The application of microscopic activity based travel demand modelling in large scale simulations

Georg Hertkorn, Peter Wagner
German Aerospace Centre, Institute of Transport Research
Rutherfordstraße 2
12489 Berlin
Germany

e-email: georg.hertkorn@dlr.de
Tel.: ++49/30/670 55 232

Abstract

The concept of microscopic modelling is intuitively understood because no artificial aggregate structures and abstract variables need to be introduced. Constraints can be incorporated in a direct, realistic way, and it is possible to analyse the results on different aggregate levels. Depending on the research or planning focus trips may be distinguished in terms of the quarter where they start or using socio-demographic or even life-style characteristics of the travellers.

One of the challenges of microscopic modelling is to address all aspects of a trip like purpose, departure time, destination and mode. Often two or more of these aspects are interrelated, and the type of interdependeny is not the same for all trips. In this paper a modelling approach is presented where destination and mode choice are combined in one modelling step. A preliminary mode choice is used as a prerequisite to determine the travel times for possible destinations. The travel times control the destination choice using the model of intervening opportunities. Travel demand is simulated for a synthetic population of the City of Cologne. Variables like total distance per traveller, trip length distribution and mode choice distribution are considered, and the impact of the geographical structure of the area under investigation and varying travel times are discussed.

Keywords: activity based travel demand modelling, time use, destination choice, intervening opportunities, mode choice

Topic: D1 Passenger Transport Demand Modelling

Introduction

In his overview over theories and models of activity patterns Timmemans characterizes microscopic models as data driven compared to models where theories about behavioural principles play a more prominent role (Arentze and Timmermans, 2000). In this line of argument the model presented here fits well into this classification as travel activity patterns are not generated from scratch according to some theory of time allocation; instead the patterns are obtained from empirical data that were collected in a nationwide study in Germany by the Federal Statistical Office. Another aspect for the classification of models are the entities being handled by the model. In this case these are inhabitants of the area under investigation rather than spatial aggregates like traffic zones. The microscopic scale seems to be the most appropriate for the implementation of an activity based approach. Furthermore, microscopic models allow to take constraints into account directly, e.g. it can be made sure that trip chains are consistent and that a vehicle is used for all trips in a chain, except for sub-tours. Other examples are the number of vehicles in a household, the number of shops within a ten minutes walking distance, the availability of public transport. Those variables influence the different aspects of a trip: departure time, destination choice and mode choice. It is a challenge in microscopic modelling to deal with the interdependency of these aspects. In addition, the actual or at least the typical conditions of the traffic network have to be considered. They lead to varying travel times, especially for trips by car. Thus, travel times depend on the time of day when the trip is made and on the available routes.

In this paper a model is presented that incorporates all these aspects. The modules that describe destination choice and mode choice are explained in greater detail, because we want to discuss the influence of travel times on these two aspects. Two scenarios for the travel times by car are considered:

a) empty network,

b) travel times of (a) multiplied by a factor of 2.

These scenarios are applied to the City of Cologne with about one million inhabitants for a typical working day.

Structure of the model

The model needs three different types of input data: a synthetic population, a synthetic town, and time use data. The synthetic population consists of individuals with home locations and socio-demographic characteristics, and it is obtained from marginal distributions using iterative proportional fitting (Beckman et al., 1996). For the test case of this paper the process is described in detail in Hertkorn (2002). To explain the algorithms of the model we call a member of the synthetic population representing an individual in reality an *agent*.

The synthetic town are the locations for activities, e.g. working places, shops, gyms. For activities like shopping with a great number of possible locations, the data stem from a model on a zonal basis, for other activities the coordinates of every single possible location are known, e.g. for theatres and cinemas (Rindsfüser et al., 2002). In general, a scenario may refer to a point in time in the future and the synthetic town would be the outcome of some land use model.

