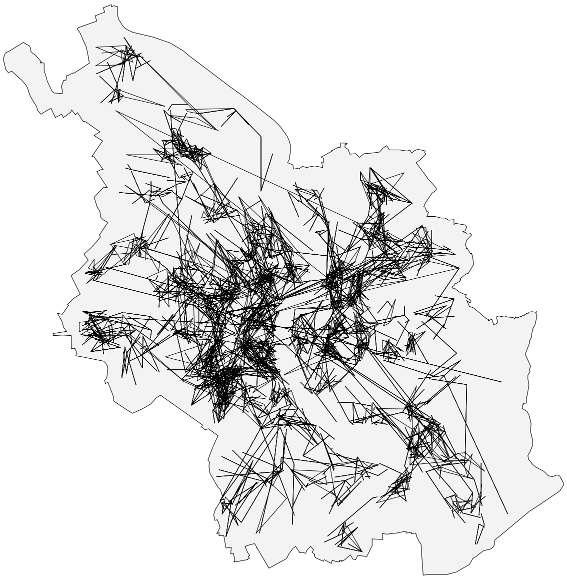


Travel Demand Modelling Based on Time Use Data

Georg Hertkorn, German Aerospace Centre
Matthias Kracht, German Aerospace Centre
Peter Wagner, German Aerospace Centre

Conference paper
Session VI

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

Travel Demand Modelling Based on Time Use Data

Georg Hertkorn
Institute for Transport Research
German Aerospace Centre
Berlin

Phone: 049-030 670 55 232
Fax: 049-030 670 55 202
eMail: georg.hertkorn@dlr.de

Abstract

Several approaches have been proposed to model travel demand on the basis of activity-patterns. There are modelling concepts where travel behaviour is regarded as the result of discrete choices for the type of activity, departure time, destination, and mode. Other researchers emphasise the interdependencies of different activities and the scheduling process. The idea of the modelling approach which is presented in this paper is to estimate the travel demand in large cities with as few parameters as possible. This ensures that the mechanisms of the simulation do not become too complicated. Very complex models tend to hamper the analysis of the impact of the different influences. Furthermore, efficient computation makes the integration within larger modelling frameworks easier, and different scenarios can be compared in reasonable time. The concept is based on a detailed analysis of diary data. The data where collected in a German nationwide time use survey in 1991/1992. The set of 30,700 diaries is classified by a modified cluster algorithm. This modification allows to handle large data sets. It is possible to distinguish clusters with different levels of heterogeneity. The distances between pairs of diaries, which are needed as input for the clustering procedure, are computed using a sequence alignment algorithm. The sequence alignment method is sensitive not only to different activity budgets but also to the time of the day, when the activities are executed. The classification process reveals which activities cause the differences in the time use patterns, and it turns out that the diary classes possess interesting features when cross tabulated with categories of administrative districts. Especially in rural areas there are pronounced preferences for specific time use patterns. The classes with higher shares of diaries that stem from those areas are different with respect to timing rather than total activity budgets. A new time use survey is conducted in 2001 and 2002, and it is planned to apply the analysis procedures to the new results. As such surveys can not be conducted in every planning or research project, a link between activity-pattern and socio-demographic variables is needed in order to take into account the peculiarities of the area under investigation. This link is established by the correlation of diary classes and socio-demographic groups. The formation of the socio-demographic groups shows which variables are the most important for the estimation of the frequency of time use patterns. The simulation computes activity-travel-patterns for the members of a synthetic population, which serve as input for a traffic flow simulation. The travel times are fed back to the module for the activity-travel-pattern generation and influence the choice of locations, modes and departure times. The simulation model is applied to the city of Cologne in Germany which has about one million inhabitants. The city is the test bed for a conjoint traffic research project, and therefore a good data basis for the generation of a synthetic population is available.

Keywords

classification of diaries, time use, urban travel demand, microscopic modelling, International Conference on Travel Behaviour Research, IATBR

Preferred citation

Hertkorn, Georg, Matthias Kracht and Peter Wagner (2003) Travel Demand Modelling Based on Time Use Data, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. Introduction

Why do people make trips? In activity based travel demand modelling the answer to this question is that people need to change places in order to participate in different activities. As a consequence, the task of travel demand estimation is shifted to the task of modelling the time allocation of individuals. It is now possible to capture the interdependencies between different trips and to study the mechanisms of trip generation on a microscopic level. There are two reasons why this is a complex problem. First, depending on the level of description, the number of possible time use patterns or schedules is huge. Even if only a modest number of activity types are differentiated they can be combined in many different orders. For each of these sequences there are many possible combinations of starting times which again have to be multiplied with the combinations due to different locations where out-of-home activities are performed. The second reason are the interdependencies of the various aspects of a schedule. Which activity is performed at a given point in time may depend on the time of day, on other activities, or on the locations that are accessible. On the other hand, there are activities which control the time when they are executed, and the location, and other aspects.

