network loading outputs traffic flow attributes such as density or speed. The outcomes are calculated for small segments of a link and stored for every couple of minutes to meet the higher degree of accuracy. In addition to this table representation, the results can be visualized as animated video.

4 DELINEATION OF STUDY AREA

4.1 INTRODUCTION

This chapter covers the first research issue mentioned in the introduction part of this report. The study area is the area within which transport flows are of interest. In existing transport models, the boundary of the study area is defined by what is called as external cordon or simply the cordon line. The area within the cordon is subject to explicit modeling and analysis. On the contrary, outside the study area boundary will not be considered as relevant for the problem at hand, describing traffic networks is limited only within the study area. In general, it is assumed that there is no change in travel demand interacts with outside the study area when scenarios are considered. Albatross is a national model which simulates activity-travel patterns of the entire Dutch population. The spatial scope of Albatross and MaDAM matters to link the two models. While Albatross so far has only been used for whole of the Netherlands, dynamic traffic simulation is computationally not feasible for a nationwide area. That is to say, small size of network is typically used for traffic simulation, implying that a smaller local study area needs to be selected to make the two models compatible. Therefore, this requires adaptation of the current version of Albatross on a national level to handle travel demands on a local level.

In turn, delineating a certain part of the entire region as a study area always accompanies problems of external trips, which have at least one trip end outside the boundary of the study area. Even though external trips influence the traffic patterns within the cordon, importance of them are often overlooked. They are usually assumed to be constant. That is, only internal trips respond to changes under scenario situation, while external trips don't. It becomes problematic when scenarios are considered for prediction of travel patterns of study area. External trips indeed, however, are influenced by changes in environment. The assumption is unrealistic in the sense that traffic movements are dynamic in nature and needs to be relaxed to achieve accuracy of prediction. Since the primary purpose of this research is to capture the very aspect, thus, changes of external travel patterns must be taken into account in its prediction. To handle this, a new algorithm is developed and embedded in Albatross application. This chapter is structured as follows. First, the local study area used in this project is presented. Second, the chapter explains definition of traffic zones and trip categories relevant to this problem. Third, an existing approach to deal with external trips, specifically the method used by Omnitrans, is introduced. Lastly, the chapter explains the new algorithm in detail and discusses the performance of the algorithm.

4.2 THE STUDY AREA

Due to the computational feasibility of running of the linked model system, a smaller local study area needs to be selected. The study area is most of Noord-Holland, a province situated at the North Sea in the northwest part of the Netherlands as depicted in Figure 4.1. The major cities and towns of the province are Amsterdam, Haarlem, Hilversum, Den Helder, Alkmaar, Zandaam, and Hoorn. The island of Texel is also part of the province. Amsterdam, which is the financial and cultural capital of the Netherlands, is located in Noord-Holland. Many large Dutch institutions have their headquarters there. Moreover, the city is one of the famous sightseeing places as it draws 4.2 million tourists every year. As the region is characterized as a city of finance and tourism, the Netherland's main airport, Amsterdam Airport Schiphol is located southwest of Amsterdam. The study area consists of 436 postcode areas. There are 25 cordoned postcode areas. These cordoned areas, called as edge zones, are created to handle external trip flows.

The size and geographical characteristics of a region influences trip patterns of the region and adjacent area as well. If the region is geographically important with key land-use functions such as central business district, more trips are attracted to the area. In turn, this implies that external trips account for considerable proportion of total travel of the study area. Likewise, Amsterdam attracts a large amount of traffic from other regions. In particular, the impact of external trips to total travel of the northern part of the Netherlands is expected to be large enough to affect traffic patterns of the study area as the city is located onto southeast of the province where the boundary of the study area lies. Therefore, external trips must be dealt with some degree of discretion in this research.

4.3 INTERNAL AND EXTERNAL TRIPS

In order to predict travel patterns of the study area, network links continuously extending towards outside a study area are cut off by the cordon line. Encircling a geographical region with a cordon line, two types of traffic zones can be created: external and internal zones. External zones exist outside the external cordon. On the other hand, inside the external cordon are termed internal zones. Based on the two types of traffic zones, three types of travel patterns apply. According to where trip ends lie in terms of the boundary of study area, travel pattern can be divided into three types (Ghareib, 1996; Martin and McGuckin, 1998; Anderson, 1999; Ortúzar and Willumsen, 2001). Trips that both begin and end inside the area bounded by the cordon are called internal-internal (II) trips. External trips have at least one end outside the cordon line. This trip category can be further classified into internal-external (IE) (or, external-internal) and external-external (EE) trips. When one trip end is outside the study area and another is within the cordon, it can be either external-internal or internal-external trip. The only difference between the two is direction of the movement. Whereas external-internal trips are inbound trips, internal-external trips are outbound trips in relation with the study area. When both origins and destinations are outside the cordon, they are termed through trips or external-external trips. Traffic zones and trip categories are depicted in Figure 4.2.

In existing transport models, external trips are usually modeled at external stations. Intersections where network links meet the cordon are defined via external stations which effectively serve as doorways to trips, into, out of, and through the study area. Circumscribed study area with discontinuing transportation network results in a finite set of origin and

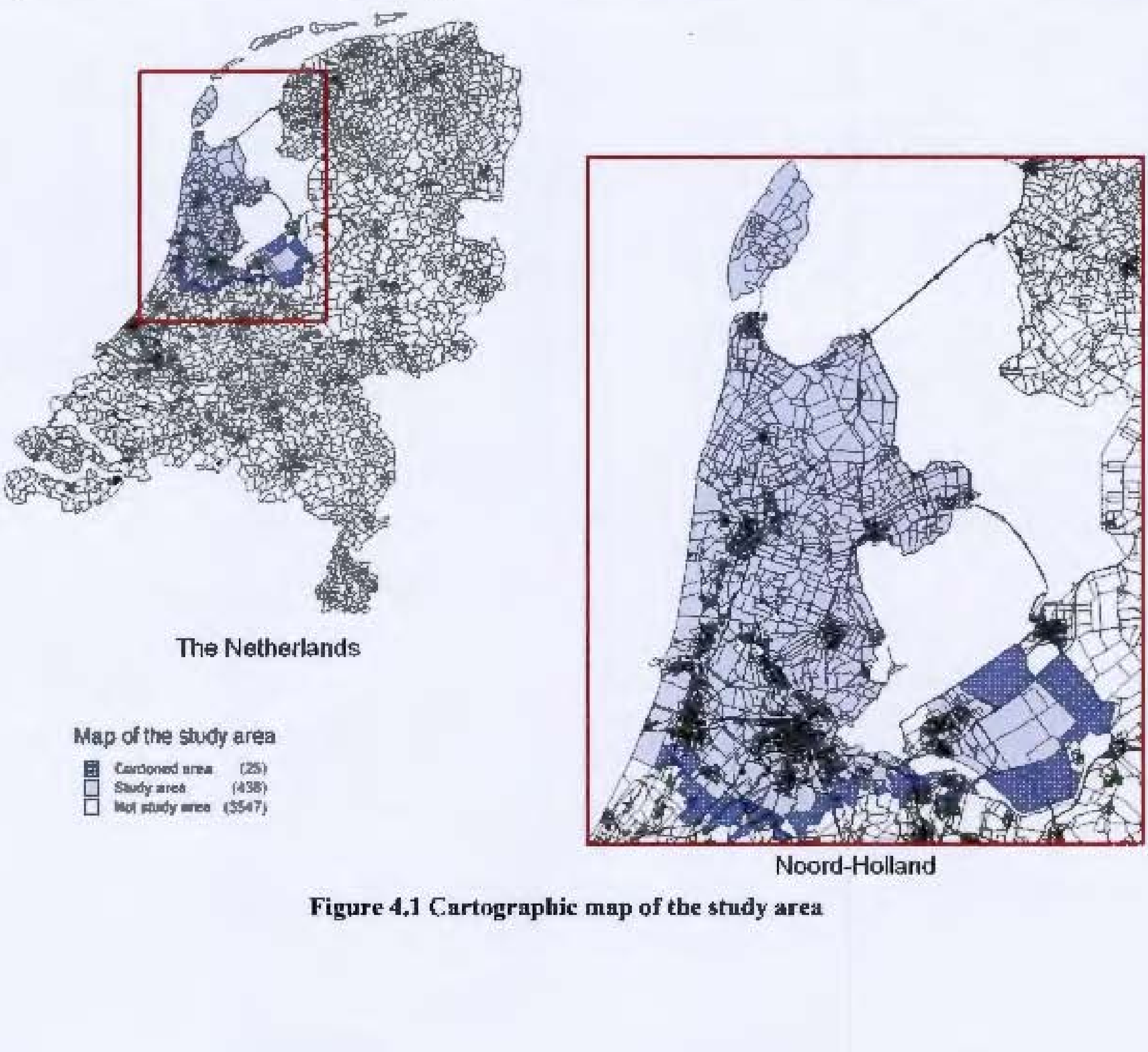


Figure 4.1 Cartographic map of the study area

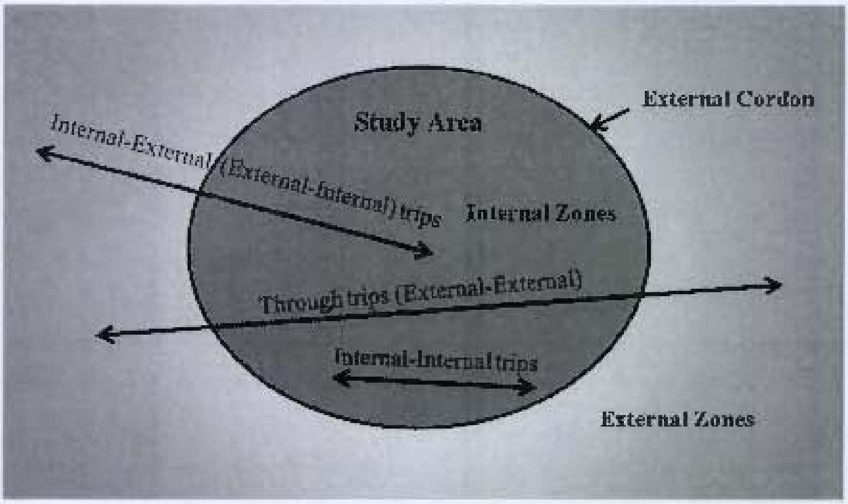


Figure 4.2 Traffic zones and trip categories

destination pairs in trip matrices. Nodes that contain the external stations are appended to study area O-D trip table as origins and destinations of external trips. All trip categories introduced above can be illustrated in a matrix format, resulting in a complete study area trip table (see Figure 4.3). In addition to internal origin and destination zones which account for internal trips, nodes composed of external stations are appended to a study area O-D trip table. The matrix can be divided into four parts that explains each trip category. Blue and red colored part accounts for internal trips and external trips respectively.

4.4 EXISTING APPROACH TO DEAL WITH EXTERNAL TRIPS

Because of the small proportion of external travel relative to total travel, the effort on measuring and modeling external trips has been less intensive than for internal trips (Martin and McGuckin, 1998). In general, the models representing external travel patterns are separate from and less complex than those that describes interactions of trips within the study area, treating external trips as completely independent from internal trips.

The approach that Omnitrans uses to deal with external trips is explained here. For ease of naming of the approach, it is called the cordon line method throughout the text. The reason behind the development of the cordon line method is twofold: i) to analyze the traffic pattern of a particular region of interest and ii) to achieve computational feasibility to run dynamic traffic assignment. In a pre-processing step, Omnitrans typically predicts the travel pattern of the Netherlands applying a national model of the conventional four step type. Since traffic simulation for a large transport network as a whole country has a problem of computation burden, static traffic assignment techniques are used to preliminarily predict traffic flows for entire the

$$\text{Study area O-D matrix} = \left[\begin{array}{cccc|cc} II & II & II & II & IE & IE \\ II & II & II & II & IE & IE \\ II & II & II & II & IE & IE \\ II & II & II & II & IE & IE \\ \hline EI & EI & EI & EI & EE & EE \\ EI & EI & EI & EI & EE & EE \end{array}\right]$$

Figure 4.3 Matrix representation of trip categories

Netherlands. However, as discussed in the literature review of traffic assignment, dynamic traffic assignment has been developed and become extensively used in analysis to capture the dynamic property of traffic. Nevertheless, static traffic assignment techniques still are widely used in transportation research. In order to run dynamic traffic assignment as well as model traffic behavior of a certain region, smaller size of study area is required and it concerns delineation of study area. Here, the cordon line method is used. Study area is circumscribed by the cordon line and external stations where external trips come in and go out are determined at nodes where transport network and the cordon intersect. As explained above, nodes at intersections are appended to study area O-D trip table as part of origins and destinations (see, Figure 4.4). It is necessary to have the size of external trips to complete the study area O-D matrix. The external trip volumes are obtained from the result of static traffic assignment of the national model of Omnitrans. It is assumed that they are fixed as constant in further analysis. The completed study area O-D trip table is used for applying dynamic traffic assignment techniques or scenario analysis of study area.

Although the cordon line method enables transport analysis only for a specific area of interest and application of dynamic traffic assignment techniques, the method is limited in the sense that the results of further analysis such as scenario analysis or dynamic traffic assignment is not truly dynamic. The assumption of static external trips is critical limitation of the cordon line method. First, the assumption of constancy of travel patterns at external zones is unrealistic because external trips are dynamic in reality, instantly changing in response to changed environment and with respect to time. The assumption becomes problematic particularly when scenarios are considered. No matter what scenario is defined for study area, the result is that only internal trip patterns are responsive to changes while external trip patterns are not. Let us consider an example scenario of population growth at outside study area. If population growth rate outside study area is dramatically increasing, this attracts many people of the study area due to job opportunities, and thus, the size of external flows may increase. And even more, this may have an impact to internal trip pattern. However, the existing method does not reflect change of external travel pattern. Second, nevertheless the percentage of external travel may be small as internal trips account for a sizable part of trip patterns of a study area, decisions regarding improvements to facilities that carry high percentages of external trips must be made with some degree of confidence. For example, in case of a big shopping mall with an easy access by motor way is constructed outside adjacent to the study area boundary, there must be huge demand for

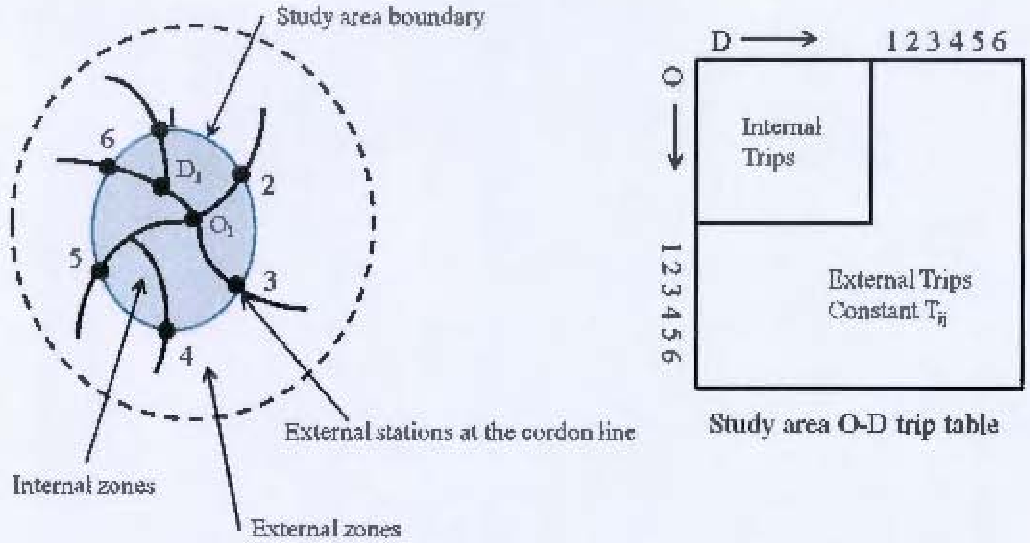


Figure 4.4 The cordon line method

the facility by residents living in the study area due to good accessibility and diversity of goods. Subsequently, shopping trip patterns would change as to this, leading to decreased shopping travel within the study area and increased outbound trips. The examples demonstrate that modeling external trips is important as much as that of internal trips. The following section presents a newly developed algorithm dealing with prediction of external trips based on Albatross because travel demand is subject to prediction by Albatross in the project.

4.5 NEW APPROACH WITH PATH FINDING ALGORITHM

The previous section introduced the cordon line method and raised the need for a new approach to deal with external trips. Not only dynamics of internal trips but also that of external trips must be incorporated in traffic prediction of study area. Although external trips occupy a relatively small proportion of total flows of a studied area, they have a significant impact on internal trip patterns and use of facilities crossing the cordon line and passing entirely through the study area. Besides, as the crucial objective of this research is to capture dynamic aspects in transport modeling, external trips needs to be modeled as they are responsive to environmental and temporal changes.

The current version of Albatross is extended to handle aforementioned aspects and enable the link with MaDAM. Albatross' original study-area, all of the Netherlands, is not changed, meaning that the activities and trips from the entire national population are predicted. Travel demand on a national level is obtained, resulting in nationwide O-D trip table. Simply sorted out cells of study area origin and destination zones from the whole O-D trip table do not include external trip counts if the smaller study area is delineated. The edge cells of the study area trip table only contain the trips that depart at the origin and arrive at the destination. External trips have their initial origins or final destinations outside study area. Therefore, in order to complete the study area O-D trip table, external trip counts should be added to edge zones where the external trips pass. Whereas Omnitrans uses static traffic assignment to obtain external trip counts, Albatross uses dynamic flows and furthermore applies a different approach because prediction depends on zonal travel time data rather than road network data. What is developed is a new algorithm which identifies entry or exit points through which external trips pass. This method satisfies a considerable amount of precision for prediction of external trips because the outside area is predicted as well whereby the algorithm directly finds the best possible external stations that can be used for making external trips. Moreover, since Albatross predicts travel patterns for a whole day of 24 hours time period as well as the whole population, time varying external trip volumes receptive to environmental changes can be obtained. It allows prediction of changing external trip patterns even under scenario settings. The following text describes a new approach with the path finding algorithm.

Above all, it is necessary to define external stations that trips are, into, out of, and through study area. For the reason that travel time data of trips in Albatross is based on postcode area matrices instead of road network directly, they are defined at centroids of postcode areas, rather than nodes where the roads intersect with the external cordon as in Omnitrans. The postcode areas where the external stations are defined are called edge zones. External stations are alternatively named as entry or exit point according to orientation of external trips. In other words, if a trip gets into study area using an external station, the station is called entry point of the trip. In contrast, the station can be exit point in case of an outbound trip. Thus, whereas internal-external trips or external-internal trips have only one entry or one exit point, external-external trips have both entry and exit point as they pass through study area. Figure 4.5 illustrates the new approach.