Time use patterns are the third type of input. They are the result of a two-step classification process of diary data collected by the Federal Statistical Office in Germany. In this survey the respondents were asked to fill in diaries for two consecutive days. The second part of the survey provides many variables about the socio-demographic situation of the respondents and their households (Ehling and Bihler, 1996). The data set comprises 30700 diaries from more than 7000 households. Each diary is a sequence of 288 activity codes (one for each five minutes interval). Sequence alignment is applied to determine the similarity of each pair of diaries (Wilson, 1998; Sankoff and Kruskal, 1983). Agglomerative clustering leads to a similarity-based classification of the diaries. The second step of classification is on episode level, where an episode is a time period continuously spent with one activity. By comparing episodes of the same type, parameters for their flexibility in starting time and duration are determined (Hertkorn and Kracht, 2002). As the simulation is done for a typical working day, the diaries selected for this analysis refer to a Tuesday, Wednesday or Thursday.

The travel demand model itself deals with one individual at a time. The first step is to select a time use pattern on the basis of the socio-demographic characteristics of the person and her household. The trips in the pattern need to be completed with destination, mode and departure time. The choice of destination and mode are strongly interrelated because it is assumed that destination choice is sensitive to the travel times which in turn depend strongly on the means of transport. In this study we discuss how the location/mode choice procedure reacts on different travel times for car trips, therefore this procedure is explained in more detail in the following section.

We call the pattern complemented with destinations and modes a schedule. For every schedule it has to be decided, if it can be realised in the environment where the individual lives. Usually the travel times that the agent experiences in the model will differ from the travel times in the activity pattern. This induces a time stress in the schedule, depending on the magnitude of the differences in travel times and on the rigidity of the episodes in the pattern. The flexibility in starting times of the episodes is exploited to shift starting times in a way that minimizes the total time stress. If after this equilibration the value of total time stress exceeds a threshold value, it is assumed that the pattern is not likely to occur in this form in reality and another pattern is selected (Hertkorn et al., 2003).

Finally, origin, destination and starting time of the car trips are written to a trip table that can be used as input for a traffic flow simulation. One of the results of the traffic flow simulation are travel times in the network. In an iterative process they are used for the next run of the travel demand model. This cycle is repeated until a self-consistent situation is obtained. First attempts to run a traffic flow simulation with the trip tables of the travel demand model were not successful. The traffic flow model did not yield useful travel times for a demand that locally may be far from equilibrium. So, travel demand for an empty network and for artificially increased travel times is discussed.

Mode Choice

Mode choice as a crucial aspect of a trip depends itself on many other aspects and the attributes of the traveller. Possible influencing variables are: purpose of the trip, number of cars in the household, sex and age of the person, distance of the trip. The full combination of all of these variables would lead to a huge number of cases to be differentiated. It would be expensive to provide an empirical data set big enough to estimate each case in a reliable way. The CHAID-Algorithm (chi-square automatic interaction detection) (Kass, 1980; Baltes-Götz, 2001) allows to reduce the number of combinations to those that are significantly different with respect to the dependent variable. It builds up a tree structure: in each step the variable leading to the most significant subdivision of the data set is determined. Often two or more variables provide a significant distinction at an arbitrarily small level of significance with respect to the χ^2-test. In these cases the value of the Goodman-Kruskal-Tau (Ludwig-Mayerhofer, 1999) is used to select the variable for a split. The data set for the estimation of the CHAID tree stems from a nationwide survey in Germany, *Mobilität in Deutschland, MiD* (Bundesministerium für Verkehr, Bau- und Wohnungswesen, 2003; Clearingstelle für Verkehrsdaten und -modelle, 2003). 100000 respondents were asked about all of their trips at a given day. The CHAID-tree is built on the basis of 44000 trips because only trips of those respondents are taken into account that live in a region of similar density compared to the area under investigation, here. Table 1 shows at which level the variables occur in the tree. The share of cases that are influenced by the corresponding variable is given in the last row of the table. This share is only a rough indicator for the importance of a variable because it reflects also that the cases are organized in the structure of a tree, e.g. the variable of the root node influences necessarily 100% of the cases. Here this variable is *number of cars in the household*. At the second level four of the five nodes are subdivided by the *distance of the trip* and one by *age of the traveller*.

