ALBATROSS: overview of the model, application and experiences[1]

Theo Arentze, Harry Timmermans

Eindhoven University of Technology
EIRASS
PO Box 513
5600 MB Eindhoven
The Netherlands
Email: eirass@bwk.tue.nl

Abstract. Albatross is an activity-based model of travel demand that has been developed over the last decade for the Dutch Ministry of Transportation. As a key characteristic, the model uses if-then rules for making decisions in a priority-based scheduling process which are empirically derived from activity diary data by using a decision tree-induction method. The model is fully operational on a national scale and first applications have been carried out. In this short paper, we give an overview of the model and discuss some application issues and experiences.

1. Introduction

Albatross, acronym for A Learning Based Transportation Oriented Simulation System, was developed for the Dutch Ministry of Transportation and Public Works, who decided to develop an activity-based model alongside their state-of-the-art tour-based model LMS. Originally, the project was part of the research program of the Ministry and not part of immediate policy application. Therefore, the project was formulated to serve two main goals. First, to explore the potential of an activity-based approach to travel demand modeling, and secondly to explore the potential of a rule-based model of individual activity scheduling decisions. The fact that a choice was made for a rule-based system as opposed to a discrete choice and hazard-type modeling approach did not reflect the fact that the people using the model were not satisfied with these modeling approaches, but simply a desire to explore alternative options.

Over the years, Albatross was developed in successive stages. In every stage, particular components of the model system were improved or generalized and several empirical tests were conducted. Version 1 (Arentze and Timmermans, 2000) was based on a limited data set, involving approximately 3,000 person-days, collected in the South Rotterdam region. This study primarily was designed to assess the potential of an activity-based approach and the Albatross framework in particular. The quintessence of the approach however never changed and most key components were developed during this first project. In Albatross 1, similar to most activity-based

---

[1] Paper prepared for the Innovations in Travel Modeling 2008 Conference, Portland, Oregon, June 22 - June 24.

models, fixed activities were taken as observed and used as anchor point of the activity skeletons.

Although this decision is acceptable if the primary aim is to assess the feasibility of the modeling approach, it is a limitation for prediction. Therefore, in Version 2, the activity skeletons were generated. Moreover, the decision rules of the model were re-induced using a national level, pooled data set of approximately 10,000 person-days. The shift from a regional to a national level also implied that a population synthesizer was developed. Also, a first application, related to congestion pricing was developed. Because such scenarios imply traveler response to external policy, an approach that linked stated response data to changes in the activity skeleton was developed. It serves as a general approach to address such problems and use the Albatross system for policies for which historical data do not exist. These extensions are described in Arentze and Timmermans (2005).

The application to congestion pricing demonstrated that rule-based models lack the detail especially for continuous variables and cannot produce elasticities satisfactory. Therefore, for Version 3, the principle of what we called Parametric Action Decision Trees was developed (Arentze and Timmermans, 2007). It meant that Albatross can compete with discrete choice models in this respect and can generate time and price elasticities and utility-based welfare measures.

In the meantime, the new travel survey in the Netherlands became available (MON). Although this is not an activity diary, but rather a conventional travel survey, the data could be transformed into an activity diary format and thus be used as input for Albatross. Albatross was therefore re-estimated on this national travel survey data set, involving 45,000 person-days.

The development of the model system went hand-in-hand with several studies and applications that allowed us to better judge the (relative) performance of the model. The first data set was used to compare the goodness-of-fit of Albatross with that of a state-of-the-art nested logit model (suite of linked models as a nested logit model was not supported by the data) and PCATS, the Japanese version of Famos. It turned out that the activity-scheduling models outperformed the competing model but that the ranking between them is unclear or dependent on the criterion considered; it also turned out that fair comparisons of completely different models are quite difficult. We also conducted a spatial transferability study, and found that the model is sufficiently robust to transfer decision rules derived from one region in the Netherlands to other regions. Having said that, we also learned that the rules driving destination choice are most difficult to transfer and over the years we developed three different approaches before a robust set of rules were derived.

Over the years, several scenarios were input to the model system, primarily to assess the sensitivity and face validity of the model system. Several of these tests were conducted by students, not familiar with the details of the system. Similarly, several external validation tests were conducted, comparing outcomes of Albatross with national trip and time-use databases. These studies provided further support to the model system.

