# Integrating a Multi-Agent Model of Land Development and an Activity-Based Model of Transport Demand: Progress and Developments[1]

Theo Arentze, Soora Rasouli and Harry Timmermans

Urban Planning Group
Eindhoven University of Technology
The Netherlands

**Abstract**
This short workshop paper gives a brief overview of recent research activities conducted by the Urban Planning Group aimed at integrating activity-based models of transport demand and models of land development. In particular, the integration of ABSOLUTE and ALBATROSS are discussed. A recent new research projects focuses on uncertainty modeling in complex model chains. The results of this project will have implications for the integration with land use as well.

**Keywords**: integrated land-use transport models, activity-based models, urban dynamics, multi-agent systems.

## 1. Introduction

A main stream of research in the Urban Planning Group is concerned with modeling travel demand, developing original activity-based models. The most important of these is Albatross (Arentze and Timmermans, 2000, 2004, 2005), a fully operational activity-based model, developed for the Dutch Ministry of Transport, which simulates for every Dutch individual which activities are conducted where, when, with whom and the transport modes involved, subject to a set of spatial, temporal, spatio-temporal and institutional constraints. This model system has been upgraded and expanded occasionally. For example, over the last couple of years, the system has been calibrated on the new national travel survey of the Netherlands, while in addition the complete model was re-engineered to better reflect all household-level decisions that impact travel behavior. The latest Albatross-related project is to make the model available on the Web and to systematically conduct an uncertainty analysis of the model (e.g. Rasouli, Arentze and Timmermans, 2010).

A new line of research on demand generation is concerned with the development of dynamic models of activity-travel demand. Whereas Albatross is concerned with a typical day, te new model system will simulate changes in activity-travel patterns along different time horizons, Most of this work will be conducted in the context of the U4IA (e.g. Timmermans et al., 2010), sponsored by the European Research Council (2 postdocs and 6 PhD students), but additional funding has been required via other channels.

Work in the Urban Planning Group related to the integration of the demand forecasting models to land use dynamics has been conducted in the context of the Albatross system. This overview paper, which is compiled and updated from a series of previous publications, describes some key concepts and models that are relevant for the workshop. In particular, the overview starts with a summary description of the latest developments in the uncertainweb enabled version of Albatross and then continues with a description of a model, called ABSOLUTE[2], that attempts to link an activity-based

---

[1] Paper for the Workshop on Integrated Travel Demand and Network Supply Modeling, Tampa, USA.

Theo Arentze, Soora Rasouli and Harry Timmermans are respectively Associate Professor, PhD candidate and Professor at the Urban Planning Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, phone +31 40 247 3315, fax +31 40 2438488, e-mail: t.a.arentze@bwk.tue.nl.

[2] Acronym for Activity-Based System of Land Use and Transport Events

model of transport demand to a model of land use change as part of a wider multi-agent system, which simulates how planning agencies, developers and facility providers develop land.

## 2. Albatross and UncertWeb

### 2.1 SUMMARY OF ALBATROSS

Albatross, acronym for A Learning Based Transportation Oriented Simulation System, was developed for the Dutch Ministry of Transportation, who decided to develop an activity-based model alongside their state-of-the-art tour- based model LMS. 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 travelling is involved, the transport mode used and chaining of trips. Albatross consists of various components that perform specialized functions in the scheduling process. Figure 1 portrays the different parts of system. A formal description of the components will be provided in the following sections.

#### The Scheduling Engine

The core component of the system is the Scheduling Engine. This component controls the scheduling process in terms of a sequence steps. Various moments in the process require decisions and information about options and conditions for decisions. The scheduling Engine identifies which condition information is relevant for the Decision Unit, activates the appropriate analytical and rule-based models in the Inference System to obtain the information, and translates decisions returned by the Decision Unit to appropriate operations on the evolving schedule. The scheduling engine uses functional specification of both Inference System and the Decision Unit but does not need to know how inference and decision functions are defined.

Figure 2 shows 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. 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 modelled 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.

#### The Decision Unit

The Decision Unit incorporates for each step in the scheduling process a set of decision rules representing conditional preferences of individuals within constraints regarding decision option. An inference engine, which selects and applies the appropriate rules to arrive at a decision upon a query of Scheduling Engine is built-in. We emphasize that only the calculation of (possibly) relevant condition variables and decision options are defined in the program code of the system. The rules connecting condition variables and decision options are external to the system and loaded from data files. This means that users have the possibility to use application-dependent rule-bases. A rule-base is the output of the second fixed component of the Decision Unit referred to as the Learning Mechanism. As the name suggests, this component derives rules from data based on principle of inductive (i.e., supervised) learning. In the present system the component is applied in a pre-processing stage external to the system. In this process a set of decision trees for making scheduling decisions is extracted from data. In the following section the procedure aimed at getting decision rules will be explained.