Having defined entry/exit points of a study area, the next step involves application of the path finding algorithm to all possible combinations of origins and destinations so as to find the best feasible entry/exit point that can be used for travel between each O-D pair. The algorithm works with a postcode O-D matrix. The matrix is not a conventional O-D trip table. This matrix defines the entry/exit point where the trips come in and the entry/exit points where the trips go out, for each combination of origin and destination. In case a trip is an internal trip or the trip that has no relationship at all with study area, entry/exit point is not relevant and the case is coded as missing value (=0) for both entry and exit points in the matrix. Otherwise, entry/exit point is coded with postcode area that the route uses for making external trip. As a result of running the algorithm, the matrix is filled with a postcode of the entry/exit point for every O-D pair.

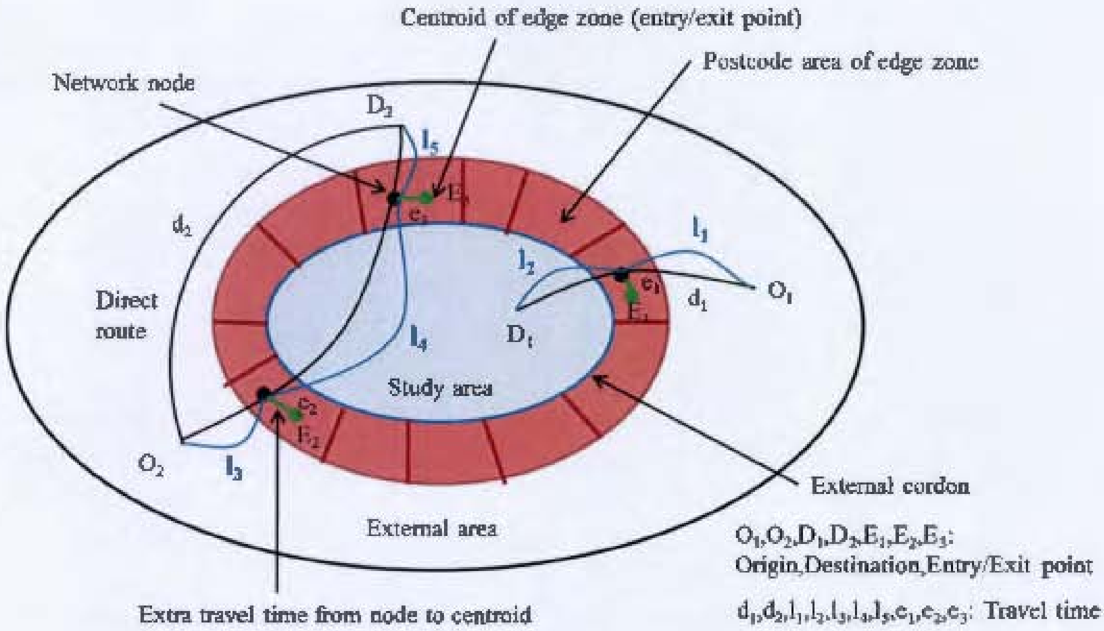


Figure 4.5 The new approach with the path finding algorithm

How the algorithm identifies entry/exit point between O-D pairs is discussed here (see, Figure 4.5). The algorithm finds an entry or exit point which minimizes total travel time if the route connecting the specific origin and destination is considered to be an external trip. It calculates travel length in total travel time of two types of paths between origin and destination. First, total travel time of the direct route between the origin and destination is obtained. The direct route travel time here is travel time of fastest route between postcode origin and destination. It is the value of a cell in the travel time O-D trip matrix which represents travel times between O-D pairs. Second, the same trip is presumably considered to be an external trip that must pass entry/exit point at edge zones. Let us call this external trip route as an artificial route. With a restriction that the path must pass the study area, the second total travel time, the artificial route travel time, is computed. For this case, the algorithm constructs the trip in such a way that it is divided into two or three legs by entry or exit point as a reference point of split. For EE trips, the artificial route consists of three trip legs: a trip from origin to entry point, from entry point to exit point, from exit point to final destination. For EI or IE trips, there are two trip legs: a trip is from origin to entry/exit point and another is from entry/exit point to destination. The artificial route travel time is the sum of travel time of every trip leg defined in the travel time O-D matrix. The entry and exit points are defined such that the sum of travel time across the legs is minimized. Constructed in that way, it should be noted that the total travel time of an external trip route is always longer than that of a direct route. This is because extra travel time between network node at edge zone and entry/exit point exists resulting from cutting a trip into pieces. Since origin and destination are defined at postcode centroids, a travel time in the travel time matrix include travel time from network node to centroid. As construction of artificial route concerns entry/exit point and it can be regarded as intermediate origin and destination within the whole trip, the artificial route travel time is aggregation of travel time of each trip leg in the matrix. This implies that there is always extra travel time from network node to entry/exit point at edge zones.

Using some tolerance of difference caused by the measurement procedure, a decision on whether the trip is considered to pass through the study area or not, is made during the process. It always holds that the artificial route travel time is longer than the direct route travel time. Here, the two travel times are compared. Within the level of tolerance, it is considered that the trip is made by through trip although the artificial route travel time is more than the direct route travel time. If the difference between the two travel times exceeds the level of tolerance, then the trip is assigned to use the route totally outside the study area which means no interaction with the study area. This procedure holds for EE-trips, i.e. trips that have both ends outside the study area. Unlike EE trips, no decision is involved for trips that have only one end in the study area because both artificial route and direct route anyhow enters or exits the study area.

In the following, it describes the algorithm in detail for two types of external trip: external-internal (internal-external) and external-external. Since the same principal of the algorithm applies to EI and IE trips, they will be explained together. And through trip case will be separately presented. Figure 4.5 illustrates the method. While internal-external trip originates from an internal zone and destined to an external zone passing through exit point, external-internal trip enters to the study area using entry point. An example is shown in Figure 4.5 of an external-internal trip. The algorithm calculates two travel times. The direct route travel time is d_1 (direct from O_1 to D_1). The artificial route travel time is $(l_1 + l_2 + 2e_1)$, a sum of travel time from O_1 to E_1 and from E_1 to D_1. $2e_1$ is the extra travel time, resulting from the construction of the artificial route. E_1 must satisfy the condition that total travel time between O_1 and D_1 is minimized. As previously explained, the travel time of the artificial route is longer than that of direct route. However, the difference between the two travel times is not much for this type of trip because the direct route any how passes through the edge zone. And what makes the difference in both travel times is the extra travel time of the artificial route. Therefore, in case one trip end is inside and another is outside the study area, comparison of both travel times is not much meaningful because the direct route is also an external trip.

It is more complicated for a trip departing from and arriving at external zones, both origin and destination is outside the study area, in other words. Similar to an EI trip described above, through trips are split into three trip legs: a trip from O_2 to E_2, from E_2 to E_3, and from E_3 to D_2. The direct route travel time is d_2. This is the fastest travel time between O_2 and D_2. The artificial route travel time is $(l_3 + l_4 + l_5 + 2e_2 + 2e_3)$. And E_2 and E_3 are the points that minimize the total travel time of the artificial route. Extra travel time is involved in every trip leg because entry/exit points do not lie on the network node. As depicted as green line in Figure 5.5, there always exists additional travel time from a node on the path to entry/exit point of the edge zone when the direct route is intentionally cut into pieces. Given the two travel times, the algorithm compares them. The tolerance level functions as a criterion of judgment which route to use between the direct route and the artificial route. The tolerance accounts for an allowable extra travel time, $2e_2 + 2e_3$, in the example. If $(2e_2 + 2e_3)$ exceeds the tolerance level, it is assumed that the trip follows the direct route, which is outside the study area. In other words, under the condition that 'tolerance level $\geq 2e_2 + 2e_3 \geq 0$' is satisfied, the trip is subject to take the through trip route.

4.6 PERFORMANCE TEST OF THE ALGORITHM

The path finding algorithm described above is tested with some predefined O-D cases to see the algorithm works properly or not. Performance of the algorithm needs to be tested in order to find suitable parameter values leading to realistic search of entry/exit point. The algorithm program is run under a set of different parameter settings. It contributes to determination of appropriate parameter values of the algorithm. It is evaluated by the extent of similarity between the route predicted by the algorithm and the route mostly suggested and used in reality. In this project, it is assumed that the directions recommended by Google Maps (http://maps.google.com) are the routes used in reality. Thus, the result of the algorithm is compared to that of Google Maps. Google Maps provides 'Get direction' feature that finds candidate routes when origin and destination are entered. It works with four digit postcode of the Netherlands as the algorithm. Comparing the algorithm result with Google result, the parameter values are determined such a way that the difference between the two is smallest.

This section first describes input and output of the algorithm run. Contents and structure of input and output are explained. Next, sample O-D cases and parameter settings to be tested are presented. Lastly, the results of algorithm run and Google Maps are discussed. The section concludes with determination of the parameter values.

4.6.1 Input and Output of the Algorithm Run

Three input files are necessary to execute the algorithm program. Each file defines fastest route car travel times of every O-D pairs (tijd-nl-car.bin), relevant spatial information (studyareaOmn-Final.dat), and sample O-D cases (testcases.txt). Fastest route travel time is identical to direct route travel time between origin and destination that is comparative travel time to that of artificial route. Spatial information concerns, for each postcode of Albatross, definitions of whether the postcode is study area or not, whether it is highway or local road entry/exit point, and corresponding centroid of Omnitrans. Prior to run the algorithm, a set of O-D pairs for testing of the algorithm needs to be defined. This is another input file. It is nothing else than a list of O-D combinations of postcodes in tab delimited text file format (testcases.txt). The first two lines of the input file account for the parameter values which relate to a preference of highway above local route for entry/exit point (P_h) and preference for external trip (i.e., tolerance) (P_t). The algorithm incorporates these parameters for identifying entry/exit point. The first parameter is a preference for entering and exiting the area at the level of a highway (rather than a local route) based on the fact that for each possible entry/exit point an assignment is made whether it is located at a highway or not. The parameter reflects the willingness to make a detour for using the highway. Especially for long distance trips, it is generally more favorable to use highway entries and exits. The parameter expresses this preference. The latter parameter is the size – in terms of a ratio – of the extra travel time that is caused by cutting a trip into three pieces. This defines a tolerance of difference in comparison between the direct travel time and the total travel time across the artificial legs. Both parameters are calibrated manually based on a sample of trip cases.

The performance of the algorithm may vary in terms of different parameter settings. Hence, the algorithm is evaluated by adjusting the parameter values. Given predefined O-D cases, the next step is to run the algorithm, executing a DOS program (rwdata.exe). The program reads each case, runs the algorithm for each case, and results are written to a predefined output file (testresults.txt). The output file is structured as follows (see Table 4.1). Similar to the input file, the parameter values are defined at the first two lines. The third line displays the output variables. 'Origin' and 'Destination' is postcode of origin and destination. 'Enter' and 'Exit' is postcode of entry point and exit point. It is coded as -1, if entry/exit point is not used for the case. Otherwise, the variable is coded as postcode. 'EnterHW' and 'ExitHW' mean whether entry/exit point is highway or local road. There are three classes for both variables: entry/exit point is highway (=1), entry/exit point is local road (=0), entry/exit point is not relevant (=-1). The last case is when no entry/exit point is used. 'Dist' is total travel time of artificial route. 'Dmin' is the fastest route travel time calculated directly from origin to destination. For example, the first case in Table 4.1 is a trip between postcode area '2421' and '2101'. It is external-internal trip that passes through the highway entry point located at postcode area '1435'. Whereas the direct route takes 28 minutes, the artificially constructed route takes 34 minutes. A value of 1.20 for P_h means that an extra 20% of travel time is still accepted for using a highway rather than a local road to enter or exit the study area. And a value of 1.25 for this parameter indicates that an extra 25% travel time is still accepted to go through the study area (Dist is 25% larger than Dmin). Note that there is no decision involved in the path of trips that have the O or the D inside the study area. This is because the direct route is anyhow E-I or I-E trip. Thus, direct route and artificial route have the same route but the only difference is total travel time. Only for trips that have both ends outside the study area, however, there is a decision to be made whether the trip does or does not pass through the area. For instance the trip between '3812' and '9725', the algorithm considers the through trip route entering '1261' and exiting '8226'. Tolerance of extra travel time for this case is 23 minutes, calculated by multiplying 25% to the direct route travel time, 92 minutes. Since the artificial route travel time is less than 115 minutes, the algorithm results in the through trip route. Especially for those cases the parameters need to be adjusted and performance evaluated.

As it is also explained in the previous section, the external trip always takes longer travel time than the direct route.

Table 4.1 The structure of the output file

1.2	preference highway						
1.25	preference through trip						
Origin	Destin	Enter	Exit	EnterHW	ExitHW	Dist	Dmin
2421	2101	1435	-1	1	-1	34	28
1141	2614	-1	1435	-1	1	62	53
3812	9725	1261	8226	1	1	105	92

4.6.2 Sample Test Case Selection and Parameter Settings

Origins and destinations to be tested are chosen from visual inspection of cartographic map of the Netherlands. Relying on eye measurement, trips that are likely to have several possible routes with different patterns are selected as test cases: i) whether the route pass through the study area or not ii) whether entry/exit point on the route is highway or local road. 15 test cases are defined in the Table 4.2. Origins and destinations are represented in two formats: postcode and name of the place. The first five test cases departing from Amsterdam and Amstelveen concern E-I trips. These five cases are just to see whether the algorithm works or not. The following six cases from case 6 to 11 are possibly can be made by through trip or direct fastest route that has nothing to do with the study area. Regarding these cases, it is expected that the parameter settings of P_t largely influence the results of the algorithm because through trip route is considered within the tolerance level defined, even though there is direct route which is faster. The trips from Alkmaar, Noord-Holland to Heerenveen, Friesland and Dronten, Flevoland are chosen to see which way (direction) the trips exit the study area. There are two possible highway exit points and one local road exit points: northern 'A7', 'A6' passing Almere, and Enkhuizen-Lelystad dike. Lastly, trips between Abcoude – Hilversum and Amsterdam Zuidoost – Wijdemeren are rather shorter distance trips. That is, those trips are likely to take a local route rather than highway detour. All of them are tested under different parameter settings. Not all but remarkable cases will be explained in detail.

Four sets of parameter values are defined in Table 4.3. The first set of parameter values does not allow extra travel times that can be caused by highway and through trip route detour, implying that the resulting route is always the fastest one. Even though parameter sets are labeled in terms of smallest to biggest in the table, the set 3 is defined prior to set 2 and 3. Parameter values of the set 3 mean that 20% of extra travel time (detour) for use of highway and 25% of extra travel time to pass though the study area are acceptable. Based on this parameter setting, the parameter values of set 2 and 4 are determined. The set 2 and 4 are assumed to permit respectively half and twice extra travel time percentage of the set 3.

Table 4.2 Postcode and place name of test O-D cases

	Origin		Destination	
	Postcode	Plaats	Postcode	Plaats
1	1011	Amsterdam	6546	Nijmegen
2	1011	Amsterdam	8448	Heerenveen
3	1011	Amsterdam	1223	Hilversum
4	1011	Amsterdam	8011	Zwolle
5	1181	Amstelveen	2312	Leiden
6	3563	Utrecht	3812	Amersfoort
7	3329	Dordrecht	9725	Groningen
8	2628	Delft	7545	Enschede
9	2517	Den Haag	8912	Leeuwarden
10	4837	Breda	9411	Beilen
11	5236	Den Bosch	9725	Groningen
12	1822	Alkmaar	8448	Heerenveen
13	1822	Alkmaar	8252	Dronten
14	1391	Abcoude	1218	Hilversum
15	1101	Amsterdam Zuidoost	1243	Wijdemeren

4.6.3 Results of the Algorithm Run and Google Maps

Table 4.4 shows the results of the algorithm run and of Google maps. For the ease of comparison, the original results of Google map is converted into the result format used by the algorithm because Google Map outputs are represented as general direction between origin and destination. Following the direction, entry/exit point used is written as postcode if it is relevant. Google Maps searches at maximum three possible routes between the O-D. 'Priority' is the order of recommendation in terms of fastest route. The fastest route means either shortest travel time or shortest travel distance, case by case. In other words, shorter travel time or shorter distance route is given priority over longer travel time or longer distance routes. In most cases, the order is based on route travel time rather than route distance.

The higher the tolerance level, trip tends to use through trip route and highway entry/exit point. Use of local road entry/exit point is more frequent when highway and through trip detour are not allowed (P_h=1.0, P_t=1.0). While the results slightly vary among set 1, 2, and 3, it is the same between the results of set 3 and 4 for all test cases. Although it is not documented here, the results of trial runs do not change anymore with higher values of parameters such as 1.75 or 2.0 for both. It can be said that there is no difference in the performance of the algorithm with the parameter values that exceeds 1.2 for P_h and 1.25 for P_t. Now the problem becomes to find the minimum threshold of parameter values that performs the best. In other words, parameter setting which identifies entry/exit point as similar as ones used in reality is determined as default to the

Table 4.3 Setting of parameter values

	Preference of highway to local road (P_h)	Tolerance of extra travel time for through trips (P_t)
Set 1	1.00	1.00
Set 2	1.10	1.13
Set 3	1.20	1.25
Set 4	1.40	1.50

path finding algorithm for prediction of external trips. Seeing that the total travel time calculated by algorithm is always smaller than that of Google Maps for all cases, different calculation methods seem to be used. Moreover, postcode centroid of the algorithm and Google Maps may be different. Therefore, comparison of the algorithm and Google Maps is based on which route the algorithm decides between the through trip route vs. the direct route and which entry/exit point is used for a route between the O-Ds. For E-E trips, it matters that how well the resulted route of the algorithm and Google Maps matches each other. The following text describes the results of some test cases.

- Alkmaar, Noord-Holland – Dronten, Flevoland: This is internal-external trip. Looking at the map, the trip from Alkmaar to Dronten may take the route crossing northern Enkhuizen-Lelystad dike (local road) or southern highways passing Harlem – Almere. As expected to the result of the algorithm under the set 1 (P_h=1.0, P_t=1.0.), the algorithm finds the exit point as the dike which is local road. This local route is the third suggestion of Google Maps. Under the other settings, the route considered by the algorithm is the same as the first recommended route by Google Maps. The route passes highways through Harlem, Amsterdam, Almere and exits the highway A6 (postcode 8226) located near Lelystad, Flevoland.
- Amsterdam Zuidoost, Noord-Holland – Wijdemeren, Noord-Holland: This is internal-external trip. The origin and the destination are closely located to each other. The algorithm finds two routes. A route totally uses a local road N236 and exit from it (1218) under the set 1 and 2. Another resulted route under the set 3 and 4 first exits the study area via A2 (1391) then arrive at the destination using a different local road N201 which is not relevant to entry/exit point. Google Maps also suggests the two routes but the first priority is using N236.
- Delft, Zuid-Holland – Enschede, Overijssel: Both the origin and the destination are located outside the study area. The direct route on a map from Delft to Enschede seems to pass

highways through Utrecht, Amersfoort, and Apeldoorn. The possible through trip route seems to require quite much longer detour than the direct route. This type of cases is notable that what route the algorithm results in between through trip route and direct route. Likewise, Google Maps recommends the direct route as the first choice and the through trip route as the next. This test case is expected to definitely show that the parameter setting of P_t largely influences the way that the algorithm identifies the entry/exit point. Under the set 1 and 2, the algorithm searches the route that has no interaction with the study area. The assumption that the route is intentionally to pass through the study area results in through trip route despite 25 minutes longer travel time. The through trip route detours through Leiden and Amsterdam until it reaches Amersfoort. Thus, the route enters the study area using A4 (1435) towards Amsterdam and exits using A1 (1412) heading to Amersfoort.