The purpose of this short paper is to give an overview of the model and discuss our experiences with applications. The paper is structured as follows. Section 2 describes the main characteristics of the model. Section 3 focuses on more operational issues and the broader model system. The last section highlights some experiences with the model and ongoing work.

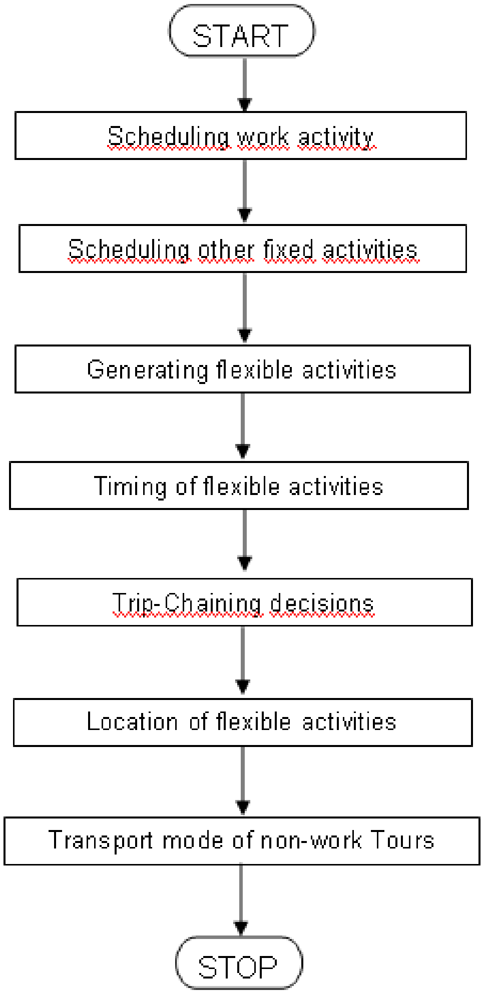


Figure 1. Main steps in the scheduling process

2. Main characteristics of the model

2.1. The scheduling process

Albatross uses a sequential decision process to generate daily activity schedules of individuals in the context of a household. Generated activity schedules describe for a given day which activities are conducted, when (start time), for how long (duration), where (location), with whom, and, if traveling is involved, the transport mode used and chaining of trips. Figure 1 shows in main lines the structure of the assumed scheduling process. As the scheme shows, Albatross uses a priority-based scheduling process where mandatory activities are scheduled first and discretionary activities are scheduled next. Furthermore, timing and trip-chaining decisions have priority over location decisions and location decisions in turn have priority over transport mode decisions. Albatross uses a relatively detailed classification of out-of-home activities (Table 1), whereas in-home activities are not further differentiated.

Table 1. Classification of out-of-home activities

Versions 2 and 3	Version 4
Work	Work
Voluntary work	Business
School	Bring or get
Bring or get	Shop one store
Daily shopping	Shop multiple stores
Non-daily shopping	Service
Service	Social
Social	Leisure
Leisure	Touring
Other	Other

Although group decision making, which may play a role in allocating tasks and joint activity participation among members of a household, is not part of this process, the model does take interactions between individuals into account in several ways. First, the scheduling processes run parallel and decisions are alternated between individuals whereby each individual takes the current state of the schedule of the other into account in making a decision. Second, schedule decisions of one individual may impose constraints on choice options of other individuals. Specifically, this happens if there are more driving licenses than cars available in the household. The model makes sure that a single car cannot be allocated to different persons at the same time. Although Albatross does not schedule the activities of children, the presence of children in a household is taken into account in the sense that this may generate activities, impose constraints and/or offers opportunities to conduct activities jointly.

Thus, a day and a household is the unit of prediction in the Albatross model. Activities are scheduled on a continuous time scale and temporal constraints are respected in the sense that the sum of durations across activity and travel episodes of a same person equals 24 hours and no overlaps and gaps between consecutive episodes can occur. Timing and duration decisions are modeled as continuous choices, at least for fixed activities. For flexible activities discrete duration classes and a subdivision of the day in episodes are used. To the extent that some flexibility is left given trip-chaining decisions and space-time constraints, the exact start time of a flexible activity within a chosen episode of the day is set randomly.