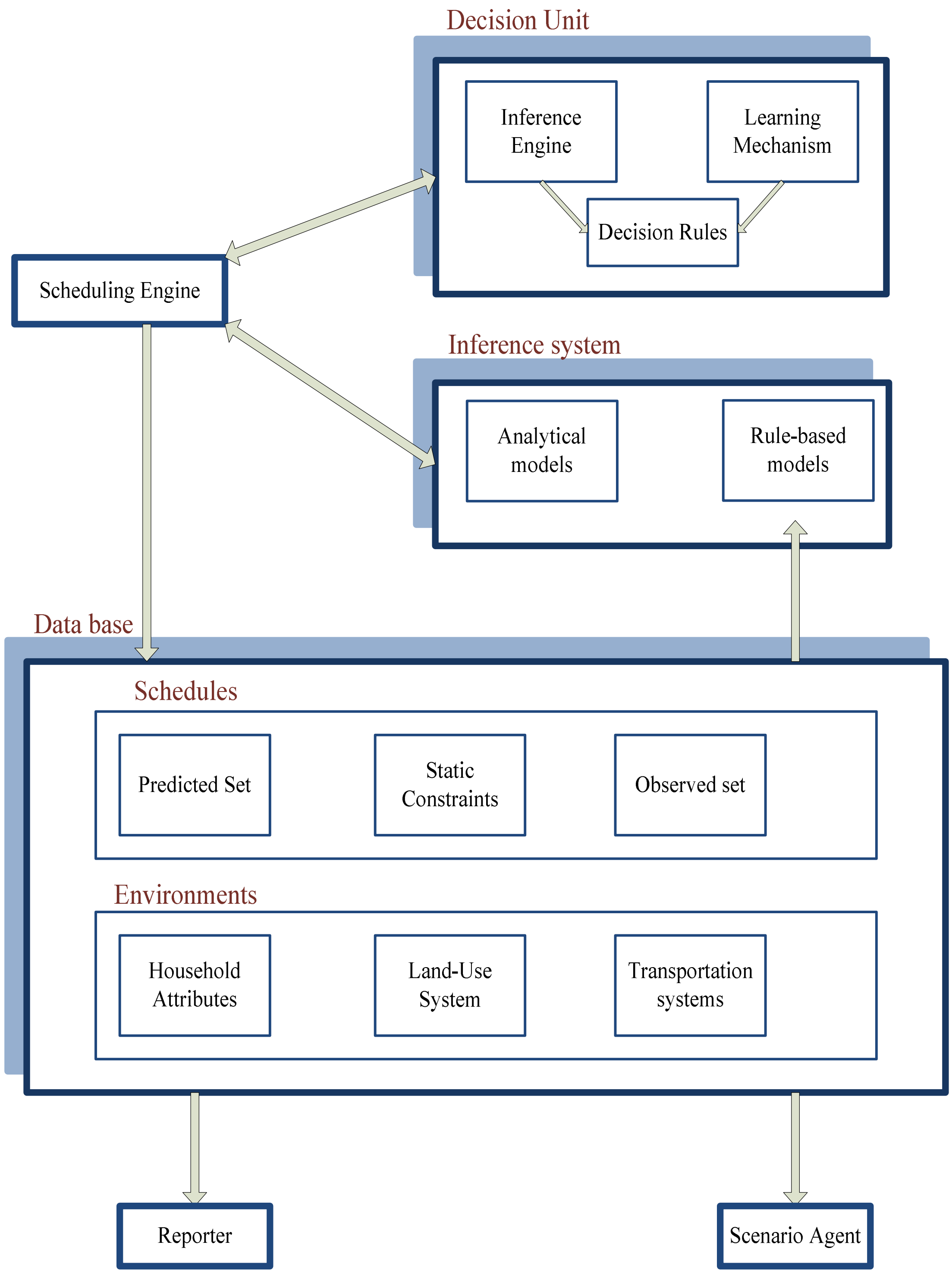


**Figure 1: Architecture of the part of the system concerned with the derivation and application of the activity-scheduling model (note: observed schedules are used only in a testing phase)**

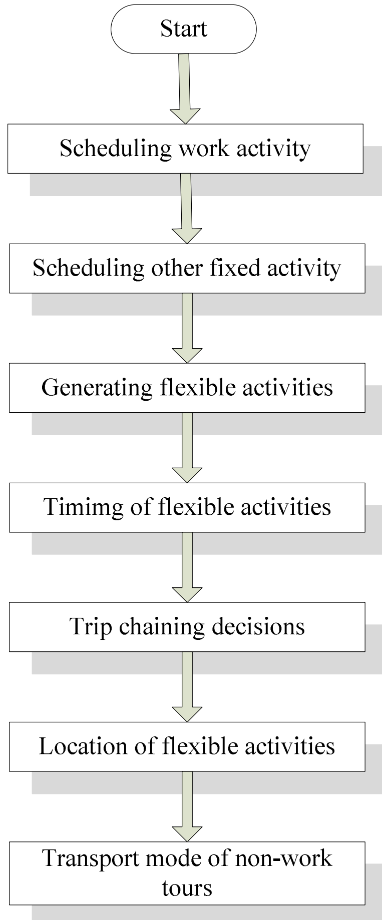


**Figure 2: Main steps in the scheduling process**

**Table 1: Classification of out-of-home activities in Albatross**

| | ***Out of home activities*** |
|---|---|
| ***1*** | Work |
| ***2*** | Business |
| ***3*** | Bring or get |
| ***4*** | Shop one store |
| ***5*** | Shop multiple stores |
| ***6*** | Service |
| ***7*** | Social |
| ***8*** | Leisure |
| ***9*** | Touring |

*Decision Trees and Decision Rules*

The rule-based approach underpinning the concept of decision tree 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. Choice behavior that emerges from learning is driven by individual dependent condition-action rules. In general format, a condition-action rule can be described as:

$$\text{if } C_1 \in CS_{1k} \wedge C_2 \in CS_{2k} \wedge \; .... \wedge C_m \in CS_{mk} \text{ then choose alternative } A_k \tag{1}$$

where $C_i$ represents condition variables, $CS_{ik}$ the condition state of the *i*-th variable in the *k*-th rule and $A_k$ the choice generated by the *k*-th rule. In this notation, a condition state is represented as a subset ofthe domain of the condition variable. If the condition variable is of a nominal measurement scale, then it may specify any subset of the domain. In the case of an ordinal or metric variable, on the other hand, the condition state specifies a certain subrange of the variable's domain.

To make sure that a rule set is able to respond to every situation, it must meet requirements of completeness and consistency. A model is considered complete if at least one rule responds and consistent if no more than one rule responds to every possible combination of values of condition variables $C_i$. These properties are guaranteed by the way the learning mechanism operates in the model. Consider an initial situation where the individual has no a-priori knowledge of the domain. Decision-making would be purely random and handled by a single 'rule':

$$\text{if } C_1 \in CD_1 \wedge C_2 \in CD_2 \wedge \; .... \wedge C_m \in CD_m \text{ then choose random} \tag{2}$$

where $CD_i$ represents the domain of condition variable $C_i$. Since every 'condition state' in this rule equals the entire domain, each variable in effect is irrelevant. Therefore, every possible state in terms of $C_i$ will trigger the rule implying that the model meets the requirements of completeness and consistency. Now assume that through interaction with the environment the individual has learned to discriminate between states on some condition variable. This can be represented by splitting the domain of that variable into two states so that the initial rule is replaced by two new rules:

$$\text{if } C_1 \in CD_1 \wedge C_2 \in CD_2 \wedge \; .....\wedge C_j \in CS_{j1} \wedge.... \; \wedge C_m \in CD_m \text{ then choose A1} \tag{3}$$
$$\text{if } C_1 \in CD_1 \wedge C_2 \in CD_2 \wedge \; .....\wedge C_j \in CS_{j2} \wedge.... \; \wedge C_m \in CD_m \text{ then choose A2} \tag{4}$$

Because the new condition states were achieved by splitting a domain, $CS_{j1} \cup CS_{j2} = CD_j$ and $CS_{j1} \cap CS_{j2} = \varnothing$, the new model still meets properties of completeness and consistency. This process of splitting could be repeated endlessly resulting in increasingly complex models while maintaining the required properties. Any set of rules that can be obtained by recursively splitting condition states starting with an initial rule of above format meets the formal definition of a decision tree as formulated by Safavian and Landgrebe (1991) and vice versa.