- Amstelveen, Noord-Holland – Leiden, Zuid-Holland: As the origin is within the study area and the destination is outside of it, it is external-internal trip. The algorithm only concerns a decision of using highway or local road. The algorithm identifies local road exit point (1431) when 10% or less extra travel time is allowed. However, A4 (1435) is used to exit the study area when more extra travel time is defined. Google Maps suggests only one route identical to the highway route.

Comparing the results of the algorithm and Google Maps for all test cases, it can be concluded that parameter setting of set 3 (P_b=1.2, P_t=1.25) is suitable, overall. The results of the algorithm do not perfectly correspond to the first recommendation of Google Maps. However, there always is the same resulted route as the second or third suggestion. In particular, when both the trip ends are located at external region, the through trip route determined by the algorithm is the second or third best route of Google Maps for some cases. Google Maps sometimes recommends direct route which traverse the outside the cordon line. As Google Maps suggests highway route rather than local route in most cases, drivers are willing to make a detour even it requires more travel time to use highway in reality. In turn, this supports the validity of the parameter setting of the algorithm.

4.7 CONCLUSION

Delineation of smaller study area raises the question of how to handle external trips. Omnitrans predicts external trip flows using static traffic assignment and these predicted flows are assumed to be constant even the scenario or dynamic traffic assignment is considered. The method Omnitrans uses is limited in the sense that external trip pattern influence the internal trip pattern. Extension of Albatross application solved the problems. Existing national model of Albatross now can handle smaller local study area by predicting external trips. Even though Albatross is able to deal with local study area, it isn't changed that Albatross predicts activity schedules for the entire Dutch population. To predict external trip flows, what is developed is the algorithm which identifies the route between the O-Ds, based on fastest travel times. In specific, the algorithm finds entry/exit point that functions as external stations in general. If it is known which entry/exit point is used between the O-Ds, these external trip counts are added to internal trip counts, completing the study area O-D tip table. The algorithm compares total travel times of two routes: the direct route and the artificial route, intentionally constructed to pass entry/exit point, external trip in other words. The determined entry/exit points satisfy minimized travel time. The parameter values are involved in comparison of the two travel times. This results in which route to use between external trip route and direct route. Significance of this chapter is to find the best suitable parameter values because the resulted route varies according to parameter settings. In order to draw the conclusion, outcomes of the algorithm are compared to Google Maps search results. Consequently, the parameter values which replicates as similar as reality are determined for the algorithm and they are further used for prediction of external trips of the study area.

Table 4.4 The results of the algorithm and Google Maps

Origin		Destination		Path finding algorithm								Google					
Postcode	Plaat, Provincie	Postcode	Plaat, Provincie	P1	P2	Enter	Exit	EnterHW	ExitHW	Dist	Dmin	Priority	Enter	Exit	EnterHW	ExitHW	min
1011	Amsterdam, Noord-Holland	6546	Nijmegen, Gelderland	1.00	1.00	-1	1411	-1	0	76	71	1	-1	1412	-1	1	82
				1.10	1.13	-1	1391	-1	1	76	71	2	-1	1391	-1	1	85
				1.20	1.25	-1	1391	-1	1	76	71	3	-1	1391	-1	1	86
				1.40	1.50	-1	1391	-1	1	76	71						
1011	Amsterdam, Noord-Holland	8448	Heerenveen, Friesland	1.00	1.00	-1	8226	-1	1	77	72	1	-1	8226	-1	1	83
				1.10	1.13	-1	8226	-1	1	77	72	2	-1	8752	-1	1	93
				1.20	1.25	-1	8226	-1	1	77	72	3	-1	1412	-1	1	110
				1.40	1.50	-1	8226	-1	1	77	72						
1011	Amsterdam, Noord-Holland	1223	Hilversum, Noord-Holland	1.00	1.00	-1	1411	-1	0	28	24	1	-1	1412	-1	1	31
				1.10	1.13	-1	1412	-1	1	29	24	2	-1	1412	-1	1	36
				1.20	1.25	-1	1412	-1	1	29	24						
				1.40	1.50	-1	1412	-1	1	29	24						
1011	Amsterdam, Noord-Holland	8011	Zwolle, Overijssel	1.00	1.00	-1	1261	-1	1	67	64	1	1	1261	-1	1	79
				1.10	1.13	-1	1261	-1	1	67	64	2	-1	8226	-1	1	82
				1.20	1.25	-1	1261	-1	1	67	64	3	-1	1412	-1	1	85
				1.40	1.50	-1	1261	-1	1	67	64						
1181	Amstelveen, Noord-Holland	2312	Leiden, Zuid-Holland	1.00	1.00	1	1431	-1	0	28	26	1	-1	1435	-1	1	35
				1.10	1.13	-1	1431	-1	0	28	26						
				1.20	1.25	-1	1435	-1	1	33	26						
				1.40	1.50	-1	1435	-1	1	33	26						
3563	Utrecht, Utrecht	3812	Amersfoort, Utrecht	1.00	1.00	-1	-1	-1	-1	20	20	1	-1	-1	-1	-1	32
				1.10	1.13	-1	-1	-1	-1	20	20						
				1.20	1.25	-1	-1	-1	-1	20	20						
				1.40	1.50	-1	-1	-1	-1	20	20						
3329	Dordrecht, Zuid-Holland	9725	Groningen, Groningen	1.00	1.00	-1	-1	-1	-1	134	134	1	1261	8226	1	1	154
				1.10	1.13	1261	8226	1	1	142	134	2	-1	-1	-1	-1	155
				1.20	1.25	1261	8226	1	1	142	134	3	-1	-1	-1	-1	166
				1.40	1.50	1261	8226	1	1	142	134						
2628	Delft, Zuid-Holland	7545	Enschede, Overijssel	1.00	1.00	1	1	-1	-1	109	109	1	-1	-1	-1	-1	127
				1.10	1.13	-1	-1	-1	-1	109	109	2	-1	-1	-1	-1	130
				1.20	1.25	1435	1412	1	1	135	109	3	1435	1412	1	1	136
				1.40	1.50	1435	1412	1	1	135	109						
2517	Den Haag, Zuid-Holland	8912	Leeuwarden, Friesland	1.00	1.00	-1	-1	-1	-1	107	107	1	1435	8752	1	1	127
				1.10	1.13	-1	-1	-1	-1	107	107	2	1435	8226	1	1	138
				1.20	1.25	1435	8752	1	1	127	107						
				1.40	1.50	1435	8752	1	1	127	107						
4837	Breda, Noord-Brabant	9411	Beilen, Drenthe	1.00	1.00	-1	-1	-1	-1	115	115	1	-1	-1	1	-1	134
				1.10	1.13	-1	-1	-1	-1	115	115						
				1.20	1.25	-1	-1	-1	-1	115	115						
				1.40	1.50	1	1	-1	-1	115	115						
5230	s Hertogenbosch, Noord Brabant	9725	Groningen, Groningen	1.00	1.00	-1	-1	-1	-1	120	123	1	1	-1	-1	-1	143
				1.10	1.13	1261	8226	1	1	132	123	2	1261	8226	1	1	146
				1.20	1.25	1261	8226	1	1	132	123	3	-1	-1	-1	-1	146
				1.40	1.50	1261	8226	1	1	132	123						
1822	Alkmaar, Noord-Holland	8448	Heerenveen, Friesland	1.00	1.00	-1	8241	1	0	93	69	1	-1	8752	-1	1	85
				1.10	1.13	-1	8226	-1	1	97	69						
				1.20	1.25	-1	8226	-1	1	97	69						
				1.40	1.50	-1	8226	1	1	97	69						
1822	Alkmaar, Noord-Holland	8252	Dronten, Flevoland	1.00	1.00	-1	8241	-1	0	66	63	1	1	8226	-1	1	90
				1.10	1.13	-1	8226	-1	1	67	63	2	-1	1261	-1	-1	96
				1.20	1.25	-1	8226	-1	1	67	63	3	1	8241	-1	0	99
				1.40	1.50	-1	8226	-1	1	67	63						
1391	Abcoude, Utrecht	1218	Hilversum, Noord Holland	1.00	1.00	1391	1218	1	0	14	14	1	-1	-1	-1	1	30
				1.10	1.13	-1	-1	-1	-1	14	14	2	-1	1412	-1	1	28
				1.20	1.25	-1	-1	-1	-1	14	14						
				1.40	1.50	1	1	-1	-1	14	14						
1101	Amsterdam Zuidoost, Noord-Holland	1243	Wijdemeren, Noord-Holland	1.00	1.00	-1	1218	-1	0	18	16	1	-1	1218	-1	0	26
				1.10	1.13	-1	1218	-1	0	18	16	2	-1	1391	-1	1	32
				1.20	1.25	-1	1391	-1	1	20	16						
				1.40	1.50	-1	1391	-1	1	20	16						

5 MODELING SCHEME AND METHODS

5.1 INTRODUCTION

Having developed the path finding algorithm identifying external stations that trips might pass when entering or exiting the study area, Albatross now can handle a smaller local study area. Whereas the extension adjusting spatial scope of Albatross and MaDAM is a preprocessing preparatory procedure to the linked model system, the following steps concern actual modeling tasks of the integrated system. The chapter deals with the materialization of the conceptual framework of the linked model described in the introduction chapter of this thesis. The chapter is structured as follows. Section 2 introduces the overall modeling procedure of the new system in brief. Since satisfying compatibility of the two different applications matters for linkage of the models, identification of input and output data contents and format are considered important. Thus, overview of input/output data is given in section 3. Section 4 describes methods of every modeling task within the whole process, together with detailed description of data properties such as contents and file formats.

5.2 MODELING SCHEME OF THE LINKED MODEL SYSTEM

The modeling procedure of the linked model system is illustrated in Figure 5.1. Compared to the conventional four step model, the distinctive difference of the linked model system is that travel demand is modeled using completely different approach and method. The first three steps concerning travel demand modeling is substituted by a rule-based and activity-based model, Albatross. The way Albatross predicts travel demand is by derivation of travel information from activity schedules which are predicted under the consideration of various constraints. In order to predict activity schedules, travelers and households for micro simulation are necessary. The population synthesizer module of the Albatross application creates a synthetic population of individuals and households using an iterative proportional fitting (IPF) type method. The synthesized population information is stored as observed schedules, but with empty activity schedule data. Of course, export of schedule or household data is possible. Having individuals from households, Albatross simulates activity scheduling of the individuals. As a result of prediction, observed schedules now become predicted schedules with activity and travel information in terms of activity type, activity location, start time, end time, activity duration, transport mode, travel time, and travel party. The next step Albatross does in the linked model system is to create the final product of the application, trip matrices. These trip matrices are generated based on predicted schedules and further used for MaDAM in Omnitrans.

When data created from an application is being used in another application, compatibility of data matters. However, it is not very often that data used in different applications does exist in the same format and style. Likewise, the trip matrices generated by Albatross cannot be directly imported to Omnitrans because both applications use different data formats, at least for trip matrices. The output of Albatross must be converted to the format required in Omnitrans. In order to solve this compatibility problem, three adjustments are made during the whole process of the new system. First, Albatross has to generate trip matrices in database format (.db file) which is a required input format in Omnitrans. Before adding the option that generates output matrices in database format, only tab delimited text format (.txt file) has been supported in

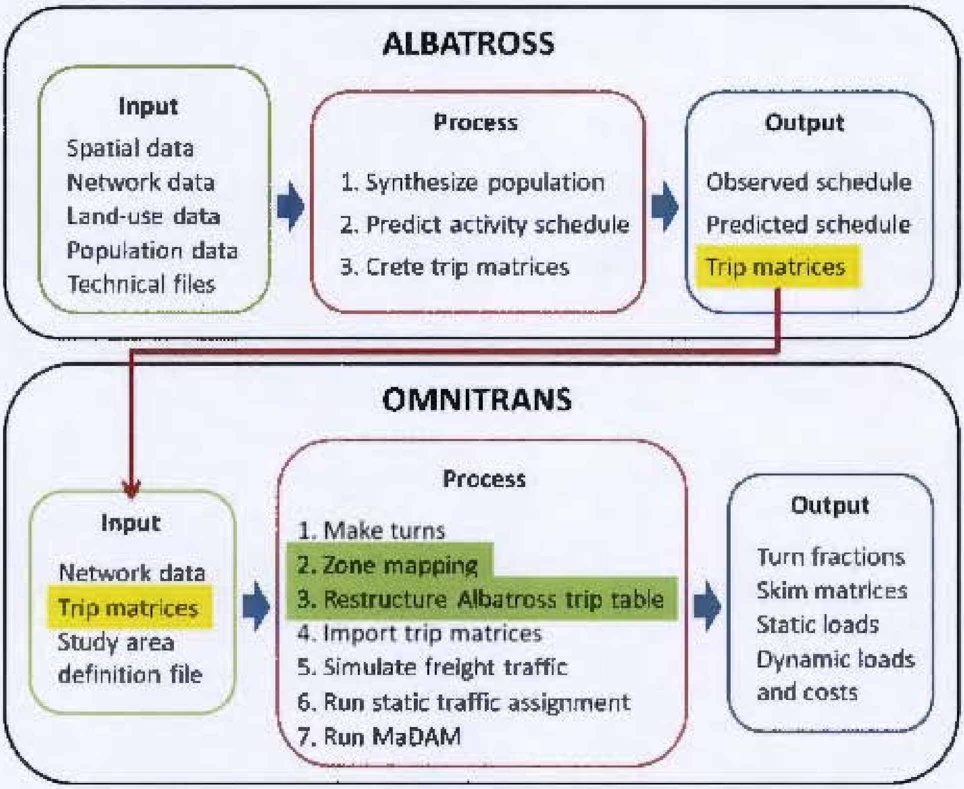


Figure 5.1Modeling scheme of the linked model system

Albatross application. Next two adjustments are involved in data conversion. The first data conversion step concerns matching the zoning systems of Albatross and Omnitrans. Although both zoning systems are based on postcode area, use of different versions of maps causes discrepancies of zone centroids at some spots on the map. Besides, one-to-one mapping failed for some postcode areas because they contain more than one zone centroids in Omnitrans and only a single centroid in Albatross. Dissimilar zoning systems become problematic when traffic assignment is done using Albatross trip table onto Omnitrans network. Hence, O-D centroid ids in Albatross trip table needs to be changed to Omnitrans centroid ids. For this, zone mapping tables, list of matching pairs of zone centroid ids of Albatross and Omnitrans, are created. Given the zone mapping tables, the second data conversion task restructures the Albatross trip matrices by transforming Albatross centroid ids to Omnitrans centroid ids.

Converted into Omnitrans format, trip matrices are imported to Omnitrans application and it is ready for traffic assignment. Although ultimate use of Omnitrans in the project is to perform dynamic traffic assignment based on trip flows predicted by Albatross, static traffic assignment is run before dynamic traffic assignment so as to calculate turn fractions. These turn fractions are required during dynamic network loading step as MaDAM adopts turn-fractions based method. At the end of the whole modeling process, one could see how traffic changes over time in the network.

5.3 DATA

Table 5.1 gives an overview of input databases that Albatross requires to perform the above mentioned modeling tasks. Spatial data defines spatial information of study area. Network data mostly concerns description of the transportation system. Travel time matrices were generated based on four-digit postcode system. Land-use data accounts for institutional context and attribute data of activity locations. Population data is used for population synthesis. Additionally, technical files relate to system parameters, PADT derivation, and study area trip matrix generation. Output data of Albatross is summarized in Table 5.2.

As ultimate use of Omnitrans in the project is to run dynamic traffic assignment given the OD trip matrices of Albatross, the main jobs relate to conversion of OD matrices to Omnitrans format and dynamic traffic assignment. For conversion, study area definition file organizing the relationships of spatial properties of study area in terms of entry/exit points and two zoning systems is necessary for building zone mapping files. In addition, both static traffic assignment and dynamic traffic assignment always requires a network for loading of traffic.

Data properties such as contents, structure, or file format are concretely explained in the next section.

Table 5.1 Overview of input databases of Albatross

Input to Albatross
Spatial data
- Codes and interrelations of postcode areas, zones, and subzones
- Distance between postcode areas, municipalities, and zones
Network data
- Subzone based travel times, distances, and costs
- Fastest route distances by car
- Fastest route distances by slow mode
- Fastest route travel times by car
Land-use data
- Attribute data at postcode area and municipality level
- Opening hours of retail, service, and leisure facilities
Population data
- Subzone based population data
- Sample household attribute data
Technical files
- Definition of system parameter values
- Data of a sample of trips for estimating travel costs and travel times
-Definition of spatial properties of study area

5.4 MODELING METHODS

5.4.1 Jobs in Albatross

A-1) Population synthesis: Although export population data is possible, it is not given in this project. Consequently, the first step is to synthesize a population for the whole of the Netherlands to simulate activity and travel patterns. The population synthesis module of Albatross application generates a synthetic population of individuals and households. The synthesis model is based on the IPF method. Basically, IPF is a mathematical scaling procedure as the initial cell values of cross-classified table from sample data are gradually adjusted through repeated calculations to fit known and fixed marginal row and column totals given by population data, finally completing multiway tables. National population statistics and a sample household attribute data are required to apply IPF method, each of them respectively defines margins and initial proportions of a multiway table. Depending on the availability of data, several multiway tables for calculation may exist for each zone. Combining the results from the multi-way tables of each zone, it constitutes a table that contains household and individual attribute data of target

Table 5.2 Output data of Albatross

Output from Albatross
Synthesizer
- Observed schedules (.obs file)
Scheduler
- Predicted schedules (.prd file)
Reporter
- Frequency tables and mobility indicators
- Trip matrices (.txt / .db file)

population of a zone. At the end of the entire process, household and individual data for all zones in the Netherlands is obtained and the result is written to empty observed schedules (.obs file).
As the method estimates a population from a sample, the sample must be representative of the target population. Even if creating the total population is possible, a fraction of the total population is synthesized due to computational burden. The synthesis module takes longer computation time as the fraction size increases. For example, a fraction 5% roughly takes 7-8 hours on a standard PC. It is reported that synthesis of a small fraction of the total population such as 5% is enough to address questions of strategic and long-term character (Arentze, 2008). In this project, however, fraction sizes are specified as high as 25%, 50%, and 100%. The reason is that the output trip matrices involved represent a relatively high degree of disaggregation especially in terms of temporal resolution, for example, Albatross defines 15 classes for time of day to describe the patterns. Setting of bigger fraction sizes is required to get reliable flows for each time of day. The fraction size can be controlled by changing the parameter level. Since investigating representativeness of the resulting activity travel patterns across different fraction sizes is one of the research issues of the project, four sets of fraction sizes are defined and independently run in four projects. Detailed descriptions of the IPF based method that Albatross uses and discussion results of the four projects defined by various fraction sizes are given in the following chapter.
A-2) Predict activity schedules: Activity schedules of individuals in each household are predicted. Inputs are the observed schedules (.obs file) which merely contain population data at household and person level. Prediction is currently executed by a DOS program (rwdata.exe) outside the system because the prediction module has not been embedded in the Albatross application yet. The program first calculates travel times and travel costs for a sample of cases stored in the PADT file of the project. And then, the prediction process starts. The DOS window displays a count of the cases (i.e., households) being processed. The bigger the fraction size is, the larger the number of cases to simulate is. In turn, this means that prediction run takes longer computation time with the increased number of cases. Outputs are the predicted schedules of a

sample fraction of the population (.prd file). For information how household activities are scheduled, see the literature review chapter of this thesis.