	Number of occurrence					Cases
Level	1	2	3	4	5	
Variable						
Cars in the household	1					100.0 %
Distance of the trip		4	5			95.7 %
Age		1	7			80.8 %
Purpose of the trip				7		30.9 %
Sex				2	7	33.3 %

Table 1: Structure of the CHAID-tree for mode choice.

According to the CHAID-tree, the distance of the trip has a strong influence on the mode choice. On the other hand, the destination choice is based on travel time, which in turn depends on the mode. The next section explains, how mode choice is integrated in the destination choice procedure.

Destination choice

The basic idea of the destination choice algorithm in this model is the ansatz of intervening opportunities (Ortúzar and Willumsen, 1999). People use the closest destination for a given purpose if there are no reasons not to do so. These reasons can be unawareness of locations or personal taste. The ansatz of intervening opportunity assumes that the closest location (usually in terms of travel time) should be preferred by the traveller. Because of the reasons mentioned above, there is some probability q that the location may not be taken into account by the traveller or that it does not meet his needs. Of course the same holds for the second closest destination and all the others. Given the destination in a list sorted by travel time, the probability for the ith location to be chosen is

$$f(i) = pq^{i-1}; \quad p = 1 - q\,. \tag{1}$$

Two tasks have to be accomplished to apply this ansatz. First, the possible destinations have to be sorted by travel times, second, the appropriate value for q has to be determined.

For the first task a preliminary mode is chosen in order to get a value for the travel time which is the criterion for the order of the list. For some activities there exists a great number of possible locations. So it would be very time consuming to perform the mode choice for all of them. Therefore, the area under investigation is subdivided into 92 zones. The zones are ordered instead of single locations using the travel time criterion. A zone is selected according to equation 1, and within this zone a location is drawn at random.

The second task requires a survey where the home location of the respondent is known as well as his working place and the location of all other possible destinations for each trip. Usually such a data set is not available. As an alternative the value of p for different groups of travellers can be calibrated indirectly using distance or travel time from usual travel surveys. As travel times and distances in turn depend on travel demand, another iteration loop has to be set up to obtain a self-consistent situation.

Distances travelled and mode choice under free network conditions

For the interpretation of the results a short description of the geographic peculiarities of the area under investigation is given. As can be seen in figure 1 the population density is quite heterogeneous within the administrative borders of the City of Cologne. The mediaevial city centre, *Altstadt*, on the left riverside of river Rhine is easily identified. The quarter boundaries follow the line of the ancient city wall. It was destroyed in 1881, and today a major road follows its course. The river flows from the south to the north. Five bridges are in the vicinity of the city centre, and there is an autobahn that crosses the river 10 km to the north of the centre.

Only in the last century some independent municipalities joined the City of Cologne. Some of them may still be identified by a relatively high population density like *Kalk* on the left hand riverside, then *Porz* in the south of *Kalk*, and *Weiden* at the western boundary of of the map. The high density in the district of *Chorweiler* in the north of the city centre

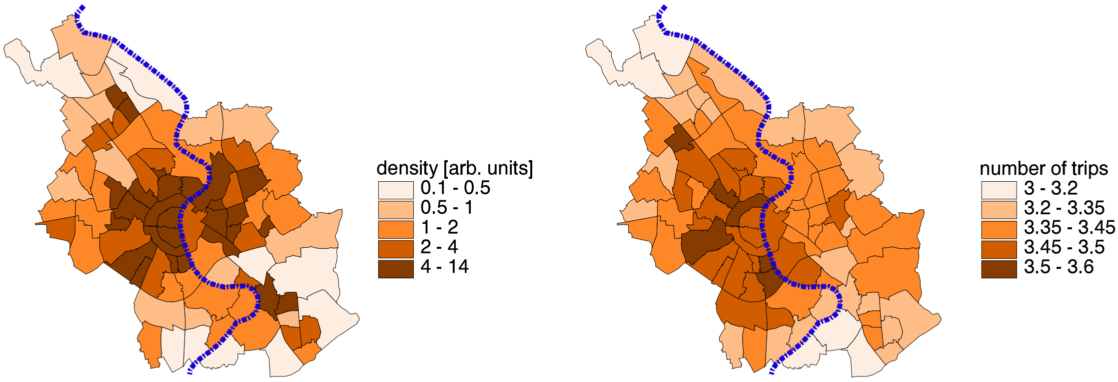


Figure 1: Cologne: population density (left) and number of trips per traveller (right).