Furthermore, an important feature of the model is that generated schedules fully meet time and space-time constraints. For each decision in the scheduling process, Albatross delineates the choice set or choice range, given all previous decisions, opening hours of facilities, travel times and availability of facilities for certain activities. In early stages of the scheduling process, little information is available to determine choice sets or ranges. Consistently, the model determines the maximally available time window for an activity/trip across possible next choices (widest time slot, nearest location, fastest transport mode, etc.). For example, if no transport mode choice has been made yet and car is still an available option, then car travel time is taken as the criterion for delineating a location choice set for an activity taking also trip-chaining into account. In sum, Albatross takes into account all scheduling constraints to the extent possible given available data.

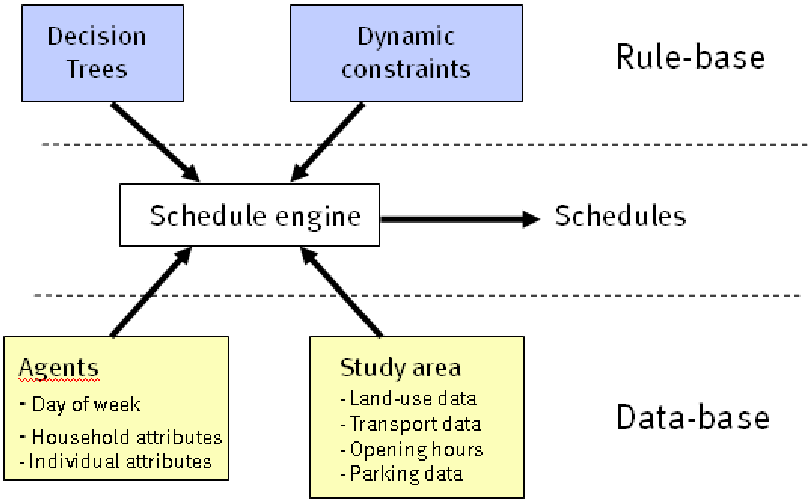
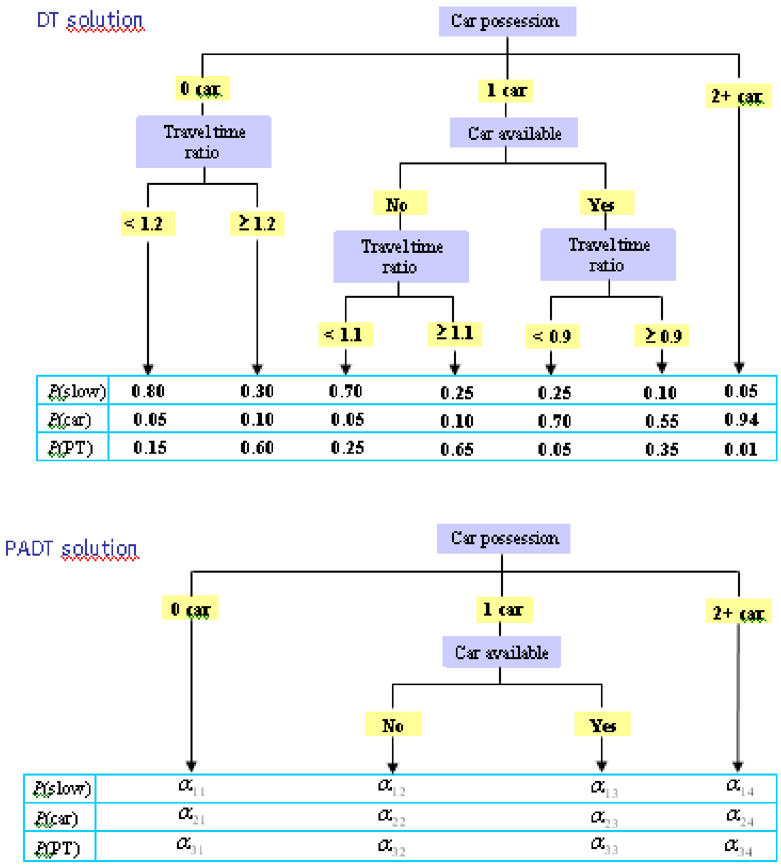


Figure 2. The scheduler

2.2. Making scheduling decisions

Albatross uses decision trees for making scheduling decisions. Figure 2 schematically shows the function the scheduler performs. The rule-based approach is based on the theory that human decision making, in case of repetitive choice behavior and large solution spaces, relies on heuristics, which are formed and continuously updated based on experiences, rather than on exhaustive evaluation of solutions. The decision trees used in Albatross are empirically derived form choice observations in activity diary data using a CHAID-based induction method. The aim of this method is to find the smallest tree that best explains a sample of choice observations by a process of recursively splitting the sample on attribute variables. A Chi-square based test of significance is used to identify in each cycle of the process the best possible way of (further) splitting across available attribute variables. The same tree construction process is also used when the choice variable is a continuous variable (i.e., start time and duration choices). Then, an F-statistic is used to evaluate possible ways of splitting. To make sure that non-systematic variance is reproduced in predictions, Albatross uses a probabilistic action assignment rule and Monte Carlo simulation to generate decisions.