A-3) Create trip matrices: Predicted activity schedules provide travel information associated with the activities. The extracted trips can be represented as trip matrices. Users can specify the design of the trip matrix. The output variables of the trip matrix are free to choose by using tab dimension of the module. However, tab dimensions do not need to be chosen in our case because selecting the option of Omnitrans trip matrices generates output trip matrices written in predefined structure that Omnitrans requires. Even though retrieval of national trip table from the predicted schedules is feasible because Albatross predicts the activity schedules of entire Dutch population, Omnitrans needs travel demand of only the Noord-Holland region. Travel demand only relevant for the study area origin and destination is extracted from national activity and travel pattern. Output trip matrices must be stored in database file format (.db file) so that Omnitrans is able to import the Albatross trip tables. The structure of Albatross trip matrix is given in Table 5.3. The columns of the table accounts for trip information in terms of six dimensions (i.e., origin, destination, purpose, mode, time, user, and volume). The rows of the table are the cells of the trip matrix predicted. 'Origin' and 'Destination' displays the id number of postcode area in Albatross. 'Purpose' refers to activity purpose and is categorized by 8 classes. 'Mode' is classified into 4 groups. 'Time' accounts for when the trip takes place and the 'time' dimension is disaggregated into 15 time slices. 'User' dimension is a required field for Omnitrans as travel demand is stored as predefined 'PMTU' structure, although it is not meaningful at all in Albatross. Hence, a single user class is specified. 'Value' relates to the size of flows, the number of trips, in other words. The classes for purpose, mode, and time dimensions are shown in Table 5.4.

Table 5.3 Structure of output trip matrix in Albatross

Origin	Destination	Purpose	Mode	Time	User	Value
105	133	8	4	12	1	4
105	133	8	4	13	1	4
105	134	8	1	13	1	8
105	136	7	1	10	1	4
105	136	8	1	10	1	4
105	136	8	1	11	1	4
105	141	2	3	7	1	4

5.4.2 Jobs in Omnitrans

O-1) Make Turns: A network merely consisting of objects such as nodes and links is not sufficient for describing traffic flows. For example, simply putting two separate links crossed on two dimensions does not mean that it is a junction. Those two links may be a separate network

Table 5.4 Classes of variables in Albatross trip matrices

	Trip Purpose		Mode		Time of day		User
1	Work	1	Car	1	3 am -< 6 am	1	User
2	Business	2	Slow	2	6 am -< 8 am		
3	Bring or Get	3	Public	3	8 am -< 9 am		
4	Shopping	4	Car Passenger	4	9 am-< 10 am		
5	Service			5	10 am -< 11 am		
6	Social			6	11 am -< 12 pm		
7	Leisure and Touring			7	12 pm -< 1 pm		
8	Other			8	1 pm -< 2 pm		
				9	2 pm -< 3 pm		
				10	3 pm -< 4 pm		
				11	4 pm -< 5 pm		
				12	5 pm -< 6 pm		
				13	6 pm -< 7 pm		
				14	7 pm -< 8 pm		
				15	8pm -< 3 am		

road on a different elevation in real world, not intersecting each other. To make it meaningful as a junction, it can be additionally defined by turn objects. Moreover, traffic movements at junctions are diverse as some traffic goes straight or make turns. The job 'make turns' creates turn objects in the network for all nodes that have more than two associated links that have either no turn objects or have not been defined as junctions. This job should be run at least once before traffic assignment step.

O-2) Zone mapping: As said, zoning systems of Albatross and Omnitrans do not perfectly corresponds to each other because both application uses different version of maps and some of postcode areas have multiple zone centroids in Omnitrans whereas Albatross have one centroids for one postcode area. To be capable of linking the two models, compatibility of the zoning system needs to be met. What relates to this compatibility is the conversion of trip tables generated by Albatross to Omnitrans format. In order to change Albatross zone system to that of Omnitrans, there must be data describing interrelationship between the two zoning systems. Specifically, it concerns matching of zoning system of Albatross and Omnitrans. Input is 'StudyareaOmn-Final.dat' text file, also used in running of the path finding algorithm, the file defines spatial information of study area. Relevant to zone mapping problem, for each postcode zone centroid of Albatross, corresponding national LMS zone centroid id is defined. The output zone mapping tables serve as an index for restructure of Albatross trip matrices in the following step.

O-3) Restructure Albatross trip table: Based on zone mapping tables created in the previous step, Albatross postcode centroid ids in Albatross trip table are converted to Omnitrans zone centroid ids. In other words, this job transforms only origin and destination ids in Albatross trip table to Omnitrans format, remaining other variables in Albatross trip table fixed. Thus, input is original Albatross trip table and output is Omnitrans formatted trip table with changed origin and destination ids.

O-4) Import trip matrices: This job imports the trip matrix converted to Omnitrans input format and stores the data to matrix cubes in PMTU (Purpose, Mode, Time, User) structure in Omnitrans. Although Albatross identifies 8 types of trip purposes and predicts travel demand for these types, Omnitrans first aggregate these purposes into a single purpose prior to import trip matrices. The reason behind the aggregation is that the purpose dimension does not have any influence at all on the result of traffic assignment. It is difficult to predict how people choose their route based on purposes. In Omnitrans, the behavior of different trip purposes appears the same for all purposes while traffic patterns in terms of different transport mode is distinguishable. Also, MaDAM cannot make a distinction between trip purposes in dynamic network loading, while traffic assignment results vary in terms of transport modes. Except trip purpose dimension, the other three dimensions (M,T, and U) are used the same as in Albatross trip table. Therefore, output of this job is 60 matrices in total (i.e. combinatorial of 1 trip purpose, 4 transport modes, 15 time slices, and 1 user).

O-5) Simulate Freight: Albatross considers only passenger traffic. Traffic pattern of individual private cars/ car passengers, public and slow transport mode are predicted. Because in reality there is also freight transport, freight traffic is simulated. What is done by this job is multiplying freight traffic factors to all cells of the existing OD matrix which contains only car driver's traffic. Freight traffic factors for different time slices corresponding to 'time' classes used in Albatross are obtained from external sources. The factors are multiplied to OD matrices of car driver's mode and the resulted 15 matrices are stored as other matrix cubes in Omnitrans, labeled purpose 'Simulate Freight'. As the result of this step, 75 matrices have been created. Only car driver's traffic is used for following steps.

O-6) Running static traffic assignment: Static traffic assignment precedes dynamic traffic assignment so that turn movement fractions are obtained for dynamic network loading. Running of static traffic assignment in this modeling scheme is nothing more than acquisition of turning proportions. The input is intermediate OD matrices (i.e., simulated freight traffic). First of all, one-hour-period OD matrices are created. As travel demand is disaggregated by 15 time slices and a day (24 hours) is not evenly divided, duration of each time slice is not equal. Under the assumption that traffic pattern in every hour within the same time slice is identical, total travel demand for time slice 't' is simply divided by the length of each time slice to create one-hour OD matrices. As it assumes identical traffic pattern within the same time slice, Omnitrans only creates 15 one-hour OD matrices and saves the data as new matrices, labeled with purpose 'OneHour'. Therefore, there are overall 90 matrix cubes in PMTU combination. And then, static traffic assignment is performed with one-hour OD matrices. Among many static traffic assignment methods, volume averaging method is used. Also known as the method of successive averages (MSA), a well-known algorithm in transportation field, it is stochastic optimization method with predetermined step sizes used for solving stochastic assignment problems. Therefore, it is mainly characterized as iterative process to find equilibrium to traffic assignment problems. Explained the method concretely, traffic volume is calculated by linear combination of traffic volume of previous iteration and results of all-or-nothing assignment in current iteration. This volume averaging process terminates when one of multiple conditions is satisfied. One condition is that the maximum number of iterations defined by user is reached. In this project, 10 times of iteration is specified. The iteration process stops also when convergence is detected. What plays a role as a criterion for convergence is user-defined value, epsilon in technical term. Epsilon indicates the change in generalized cost between two consecutive iterations. When the relative increase or decrease in the cost falls below a predefined value, the process terminates. The true convergence is when the result of iterations does not change any more. However, epsilon is usually defined as a small, non-zero value for systematic efficiency. If the difference falls below the value, it is considered as being converged even if there is a slight change in the generalized cost for iterations. Furthermore, scope of assignment can be controlled. In other words, the assignment process can be constrained to a specified set of zone centroids by limiting the path building to only the specified set of to/from zone centorids. The centroids relevant for the study area are defined in the job. As an output of static traffic assignment, skim matrices (.skm files) are generated and saved as skim matrix cube that has PMTURI (Purpose, Mode, Time, User, Result, and Iteration) dimensions. The impedance data such as total generalized cost, total distance, and total travel time calculated between zone centroids by one of the assignment process are stored in 'Result' dimension in the Skim cube. Moreover, turning movement proportions are obtained. Another important output is static loads (.db file) on the links in the current network.

O-7) Running dynamic traffic assignment: Static loads calculated in static traffic assignment step are input to dynamic network loading. The simulation results can be calculated for every second. However, storing the results on that time basis is not practical. Thus, aggregation to some extent is needed to make the results are manageable. Time step size can be controlled by user and the length of period is defined as 30 minutes. During the simulation time, a vehicle may traverse a link quicker than the duration of the time step. That is to say, a vehicle arrives at its destination before end of simulating minutes. However, the model does not allow this, implying that length of a link and the simulation time slice need to match. As a solution to this, link length is adjusted to match the simulation period by links are automatically lengthened when they are too short for simulation time period. Dynamic loads and cost on the links is the output result of this step.

5.5 CONCLUSION

The operational linked model system is described with detailed modeling steps and identification of input and output databases. Whereas Albatross predicts travel demand, eventually creating trip matrices and delivers them to Omnitrans, Omnitrans runs MaDAM for dynamic traffic assignment. Bridging the two systems practically goes with the problem of data conversion. Trip matrices are delivered from Albatross to Omnitrans. For this, some adjustments are made during the whole process. Albatross generates trip matrices structured into Omnitrans required format. Moreover, zone mapping that matches the two different zoning systems (i.e., zone centroids ids) of Albatross and Omnitrans is executed. Lastly, Albatross zone ids are changed to those of Omnitrans based on zone mapping files. The following chapters discuss the results of some modeling steps, population synthesis and traffic assignment.

6 SAMPLE SIZE AND PREDICTED TRAFFIC FLOWS

6.1 INTRODUCTION

Modeling tasks are identified in terms of methods, data contents, and data formats. Especially, conversion of data into transferrable format between Albatross and Omnitrans is the most important problem that must be solved to make the linkage. Organizing the modeling steps into a certain order, the linked model system which bridges Albatross and Omnitrans gets into operation. The chapters from now on discuss some research questions with the results obtained from runs of the linked model system. This chapter touches on the third research question of how resulting traffic pattern varies in relation to different fraction sizes of the population. Due to computation times of the population synthesizer, not whole population but only a small fraction of it has been used for simulation in Albatross. The practice of synthesizing a fraction is reported to be allowable because that small fraction of population is sufficient to address long-term strategic questions with regard to people's activity-travel patterns. Since Albatross, like any other micro simulation model uses Monte Carlo simulation, however, the size of sample fraction can have an impact on outcomes. Furthermore, the problem focus of the present research is rather microscopic as traffic patterns are described at more refined temporal and spatial resolution, for instance, every moment in time of the day and every link of the road network. Thus, a small fraction of a population may not be able to fully display a representative picture of the constantly varying traffic pattern of the day. In other words, bigger sample size may be required so that the results are reliable. Therefore, this chapter deals with the influence of sample size on the resulting traffic pattern predicted by the model. To see how traffic flows differ with respect to sample size, the linked model is independently run under different population fraction sizes as separate projects to each other. The fraction sizes are given in the following section. The influence of sample size on predicted traffic flows is analyzed by the results from Albatross OD trip matrices and Omnitrans traffic assignment output data. Section 3 discusses the results of the analysis and the chapter finishes with a conclusion.

6.2 POPULATION SYNTHESIS WITH DIFFERENT FRACTION SIZES

A set of prediction runs of Albatross for different settings of the parameter of the synthesis model is required in order to see how traffic pattern differs according to sample size of a population. The parameter defines the fraction of the population. Another extended feature of Albatross concerning population synthesis is that different levels of fraction of the population can be independently defined for inside and outside study area. As the interest of traffic behavior is only limited to the study area and the pattern outside is excluded from the traffic analysis, it is obvious that smaller sample size for outside study area than for inside study area would not be a problem. The advantage of defining different fraction levels for both areas increases computational efficiency, meaning reduced computation time. Considering that synthesis of larger fraction of population takes longer hours for its calculation, setting fraction size as small as 2% or 4% for outside study area saves time compared to the time taking for the synthesis of as big as 25%, for example, regardless of whether it is inside or outside study area. Different settings of fractions sizes, which are used for the analysis here, are given in Table 6.1. Even though as small as 5% is sufficient in practice for applications of Albatross so far, larger fraction sizes such as 25%, 50% and 100% are defined for study area because the output trip matrices involved represent a relatively high degree of disaggregation especially regarding temporal resolution. On the one hand, Albatross defines 15 classes for time of day to describe the temporal dimension of patterns. On the other hand, MaDAM calculates traffic condition for every second. Setting of bigger fraction sizes might be required to get reliable flows for each time of day. In contrast, the fraction levels of outside study area are specified as small as 2% and 4%. Combining aforementioned fraction levels, four sets of fraction sizes are suggested. This implies that four projects are defined and computed in Albatross. For each project, it follows the entire modeling steps within Albatross, synthesizing population, predicting schedules, and generating trip matrices. In order to have predicted traffic flows of the entire population, Albatross rescales output results (e.g., trip flows in an OD matrix) with the inverse of the sample fraction after prediction module. Rescaling is done only with respect to the output in reports (e.g., indicators) and trip matrix. For instance, the inverse of 25%, four, is used as multiplication factor if a fraction of 25% of population is synthesized.

Table 6.1 Fraction levels of the projects

	Inside study area	Outside study area
Project 1	25%	2%
Project 2	50%	2%
Project 3	100%	2%
Project 4	100%	4%

6.3 RESULTS

Four projects are specified with different levels of fraction of population. This indicates that all modeling steps of the linked model are executed four times under the different parameter settings. Traffic patterns can be analyzed with data obtained at Albatross level and Omnitrans level. At Albatross level, OD trip matrices are used to derive information about traffic flows. On the other hand, Omnitrans provides network property type data such as link speed or cost obtained as a result of traffic assignment. While traffic pattern is represented at road segment level in Omnitrans because traffic assignment is a process of simulating how vehicles are moving through the road network, Albatross describes traffic flows at OD relational level. This section describes the analysis which finds the answer to the main question of this chapter "Are there differences between flows on the road network (for times of day) when different fractions are used?"

6.3.1 Summary Results of Predicted Flows

In Albatross' OD trip matrices, trips between each OD relation are represented as an aggregate flow for a certain combination of origin, destination, activity purpose, transport mode, and time of day. In this section, we consider first the 'Time of day' dimension only to see how traffic flow varies over time according to sample size. Table 6.2 can be extracted from original Albatross OD trip matrix. For each project, the first column is the number of predicted trip cases of the sample. The second column is the size of total flows (the number of trips) of the predicted trips after scaling of sample cases to population. In the table, trips are aggregated by 'time of day'. The sample size is bigger with the higher level of population fraction. However, the resulting size of flow for each time of day across projects seems to be similar to each other because the flow of every project is at population level, resulted from rescaling of predicted trip cases of sample to those of population by multiplying the inverse of sample fraction. Each trip case from a sample of half population, the predicted case is doubled to be cases of population, making travel demand of whole population. Figure 6.1 is a bar graph drawn from the data in Table 6.2. The graph displays the size of flow by time of day across four projects. The overall pattern of the four projects looks alike. There seems to be little difference in the height of bars, the size of flow, of the four projects within the same class of time dimension. Table 6.3 displays the pattern in another way. Absolute differences between two projects are shown as a percentage. The row margin of the table represents the difference of the total size of flow of the day. Column margin represents the mean of the differences calculated for each time of day. Although the difference between different sample sizes is a bit bigger at trips between 3am-6am compared to that of other times of day, the difference on the whole is not much, mostly less than 1%.