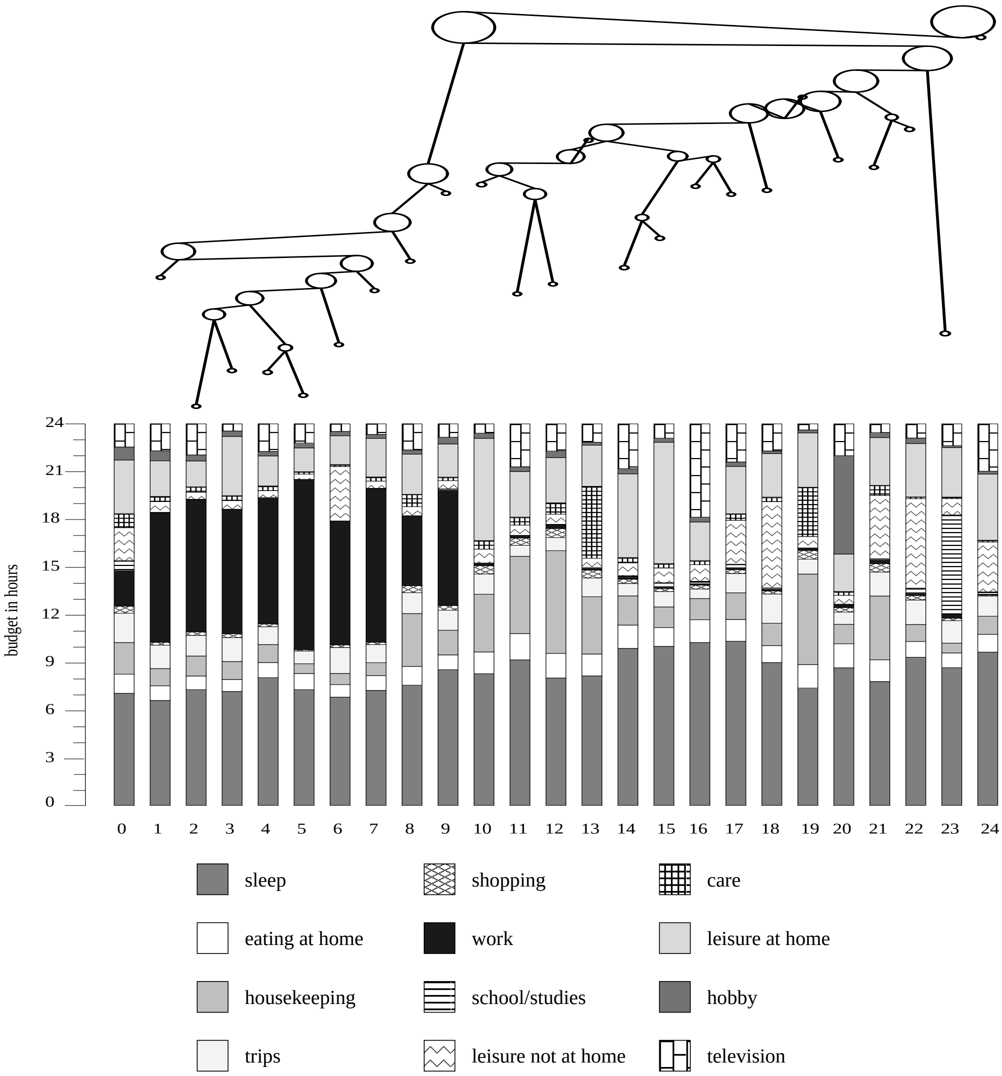
Several approaches have been proposed to handle the complexity of time use modelling. One major approach is to classifiy the schedules by building a hierarchy of the different aspects. This is done by Bowman and Ben-Akiva (2001). First, they distinguish days without a trip from days with one or more trips. If trips occur, they are grouped into tours, i.e. is a sequence of trips and stops that starts and ends at home at home. A day can contain only one primary tour which may be combined with one or more secondary tours. The primary tours are differentiated by the number of stops and the presence of a subtour. Other aspects, like time of the day or travel modes are considered in subsequent steps. In the model of Lipps (2001) the activities are ordered by priority. Then each tour is denoted by the activity with the highest priority. The empty pattern has for empty positions. First, the tour whose major activity has the highest priority is determined, and set to the second position. If other tours follow after the main tour, they are set to the third and forth position, respectively. If there are tours before the main tour, then the first of these is set on the first position. The description of diaries by patterns helps to analyse empirical data and serves as a basis for the creation of synthetic schedules. On the other hand, if the reaction of people to different scenarios is analysed, then the responses will always be confined to the pattern that guides the modelling process. In computational process models the complexity is reduced by setting up a system of rules that govern the formation of a schedule. Arentze and Timmermans (2000) developed a system where the formation of a schedule is broken down into six steps. The rules that are applied for each step are derived from the results of the previous steps, and from a number of constraints. When the rules are formulated it is not only neccessary to ensure the consistency of the system. It also has to be explained how rules or the parameters of rules can be obtained from empirical data.