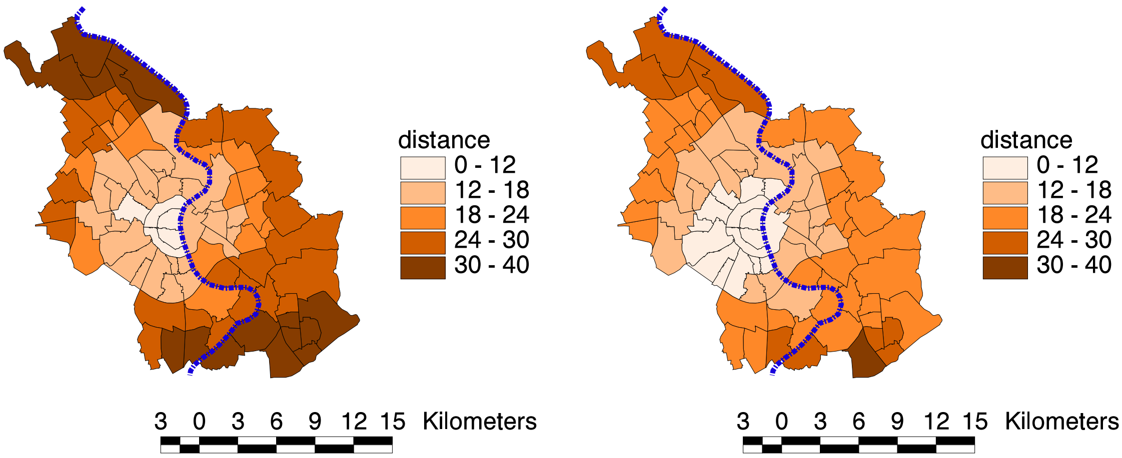


Figure 2: Distance [km] per person and day for travel times in an empty network (left), for double car travel times (right).

is due to the tower block housing that was built there in the 1960s. The right hand side of figure 1 illustrates the number of trips per person and day as estimated by the simulation for a weekday. Only persons with at least one trip are taken into account. Again the picture is heterogeneous, which is due to socio-demographic peculiarities of the quarters. Small trip rates occur in a corridor-like zone in the north-west and in some quarters in the south-east.

We present these maps to contrast them with the maps of total trip length per day, figure 2. Here the city centre plays a predominant role. Only the central quarters exhibit a mean total travel distance below 12 km. For the other quarters the value rises with growing distance from the city centre. However, the river has an impact on the travel distances as the zone of smaller travel distances extends to the north and south along the left hand river side. On the right hand side of the river the travel distances are at a higher level: 21.1 km compared to 17.2 km on the left hand side (city centre excluded). This is plausible if the city centre