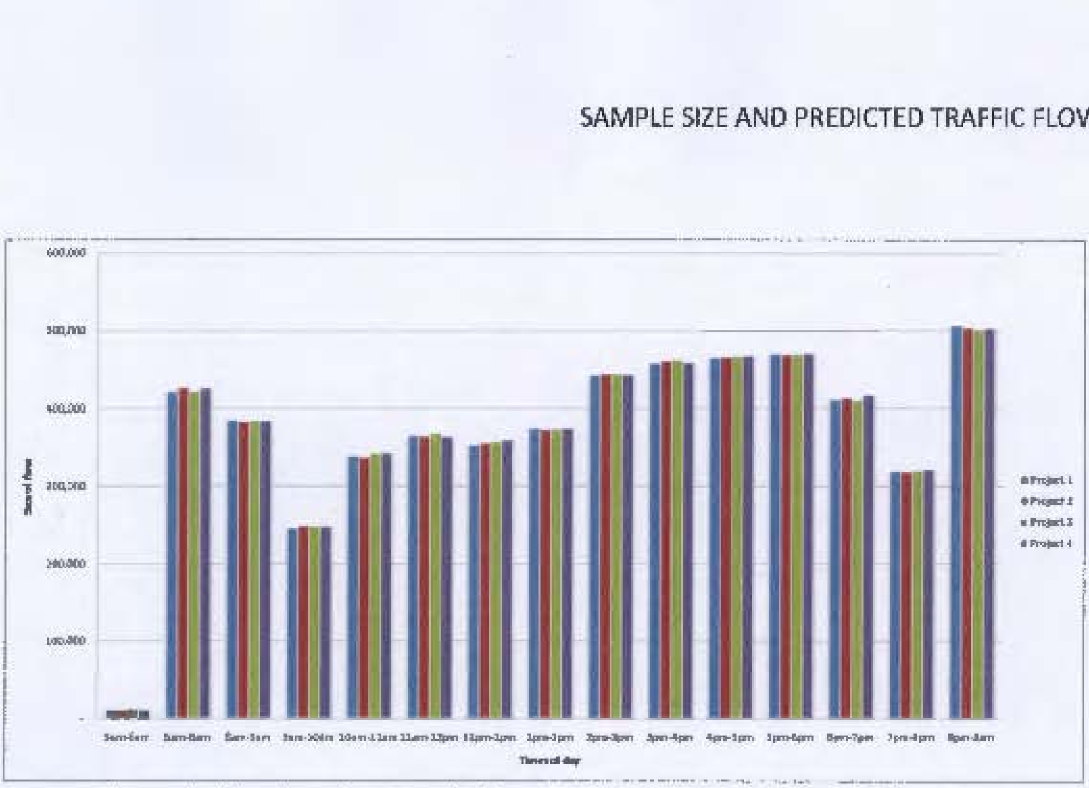


Figure 6.1 Size of predicted traffic flow by time moment for the different projects

That is to say, it can be said that traffic flows of a population on an aggregated level is robust for choice of sample size within a wide range in Albatross.

6.3.2 Results at Road Segment Level

Traffic assignment distributes trip demand at O-D relations onto the road network of the study area and lets us know how the network is used. The assignment results in values of network properties such as link loads for every moment throughout the day. Whereas static traffic assignment results are computed for 15 time slices for the day, the traffic situation is calculated for every second in dynamic traffic assignment but it is saved for every 30 minutes. The results are stored for each link in the network, and then, traffic patterns at road segment level are described. The output data can be viewed by the database inspector of Omnitrans in table format. The table displays the result of each link in PMTURI (purpose, mode, time, user, result, iteration) combinations and direction of the link along with calculated network properties such as load, coast and speed. Additionally, the results can be represented as video display of animated link bandwidths of link loads.

Here, traffic pattern is described in the context of static traffic assignment because the gridlock problem is diagnosed in dynamic traffic assignment. It refers to a phenomenon that traffic flows disappear, gradually decreasing to zero, while the phenomenon is spreading outwards across the network, after extremely severe traffic congestion occurred. It definitely appears in the center of Amsterdam and its surrounding ring road A10. The problem occurs when

Table 6.2 Predicted size of total traffic flows by times of day

		Project 1 (N sample = 386396)		Project 2 (N sample = 657,810)		Project 3 (N sample = 1,200,549)		Project 4 (N sample = 1,315,546)	
		N predicted cases	Size total flow	N predicted cases	Size total flow	N predicted cases	Size total flow	N predicted cases	Size total flow
Time of day	3am-6am	2,344	11,942	4,323	12,048	7,893	12,290	7,835	11,858
	6am-8am	47,794	422,550	75,841	427,748	117,658	422,003	118,066	426,729
	8am-9am	45,600	385,006	74,041	383,030	116,527	384,648	116,554	384,867
	9am-10am	37,555	245,980	63,636	249,202	105,940	248,040	106,453	248,523
	10am-11am	42,413	338,714	71,716	337,732	116,860	341,479	116,890	342,879
	11am-12pm	48,279	365,578	80,945	364,980	131,947	367,721	131,392	364,052
	12pm-1pm	46,007	353,412	77,491	356,612	127,154	357,935	127,421	359,645
	1pm-2pm	51,843	374,126	86,875	372,878	142,021	374,042	142,962	374,669
	2pm-3pm	54,619	442,460	91,635	444,296	148,190	443,922	148,577	443,258
	3pm-4pm	58,913	458,928	98,736	461,212	159,877	462,547	160,048	459,593
	4pm-5pm	55,858	465,756	91,208	466,294	145,845	467,508	146,434	468,281
	5pm-6pm	58,774	470,460	95,959	469,266	154,846	469,836	155,320	470,827
	6pm-7pm	49,806	411,756	82,426	413,772	133,133	410,743	134,348	417,813
	7pm-8pm	41,525	319,562	69,365	318,502	112,382	319,884	112,767	321,204
	8pm-3am	54,526	506,208	87,968	503,818	137,604	501,059	138,787	503,105
Total		**695,856**	**5,572,438**	**1,152,165**	**5,581,390**	**1,857,877**	**5,583,657**	**1,863,854**	**5,597,303**

Table 6.3 Percentage of absolute difference in predicted traffic flows between projects

	Absolute difference between project X and project Y (%)						Mean
	1_2	1_3	1_4	2_3	2_4	3_4	
3am-6am	0.89	2.91	0.70	2.01	1.58	3.52	1.93
6am-8am	1.23	0.13	0.99	1.34	0.24	1.12	0.84
8am-9am	0.51	0.09	0.04	0.42	0.48	0.06	0.27
9am-10am	1.31	0.84	1.03	0.47	0.27	0.19	0.69
10am-11am	0.29	0.82	1.23	1.11	1.52	0.41	0.90
11am-12pm	0.16	0.59	0.42	0.75	0.25	1.00	0.53
12pm-1pm	0.91	1.28	1.76	0.37	0.85	0.48	0.94
1pm-2pm	0.33	0.02	0.15	0.31	0.48	0.17	0.24
2pm-3pm	0.41	0.33	0.18	0.08	0.23	0.15	0.23
3pm-4pm	0.50	0.79	0.14	0.29	0.35	0.64	0.45
4pm-5pm	0.12	0.38	0.54	0.26	0.43	0.17	0.31
5pm-6pm	0.25	0.13	0.08	0.12	0.33	0.21	0.19
6pm-7pm	0.49	0.25	1.47	0.73	0.98	1.72	0.94
7pm-8pm	0.33	0.10	0.51	0.43	0.85	0.41	0.44
8pm-3am	0.47	1.02	0.61	0.55	0.14	0.41	0.53
Total size of flow	0.16	0.20	0.00	0.04	0.00	0.24	0.11

the capacity is too low on several links close to each other or when there is one severe bottleneck on a critical part in the network. A sort of "circular" congestion occurs in which the head of the congestion reaches the tail of the congestion, after a while it is noticed that there is no traffic at all and the affected area grows progressively. It is exceptional because traffic disappears after the severe congestion at afternoon peak hours such as 4pm-5pm, even though the congestion later in evening reaches a peak in static assignment results. We emphasize that, even though we use static assignment, predicted traffic flows are still dynamic in the sense that we can monitor the change of flows through time.

The effect of sample size on traffic pattern is examined by comparing link loads of separate links evenly distributed onto the study area rather than routes composed of multiple sequential links. Since traffic stream changes only at junctions or joints where vehicles on the current link exit from the link turning to another direction and where vehicles from another link get into the current link, the continuous links with only straight flow without junctions have the same values of link load until one of the link reaches an intersection. Especially most links on highway routes are straight links except the links for exit or entrance. Furthermore, reflecting the objective of the analysis to see whether traffic flow varies with regard to different sample sizes, it is more appropriate to select separate links and compare loads of the same links between different projects. 20 links are selected for analysis: 10 links on highway and 10 links on more local roads (see Table 6.4). The selected highway links are generally a part of major routes where traffic patterns of off-peak and peak hours are distinctive shown from animated bandwidths of static traffic loads in Omnitrans. Regarding local road links, 10 municipalities at province Noord-Holland are chosen in terms of population density obtained from CBS Netherlands. Listed from high to low population density, inner city like Amsterdam has the highest population density and relatively outer city, Schagen, has the lowest among the chosen local road links. The highway links and the local road links are analyzed separately because each of the road types obviously reveals dissimilar patterns due to different network properties such as speed limit and capacity.

6.3.2.1 Visual representation of traffic flows

Having selected the links, static loads of the links are collected from Omnitrans database for examination. The loads of each link are disaggregated by time of day and project. Since there are 10 links for each road type and 15 times of day, in total 150 scores are obtained for each project. First to get a first indication and overall insight on differences of flows according to sample size, the data is visually represented. Omnitrans provides a useful tool for this: animated bandwidths display of static loads. It can be viewed as a video by consecutive playback of whole simulation periods. Figure 6.2 displays a collection of traffic conditions of the link 113401 on A4 as an example. Each image relates to a specific time of day and project. The link load is shown as the number on the side of the link by direction. The colored band displays the percentage average calculated speed with respect to maximum allowed speed of the link. Color change of band

Figure 6.2 Animated bandwidths of highway link by time of day and project

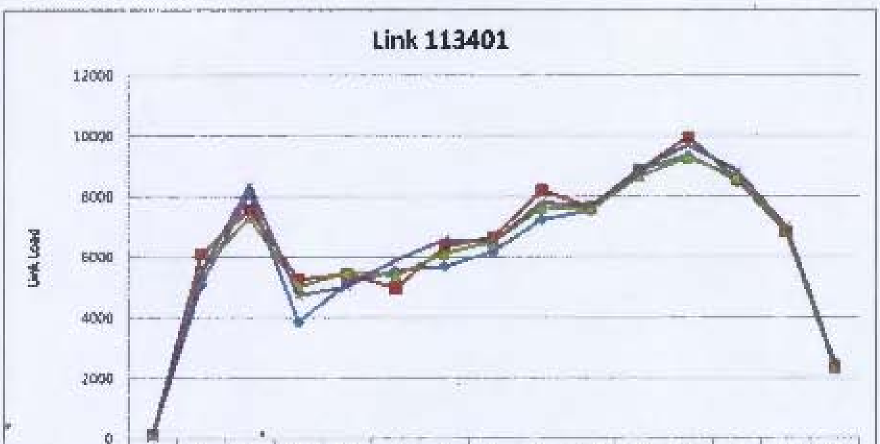
Table 6.4 Selected links on highway routes and local routes

Link Type	Link Number	Link Description
Highway	112269	A9
	112235	A200
	114315	A4
	113401	A4
	109863	A10
	109823	A7
	112443	A1
	111856	A6
	107326	A2
	117453	A5
Local road	30869	Amsterdam
	10224	Haarlem
	12439	Zaanstad
	11026	Haarlemmermeer
	11595	Alkmaar
	31785	Amstelveen
	11764	Den Helder
	11464	Aalsmeer
	35260	Diemen
	12387	Schagen

indicates the extent of congestion of the network. As the percentage level decreases, it means more congestion changing from green through yellow to red color. The images at each row illustrate traffic conditions at a specific time across projects. The pattern of each project looks similar to each other.

Alternatively to the link bandwidths representation, a line graph for each link is drawn by plotting the loads at different times of day across different projects in Excel (Figure 6.3). X and Y axis of the graph represent time of day and link load respectively. Each project is a data series of a chart. Loads of each project is depicted as an independent line of a different color for comparison of flows. There are four lines in a graph and each of them represents the loads of a project. The graphs generally display the same pattern of little difference of link loads across projects. Seeing the graph of the link on A4 (Figure 6.3), for instance, the loads of all links change with alike pattern although the actual levels of load of each project slightly varies. The overall appearance is similar as the lines overlap in large part with each other.

Both the animated bandwidths and the line graphs are helpful to get rough understandings of the traffic flow patterns. They indicate that there seems to be not much difference in traffic flows by different fraction sizes. However, the visualization techniques lack information in that they do not numerically exhibit to what extent the traffic flows differ and how significant the differences are between the projects. Therefore, we apply statistical measures to analyze the differences.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Project 1	116.77	5116.5	8121.9	3857	5125.9	5532.1	5695.7	6180.4	7243.8	7530	8895.9	9347.9	8505.7	6848.5	2409.9
Project 2	124.83	6092.7	7570.8	5276.1	5437	4965.5	6355.7	6646.6	8203.5	7634.9	8853.6	9964.2	8543.9	6894.4	2347.9
Project 3	149.74	5589	7313.7	5001.5	5500.9	5378.4	6116.2	6489.8	7640.1	7584.8	8664.8	9291	8627.1	6895.9	2323.1
Project 4	94.38	5520.6	8371.8	4719.1	5018.2	5913.7	6570.8	6468.9	7806	7662.3	8918.3	9710.4	8849.8	7000.9	2473.8

Figure 6.3 A line graph of traffic flows across project

6.3.2.2 Hypothesis testing of traffic flows using F-test

The main question to answer through the chapter is whether fraction size influences the resulting traffic flow. We expect that the traffic flows are less stable with the smaller sample size, which means more random variation in the link loads around average loads.

On the contrary to this, the larger the sample size is, the smaller the deviation between the link loads and average load of the link will be. Moreover, the extent of variation is expected to be larger in local roads than in highway. As Monte Carlo method is based on stochastic model for prediction, there is an element of randomness of it. What differentiates stochastic from deterministic simulation is the use of random predictions. Generally, the volume of trips on local roads is smaller compared to highways, which means that more fluctuation of the volume of local road trips is expected due to the stochastic nature of the model. Hence, two propositions are tested throughout the text: i) Does sample size have an impact on predicted traffic patterns? And ii) Does sample size differently influence traffic flows on highways and local roads?

Conceptually identical to the deviation described above, variance in statistics can be used to verify the prepositions. The level of variances of the projects explains the randomness of loads with regard to sample size. Hence verification is mainly based on comparison of variances. The procedure to calculate relevant statistics is described here. First, overall means (μ_{ij}, where i is link j is time of day) which are the average loads across the four projects is calculated for each link and each time slice. These mean a true value or best guess of the true value meaning the load obtained if the size of the sample is very large. For each link, the load by the time of day is summed across the projects ($\sum x_{ijk}$, where x is observed value, i is link, j is time of day, and k is project), and then it is divided by the number of groups, 4 in this case. Based on overall means, sample variance is calculated for each project. Sample variance is the sum of squares of deviation between observed score and overall mean ($\sum (x_{ijk} - \mu_{ij})^2$) divided by ($n - 1$, where n is the number of sample of the group).

Table 6.5 shows the calculated variances of highway links and local road links. Concerning variances at least, the result proves that sample size influences traffic flows since the random variation varies across different groups. However, it should be noted that variances of traffic flow obtained from every prediction run with the same fraction size always differ because Monte Carlo simulation method is used for prediction. Thus, it is unlikely to have exactly the same variance with the approach. The variances tend to decrease with increasing sample size, as we would expect. Intuitively thinking, it is expected that the larger the sample size as close as to population, the smaller the variance, lesser random variation. However, project 3 seems to be an outlier. Two trends are observed from the result. The first observation is a decreasing trend of variance from project 1 through 2 to 3. The second concerns decreasing variance from project 3 to 4. Traffic flows in highway follow the first trend but not the second. In contrast, not the first but the second one accounts for local road flows. According to the first trend, variance of project 3 on local road flows should be smaller than that of project 2. When it comes to the second trend, smaller variance of project 4 than that of project 3 is expected on highway flows considering decreasing trend from project 1 through 2 to 3. Therefore, the result of project 3 is exceptional in this research.

Nevertheless, comparing merely the size of variances does not tell the degree of difference between the projects. For instance, there is no information how much the flows differ between for example project 1 and project 4 and how significant the difference is. F-test, the test of equality of variance, demonstrates it. Hypothesis testing is based on F-statistic, the ratio of variances from two independent samples with normal distribution. As the statistical test departs from normality assumption, this condition needs to be met so that the F-test is meaningful. Normality of the data is checked by drawing a so-called normal Q-Q plot. It is a graphical technique for determining if the data sets come from populations with a normal distribution. The plot charts observed values against a normal distribution. Figure 6.4 displays the result of a normality test of SPSS. The straight line represents the data when it is perfectly normally

Table 6.5 The level of variances

Link Type	Project	Fraction size (inside/outside)	Variance
Highway	1	25% / 2%	65856.40
	2	50% / 2%	53858.21
	3	100% / 2%	38389.11
	4	100% / 4%	42788.15
Local road	1	25% / 2%	3693.89
	2	50% / 2%	2156.38
	3	100% / 2%	2873.98
	4	100% / 4%	1605.34

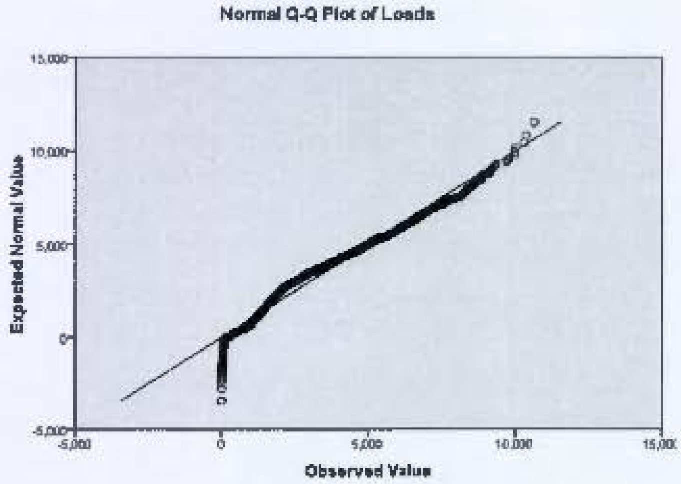


Figure 6.4 Normal Q-Q plot of loads

distributed and the line of dots is the observed values from the data. Thus, as the plots are closer to the 45-degree reference line, the more likely it is that the data are normally distributed. Although the plot shows the distribution deviates somewhat from normality at the low end, it overlaps the reference line in large part. Therefore, it can be concluded that the data is close to normal, implying that the F-test can be applied to this data. The hypothesis testing procedure is illustrated below.

First of all, hypotheses to be tested needs to be formulated. The logic of hypothesis testing always evaluates an empirical result against the null result. The empirical data exhibits different traffic flows in relation to sample size. Since it is a test of equality of variance, the null hypothesis is H_0: $\sigma_i = \sigma_j$ (The variance of project i and project j are equal). Thus, the test statistic F should equal 1 if the null hypothesis is true, because the variances are the same. As mentioned previously, variances of two groups cannot be identical because the prediction in Albatross takes Monte Carlo simulation approach that the outcomes always differ from run to run under the same parameter setting, the same fraction size here. Thus, the variances from two identical size samples are assumed to be the same for the hypothesis testing here. However, the formulated equality of variance hypothesis is not problematic in that no multiple runs of the same sample fraction size are executed and it focuses only on the effect of different sample fraction size. The alternative hypothesis is H_1: not H_0. If the null hypothesis is true, then there is no difference of flows with regard to sample size. If the alternative is true, then traffic flows differ in accordance with sample size.