The decision trees used in Albatross are empirically derived from 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. 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.

The data used for obtaining decision trees in the version of Albatross we currently use in Uncertweb project are from a travel-survey in 2004 named MON (MobiliteitsOnderzoek Netherlands). This national travel survey data set involves 45000 individuals.

Having derived a decision tree for each choice facet, the next question becomes how to derive decisions from trees for prediction. Consider a response variable that has $Q$ levels and for which

CHAID produced a tree with $K$ leaf nodes. In the prediction stage, the tree is used to classify new cases to one of the $K$ leaf nodes based on attributes of the case. A response-assignment rule needs to be specified that defines a response (decision) for each classified case. In many other applications of decision trees, a plurality rule is used. This rule assigns the modal response among training cases (i.e., the sample used for developing the tree) at a leaf node. A deterministic rule like this may yield the best predictions at an individual level, but fails to reproduce residual variance (if any) at leaf nodes in predictions. Given our modeling purpose, we, therefore, use a probabilistic assignment rule instead. According to this rule, the probability of selecting the $q$-th response for each new case assigned to the $k$-th node is simply given by:

$$p_{kq} = \frac{f_{kq}}{N_k} \tag{5}$$

where $f_{kq}$ is the number of training cases of category $q$ at leaf node $k$ and $N_k$ is the total number of training cases at that node. This rule is sensitive to residual variance, but fails to take scheduling constraints into account. Scheduling constraints entail that dependent on individual attributes and the state of the current schedule some choice alternatives for the decision at hand may be infeasible. If such constraints are represented in the decision tree, the probabilistic rule would assign zero probability to infeasible categories and the response distribution should not be biased. However, even though it is likely, it is not guaranteed that the induction method discovers constraint rules in data. Therefore, to cover the general case we need to refine rule (5) as:

$$p_{kq} = \begin{cases} 0 & \text{if } q \text{ is infeasible} \\ \dfrac{f_{kq}}{\sum_{q'} f_{kq'}} & \text{otherwise} \end{cases} \tag{6}$$

where $q'$ is an index of feasible alternatives for the decision at hand. Even though this rule may work well in practice, it may produce slightly biased patterns at an aggregate level that should be noted. That is to say, the rule tends to over predict responses that are feasible in the majority of cases (at that leaf node), because the probability of these responses is increased by rule (6) in constrained cases and stays the same as (5) in unconstrained cases.

**Inference System**

To compute possibly relevant conditions and availability of decision options, the Inference System consists of a collection of built-in analytical and logic-based models representing basic knowledge about the transport and land-use system and scheduling constraints. Unlike the decision rules, the inference model is a fixed part of the system reflecting the assumption that the knowledge it conveys is basic and does not vary across individuals or environments. In particular, the incorporated models implement dynamic constraints to determine the availability of decision options in each step of the process, such as for example whether or not an activity fits in a given time slot given required travel times, minimum duration of the activity, possible locations for the activity, opening hours of facilities and so on. Information flows between the Inference System and Decision Unit pass through and are controlled by the Scheduling Engine. Hence, both the Decision Unit and Inference System do not need scheduling-process knowledge. The following section describes different types of constraints which Albatross takes into account in the scheduling process by the Inference System.

(1) *situational constraints* impose that a person, transport mode and other schedule resources cannot be at different locations at the same time.

(2) *institutional constraints,* such as opening hours, influence the earliest and latest possible times to implement a particular activity.

(3) *household constraints*, such as bringing children to school, dictate when particular activities need to be performed and others cannot be performed.

(4) *spatial constraints* also have an impact in the sense that either particular activities cannot be performed at particular locations, or individuals have incomplete or incorrect information about the opportunities that particular locations may offer.

(5) *time constraints* limit the number of feasible activity patterns in the sense that activities do require some minimum duration and both the total amount of time and the amount of time for discretionary activities is limited.

(6) *spatial-temporal constraints* are critical in the sense that the specific interaction between an individual's activity program, the individual's cognitive space, the institutional context and the transportation environment may imply that an individual cannot be at a particular location at the right time to conduct a particular activity.

The implementation of situational, household and temporal constraints is straightforward. So here an example of spatial-temporal constraint determining location choices is explained: A location $l$ is considered feasible if the following two conditions are met:

$$\exists g \in G_l, g \in G\{\, a(\tau) \,\} \tag{7}$$

$$T_{l_g}^{f\max}(\tau) - T_{l_g}^{s\min}(\tau) \geq v^{\min}(\tau) \tag{8}$$

where, $\tau$ is an index of activities in a given schedule $S$, $G_l$ is the set of known facility types at location $l$, $G\{\, a(\tau) \,\}$ is the set of facilities compatible with activities of type $a(\tau)$, $v^{\min}(\tau)$ is the minimum duration and $T_{l_g}^{s\min}$ and $T_{l_g}^{f\max}$ define the time window for the activity dependent on the current schedule and opening hours of facilities. The latter terms are formally defined as:

$$T_{l_g}^{f\max}(\tau) = \max\{{}_d t_{l_g}^{\min}, T^{f\min}(\tau - 1) + t_l^t(\tau)\} \tag{9}$$

$$T_{l_g}^{s\min}(\tau) = \min\{{}_d t_{l_g}^{\max}, T^{s\max}(\tau + 1) - t_l^t(\tau + 1)\} \tag{10}$$

where, ${}_d t_{l_g}^{\min}$ and ${}_d t_{l_g}^{\max}$ are the known opening and closing times of facilities of type $g$ at location $l$ on day $d$, $T^{f\min}$ is the earliest end time and $T^{s\max}$ the latest start time of the previous and next activity respectively and $t_l^t$ is travel time to the activity location using the mode chosen in a previous step. Earliest start times and latest end times of activities are calculated by shifting previous activities as far as possible to the right on the time scale and next activities as far as possible to the left within temporal constraints.