Decision tree induction allows one to take a large set of attribute variables into account in each scheduling decision. In Albatross, attribute variables used for each decision relate to the individual, the household, the space-time setting, the current state of the schedule (also the one of the partner, if any) and choice alternatives. This means that decision trees represent segmentations in terms of socio-economic variables and space-time setting variables simultaneously with decision rules used within these segments. To put it another way, in the decision tree induction process, segmentation and derivation of decision rules are carried out simultaneously. Furthermore, since all earlier decisions are considered as attribute variables consistently for a current decision, the model is able to take interactions between activity-travel choices into account both within and between individuals.



$$P_{ijk} = \frac{\exp(\alpha_{jk} + \beta T_{ij} + \delta C_{ij})}{\sum_{j'} \exp(\alpha_{j'k} + \beta T_{ij'} + \delta C_{ij'})}$$

Figure 3. The PADT approach: simple example of a mode choice decision

The decision tree is a particularly powerful technique to represent discontinuous effects and effects of discrete attribute variables on choice behavior. However, sensitivity to small changes in price variables and travel time variables over a continuous range is a concern in conventional rule-based approaches. To better achieve such sensitivity, we proposed the so-called Parametric Action Decision Tree (PADT) and use this extended form of the conventional Decision Tree (DT) in the last versions of Albatross. Figure 3 shows an arbitrary and strongly simplified example of how a PADT relates to a conventional DT in a model concerned with the transport mode choice for a given trip. As the example shows, travel-time and travel-costs consequences of choice alternatives are no longer considered as attribute variables in the tree, but instead are used in the action assignment phase. Thus, choice probabilities associated with leaf nodes, which are static in a conventional DT, are generated using a discrete choice model in a PADT. In Albatross, a conventional DT is replaced by a PADT for each decision that has travel-costs and travel-time implications. This includes activity, location, trip-chaining and transport-mode choices. As a result, Albatross is able to reproduce price and time elasticities for a wide range of choice facets of individuals' activity patterns. The parameters have been estimated based on data collected in stated adaptation experiments.

3. The model system and operational issues

3.1. Steps in an application and computation times

Albatross is a micro-simulation model meaning that it generates the schedules of the individuals (excluding children) in each household of a studied population. To achieve this, a population needs to be synthesized and, since the whole of the Netherlands is the study area of Albatross, this usually involves the entire Dutch population (or a sufficient fraction of it). A prediction run thus results in a large set of activity schedules and some post-processing is needed to derive useful information from this set. Figure 4 schematically shows the components of the model system supporting these steps and purposes. The following techniques and computation times are involved.

To synthesize a population, Albatross, like many other micro-simulation models, uses Iterative Proportional Fitting. The goal of IPF is to specify a multi-attribute table, where initial cell proportions are determined based on a sample of the population and marginals of the table are given by demographic data. For Albatross applications demographic data are available at the level of so-called LMS subzones; there are 1308 such zones in the Netherlands. On the other hand, the unit of location in the scheduler is the 4-digit postcode area; there are 3987 of such areas in the Netherlands. Albatross generates a synthetic population for each of the 1308 zones and next allocates the synthetic households to these smaller postcode areas. The sample of households used to determine initial cell proportions is segmented based on urban density and province. The segmentation is optimized to ensure sufficient filling of the multi-way table, on the one hand, and to reproduce as much as possible existing spatial differences in odd ratios between attributes, on the other.

Although IPF is a well-established technique, it cannot be readily used for synthesizing households given the fact that the available demographic data relate to individuals. To solve this problem, we developed a two-staged procedure where counts of individuals are first transformed into statistics on household level before standard IPF is used to fit a multi-attribute table of households (Arentze et al., 2008). This pre-processing step makes use of so-called relation matrices that describe the composition of households regarding some attribute. The synthesizer uses a relation matrix for age and work status. Provided that the total number of households is known for each zone, a relation matrix can be fitted by a same IPF procedure if two additional parameters are a-priori set: the ratio of single females and the ratio of females living in a household (as a child) of the total number of females.