Given a null and an alternative hypothesis, the next step concerned calculating test statistics and p-values at specified level of significance. We use the level, $\alpha = 0.05$ by convention. F-statistic is the ratio of variances of two independent samples. The critical value f_c that corresponds to the level of significance (5%) that we have chosen is 1.31 for a F-distribution with 149 degrees of freedom for both numerator and denominator because all groups (projects) have the same number of cases. The degrees of freedom is calculated by (the number of sample – 1). Thus we reject the null hypothesis when the value of the F-statistic calculated is greater than or equal to the critical value, 1.31. Otherwise, the null hypothesis is true. The four projects in each road type are compared in pairs and hence six ($_4C_2$) F-statistics (σ_i / σ_j, where i and j are project) can be obtained (see Table 6.6). Based on F-statistics, p-values are calculated to assess statistical significance of the results.

The last step of hypothesis testing is interpreting the results calculated in the previous phase. The ultimate use of test statistic is for testing of hypothesis nevertheless, F-statistic itself is useful if F-statistics are compared to each other, explicitly displaying the extent of differences between the two groups. Since we want to examine the effect of sample size in our case by comparison of two samples with different sizes, the ratio of variances is anticipated to increase as the extent of difference in sample size increases. That is to say, the result of F-test between 25% and 100% sample is larger than that of between 25% and 50%. In order to see this, the results are ordered from a comparison with the least difference of sample sizes to the most difference of sample sizes, as shown in Table 6.6. As explained in the section above, the size of the test statistics support that project 3 is obviously an outlier. Although project 4 has more samples and also the bigger predicted flows than project 3 in Albatross (see, Table 6.2) and thus one can expect that, for instance, F_{24} is bigger than F_{23}, F_{24} is smaller than F_{23} and F_{13} is bigger than F_{14} regarding highway flows. When it comes to local road flows, F_{34} and F_{23} shows unexpected result as the test statistic of samples with the least difference in size is smaller than the samples with almost twice difference, unlikely to highway flows. Therefore, the outlier case, project 3, is excluded in that it is difficult to draw clear conclusion with it because the project exerts dissimilarly to road types.

Table 6.6 The result of F-test

Link Type	Fraction Sizes (inside,outside/)	Test Statistic	p-value	Description
Highway	100%,2% / 100%,4%	$F_{34} = 0.90$	0.746	Not statistically significant
	50%,2% / 100%,2%	$F_{23} = 1.40$	0.02	Statistically significant
	50%,2% / 100%,4%	$F_{24} = 1.26$	0.081	Not quite statistically significant
	25%,2% / 50%,2%	$F_{12} = 1.22$	0.11	Not statistically significant
	25%,2% / 100%,2%	$F_{13} = 1.72$	0.001	Extremely statistically significant
	25%,2% / 100%,4%	$F_{14} = 1.54$	0.004	Very statistically significant
Local road	100%,2% / 100%,4%	$F_{34} = 1.79$	0.0002	Extremely statistically significant
	50%,2% / 100%,2%	$F_{23} = 0.75$	0.96	Not statistically significant
	50%,2% / 100%,4%	$F_{24} = 1.34$	0.036	Statistically significant
	25%,2% / 50%,2%	$F_{12} = 1.71$	0.001	Extremely statistically significant
	25%,2% / 100%,2%	$F_{13} = 1.29$	0.063	Not quite statistically significant
	25%,2% / 100%,4%	$F_{14} = 2.30$	0.0001	Extremely statistically significant

Further, significance testing is described here. Two values can be used for judgment of whether to reject the null hypothesis or not: F-statistic compared to the critical value and p-value compared to the level of significance. It must hold that F-statistic $\geq f_\alpha$ or p-value $\leq \alpha$ to reject the null hypothesis. P-value tells us how likely it is that we could get this empirical result simply due to chance. If the value is less than 0.05, then we are confident that the result is not due to random error, thus not rejecting the alternative hypothesis. Furthermore, p-value can be understood as the degree that the test statistic supports the null hypothesis. The smaller the p-value, the bigger the null hypothesis is rejected. When F-statistic is larger or equal to 1.31, or when p-value is less or equal than 0.05, the alternative hypothesis is accepted, meaning that variance of traffic flows differs in accordance with sample size. It is interpreted like, for example, F_{14} (2.30) of local road by far exceeds 1.31 and the p-value (0.0001) is far less than 0.05 in the comparison between project 1 and 4. We reject the null hypothesis and accept the alternative that variance of traffic flows differ in accordance with sample size. And the difference between the projects is extremely statistically significant.

The highest F-statistic and the lowest p-value in both link types result from the comparison between the smallest sample size with 25% fraction size and the biggest with 100%. It strongly supports our expectation that sample size have an impact on variance (or sample error) of resulting traffic flows. Furthermore, regarding the local road, all the F-statistics without comparison with project 3 are bigger than the critical value and all the p-values are smaller than the significance level, thus, leading to rejecting the null hypothesis. The value of local road links is generally bigger than that of highway links. Referring to the verbal description of interpreting results from the statistical test shown in the column 'description' in Table 6.6, the extent of difference of flows is generally stronger in local road links than highway links as four out of six tests conclude as statistically significant, rejecting the equality of variance assumption. From this, the second proposition that more variation exist in local road links than highway links is demonstrated. Consequently, the conclusion of the statistical inference is stated as twofold: i) Sample size influences traffic flows. And ii) The traffic flows of local road links are more sensitive to sample size than that of major routes like highway links.

The result from hypothesis testing supports our expectation derived from empirical data that sample fraction size influences resulting traffic flows. When it comes to hypothesis testing, however, the result should be interpreted with caution for some reasons. The data collected might be erroneous, reflecting meaningless variation due to sampling. Moreover, formulated hypothesis can be incorrect. Lastly the most important aspect to be considered, the logic of hypothesis testing is always conditional. Adopted in many texts, common approach to hypothesis test is a problem in which one of the null or the alternative will be asserted. Though, it should be noted on the distinction between accepting the null hypothesis and failing to reject it. Although rejection of the alternative conversely means acceptance of the null in common approach, this can be alternatively understood that the result supports or fails to support the hypothesis. Pointed out by Dallal in his webpage about "The Little Handbook of Statistical Practice" (http://www.tufts.edu/~gdallal/LHSP.HTM), null hypothesis are never accepted. Rather, we either reject or fail to reject them. Indeed, the hypothesis test never proves anything, only hypothesis that we formulated based on empirical data. Simply, it can be recognized that the null hypothesis is a possibility that could be observed. Supportive to this discussion, Fischer did not use an alternative hypothesis, and therefore there was no concept of "accepting" an alternative in his construction of significance tests, implying imprecise use of H_0 and H_1. To apply this discussion to our problem, statistical significance does not necessarily mean that it is also significant in practice. The results of hypothesis testing is may vary in accordance with how hypothesis is defined and how big the number of groups considered.

6.4 CONCLUSION

The chapter examined the impact of sample size on traffic flow. It should be noted that the sample size refers to sample fraction of population used for prediction of traffic behaviors, not the number of samples considered in statistical testing. It is described at two levels: O-D relational level of Albatross and road segment level of Omnitrans. Albatross result at least shows little difference of flows according to sample size. Visually represented as line graph, it is compared that total number of trips by times of day across projects with varying number of sample sizes. On the contrary, the traffic flow at more microscopic level of road segment displayed variation of flows with regard to sample size. Since link load data is available as a result of traffic assignment, link loads across different projects is compared. As it is general that randomness of data decreases with increased number of samples as close as to population, random variation with regard to sample size is tested. Variance accounts for the randomness of data. Merely variances tell that sample fraction size influences traffic flow. However, it is not enough to examine what extent traffic flow differs in accordance with sample fraction size of population, and therefore F-test is used for the analysis. The result of F-test has proved our expectation derived from empirical data which is the output of Omnitrans in twofold: i) Predicted of traffic flow is affected by sample fraction size. And ii) Traffic flows of local road links are more sensitive to sample size than that of major routes like highway links. Even though the result of hypothesis test supports empirical data, the result of significance test should be interpreted with discretion due to possibility of sampling error, incorrect formulation of hypothesis, and the logic of hypothesis testing itself. A distinction between rejection of the null hypothesis and failing to support the null hypothesis is the critical aspect considered in the interpretation of result. Even though a significant difference is found, this refers to significance in a statistical sense, only valid within the hypothesis testing. The difference might be small, and thus, not be meaningful as it can be ignored in practice.

7 ILLUSTRATION OF TRAFFIC FORECAST IN 2020

7.1 INTRODUCTION

Making technical linkage between Albatross and Omnitrans concerning data transformation for compatibility is successful since no execution error has diagnosed. Every predefined job within the linked model system has completed its task without any error during process. However, making software linkage is not enough to say the linked model system is valid. The system must be truly operational meaning that the output result makes sense. The chapter mainly discusses demonstration of the linked model system. And what helps model validation is scenario analysis. It is a test of the system and an essential ingredient of model building and quality assurance of the model. The analysis is served for validation purpose in the research. In particular, the most important aspect to check with this analysis is whether the operational linkage is established. Since Albatross output is used as input to jobs in Omnitrans, Omnitrans output has to be consistent with Albatross result, if the model system works properly. In other words, it is significant that predicted travel demand of Albatross is appropriately reflected in network loading in Omnitrans, and consequently, resulted link loads or speed is in line with Albatross result.

To evaluate validity of a model, base scenario is always involved to see changes under scenario situation. A scenario is considered and the changes in terms of the scenario are described in the chapter. As noted earlier, however, the scenario used in the research is a just particular case for system demonstration purpose, not to give practical implications of the scenario in transport planning perspectives. What want to show with this example case is how this linked model system can be used in practice by identifying useful information drawn from the output.

Scenario used in the research is aging society in 2020, literally meaning that older population would increase in future. There has been a rapid growth in the elderly population in the Netherlands like many other developed countries as a quarter of the Dutch will be over 65 years old by 2030. As elderly occupies a larger and larger proportion of entire population, their influences on society have become dominant and more important than ever. Lately, much has changed with regard to the position of older people in society. Not only do the elderly account for a higher proportion of the Dutch population but their income position and health has also improved in recent decades. Contributed from social developments, they are richer and healthier. Higher income status of the elderly encourages car possession possibly discouraging use of public transport. Moreover, they have more free time than young people and are flexible in time use because their labor participation still remains remarkably low. Those changes induce increased participation of out-of-home activities and in turn, the participation involves travel. As their mobility is growing, the aged are expected to exert significant impact on traffic patterns in the Netherlands. By comparison of traffic behavior of base year and scenario year, effects of mobility consequences of aging population can be explained in detail. And implications drawn from scenario analysis would contribute to transport planning. As emphasize in the previously, however, it focuses more on validation of the linked model system.

The chapter is structured as follows. First, assumptions and definitions of the scenario are described. The second section discusses the results. Since the scenario is involved, the analysis concerns comparison of baseline scenario and forecast scenario. Comparison is possible at two levels, Albatross activity-travel schedules and Omnitrans traffic network loading. The section mainly describes traffic pattern changes in relation to the scenario. Section 4 deals with the implications of the linked model system which is drawn from the description in section 3. This last section is to literally evaluate to what extent the objective of the linked model system mentioned in introduction chapter has been accomplished. It concerns answering to questions such as "Has operational linkage established between Albatross and Omnitrans?" or "How the linked model system can be used in practice?" The first question is answered by checking consistency of the results from Albatross to that of Omnitrans, whether Omnitrans traffic pattern of network loading is in line with Albatross traffic pattern of activity-travel demand. The latter is demonstrated by the analysis in section 3. Information drawn from the output shows the possibility of the linked system application.

7.2 AGING POPULATION SCENARIO IN 2020

As scenario is considered, traffic of base year and forecast year are predicted. Scenario analysis always involves baseline scenario because it serves to show what changes are predicted by the scenario. The scenario is increase of aging population in the Netherlands and the year 2000 and 2020 are taken for the baseline and the future scenario respectively (Arentze, 2008; Arentze, et al., 2008). Thus, the description of results is about mobility consequences of aging population.

The scenario 2020 considered here is a limited-growth scenario, called Regional Community scenario, which was developed in a joint study of several Dutch national planning agencies. The scenario considers several changes in terms of demography, transportation system, land-uses, and economic status. Reflected those changes in input data and parameter settings of Albatross, they are implemented in Albatross system for micro-simulation of activity and travel in 2020. For population synthesis of 2020, demographic changes, work status, income, and car possession (incl. driving license) are incorporated by manipulating sample or population census data of the base year. The population growth are expected to be roughly 4%, the senior citizens over 65 years are going to significantly increase and the population less than 35 years will decrease. These are taken into account in demographic data for synthesis. Two types of ratios, the ratio single female household head and the ratio living in female (position not as a household head) are applied for adjustment of distribution of households across household types. In 2020, proportion of single female household is reduced and females not being a head of household is increased. The sample data for synthesis of scenario 2020 are obtained by applying transitional probabilities, the shift of individuals or households from a class to another class, to the baseline sample of work status, income (economic growth), and car possession. Changes in land-use and transportation system are made by direct change of input data of Albatross. Ratios between 2020 and 2000 for each subzone by land-use sector are multiplied to the baseline land-use data. Travel times or distance matrices are also changed. General variable costs for transport mode are given by price indices and defined for system parameter settings of Albatross (see Table 7.1).

Table 7.1 Parameter setting of base year and forecast year

Indices settings	Baseline scenario (2000)	Scenario 2020
Car costs Off-peak hours	100	86.2
Car costs peak hours	100	86.2
Train costs before 9 am	100	119
Train cost after 9 am	100	114
BTM costs younger 65 yrs	100	108.4
BTM costs 65yrs or older	100	108.4
Car travel time off peak hours	100	100
Car travel time peak hours	100	100
Public transport travel time	100	100
Fraction of population	100% inside 4% outside	100% inside 4% outside
Ratio living in females of total	0.324	0.325
Ratio single females of total	0.172	0.164

It is expected that car use will become more efficient because the average fuel prices and the average fuel use per kilometer of car will be reduced, leading to decrease in car travel costs. Thus, car costs for off-peak and peak hours are defined by smaller value than the baseline scenario. According to price policy of the Netherlands, public transport costs will be increased.

Tariffs of train will be increased more during morning peak hours than rest of day. Lastly, the total population of study area is synthesized for the baseline and the 2020 scenario.

7.3 RESULTS

The sensitivity of the linked model is analyzed at two levels. First, Albatross application has a function of comparing activity-travel pattern from predicted schedules of multiple scenarios. The comparison of two scenarios (the baseline and the future scenario) is based on the reports that frequencies and mobility indicators are summarized for each scenario. The output of comparison contains information of percentage difference between the base year and the forecast year scenarios with significance of differences with t-statistics. Second, Omnitrans output database and animated bandwidths resulted from traffic assignment can be used to see how the transport network is used in different times of the day. Particularly attractiveness of traffic assignment is that congestion can be diagnosed by animated bandwidths and calculated loads and speeds during network loading.

7.3.1 Results from Albatross Activity-Travel Scheduler

The fraction levels (as shown in Table 7.1) for population synthesis create 1,315,546 households for the baseline and 1,416,878 households for the 2020 scenario (see Table 7.2). The number of households increases 7.7 % and the size of flows (the number of trips) increases 3.7%. The outputs of Albatross's prediction run can be represented in frequency tables summarizing synthetic population, mobility indicators, and activity-travel choice. Since mobility effect of the scenario is of the critical interest, however, the results of synthetic population and activity-travel choice are briefly described as summary and the mobility consequences are described in detail.
Assumed population changes in 2020 described in the previous section are well reflected in the results. Both household and individuals in household increase roughly 7%. At household level, single and double head household with no work status extensively increase. Increase in no work household implies growth of the older age groups. As it appears age year group older than 55 years increase while the young groups and household with children decrease. Economic status of household is improved as medium and high income classes mainly increase. Change in car possession of household is modest as population growth. At individual level, increased proportion of individuals are generally bigger than that of household, implying that 2020 is aging society in an absolute sense as age of living in household member also becomes older as household head does. There will be twice more male population than female population. In terms

of activity-travel choice, people in 2020 participates out-of-home activities less considering that population growth is bigger than increase of the frequency of trips. However, the out-of-home activities for the elderly strongly increase. While work and bring or get activities decreases in entire population and other types increase, the activity participation of the elderly particularly involves work and business purpose. And activity begin time tends to avoid peak hours as trips between 10 a.m. and 4 p.m. increase while before 10 a.m. largely decrease and it is modest for evening peak hours. However, the elderly begins there activity particularly before 10 a.m. and this may be due to their main activity participation of work or business. Activities are taken place at bigger city than home municipalities implying increase of travel distance.

Table 7.2 Predicted trip cases and size of flow

Time of Day	Base year (N sample = 1315546)		Scenario year (N sample = 1416878)	
	N Predicted cases	Size of total flow	N Predicted cases	Size of total flow
1	7835	11858	7000	10404
2	118066	426729	114499	400882
3	116554	384867	116935	379400
4	106453	248523	110439	261605
5	116890	342879	123013	374644
6	131392	364052	137495	394434
7	127421	359645	131639	377947
8	142962	374669	148447	399317
9	148577	443258	155207	479271
10	160048	459593	165122	485944
11	146434	468281	147609	476719
12	155320	470827	156404	478844
13	134348	417813	134864	426978
14	112767	321204	114523	333600
15	138787	503105	141354	526544
Total	1863854	5597303	1904550	5806533

Table 7.3 displays the mobility indicators of all cases. The first column represents the results for the reference scenario (m0) and the second for the scenario considered (m1). The third column shows the difference between the scenario considered (m1) and the reference as a percentage of the reference. m0 and m1 are the average of three randomly divided groups of entire data. The last column displays the significance level of the difference. Significance levels are relevant because predictions in Albatross are based on Monte Carlo simulation. The number of stars indicates the significance level of the t-value of an independent samples t-test. One star means the difference is significantly different from zero on a 5% alpha level and two stars means that the difference is also significant on a 2.5% alpha level. In 2020, people make less out-of-

home activities as the number of trip less increase than the number of household. However, the mobility is improved because the number of trip and total travel time with reference to total travel distance implies that trip length and travel speed are increased. Moreover, that increase of car travel distance is bigger than that of car travel time indicates that road infrastructure is added or improved. People use private car more frequently than public transport reflecting increased tariffs of public transport price policy discourage the use of public transport while car cost becomes cheaper as they were defined in the parameter settings.