### 2.2 UNCERTAINTY-ENABLED ALBATROSS MODEL

The original rule-based Albatross does not consider uncertainty. However, it has become clear that uncertainty analysis is important from a policy making point of view; especially for complex comprehensive models. Uncertainty in forecasting error can be attributed to two basic sources: input uncertainty and model uncertainty. Input uncertainty is concerned with the effect of uncertain input data, due to measurement error or to scenario uncertainty, on the ultimate forecasts of the model. In contrast, model uncertainty consists of two types of error: specification error and calibration (or estimation) error. Specification errors result from a failure of the researcher to identify the true model, a simplification of the model or the statistical distribution of random components. Estimation error involves error in estimating the values of various constants and parameters in the model structure. If we have some confidence in the specification of the model, estimation error can be determined by standard statistical procedures. Assessing the effects of specification errors is more challenging.

Considering input uncertainty for all inputs involved in Albatross is not possible since there are huge numbers of inputs involved (i.e., data of the transport and land-use system and population). The approach which was chosen is considering uncertainty in inputs which are believed to be more important and contributed the most in variability of the results. Considering all types of model uncertainty is also not necessary because, for example, the probability of model failure to predict travel is not considered because it was already tested by goodness-of-fit analysis at the national level. So in terms of model uncertainty the structure of Albatross in terms of decision trees is considered certain and uncertainty analysis is performed with regard to the randomness involved in predictions obtained from the probabilistic decision tree choices.

Albatross has 27 decision trees for deciding about different facets of activities. Starting from scheduling work and other fixed activities to the transport mode of flexible activities. Albatross uses Monte-Carlo simulation for getting the response of each decision of each individual. That is why each case might choose different responses in each decision tree in different runs. Accordingly, different runs of the model would result in different outputs also on the level of the more aggregate indicators. Analyzing Albatross several times to compare the results and report the confidence interval or standard error is necessary. This type of model uncertainty or variability will be considered in Uncertweb.

Input uncertainty is related to data fed into the Albatross as input. The amount of information input is not bigger than in other large-scale transport models, but is nevertheless large, so it has been categorized into 3 main groups:

*Sampling bias and uncertainty in travel data*

The first type of input uncertainty in Albatross might arise due to the sampling bias. The model uses the 2004 national travel survey MON to obtain the decision trees. These data are obtained from questionnaire administered through the whole of the Netherlands. In these paper-based questionnaires, the people were asked to first write down their individual and socio-economic characteristics and then report their activities during two consecutive days. The first issue here is that experience shows in such surveys some groups might be under reported due to their unwillingness to take part in the survey or not having the time to complete it. The other issue is the mistakes the respondent might make in reporting characteristics or activities, or also the people turning the paper-based questionnaire into computer-based might introduce some mistakes. However an interactive computer system was developed to test the logical consistency and completeness of activity diaries, still some types of uncertainty analysis based on sampling bias seems attractive. The reason is the program could identify errors that incur logical inconsistency (for example a person younger than 16 years of age reporting to have driving license). Therefore random sampling from the original sample (MON) using the bootstrapping method with new samples in the analysis to see the effect of changes in sample on the output would be useful. To discriminate between model and this input uncertainty, the simulator random number seed should be kept constant. However at some stage it would be interesting to consider both types of uncertainty together.

*Uncertainty in Land-use data*

Land-use data is input as an indicator of destination attractiveness in Albatross. These data are obtained from other sources and have some uncertainty in it. All of these data could be considered for uncertainty analysis. However uncertainty in some of them is believed to induce more variability in results. One of them is number of employees in daily goods retailing. The reason is that in most cases this activity appears in the schedule and consequently variability in it would significantly affect the travel-demand prediction results. The probability distribution for this data could be obtained by an expert elicitation tool. With Monte-Carlo simulation, a set of new land-use data base would be developed and each of them is used separately in the Albatross process. To prevent the effect of model uncertainty in different runs, the random number seed needs to be kept constant across different runs.