Table 2 shows which attributes at individual and household level are used by the scheduler and, hence, need to be synthesized. To give an indication of size, the multi-way table of households includes 552,960 cells in Version 3 and even a larger number of cells in Version 4 (due to a refinement of age classification). Because the

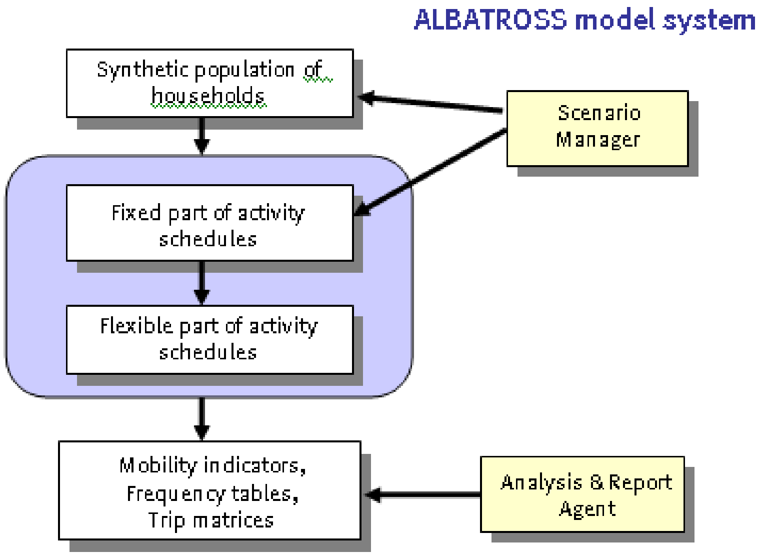


Figure 4. Structure of the model system

primary use of Albatross is to address questions of a strategic, long-term character, in most applications it suffices to synthesize only a fraction of the total population. For example, a fraction of 10% would suffice to reveal the mobility effects on a national level of even a small increase in fuel price. Depending on the scenario a 10% fraction comes down to 680,000 – 860,000 households and 1,090,000 – 1,280,000 individuals. Generating a synthetic population of that order of magnitude roughly takes 9 hours computation time on a standard PC.

Table 2. Attributes at household and person level considered in Albatross (Version 4)

Attribute	Levels
Household composition	Single, no worker; Single, one worker; Double, one worker; Double, two worker; Double, no worker
Household income	Minimum, Low, Medium, High
Household age oldest member	< 35 yr; 35 – 44 yr; 45 – 54 yr; 55 – 64 yr; 65 – 74 yr; 75+ yr
Household children	No children; < 6 yr; 6 -< 12 yr; 12-< 17 yr
Household number of cars	No car; One car; Two or more cars
Person, gender	Male; Female
Person, work status	No, Full time, Part time
Person, age	< 35 yr; 35 – 44 yr; 45 – 54 yr; 55 – 64 yr; 65 – 74 yr; 75+ yr

Having generated a synthetic population, the next step is to generate the activity schedules of the individuals in each household. Currently, Albatross does not simulate the implementation of activity schedules in space and time and, hence, there is no feedback of travel-time realizations to the scheduler. This also means that there are no iterations involved in a prediction run. The scheduler processes approximately 85,000 households per hour on a standard PC. Thus, when a fraction of 10% of the population is used, a prediction run takes 8 – 10 hours depending on the (population) scenario considered.

The Albatross program allows users to further process generated activity schedules in various ways. The program can generate a trip matrix possibly segmented on some variable chosen by the user (e.g., activity at the destination, time of day, transport mode, etc.), calculate a set of mobility indicators and produce frequency tables of maximally three dimensions defined by the user. Furthermore, Albatross can generate a standard report consisting of a standard set of frequency tables and mobility indicators that are generally of interest to assess and explain consequences of a scenario. Given the fact that Monte Carlo simulation is used by the scheduler, multiple runs of the model will lead to different outcomes even when conditions stay the same. To support the computation of t-statistics, Albatross calculates a mean and standard deviation for each output variable (be it a frequency or an indicator) based on a random segmentation of the data set. To give an indication of computation time, the generation of a standard report for a 10% fraction of the population roughly takes 2 hours on a standard PC. When the report of a scenario is compared to a report of a baseline situation, Albatross reports significance levels of differences for each output variable.