Table 7.3 Mobility indicators of total flow

	m0	m1	m1-m0 (%)	sign
Total travel time	31695464	32461602	2.42	**
Travel time car driver	14838577	15806159	6.52	**
Travel time public	3736376	3304362	-11.56	**
Travel time slow	10019187	10300833	2.81	**
Travel time car passenger	3079393	3029239	-1.63	**
Number of tours	887823	927105.3	4.42	**
Number of trips	2016303	2103995	4.35	**
Ratio trips-tours	2.271	2.269	-0.07	**
Ratio single stop tours - all tours	0.803	0.803	0	
Total travel distance	26715494	28792302	7.77	**
Distance car driver	19653608	21870482	11.28	**
Distance car passenger	3813875	3756966	-1.49	**
Distance slow	1406805	1446802	2.84	**
Distance public	1841279	1718023	-6.69	**

7.3.2 Results from Omnitrans Network Loading

Predicted travel demand of Albatross is used for network loading in Omnitrans. Whereas travel information of Albatross is available at O-D relational level, it is at road segment level, for instance, the amount of car traffic flow on the transport network by times of day. This section discusses results of the scenario with Omnitrans output. The description is made for several routes reported to be congested road by Dutch organization. Route selection and general characteristics of the routes are presented, first. Omnitrans output relevant to the chosen route is analyzed. Useful figures for illustration of the scenario are derived from it. Having the figures, the last section discusses traffic changes in 2020.

7.3.2.1 Routes

While comparison between discontinuous independent links is appropriate to describe the effect of sample size on resulting traffic flow dealt in the previous chapter, the approach is not suitable for describing the differences that would occur under scenario situation. Since more behavioral aspect of traffic pattern is important in scenario analysis, say changes in traffic patterns of individuals in 2020, a route consisting of multiple links continuously connected to each other is unit of the analysis here. Investigating traffic flow changes of continuous links provides richer information than discontinuous independent links do in that traffic flow entering to the link on the route and exiting from the link can be implicitly known from the differences of link load between two adjacent links. In other words, we could know the amount of incoming or exiting traffics at junctions and interchanges. With the approach, moreover, bottleneck can be easily found where the link speed is significantly reduced and where the number of lane decreases, given speed and the number of lane variation of continuous links.

Routes are selected based on the traffic congestion map of the Netherlands provided by ANWB (The Dutch organization for drivers). Figure 7.1 illustrates the expected congested road of morning and evening peak hours. Morning peak hours are from 6:30 a.m. to 9:30 a.m. Evening peak hours are from 3:30 p.m. to 7:00 p.m. Congested road segment is displayed by colored band for the direction of flow. Red band for one side of the road means high chance of traffic congestion for a direction. Purple means very high chance of congestion.

Figure 7.1 Morning and Evening Peak Traffic Congestion of the Netherlands (from http://www.anwb.nl)

There are total 14 congested areas for both morning and evening peak hours relevant to the study area. The maps show dissimilar pattern of congested route between morning peak and evening peak hours with regard to congested location and direction of traffic flow. In the morning, the traffic flows mostly towards Amsterdam and Schiphol airport region. They are mostly from northern part of Noord-Holland such as Alkmaar by A9, Hoorn by A7, Zaandam by A8, from south eastern part, Almere by A6 and A1, from west, Haarlem by A9 to Schiphol airport, and southern part of Amsterdam by Ringweg A10. On the other hand, majority of evening peak flows are opposite direction to the flows during morning hours mostly at A1, A2, A9 and A10. Congested segments at A9 are from Schiphol to Haarlem and from Amstelveen to Diemen, Amsterdam-Zuidoost. At Ringweg A10, severe traffic congestion would occur at western A10, Einsteinweg passing Coentunnel, from Amsterdam-zuid to Coenplein, at southern A10, Ringweg Zuid and at southern east A10, a part of Ringweg Oost, up to intersection A1. A2 is very congested because traffics from A9 and A10 join A2 then moving towards Utrecht. A1 from interchange A10 to A9 is congested also. Each congested routes are summarized in Table 7.4 in terms of the number of links consisting, total distance, highway number, and short description of the route. Total distance is sum of link lengths of the consisting links in Omnitrans.

Table 7.4 Congested road segments of study area

Peak hours		Nr. Links	Distance	Highway	Route Description
Morning peak (6:30 – 9:30)	1	2	7.06 km	A9	Alkmaar to south
	2	8	11.18 km	A9	A22-A200-A5
	3	7	4.99 km	A10	Ringweg Zuid to east
	4	5	4.43 km	A10	Ringweg Zuid to west
	5	9	4.25 km	A8	Joint A7 - Coenplein
	6	11	15.41 km	A7	Hoorn - Purmerend
	7	14	14.53 km	A6	Almere to west
	8	20	15.79 km	A1	Naarden to north
Evening peak (15:30-19:00)	9	11	14.66 km	A9	Schiphol - Haarlem
	10	13	9.27 km	A10	Einsteinweg, Ringweg 10
	11	15	8.75 km	A10	Ringweg Zuid - Ringweg Oost
	12	20	14.72 km	A9	Amstelveen – Amsterdam-Zuidoost
	13	9	11.42 km	A2	Amsterdam to south
	14	6	4.44 km	A1	Amsterdam to south

7.3.2.2 Analysis of Omnitrans output

Omnitrans output can be summarized in terms of link load and speed for each link and for each times of day for a selection of routes/links and times. Table 7.5 shows an example output table of one of the routes. Let's denote the load and speed score in the table as T_{ijk} and V_{ijk}, respectively, where i refers to the link, j to time of day, and k to the scenario. Based on this original Omnitrans output, several figures useful to derive implication are available by simple calculation.

First, the figure describes traffic changes on average, say, difference of average traffic flow by times of day within the scenario and between the two scenarios. Summing up link load of all links on the route for each time of day for baseline and scenario year and then dividing the total loads of the route by the number of links, the average load is calculated. The average speed can also be obtained with the same approach. The average flow data is available by time of day dimension and thus temporal distribution of average flow can be known. In addition, the percentage increase or decrease of average load and speed between 2000 and 2020 can also be calculated. It is expected that the amount of traffic flow change between 2000 and 2020 would varies according to times of day as shown in the distribution of activity begin time in Albatross. Moreover, regional characteristics along the route such as functional location (e.g. land-use of the area) may represent dissimilar pattern in different routes.

While the first one concerns change of average flow between 2000 and 2020 by different times of day at a route level, the other two figures enable more detailed description of the scenario. They account for changes between the two connected links in 2000 and 2020, based on differences between individual scores of the data. As can be seen in Table 7.5, speed variation of links (V_{ijk}) along the route appears. Speed variable is very useful to diagnose traffic congestion. Especially where travel speed reduces a lot, reasons can be found from several factors. Decrease of the number of lanes causes congestion as bottlenecks on road. In addition to the number of lanes of link, the congestion would occur when many vehicles enter to the link of the route. Characteristics of surrounding area, for instance, land-uses, have influence on the amount of traffic using the route.

Table 7.5 Summary of route data from Omnitrans output (Example route: Ringweg Zuid)

Ringweg Zuid to west						
Link id		107321	112396	109863	107300	107282
2020						
Load (T_{ijk})	6am-8am	4331.32	2912.87	4769.13	4589.85	5160.35
	8am-9am	4723.11	2856.96	5241.92	4976.4	5995.99
	9am-10am	3556.9	2211.41	3858.12	3508	4121.4
Speed(V_{ijk})	6am-8am	98.19	102.65	99.8	99.73	96.83
	8am-9am	97.07	102.7	97.98	98.42	92.67
	9am-10am	99.7	103.06	102.16	102.08	100.06
2000						
Load	6am-8am	5725.31	4099.27	6099.35	5729.34	6182.49
	8am-9am	6286.8	4354.57	6951.72	6559.87	7340.32
	9am-10am	3601.06	2347.83	4070.85	3799.64	4269.24
Speed	6am-8am	92.88	100.94	93.47	95	91.53
	8am-9am	89.61	100.32	87.39	89.8	82.85
	9am-10am	99.64	103	101.73	101.61	99.72

Lastly, the difference of link load between two consecutive links is available from the Omnitrans output. Since the chosen route consists of straight links without any direction changes of the route, the difference of loads only can be found at junctions, interchanges, or intersections. The percentage difference between the following link and the current link $((T_{2jk} - T_{1jk}) / T_{1jk}*100)$ is actually a proportion of the number of outgoing or incoming flow to the route flow. Let'sdenote the calculated difference as D_{njk} where j is time of day, k is scenario year, and n refer to n-th junction on the route. We can calculate the percentage difference between the links for 2000 and 2020. Simply subtracting the value of 2000 (D_{nj2000}) from that of 2020 (D_{nj2020}), the difference by the year can be known. The bigger the absolute level of the difference ($|D_{nj2020} - D_{nj2000}|$), the more changes in proportion of incoming or exiting flow would take place between 2020 and 2000. As in the speed variation of connected links, the explanation is possible in relation to surrounding region of the route.

Even though the above mentioned three figures seem to be useful as each of them implies different aspects of traffic pattern, only the first one, the average traffic flow on a route level, is sufficient to explain the system sensitivity. As noted earlier, the purpose of this section is the demonstration of the linked system rather than having interest in mobility consequences for practice. The description on link level concerning the second and the third figures are too much in detail beyond the scope of the analysis because the figures are more case-specific than the first one as it is largely affected by circumstances such as location and characteristics of roads. Thus, regarding the two figures, it is just enough to mention what they mean and how they can be calculated.

Changes in average flow and speed of the routes are described below. Table 7.6 and Table 7.7 represent the average load and the average speed across all links on the route and percentage difference between 2000 and 2020 by times of day for morning and evening peak hours. Since different segments of road are congested during morning and evening, they are separately explained. Similar to the peak hours in the ANWB map, the output of classes between 6 a.m. – 10 p.m. for morning and 3p.m. – 7p.m. for evening are used.

Traffic simulated in Omnitrans supports activity begin time pattern of aging population predicted in Albatross. They are closely related. However, activity begin time distribution and traffic flow distribution in Omnitrans are not the same in the sense that they are activity timing and traffic timing. Furthermore, neither origin nor destination is known for vehicles on the transport network. And thus, departure time of vehicles is unknown in Omnitrans.

7.3.2.2.1 Morning peak hours traffic pattern

Table 7.5 is summary table of average load and speed of baseline and scenario. Total average flow of entire route is summarized in table margin. The figures shown in the last couple of rows of the table are summation of average load by time of day across routes and the total during morning peak hours, describing the overall traffic flow. Since load is the number of vehicles on the route, summated value can say the amount of traffic increase or decrease. However, average speed of all routes by time of day is proper. The overall traffic flow decreases about 2.5%. According to temporal distribution of traffic flow, traffic before 9am when commuting trip is mostly taken place decreases while afterwards hour traffic increases. The decrease of loads is bigger during the earliest morning peak hour, 6am-8am (-7.88%) than the decrease during 8am-9am (-4.97%). Traffic after 9am increases about 11%. Speed variation by times of day inversely related to load. If average load increases, average speed tends to decrease. The extent of average speed change is not as much as load changes and there is no linear relationship between the two since density of link relating link capacity and load concerns. For example, A6 passing Almere, speed change is moderate even though traffic flow extensively increases almost 80% during 9am-10am. The density of the route is not high as the load is not enough to capacity of road. As average speed is inversely related to load, speed of flow increases before 9am and slightly decreases after 9am.

According to the output data, it is difficult to conclude that traffic change would generally the same to all routes although description of the overall pattern is made above. Traffic change is different from the route to the route. Traffic change on A6 is significantly different from the overall pattern. Traffic increases as big as 30% during all peak hours. Traffic flow on A9 is similar to the overall pattern. However, the amount of traffic increase is bigger in Haarlem than in Alkmaar as the city Haarlem has more population and serves as bedroom community where many workers to center of Amsterdam reside. Traffic on A8 and A7 is in line with the overall pattern. On the southern part of ring road A10, the data is available for both directions. Traffic flow of each direction is different. While the flow from intersection A4 to A2 is similar to the overall pattern, the opposite direction of flow decreases a lot. Dissimilarity of traffic change of route may due to characteristic of route such as route direction and location. When the scenario is considered for what-if analysis, the individual differences among routes need to be investigated with factors influence it.

Table 7.6 Summary of Traffic Morning Peak Hours

Route	Time of Day	Average Load and Difference			Average Speed and Difference		
		2000	2020	(%)	2000	2020	(%)
A9 Alkmaar	6am-8am	4223.21	4017.28	-4.88	89.42	93.75	4.84
	8am-9am	3495.12	3579.62	2.42	103.41	102.00	-1.36
	9am-10am	2150.21	2389.32	11.12	117.30	115.94	-1.16
	Total	9868.54	9986.22	1.19			
A9 A22 – A5	6am-8am	4592.00	4410.92	-3.94	94.07	96.82	2.93
	8am-9am	3799.68	3997.98	5.22	106.10	103.76	-2.20
	9am-10am	2486.90	2887.31	16.10	116.32	114.17	-1.85
	Total	10878.57	11296.21	3.84			
A10 Ringweg Zuid A4 – A2	6am-8am	3600.56	3581.38	-0.53	98.31	98.47	0.16
	8am-9am	4568.02	4126.17	-9.67	94.94	96.75	1.91
	9am-10am	3032.41	3197.07	5.43	99.20	99.06	-0.14
	Total	11201.00	10904.62	-2.65			
A8 Coenplein	6am-8am	5867.38	5489.04	-6.45	77.61	81.14	4.55
	8am-9am	4840.96	4740.24	-2.08	84.74	85.50	0.90
	9am-10am	2975.06	3297.05	10.82	95.05	93.94	-1.17
	Total	13683.40	13526.33	-1.15			
A7 Hoorn – Purmerend	6am-8am	4112.29	3755.34	-8.68	89.75	96.54	7.57
	8am-9am	3352.30	2952.01	-11.94	103.04	108.17	4.98
	9am-10am	1906.38	2158.06	13.20	115.87	114.82	-0.90
	Total	9370.97	8865.40	-5.40			
A10 to west A2-A4	6am-8am	5567.15	4352.70	-21.81	94.76	99.44	4.93
	8am-9am	6298.66	4758.88	-24.45	89.99	97.77	8.64
	9am-10am	3617.72	3451.17	-4.60	101.14	101.41	0.27
	Total	15483.53	12562.75	-18.86			
A1 Naarden – Amsterdam	6am-8am	5387.03	4238.14	-21.33	86.07	92.92	7.96
	8am-9am	5460.18	4941.79	-9.49	86.02	89.06	3.54
	9am-10am	2895.36	2969.31	2.55	98.62	98.37	-0.25
	Total	13742.57	12149.25	-11.59			
A6 Almere	6am-8am	3528.35	4128.81	17.02	109.47	105.07	-4.01
	8am-9am	3290.59	4263.28	29.56	110.91	104.30	-5.96
	9am-10am	1263.49	2271.57	79.79	114.89	114.03	-0.75
	Total	8082.43	10663.66	31.94			
All routes	6am-8am	36877.97	33973.61	-7.88	92.43	95.52	3.34
	8am-9am	35105.50	33359.97	-4.97	97.39	98.41	1.05
	9am-10am	20327.54	22620.85	11.28	107.30	106.47	-0.77
	Total	92311.01	89954.43	-2.55			

7.3.2.2.2 Evening peak hours traffic pattern

Table 7.7 contains the same information as Table 7.6. Unlikely to morning peak hour's traffic, traffic during evening peak hours generally increases in 2020 by 1.3%. The temporal change (times of day) of traffic flow in general, there is decreases in traffic during 4pm-6pm when the most commute to home takes place while traffic increases before 4pm and after 6pm. The temporal distribution demonstrates the Albatross result that people tend to avoid trips at off-peak hours in 2020. As the load has inverse relationship with speed, traffic flows at decreased speed than in 2000. The average speed of route during the evening peak is lower than during the morning peak, meaning more traffic congestion in the evening.

Evening peak hour pattern is different from the route to the route. However, the extent of deviation among the routes is not so serious during the evening peak hours than the morning peak hours. A probable reason to this is that traffic flow is influenced by spatial location. Large amount of traffic in the morning much more involves travel for obligatory activities that destination of trips are rather limited, for example, business district at the core of the city. Since diverse activities are conducted in the evening in addition to commuting to home, such as grocery shopping or social meeting with friends, traffic change during the evening may not be affected by location as much as in the morning. Congested routes are generally found close to center of Amsterdam showing that Amsterdam is the place where many out-of-home activity facilities are densely located. Although the overall pattern displays decreased traffic flow during the most peak hours between 4pm and 6pm, traffic flow during that hour increases for most of the routes. Especially, A2 from Amsterdam to south towards Utrecht there is quite large decrease of flow during 4pm-6pm. A1 from Amsterdam to south towards Muiden, traffic also decreases after 6pm while all other routes have increased traffic after 6pm.