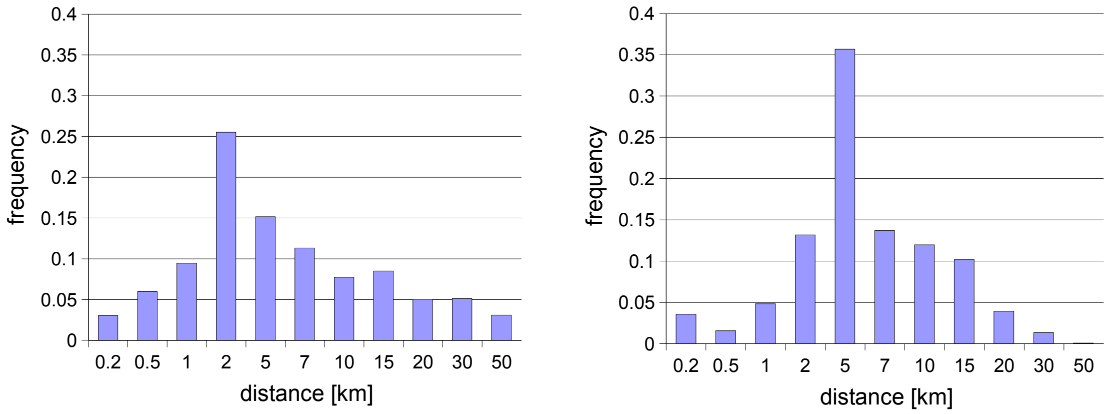


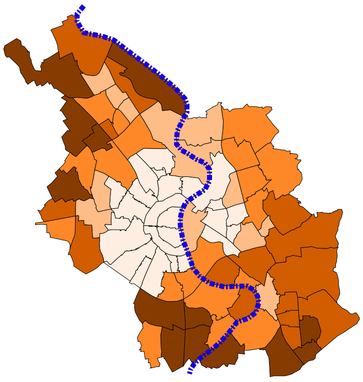
Figure 3: Distance per trip; empirical data (left) and simulation (right).

attracts many trips. Some extra distance has to be covered for inhabitants of the right hand river side to get to one of the bridges.

The distance travelled per traveller and day in the survey *MiD* is 24 km and somewhat above the values obtained in the simulation. However, a look at the trip length distribution for the single trips adds some interesting aspects. Very long trips above 20 km are very rare in the simulation because of the finite size of the simulated area. But the mode in the distribution of the simulated trips is between 2 and 5 km whereas it is between 1 and 2 km in the survey data (see figure 3). It is not clear whether this is caused by the spatial structure of Cologne or by the propensity of the agents in the model to travel by bicycle or car. In order to study how the destination choice and mode choice reacts on rising travel times we run the simulation with car travel times multiplied by a factor of 2. There is a noticeable drop in the overall level of distances travelled (figure 2), but the monocentric picture prevails. In the distribution of trip lengths the mode remains at 2 to 5 km, and is even more pronounced (38.5%) at cost of the classes of longer distances.

The relatively short travel times in the empty net promote the usage of the car (figure 4). The fact that the number of trips as car passenger is much lower than in the survey data can be explained by the distribution of household sizes. In Cologne 45% of the households are single households, compared to 19% in *MiD*. For non-motorized trips the bicycle seems to be very attractive. There is a big university in Cologne and several colleges, and the share of students among the inhabitants my be the reason for the high number of bicycle trips. But it has to be checked if the willingness to use the bicycle is overestimated by the mode choice procedure, because the bicycle allows to reach much more locations in a given amount of time compared to walking. The share of trips by private car is different for different quarters. It is low in the city centre on both sides of the river and in some quarters in the north-west and south-east. The map of the number of cars per person shows the same pattern, which is not surprising as the number of cars per household is used in the model for mode choice. However, this correspondence could indicate that car ownership is a function of accessibility.

The second scenario with an increase of car travel times by a factor of two makes the



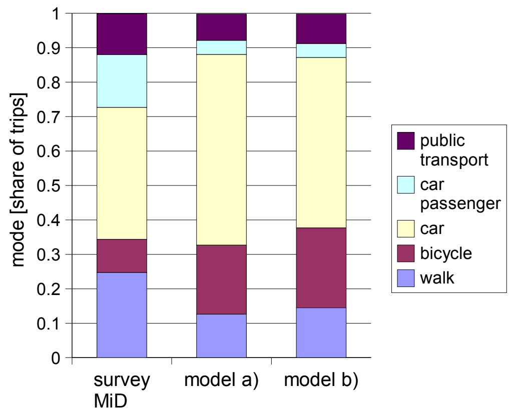


Figure 4: Share of car trips (left), mode distribution (right).

car less attractive. This is an indirect effect: far locations get a higher index position in the destination choice list if the preliminary mode choice suggests a car trip to get there. Closer locations within walking or cycling distance are chosen with higher probability. It is interesting that public transport does not benefit very much from the increase in car travel times. This could indicate that public transport is used in specific occasions or for specific relations, and that it is difficult for the travellers to substitute car trips with trips by public transport.