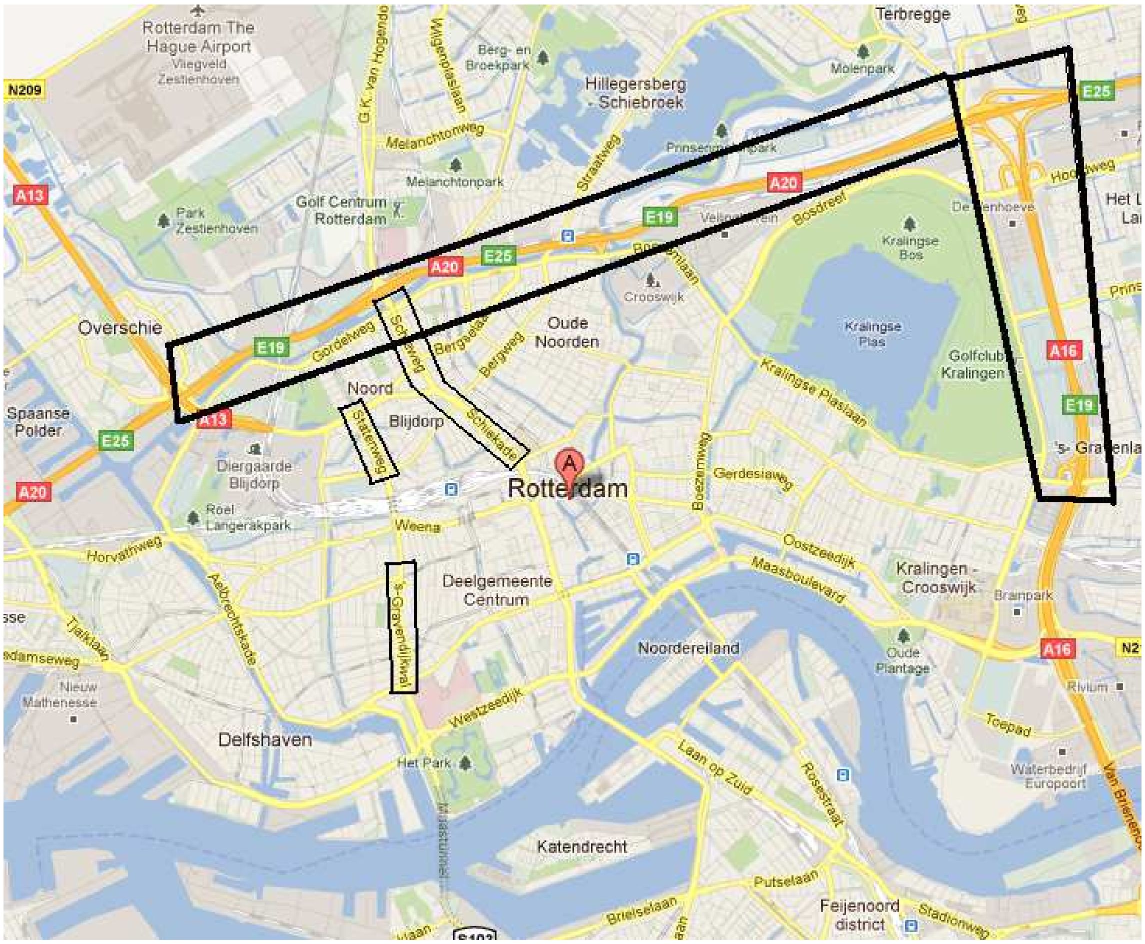


**Figure 3: Crucial links in Rotterdam for travel time uncertainty**

*Uncertainty in Travel time*

Travel time data related to private forms of transport (car and bicycle) is extracted from an existing road network datafile. In this file, link speed represents an average flow speed on the concerned link throughout the day according to expert assessment at the level of individual links for main roads. Albatross also uses travel delay factors for the morning peak and one for the evening peak, both based on the national model system. Considering uncertainty for average speed at which most travels take place would be necessary. The expert elicitation tool again could be used for getting the probability distribution. However considering all links for this type of uncertainty is impossible. In case of Rotterdam, 6 main links were selected after consulting with some experts. These links are shown in Figure 3. Similar to the land-use process, thus sets of new travel-time data will be developed and each of them will be used as input in different runs of the model. Again the random number seed will be kept constant.

The latest development in the Albatross system is to allow a systematic uncertainty analysis. Results will become available at the level of (i) travel-demand indicators, (ii) OD matrices, (iii) network traffic flows, and (iv) individual space-time trajectories.

## 2. Integrating a facility location model and a model of activity-travel patterns

Work has been conducted to integrate the Albatross system with a facility location model, called ABSOLUTE.

*Scope of ABSOLUTE*
To define the scope of the model system it is helpful to make a distinction between three types of urban facilities. These include housing facilities, productive facilities and consumptive facilities. Whereas productive facilities, such as offices and industrial plants, provide places for the work activities of individuals, consumptive facilities, such as stores, parks, schools, post offices, etc., provide the locations where individuals can conduct consumptive activities. Below, we describe some ideas used in modeling consumptive facilities. Other ongoing work is focusing on simulating the dynamics of other facilities from a lifecycle perspective.

On the demand side, individuals schedule their activities and execute their activity schedules on a daily basis. The schedule determines which activities are conducted where, when, for how long and the transport mode used, on a continuous time scale. On the supply side, firms and/or the planning authority decide on where to develop how much of which (consumptive) facilities to support those activities. Whether a firm or the planning authority operates the facility depends on the type of facility.

To model and micro-simulate the dynamics we use a multi-agent system approach. The agent model does not represent the supply-side actors individually. Rather, the objective of the model is to represent behavior accurately at the group level. As a consequence, the rules used by the agents do not necessarily represent objectives of each actor individually (such as, for example, profit maximization) and should be evaluated on their ability to predict outcomes related to each facility type. The objective of the integrated system is to predict emergence of facility location patterns and facility usage patterns that are consistent with behavior of all the actors involved. This means that a time path is not predicted.

*Key principles and notation*
Let there be given a plan area, represented as a regular grid of cells that are considered the units of location, a zoning plan for the area determining the allowed facility types in each cell and a population with known home and work locations. The system uses the following classifications and allocation variables:

| | |
|---|---|
| $U$ | is the set of land-use types distinguished in the given zoning plan; |
| $A$ | is the set of activity types distinguished by individuals in scheduling their daily activities; |
| $G$ | is the set of demand types requiring a (consumptive) facility of some sort; |
| $H$ | is the set of (consumptive) facility types; |
| $B$ | is a smaller set of more aggregate demand types, in the following referred to as demand *sectors*. |
| $P_g^a$ | is the probability that activity $a$ involves a demand of type $g$; |
| $P_e^g$ | is the probability that individuals use heuristic $e$ to select a location for purpose $g$; |
| $U_b$ | is the subset of land-uses allowing facilities of type $b \in B$ ($U_b \subseteq U$); |
| $G_h$ | is the subset of demand types covered by facility $h \in H$ ($G_h \subseteq G$); |
| $H_b$ | is the subset of facilities belonging to sector $b$ ($H_b \subseteq H$). |

Throughout the text we will use the same symbol to refer to demand and supply as they both relate to a same service provided by the facility. As implied by these definitions, we assume that there does not necessarily exist a one-to-one correspondence between scheduled activities, on the one hand, and facility types, on the other. Rather allocation of activities to facility types is conducted in the system by Monte Carlo drawings from probability distribution $P_g^a$ for a given activity $a$. We emphasize that not all activities in $A$ require a (consumptive) facility. Examples of activities that do not require a consumptive facility include social activities conducted at the own or someone else's home and work activities. Also worth mentioning in this regard are picking up/dropping off persons or goods. These activities may involve a visit to a (consumptive) facility location without occupying floor space of the facility and therefore do not classify as a consumptive activity meaning that $P_g^a$ is equal to zero for each $g$ for those activities. On the other hand, an activity such as green recreation *is* considered consumptive even though it may not require a facility in a more narrow sense.