Table 7.7 Summary of Traffic Evening Peak Hours

Route	Time of Day	Average Load and Difference			Average Speed and Difference		
		2000	2020	(%)	2000	2020	(%)
A10 Einsteinweg to Coenplein	3pm-4pm	4866.34	5137.39	5.57	87.28	85.24	-2.34
	4pm-5pm	5668.65	5825.95	2.77	80.18	79.07	-1.38
	5pm-6pm	5828.25	5880.94	0.90	79.66	78.96	-0.89
	6pm-7pm	5169.44	5305.61	2.63	87.13	85.03	-2.41
	Total	21532.69	22149.88	2.87			
A10 Ringweg Zuid - Ringweg Oost	3pm-4pm	5111.12	5473.59	7.09	93.79	92.08	-1.82
	4pm-5pm	5916.05	5795.51	-2.04	89.23	90.23	1.11
	5pm-6pm	6220.01	6254.31	0.55	87.29	87.26	-0.03
	6pm-7pm	5117.13	5396.53	5.46	93.72	92.14	-1.69
	Total	22364.32	22919.94	2.48			
A1 Amsterdam – Muiden	3pm-4pm	6697.27	7554.27	12.80	77.82	68.81	-11.57
	4pm-5pm	7293.98	6821.69	-6.48	71.63	77.06	7.57
	5pm-6pm	7332.15	7518.89	2.55	71.10	69.59	-2.11
	6pm-7pm	6413.50	6011.38	-6.27	80.21	84.43	5.26
	Total	27736.90	27906.22	0.61			
A9 Schiphol to north	3pm-4pm	4044.98	4132.73	2.17	94.67	93.44	-1.30
	4pm-5pm	4818.48	4955.58	2.85	83.77	81.87	-2.26
	5pm-6pm	4934.29	5119.45	3.75	82.26	79.40	-3.48
	6pm-7pm	4287.29	4673.55	9.01	91.71	86.10	-6.12
	Total	18085.04	18881.31	4.40			
A2 to south	3pm-4pm	8271.13	9246.94	11.80	78.70	71.03	-9.74
	4pm-5pm	9334.15	8282.42	-11.27	70.31	78.47	11.60
	5pm-6pm	8927.86	8013.51	-10.24	73.65	80.53	9.35
	6pm-7pm	8437.21	8731.19	3.48	77.34	75.04	-2.97
	Total	34970.34	34274.06	-1.99			
A9 Amstelveen – Amsterdam Zuidoost	3pm-4pm	2844.12	3229.38	13.55	101.95	97.72	-4.15
	4pm-5pm	3448.98	3440.74	-0.24	93.24	94.98	1.87
	5pm-6pm	3691.06	3801.93	3.00	89.33	89.77	0.49
	6pm-7pm	2911.62	3139.41	7.82	100.71	98.74	-1.96
	Total	13259.64	13611.45	2.65			
All routes	3pm-4pm	31829.30	34774.29	9.25	88.99	84.72	-4.79
	4pm-5pm	36559.30	35121.88	-3.93	81.37	83.61	2.76
	5pm-6pm	37068.32	36589.02	-1.29	80.44	80.92	0.60
	6pm-7pm	32492.00	33257.65	2.36	88.29	86.91	-1.56
	Total	137948.92	139742.85	1.30			

7.4 IMPLICATIONS AND DISCUSSIONS OF RESULTS

The main objective of the chapter is to see whether the linked model system is valid. First, the linkage as an operational model has established. The scenario is applied and the output varies reasonably to the assumptions of the scenario. Not only the output looks like the expectation in light of the assumptions but also Omnitrans result makes sense based on Albatross result. Moreover, the connection between activity timing in Albatross and traffic timing in Omnitrans has found, and consequently demonstrating the validity of the linked model system. Since traffic loads by time of day do not have any clue when the vehicle departed and from where, activity begin time is not directly linked to Omnitrans output. According to Albatross output, however, more than half of trip cases have their activity location near home or within home municipality, implying that travel time is not so long. Activity begin time is quite close to the time when trip has made. Therefore, it can be said they are closely related and the output data verifies it. Second, future application of this linked system is discussed. Several figures explaining diverse aspect of traffic pattern are introduced in the result section. The description of the results is made only with one of the figures because the purpose of this chapter is to reveal the system works properly and less interests in consequences of the scenario. Among unexplained figures, investigation of speed variation of continuous link will be very useful for planning perspectives, in particular. That speed variation is known is the most attractive feature of traffic assignment in the sense that traffic congestion by times of day can be diagnosed. Likely to animated bandwidth displaying the extent of congestion by ratio of calculated speed and free flow speed, congestion can be analyzed if there is any predefined value representing the extent of congestion. The variation of traffic flow at the link level enables very sophisticated explanation and therefore it will be valuable to transport policy making or designing in practice.

7.5 CONCLUSION

The chapter concerns validation of the integrated model and illustration of the system for application. Scenario analysis is adopted here to test the system. The test was to evaluate to what extent the objectives of building the linked model system introduced in the introduction chapter of the thesis are achieved. The consistency of Albatross travel demand with Omnitrans network use of vehicles verifies the establishment of the operational linkage between the two models and applications. Illustration of result shows the possibility of the system application in practice. Useful information can be derived from the original Omnitrans output. Especially speed of links can explain traffic congestion. In overall, the linked model system works properly as the system is sensitive to scenarios.

8 CONCLUSIONS AND DISCUSSIONS

8.1 RESEARCH OBJECTIVES

The research objectives are briefly mentioned again to discuss what this research has contributed, which goals have been accomplished, and to what extent this research has added value in transport. The central aim of the research is identical to the thesis title which is, "Embodying dynamics into transport modeling: Linking Albatross and MaDAM" As the recognition of traffic is truly dynamic in nature, the research attempted to add dynamics to transport modeling by bridging the models. Making the comprehensive operational system beyond conceptual integration of the two approaches always accompanies system compatibility problems as the major essential component to solve. Many chapters of the thesis dealt with the methodological way to achieve that compatibility. Having provided the solution for compatibility, examination of system performance is a critical step in model building. System performance was discussed in terms of variety aspects of the linked system.
Research objectives generated several research questions aiming at satisfying the objectives. The specified research questions are:

- How to handle study area travel demand in Albatross national model?
- What are data requirement for each modeling component of the linked system?
- Is predictability of the model affected by sample fraction size?
- How does the linked model system perform?

8.2 RESEARCH CONTRIBUTION

The aforementioned research questions were dealt with in separate chapters. Solutions were developed for the compatibility problem and the analysis was conducted for the system validation. This section summarizes the major conclusions of each chapter, discusses whether the research questions are solved, and what the added value of the linked model system is.

The major conclusions can be summarized as follows.

i) First, problems accompanying the study area delineation were discussed. Albatross was extended to be able to handle travel demand of a smaller, local study area. Originally, Albatross is a national model which predicts activity-travel behavior of the whole Dutch population. However, traffic assignment is conducted for a smaller study area to reduce computational demands. Omnitrans requires travel demand of a study area as input for traffic assignment. Traffic of a study area always includes external trip flows and those external flows are assumed to be constant by convention in Omnitrans even when scenarios of change are considered. Albatross is extended to be able to delineate trip matrices for smaller local areas where trip flows entering and leaving the area are defined as the margins of the table. This is not modification but rather an adaptation of the form of the output because external trip prediction is a post-processing step after travel demand is predicted at a national level by Albatross. A path finding algorithm, which identifies where the route of trips between predefined external origin and destination would pass the area, is developed. Thus the entry and exit points of external trips are predicted and incorporated in the study area O-D trip table. The performance of the algorithm was examined by comparison of identified entry or exit points by the route planner in Google maps. Under the assumption that suggested routes by Google is used in reality by most drivers, it turned out that the path finding algorithm works reasonably well and therefore, the algorithm is used to generate study area O-D trip tables. By the help of the algorithm, Albatross now can handle local study areas.

ii) Second, implementation of the linked model system is discussed. Making an operational model concerns identification of modeling tasks and databases and data conversion of output to new input. Albatross predicts travel demand and Omnitrans does traffic assignment. Albatross first generates a synthetic population for simulation and then activity-travel behavior of households is predicted. Another extension of Albatross concerning population synthesis is setting of different fraction sizes for study area and outside study area for saving computation time. What Albatross delivers to Omnitrans is O-D trip matrices extracted from the predicted activity schedules. Data conversion or formatting mainly takes places here. First, Albatross generates trip matrices structured into Omnitrans required format. Moreover, zone mapping that matches the two different zoning systems (i.e., zone centroids ids) of Albatross and Omnitrans is executed. Lastly, Albatross zone ids in Albatross trip table are changed to those of Omnitrans based on zone mapping files before importing to Omnitrans. The adjustment was necessary to link the systems technically.

iii) Whereas the previous two chapters concern preparatory process to build the operational system, the remaining two chapters discussed the system performance based on results from running the linked model system. The impact of sample fraction size on resulting traffic patterns was examined. It is hypothesized that a small fraction size, which is reported to be not problematic for describing long-term characteristics of travel demands, may not be able to generate a representative picture of traffic flows on the road network. The reason behind the doubt is that traffic is analyzed at more microscopic and disaggregated level as temporal change of traffic behavior within the day is the focus of the approach adopted in the research. To see how resulting traffic patterns vary with regard to sample size, multiple projects were run with a set of different fraction sizes. It is expected that more random variation exists for smaller sample size. Thus, Omnitrans output, particularly link loads, is analyzed by a test of equality of variance (F-test) as the extent of variation was of interest. The results of the statistical testing confirmed that sample size influences traffic flows and that the traffic flows of local road links are more sensitive to sample size than that of major routes like highway links. Although significance testing was conducted, the interpretation of results must still be cautious because the difference in size of variance, even though statistically significant, may still be small and acceptable in practice.

iv) Lastly, the performance of the linked model system is validated. Model validation is a crucial component of model building. A scenario analysis was conducted to test the face-validity and sensitivity of the linked model system. It is emphasized that the purpose of the scenario analysis is to examine whether the system responds properly to varying assumptions rather than to describe mobility consequences of the scenario. The linkage is established as it is shown that Albatross activity begin time distribution and Omnitrans loads by time of day are closely linked: travel-demand of Albatross is appropriately reflected in the transport-network use resulting from traffic assignment. Moreover, a possible application of the linked model system is presented with useful information that can be drawn from the original Omnitrans output. In particular, analysis of traffic congestion by time of day is one of the attractive features of the linked model system that users can entertain.

Overall, each chapter reasonably answered the research questions specified in the introduction chapter. Even though the constructed linked model system still has some limitations as described in the next section, the research fairly well solved the problems. The most important objective of the research, building the linked model system, was achieved. The modeling schemes with data requirements including conversion method were formalized. The structured system can be further elaborated. Moreover, the illustration of the application of the linked system demonstrated the face validity of system and demonstrated the potential use of the system.

8.3 SUGGESTIONS FOR FUTURE RESEARCH

The discussion about accomplishment of the research suggests future research issues. First, individual differences in choice need to be considered. We may question that people always use shortest travel time path for their travel. Travel time matrices are involved for scheduling activities in Albatross and route choice mechanism adopts its methodology from fluid mechanics. For determining entry and exit points it was assumed that suggested routes in Google maps are the routes that drivers follow in reality. As the suggested routes literally mean a recommendation, actual route choice may not be guided by minimizing travel time which is the assumption of the path finding algorithm. More importantly individual's route choice is not as simple as the water flows through water pipe, as assumed in the traffic assignment model. For more sophisticated modeling, individual differences should be considered.

Second, dynamic traffic assignment appeared to be problematic as gridlock problems occurred. Therefore, static traffic assignment by time moment of the day was used to generate Omnitrans' output of link loads and link speeds. However, this does not mean that the research fails to add dynamics to transport modeling. Static traffic assignment is disaggregated by time of day and thus variation of traffic conditions on the transport network within the day is described. Moreover, the linkage has been established by solving compatibility problems, satisfying the central objective of the research. Although it is not a problem of the linkage of the two models, finding the solution for the gridlock problem within the traffic assignment model is suggested for further research.

Traffic modeling is generally an iterative process where travel demands are constantly updated based on the use of transport network and vice versa. Since building the linked model system by identification of modeling components and by data conversion was the primary interest of the research, iterative loops have not been considered in model building. The feedback loop has to be added to the linked model system so that the system becomes truly dynamic. The reverse way of data conversion from Omnitrans to Albatross is involved for adding the iterative loop.

Lastly, the linked model system currently is merely an integration of several model components by putting them in a certain sequence of execution. The linked model system is now operational but it is not a complete system where each model component is executed with easy clicking of icon in single application software. The prediction module of Albatross is run outside as a stand-alone application. Since it is not embedded in the application, the connection is realized through input and output files. The users not familiar with the technical details of these files will have difficulty in performing the modeling task. Furthermore, detailed Omnitrans traffic assignment output can only be seen with a database inspector which is not a documented feature of the application. The general form of displaying the traffic assignment result is animated bandwidths. Although the visual representation helps understanding the traffic patterns, detailed investigation of the patterns requires more specific numerical data as well. Exporting results of interest is guided by SQL queries. For general use of the linked model system, the system should be easy to use. Therefore, the system should be executed in a single application or every task must be done with an overall system interface. As the next step of the research, making such a system-user interface is proposed.

9 Bibliography

1. Anderson M.D. (1999) Evaluation of Models to Forecast External-External Trip Percentages, *Journal of Planning and Development*, Vol. 125, Issue 3, 110-120

2. Arentze T.A. (2008) *ALBATROSS 4.0 Manual*, European Institute of Retailing and Services Studies, Eindhoven, The Netherlands

3. Arentze, T.A. (2008) Application of ALBATROSS to aging and mobility scenarios for the year of 2020. Research Report, Eindhoven: European Institute of Retailing and Services Studies, September, 2008

4. Arentze T.A. and H.J.P. Timmermans (2000) *Albatross: A learning-based Transportation Oriented Simulation System*. European Institute of Retailing and Services Studies. Eindhoven, The Netherlands

5. Arentze T.A. and H.J.P. Timmermans (2001) Creating Synthetic Populations: Approach and Empirical Results, *Association for European Transport*

6. Arentze T.A. and H.J.P. Timmermans (2005) *Albatross version 2: A learning-based Transportation Oriented Simulation System*. European Institute of Retailing and Services Studies. Eindhoven, The Netherlands

7. Arentze T.A. and H.J.P. Timmermans (2008) ALBATROSS-4: Description of the Model and Empirical Derivation from MON data, Research report, Eindhoven University of Technology, The Netherlands

8. Arentze, T.A., H.J.P. Timmermans, P. Jorritsma and M-J. Olde Kalter (2008) More Gray Hair – But for Whom? Scenario-Based Simulations of Elderly Activity Travel Patterns in 2020. *Transportation*, 35, pp. 613-627.

9. Axhausen K.W. (1995) Data Needs of Activity Scheduling Models, *Activity Based Approaches to Travel Analysis*, Pergamon

10. Beckman, R.J., Baggerly K.A. and McKay M.D. (1996) Creating Synthetic Baseline Populations, *Transportation Research A*, 30, 415-429

11. Ben-Akiva M.E. and J.L. Bowman (1996) Activity Based Travel Demand Model Systems, Proceedings of *Equlibrium and Advanced Transportation Modelling*, Kluwer, Montreal, Quebec, 27–46

12. Bowman J.L. (1998) The day activity schedule approach to travel demand analysis, Ph.D Thesis, Massachusetts Institute of Technology, Cambridge, The United States

13. Bowman J.L. and M.E. Ben-Akiva (1996) Activity-based travel forecasting, Paper presented at the Activity-Based Travel Forecasting Conference, New Orleans, Louisiana

14. Bowman J.L. and M.E. Ben-Akiva (2001) Activity-Based Disaggregate Travel Demand Model System with Activity Schedules, *Transportation Research Part A*, 35(1), 1-28

15. Bliemer M.C.J. (2001) Analytical Dynamic Traffic Assignment with Interacting User-classes: Theoretical Advances and Applications using a Variational Inequality Approach, Ph.D thesis, Delft University of Technology, The Netherlands

16. Bovy P.H.L, M.C.J. Bliemer and P.C.H. Opstal (2006) CT 4801 Transport Modeling, Lecture notes, Delft University of Technology, The Netherlands.

17. Cullen I. and Godson V. (1975) Urban networks: The structure of activity patterns, *Progress in Planning*, 4, 1-96

18. Daly A.J., van Zwam, H.H.P., and van der Valk, J. (1983) Application of Disaggregate Models for a Regional Transport Study in the Netherlands, *World Conference on Transport Research*, Hamburg

19. De Romph E. (1994) A Dynamic Traffic Assignment Model: Theory and Applications, Ph.D Thesis, Delft University of Technology, the Netherlands

20. Deming W.E. and Stephan F.F. (1940) On a least squares adjustment of a sampled frequency table when the expected marginal tables are known, *Annals Mathematical Statistics*, 11, 427-444.

21. Ettema D. (1996) Activity-Based Travel Demand Modeling, Ph.D Thesis, Eindhoven University of Technology, The Netherlands

22. Ettema D. and H.J.P. Timmermans (1997) *Activity Based Approaches to Travel Analysis*, Pergamon

23. Ghareib A.H. (1996) Different Travel Patterns: Interzonal, Introzonal, and External Trips *Journal of Transportation Engineering*, Vol.122, 67-75

24. Hensher D.A. and K.J. Button (2002) *Handbook of transport modelling*, Pergamon

25. Jansen G.R.M. and T. van Vuren (1989) Travel Patterns in Dutch Metropolitan Cities: The Importance of External Trips, *Transportation* 15, 317-336

26. Jones, P.M., Koppelman, F.S. and Orfueil. J.P (1990) Activity analysis: State-of-the-art and future directions. *New Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Gower Publishing, Aldershot, England, pp. 34–55

27. Martin W.A. and N.A. McGuckin (1998) NCHRP Report 365: Travel Estimation Techniques for Urban Planning, *Transportation Research Board*

28. McNally M.G. and C. Rindt (2008) The Activity-Based Approach, Center for Activity Systems Analysis, Paper UCI-ITS-AS-WP-07-1

29. Merchant D.K. and G.L. Nemhauser (1978a) A Model and an Algorithm for the Dynamic Traffic Assignment Problems, *Transportation Science*, 12(3), 183-199

30. Merchant D.K. and G.L. Nemhauser (1978b) Optimality Conditions for a Dynamic Traffic Assignment Model, *Transportation Science*, 12(3), 200-207

31. Miller H.J. (2004) "Activities in Space and Time" P. Stopher, K. Button, K. Haynes and D. Hensher (eds.) *Handbook of Transport 5: Transport Geography and Spatial Systems*, Pergamon/Elsevier Science.

32. Miller E.J. (2005) Propositions for Modeling Household Decision-making, Integrated Land-Use Transportation Models: Behavioural Foundations, Elsevier, 21-59

33. Norman, P. (1999) Putting Iterative Proportional Fitting on the Researcher's Desk. Working Paper. School of Geography , University of Leeds.

34. Kitamura R., C.V. Lula and E.I. Pas (1993) AMOS: An activity-based, flexible and truly behavioral tool for evaluation of TDM measures. Proceedings of the 21st Summer Annual Meeting: Transportation Planning Methods, PTRC Education and Research Services, Ltd., London, 283-294

35. Kitamura R. and S. Fujii (1998) Two Computational Process Models of Activity-Travel Behavior, Proceedings of Theoretical Foundations of Travel Choice Modeling, 251-279

36. Lam W.H.K. and Y.Yin (2001) An Activity-based Time-dependent Traffic Assignment Model, *Transportation Research* Part B, 35, 549-574

37. Lin DY, N. Eluru, S.T. Waller and C.R. Bhat (2008) Integration of Activity-Based Modeling and Dynamic Traffic Assignment, *Transportation Research Record*, (forthcoming)

38. Ortúzar J.D. and L.G. Willumsen (2001) *Modelling Transport*. John Wiley and Sons

39. Peeta S. and A.K. Ziliaskopoulos, (2001) Foundations of Dynamic Traffic Assignment: The past, the Present and the Future, Networks and Spatial Economics, 1, 233-265

40. Smith L., R. Beckman, K. Baggerly, D. Anson and M. Williams, (1995) TRANSIMS: Transportation Analysis and Simulation Systems, Proceedings of the 5th National Conference on Transportation Planning Methods Application, 2, Transportation Research Board, Washington, D.C.

41. Timmermans H.J.P, T.A. Arentze and C.H. Joh (2002) Analysing Space-Time Behaviour: New Approaches to Old Problems, *Progress in Human Geography*, 26(2), 175-190