Conclusion

The activity based, microscopic modelling approach offers the opportunity to study the impact of policy measures not only on a detailed spatial level but also differentiated by groups of travellers. It has been shown how the interplay of mode choice and destination choice can be incorporated in a microscopic traffic demand model. The application of the model to the City of Cologne shows that both, mode choice and destination choice are sensitive to car travel times. Whereas the population density and car ownership are heterogeneous in the area under investigation, the average distance travelled per day increases monotonically with the distance from the city centre. It will be investigated whether substructures become visible when activity types are analysed separately. Furthermore, as the model allows for fast computation of travel demand, it is possible to feed back travel times from dynamic assignment and traffic flow simulation in order to study the influence of bottlenecks in the network and differences in the accessibility of the quarters.

References

Arentze, T. and Timmermans, H. (2000). *Albatross — A Learning Based Transportations Oriented Simulations System*. EIRASS, European Institute of Retailing and Services Studies, Technische Universiteit Eindhoven.

Baltes-Götz, B. (2001). Segmentierung und Klassifizierung mit AnswerTree 2.1. www.uni-trier.de/urt/user/baltes/docs/at/v21/at21.pdf, access: Dec 19th 2003.

Beckman, R. J., Baggerly, K. A., and McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A*, 30(6):415–429.

Beckmann, K. J. (2002). SimVV Mobilität verstehen und lenken — zu einer integrierten quantitativen Gesamtsicht und Mikrosimulation von Verkehr. Abschlußbericht, Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (MSWF), Düsseldorf.

Blanke, K., Ehling, M., and Schwarz, N. (1996). *Zeit im Blickfeld : Ergebnisse einer repräsentativen Zeitbudgeterhebung*, volume 121 of *Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend*. Kohlhammer, Stuttgart.

Bundesministerium für Verkehr, Bau- und Wohnungswesen (2003). Mobilität in Deutschland 2002 — Kontinuierliche Erhebung zum Verkehrsverhalten. Endbericht 70.0681/2001, Bundesministerium für Verkehr, Bau- und Wohnungswesen der Bundesrepublik, Berlin.

Clearingstelle für Verkehrsdaten und -modelle (2003). KONTIV/Mobilität in Deutschland. http://www.clearingstelle-verkehr.de access: Nov 17th 2003.

Ehling, M. and Bihler, W. (1996). Zeit im Blickfeld. Ergebnisse einer repräsentativen Zeitbudgeterhebung. In Blanke et al. (1996), chapter 11, pages 237–274.

Hertkorn, G. (2002). Synthetische Bevölkerung. In Beckmann (2002).

Hertkorn, G. and Kracht, M. (2002). Analysis of a large scale time use survey with respect to travel demand and regional aspects. In Elsa Fontainha and Heloísa Perista, editor, *Work Time and Leisure Time — Dynamics and Convergence in Changing Contexts*, Lisbon, Portugal. International Association for time use research (IATUR) - Conference 2002. http://ivf.dlr.de.

Hertkorn, G., Kracht, M., and Wagner, P. (2003). Travel demand modelling based on time use data. Paper presented at the 10th International Conference on Travel Behaviour Research, 10-15 August 2003, Lucerne.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2):119–127.

Ludwig-Mayerhofer, W. (1999). Internet-Lexikon der Methoden der empirischen Sozialforschung, Tau (Goodman und Kruskal). http://www.lrz-muenchen.de/~wlm/ilm_t2.htm access: Dec 5th 2003.

Ortúzar, J. and Willumsen, L. G. (1999). *Modelling Transport*. John Wiley & Sons, Chichester.

Rindsfüser, G., Ansorge, J., and Mühlhans, H. (2002). Aktivitätenvorhaben. In Beckmann (2002).

Sankoff, D. and Kruskal, J. B., editors (1983). *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Addison-Wesley, Reading, MA.

Wilson, W. C. (1998). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A*, 30:1017–1038.