Once the demand is established for a given activity of a given individual, the optional facilities for the activity are defined by sets $G_h$. Note that the model does not assume that a one-to-one correspondence necessarily exists between demands and facilities. Rather, a single facility may combine different types of supply and supply sets $G_h$ may overlap between facilities. This provides the flexibility to represent a large variety of existing facility types in urban systems. The allowed types are discussed in more detail later.

In addition, the system uses the following variables to describe the study area at any moment in time:

$P_a$ is the probability that activity $a$ is conducted on a given day by a given individual;
$P_g$ is the probability that a given individual has demand $g$ on a given day;
$P_i^g$ is the probability that a given trip with purpose $g$ originates from $i$ (home, work, other activity locations);
$U_l$ is the land use of cell $l$ (defined by the zoning plan)
$S_{lh}$ is the size (square meter floor space) of facility $h$ in cell $l$;
$V_{lg}$ is the average number of visits per day to cell $l$ for satisfying demand $g$;
$N_{li}$ is the number of individuals visiting location $l$ for purpose $i$ (residential, working, visitors of certain facility types) per day;

Probabilities $P_a$ are not set by the user, but rather follow from the activity-scheduling model. Note that the following relationship exists:

$$P_g = \sum_a P_g^a \tag{11}$$

Furthermore note that the $P_i^g$ probabilities sum up to one across $i$ for each $g$. Where $U_l$ is given by the zoning plan, $S_{lh}$ represent the outcomes of suppliers' decisions and $V_{lg}$, $N_{li}$ represent the aggregate result of individuals' decisions. As implied by the definitions, the present system uses the period of a day as the unit of time and, consequently, does not take congestion conditions into account. The continuous time scale on which schedules are defined would allow a further (unlimited) disaggregation of time so that an extension in that direction is possible.

*The supply structure of facilities*

Before describing the behavior of individuals and suppliers it is useful to make a distinction between several facility types regarding their supply structure as defined by $G_h$. *Elementary* facilities support only one activity. *Mixed* and *higher-order* facilities are facilities that incorporate multiple elementary facilities. They differ regarding whether or not they support also activities not covered by elementary facilities. A higher-order facility does have unique supply, whereas a mixed facility does not offer more than a collection of elementary facilities could offer (it just brings them under the same roof). A typical example of a mixed facility is a health care center combining elementary facilities such as a physicist, pharmacy, physiotherapist and possibly other medical services that can also be provided by individual facilities. A typical example of a higher-order facility is a district shopping center that includes the supply of a neighborhood shopping center and in addition offers supply not included in this facility.

*Behavior of individuals*

Each individual of a study population is represented as an agent in the model. The behavior of each individual agent consists of scheduling activities and executing the schedule on a daily basis. In a prototype of the system, a version of ALBATROSS is incorporated as a method in each individual agent to accomplish this. As explained earlier, for a given day, the model determines which activities are conducted where, for how long, when, and, if travel is involved, the transport mode used taking into account schedule constraints, some socio-economic variables, day of the week and residential location. In-home activities are not further differentiated. Thus, an activity generated can be described as:

$$e = (a, t^s, v, t^t, m) \tag{12}$$

where $a \in A$ is the activity type, $t^s$ the start time, $v$ the duration, $t^t$ the travel time and $m$ the transport mode of episode $i$. The activities conducted by the same person must meet the temporal constraints:

$$\sum_{i \in E_{dp}} (v_e + t_e^t) = 24 \times 60 \qquad \forall dp \tag{13}$$

$$t_e^s + v_e + t_e^t = t_{e+1}^s \qquad \forall idp \tag{14}$$

where time is expressed in minutes, $E_{dp}$ is the schedule of a person $p$ on day $d$, defined as an ordered set of activity episodes, and $e$+1 is the activity succeeding $e$ in $E_{dp}$. We emphasize that constraint (13) is a logical constraint: since the activity-based model predicts the activities for a day, they should fit into the time frame of a day.

The agents make location choices for out-of-home activities in the sequence in which they occur in the schedule. The schedule defines for each activity the transport mode used for the trip to the activity location and the travel time. Predicted travel times refer to the duration of the trip, but are interpreted here as the time the individual is *willing* to travel. The origin location is given by the location of the previous activity in the sequence. As a consequence of trip chaining, the origin location of an activity does not need to be the home location. As it appears, in a substantial portion of trips the origin location is not the home location.

The combination of origin location, maximum travel time and transport mode determines the locations that are within reach. Before a choice set can be delineated for a given activity $a$, the demand type is determined by drawing $g$ from probability distribution $P_g^a$. Then, the choice set is defined as all locations (cells) within reach containing supply $g$. For making a choice, the agent then determines a location selection heuristic. In the prototype system, only two simple heuristics are considered as alternatives, namely choosing a cell at random and choosing the nearest cell. We emphasize that in combination with the choice-set delineation and demand-selection rule, the heuristics give rise to more complex behavior than one would expect at first sight. For example, the model could select the nearest highest order location within a maximum travel time as a result of a certain combination of rules.

Selecting a location in this way may fail, however, namely when a facility of the given order is not within reach. If selection fails then the agent tries several ways to overcome the impasse. First, it lowers the demand by accepting also facilities that offer the lower-order service. If this fails, it then relaxes the travel time constraint and searches in a wider area for facilities. In this way, the agent will always find a location for the activity, unless a facility of the demanded type or a lower-order of that type does not exist.

*Behavior of suppliers*

For each demand sector ($b \in B$) the system implements an agent, which is concerned with developing and maintaining a network of facilities of types $H_b$. In turn, each of these agents incorporates one or more subagents specialized in a specific facility type involved in that sector ($h \in H_b$). An agent at the sector level co-ordinates, where needed, the actions of it's subagents, but leaves all tasks involved in developing and maintaining the network to the specialized subagents.