3.2. Databases

Table 3 gives an overview of the databases of the study area that the scheduler uses. Data about land-use are available at the level of postcode areas and involve employment by sector. These data are used by the scheduler to determine feasibility and attractiveness of locations for conducting certain activities. Data on the transport system are available partly on the level of postcode-areas and partly on the level of LMS-subzones. On the postcode area level, the data describe fastest-path distances and travel times by slow mode and car mode under free-floating traffic conditions. On the level of LMS-subzones, the data relate to several variables of OD relations that are used to adjust travel times and travel distances and assess travel costs by car taking into account traffic conditions during morning peak and afternoon peak (delay ratios), adaptive route choice (detour ratios) and congestion charges. The latter charges may be relevant for future scenarios, but are zero for the baseline condition, as congestion pricing does not exist in the Netherlands to date. At the level of LMS subzones also travel times and travel costs data for train and Bus/Tram/Metro are available. Finally, the scheduler uses data on parking facilities and opening hours. Apart from these databases, several system parameters may play a role in scenarios. These include among others price indices for using car (e.g., fuel price, flat road price) and train (split up by time of day).

Table 3. Overview of databases of the study area

Postcode-area file

- Number of employees by employment sector
- Number of paid and free parking places, average price of paid parking places

Postcode-area by postcode-area file (3987 postcode areas)

- Travel distance by car
- Travel time by car
- Travel time by slow mode

An LMS-subzone by LMS-subzone file (1308 LMS subzones)

- Car travel time delay ratios by trip purpose for morning peak and afternoon peak
- Car distance detour ratios by trip purpose for morning peak and afternoon peak
- Car congestion charge by trip purpose for morning peak and afternoon peak
- Bus/tram/metro travel time and tariffs
- Train travel times for access, in-vehicle and egress stages
- Variable train costs

Opening hours of daily and non-daily shopping facilities

4. Conclusion, experiences and ongoing work

In this short overview paper we gave an overview of some key characteristics of the Albatross model and highlighted some operational issues. Albatross is an exponent of the computational process modeling approach in the sense that it assumes a priority-based activity scheduling process and uses if-then rules in the form of decision trees to make decisions in this process. The model uses a rather detailed classification of activities and takes a wide set of scheduling constraints into account. The latest version of the model uses parametric action decision trees to combine the strengths of rule-based and (parametric) discrete choice models. In this way, sensitivity of the model to price and travel time changes is implemented in all decisions that have travel costs and travel time consequences, such as activity, location, trip-chaining and transport-mode choices.

Activity-based models have only recently started to make the transition to practice implying that experiences with applications of these models are still very limited to date. The first ‘real’ applications of Albatross have been carried out. Experiences reveal specific strengths of the activity-based approach. Predictions tend to show clear activity scheduling effects implying that a behavioral change on one facet often generates changes on other facets of activity patterns. For example, in predictions we often see that increases of activities in one category are partly or fully compensated by decreases of activities in other categories with as a result that the number of trips displays a tendency to stay more or less constant. Although the number of out-of-home activities is rather stable, predicted distributions across activity categories tend to be very sensitive, in particular, to demographic variables. Shifts in activity choice in turn tend to have strong effects on choices of implementing an activity in terms of start time, location, trip-chaining and transport mode. Due to links between choice facets within patterns these changes may incur secondary effects. For example, a shift in start time from early to later times of a day tends to increase opportunities for car sharing giving rise to a shift in mode choice. Such secondary responses would not be revealed when a trip-based or tour-based model is used. Finally, we note that the detailed information that an activity-based model such as Albatross provides about activity patterns helps building a comprehensive and coherent view of the behavioral changes underlying changes in mobility patterns. A better understanding of mechanisms greatly helps formulating effective policies.

Ongoing work is focused on ways of extending and refining the model. This includes incorporating the activity schedules of children and better representing within-household interactions between individuals in the contexts of task allocation and joint activity participation. Furthermore, applications of the model in other countries and linkages with emission models and dynamic traffic assignment models are under development.

References

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.

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.

Arentze, T.A. and H.J.P. Timmermans (2007), Parametric action decision trees: Incorporating continuous attribute variables into rule-based models of discrete choice, *Transportation Research B,* 41, 772-783.

Arentze, T.A., H.J.P. Timmermans and F. Hofman (2008), Creating synthetic household populations: problems and approach, *Transportation Research Record*, 2014, pp. 85-91.