The methods available to each subagent address the problems of assessing the value of a given cell for developing a facility unit and making location decisions based on this assessment. If the (sub)agents would have unlimited knowledge of the activity and location choice of individuals, they would be able to predict exactly the amount of demand a new facility in a specific cell would attract and how much demand would be distracted from existing competing facilities. However, in the model as well as reality, suppliers have only limited knowledge about the behavior of individuals and, therefore, have to rely on estimation methods. In the model it is assumed that, regardless of facility type, all agents perform a catchment area analysis. In this method, a primary and secondary catchment area is defined by drawing concentric circles around the site. Next, the demand attracted from each cell within the circles is estimated taking into account the location of existing competing facilities. Catchment area analysis is a well-established method in the retailing sector We emphasize, however, that the assumption here is more encompassing. We assume that location choice of supply-side agents in general can be modeled as if the decisions are based on a catchment area analysis. This is exemplified by the fact that parameters of the method can be estimated based on commonly used facility-planning guidelines and statistics..

The specific characteristics of facilities are acknowledged by a set of parameters including the following:

| | |
|---|---|
| $\chi_h^+$ | is the maximum cannibalism tolerated for a facility of type $h$; |
| $S_h^-$ | is the minimum normative floor space size (square meter) for facility $h$ to be viable; |
| $\sigma_h$ | is the normative floor space size (square meter) per unit demand for added supply of $h$; |
| $r_h^{1,i}$, $r_h^{2,i}$ | is the radius of the primary and secondary catchment area of added supply of $h$ related to segment $i$; |
| $\pi_h^{1,i}$, $\pi_h^{1,i}$ | is the penetration rate of added supply of facility $h$ in the primary and secondary catchment area related to segment $i$; |
| $V_{hj}^r$ | is the extra demand attracted to $h$ if the nearest main road lies within the $j$-th distance band from $h$; |
| $V_{hj}^c$ | is the extra demand attracted to $h$ if the city center lies within the $j$-th distance band from $h$. |

Given these parameters, a catchment area analysis establishes the value of a given location based on the following equation:

$$Q_{lh} = \Delta Q_{lg_h} + Q_{lh'} \quad (15)$$

where $Q_{lh}$ is the estimated value of cell $l$ for facility $h$, $\Delta Q_{lg_h}$ is the estimated value of the added supply, $g_h$, of facility $h$ and $h'$ is the (next) lower-order facility of $h$. Note that $Q_{lh}'$ is recursively defined by the same equation, so that, for example, for a three-order facility the equation would read as:

$$Q_{lh} = \Delta Q_{lg_h} + \Delta Q_{lg_{h'}} + \Delta Q_{lg_{h''}} \quad (16)$$

As implied by this equation, the market value equals the sum of market values across the constituent supply layers. In case $h$ has no lower order, the second term on the RHS of equation (15) is simply set to zero. The value of added supply ($g = g_h$) is defined as:

$$\Delta Q_{lg} = \sum_i V_{lg}^i + V_{lg}^r + V_{lg}^c \quad (17)$$

where the $V^i$ terms represent the estimated number of visitors that would be attracted to location $l$ for $g$ from home locations ($i$ = 1), work locations ($i$ = 2) and other activity locations ($i \in B$) within catchment areas and the last two terms represent estimates of additional flows from a main road ($r$) and city center ($c$), respectively. The last two terms are determined in a straightforward way based on parameters $V_{hj}^r$ and $V_{hj}^c$. On the other hand, the $i$-terms are all estimated based on a catchment area analysis. The method for this can be described as:

$$V_{lg}^i = P_i^g P_g \sum_{l'} \pi_{lg}^i (l') N_{l'i} \quad (18)$$

where, as before, $P_i^g$ is the probability that a given $g$-trip originates from activity location $i$, $P_g$ is the probability of a $g$-demand by an individual on a day, $N_{l'i}$ is the number of people present in cell $l'$ for purpose $i$ on a day and $\pi_{lg}^{i}(l')$ is the probability of attracting those people to $l$ for satisfying demand $g$. The $\pi$ term represents estimated penetration rates in cells and is defined as follows:

$$\pi_{lg}^{i}(l') = \begin{cases} \pi_g^{1,i} & if \;\; l' \in PCA_i \wedge l' \notin PC_i \\ \chi\pi_g^{1,i} & if \;\; l' \in PCA_i \wedge l' \in PC_i \wedge d(l') < d^- \\ (1-\chi)\pi_g^{1,i} & if \;\; l' \in PCA_i \wedge l' \in PC_i \wedge d(l') > d^- \\ 0.5\pi_g^{1,i} & if \;\; l' \in PCA_i \wedge l' \in PC_i \wedge d(l') = d^- \\ \pi_g^{2,i} & if \;\; l' \in SCA_i \\ 0 & otherwise \end{cases} \tag{19}$$

where $PCA_i$ and $SCA_i$ are sets of cells defining the primary and secondary catchment area of an $lg$-facility, $PC$ is the union of the primary catchment areas of all competing facilities (for $g$), $0 \le \chi \le 1$ is a loss in penetration due to competition, $d(l')$ is the distance between $l$ and $l'$ and $d^-$ is the current nearest distance to a competitor (if any) from $l'$ ($l$ not included). In this equation, the catchment areas are defined based on radius parameters specific for the segment $i$. This represents the notion that, for example, people may consider a larger radius for a destination when they are at home than when they are at the work location, and so on. The equation further represents several simplifying assumptions typically made in catchment area analysis: penetration outside the secondary catchment area is zero, the presence of specific competitors is taking into account only in overlapping areas of *primary* catchment areas and penetration does not decay with distance within the same catchment-area shell (primary or secondary).

As implied by equation 19, cannibalism accounted for occurs only in overlapping areas of primary catchment areas. Consistent with equation 19, the estimated size of cannibalism is calculated as:

$$V_{lg}^{i,ca} = P_i^g P_g \sum_{l' \in PCA_i \cap PC_i} \pi_g^{1,i} N_{l'i} \tag{20}$$

However, in site assessment, the value of cannibalism relative to a total market potential of the site is relevant rather than the absolute value. Total market potential is conceptualized as the size of demand attracted according to (18) if there were no competition. Hence, it is calculated as:

$$V_{lg}^{i,po} = P_i^g P_g \left( \sum_{l' \in PCA_i} \pi_g^{1,i} N_{l'i} + \sum_{l' \in SCA_i} \pi_g^{2,i} N_{l'i} \right) \tag{21}$$

Then, the cannibalism ratio is calculated as:

$$\chi_{lg} = \frac{\sum_i V_{lg}^{i,ca}}{\sum_i V_{lg}^{i,po}} \tag{22}$$

Equations 15 – 22 reflect our assumptions about the limited knowledge suppliers have about the system and in particular the behavior of individuals. The radius and penetration parameters all refer to location choice of individuals. Yet, the choice of the values of these parameters is based on expertise of the suppliers rather than on direct information on the decision rules of individuals. The chosen values are best guesses based on long time experience and recordings of actual penetration rates. On the other hand, the $N$ and $P$ terms represent aggregated results of micro-level decisions of individuals. We assume that knowledge regarding $P_g$ terms is accurate, i.e. consistent with actual activity choice probabilities in the population, whereas knowledge regarding $P_i^g$ is tentative, as data of the latter variables is more difficult to obtain. Finally, regarding the $N$ variables we assume that the agents have accurate data about the residential population, the population at work locations and the size of existing facilities in all demand sectors. So, where they can observe $N$ for housing and employment, they derive estimates of facility visitors volumes based on actual facility sizes.

The optimal size of a (possible) facility is derived from the (estimated) number of visitors. The method can be written as:

$$S^*_{lh} = \Delta S^*_{lg_h} + S^*_{lh'} \tag{23}$$

where $S_{lh}{}^*$ is the optimal size of facility $h$ in cell $l$, $\Delta S^*_{lg_h}$ is the optimal size of added supply, $g_h$, of facility $h$ and $h'$ is the (next) lower-order facility of $h$. This equation has the same structure as equation 15. Similarly, the optimal size is found as the sum of optimal sizes across the supply layers constituting the facility. The optimal size of added supply is defined as ($g = g_h$):

$$\Delta S^*_{lg} = \sigma_g V_{lg} \tag{24}$$

where $\sigma_g$ is a normative floor space size per visitor per day and $V_{lg}$ is the number of visitors per day as observed or estimated (using equation 18) depending on whether the facility is currently operating or not.

The general objective of each agent is to develop as many facilities as possible within several feasibility constraints. First, a facility must be viable according to the rule:

$$\Delta S^*_{lg} \geq S^-_g \qquad \forall g \in G_h \tag{25}$$

For mix facilities this constraint is applied to the overall facility level instead of the added-supply level. Consequently, a mix facility may be feasible even if one or more of the elementary facilities offering the constituent services would not be feasible. In this way a mix facility can benefit from economies of scale. Second, the cannibalism incurred by the facility must be tolerable according to the rule:

$$\chi_{lg} \leq \chi^+_g \qquad \forall g \in G_h \tag{26}$$

where $S_g{}^-$ is the minimum normative size of added supply $g$.

The sector-level agents ($B$) are responsible for making location decisions. These agents continuously monitor the study area for opportunities to open new outlets. Hereby, they consider facility types (in their sector) in a certain order of priority, which is a parameter in the model. Giving priority to higher order facilities, generally, would be in line with the objective to develop the largest possible facility network in a sector. Having selected the facility type, an agent considers a cell for possible development only if the land use in the cell allows the type of development and there are no existing facilities competing for the same demand in the cell. Given these cells, the agent selects the cell maximizing $V_{lg}$ within constraints (15) and (16) and opens a facility of optimal size according to (13) at the site. This process is repeated until none of the cells turns out to be feasible. Then, the next facility type in the priority order is considered and the same process is repeated for this facility, and so on.

In addition to a priority order on subagent level, the system also uses a priority order on the higher demand-sector level. In a monitoring stage, the visitors flows are known and each (sector level) agent reconsiders the size and evaluates the viability of existing facilities within their sector. A closure is indicated when the optimal size has dropped below the minimum. The agents do not close more than one outlet at a time, to avoid the risk of closing more outlets than is necessary. Closing an outlet generally improves the market conditions for other outlets competing for the same demand. Therefore, a facility that is not viable in the current time step may become viable in the next time step if a competing facility is closed. Giving higher priority to maintaining larger outlets, an agent ranks currently unviable outlets first on their order and next on their performance, and, then, closes the outlet with the lowest rank. The agents of the different sectors implement such adaptations simultaneously assuming that possible cross effects between sectors can be ignored.

## 3. Conclusions and discussion

This compilation of papers has summarized some recent progress in the development of the Albatross model system and how the simulated predicted activity-travel patterns can be linked in the context of an agent-based model with a model of land development. The focus of the latter has been on consumptive facilities, but similar research on office development has been completed recently. Future work will (i) complete the uncertain-web enabled version of Albatross, (ii) allow a link with the dynamic based activity-travel demand system currently under development, (iii) also develop this model under conditions of uncertainty, and (iv) integrate (dynamic) activity-travel demand with agent-based simulations of land development and land use changes, expanding the system to various types of land use.

**Acknowledgements**
Part of this work has been funded by the European Commission, under the Seventh Framework Programme, by Contract no. 248488 within project The Uncertainty Enabled Model Web. The views expressed herein are those of the authors, and are not necessarily those of the European Commission.

## 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 (2004), A learning-based transportation oriented simulation system, *Transportation Research B*, 38**,** 613-633.

Arentze, T.A. and H.J.P. Timmermans (2005) Albatross - II: 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 (2004), Multi-agent models of urban land development: Theory and numerical simulation. In: *Proceedings 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., USA, USA (CD-Rom: 19 pp).

Rasouli, S., T.A Arentze and H.J.P. Timmermans (2011), Error propagation in complex large-scale computational process models of activity-travel behavior. In: W.Y. Szeto, S.C. Wong and N.N. Sze (ed.), *Transportdynamics*, HKSTS, Hong Kong, China, pp. 291-298.

Timmermans, H.J.P., T.A. Arentze, S. Cenani, H. Ma, A. Pontes de Aquino, F. Sharmeen and D. Yang (2010), U4IA: Emerging urban futures and opportune repertoires of individual adaptation. In: *Proceedings of the Bi-Annual Design and Decision Support Systems Conference* (CD-Rom, 19 pp.)