In the approach that is presented here, the question how time is organised by individuals under certain conditions is reduced to the task of finding a reasonable time use pattern for each person and to adapt it to the specific environment where the person lives. To do this, a mechanism has to be found how patterns can be derived from diary data with a minimum of *a-priori* assumptions of how a schedule is structured. Then it has to be determined how patterns are assigned to the individuals in the simulation, and which characteristics of the individuals control this assignment. Finally, it has to be decided, if the environment where a person lives allows to complete the pattern with locations that are accessible in reasonable time.

2. Data

The model was developed using diary data from the time use survey of the Federal Statistical Office of Germany in 1991/92. In this survey the respondents were asked to fill in diaries for two subsequent days. The forms for the diaries provided one line for every five minutes interval, and the names of the activities were chosen by the respondents. The activities were coded afterwards using a catalogue of 231 activities. 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.

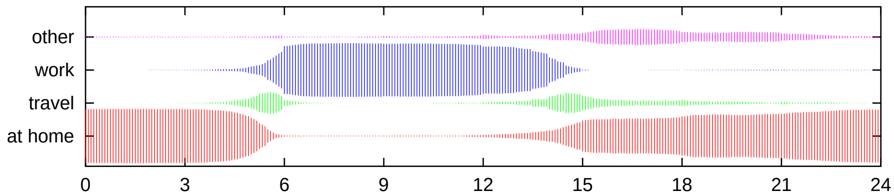
Figure 1: Classification of diary data and budgets by classes



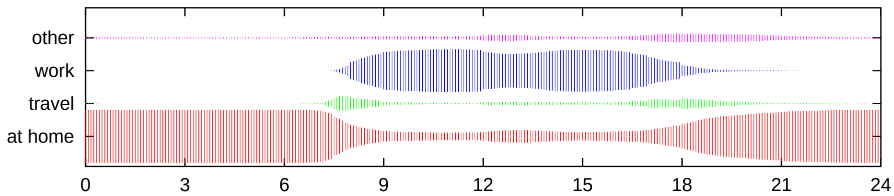
The aim of the first step of the analysis is to distinguish different categories of diaries in order to reveal the structure of the data set. This is done with an agglomarative clustering algorithm. Clustering is based on a distance measure of the elements of the set. Wilson (1998) proposed to compare diary data by sequence alignment methods. Joh *et al.* (2002) extended the comparisons to several dimensions, e.g. location and transport mode are taken into account in addition to activity codes. In our case, the diaries are stored as sequences of fixed length with 288 elements. If the diaries are not identical, the alignment is done by inserting additional elements in the sequences or by replacing some elements. The number of such operations is used as a measure of the dissimilarity or distance of the sequences. A detailed description of the method is given by Sankoff and Kruskal (1983). It turns out that the distinction of 231 activity codes

Figure 2: Sequencing of activities in two diary classes with full-time working hours

Diary class No. 2: Full-time 2



Diary class No. 4: Full-time 4



leads to large distances between the diaries. The acticity catalogue was reduced to only 12 activity types by focussing on the travel generating effect of the activities and by taking into account their impact on the temporal structure of the diaries. E.g. *watching TV* is an activity that is often embedded in a period of time that is spent at home anyway, but if people like to watch a specific programme this has an impact on the schedule of the day. Furthermore, the relatively large total budget of this activity justifies to include it into the activity catalogue.

The clustering shows that the range of values for the variance in the clusters is very big. Hence, using the variance as a general criterion for the number of clusters would lead to clusters of very different size. The way the clusters are used in the model makes a balanced partition more preferable. Therefore, the procedure is split into two steps. First, a complete clustering is performed until the whole set is contained in a binary tree. Then the tree is traversed from the root to the branches and nodes are marked if the size of both of their branches lies above a given threshold. In the second step the leaves of those nodes are used as nuclei for the final classification: each diary is assigned to it's closest nucleus. If the threshold for the minimum size of a nucleus is set to 210, then the ratio of the sum of squares within the classes to the total sum of squares is 51%. This value increases only slowly if the threshold is reduced further.

Figure 1 shows the classification tree that is built during the clustering and the budgets for the twelve activities that where used for the comparison of the diaries. The budgets are the average time that was spent in all diaries in the corresponding class for a given activity. Activities like *sleeping*, *eating at home*, *housekeeping*, *trips*, *leisure at home*, and *watching TV* can be found in all classes, but with varying shares. On the other hand, *school attendance*, *care*, *hobby/work at home*, and *leisure out of home* are characteristic for single classes. *Payed work* is responsible for the splitting of the main branches during clustering. The branch containing working days ends in very homogeneous leaves. The right hand side of the figure shows a greater variety: Almost any of these classes contains an activity that is peculiar to the class, and the variation within the classes is bigger than for the classes on the left hand side. All of the

Table 1: Person categories used to determine the frequency of diaries with respect to diary classes

Group	Av. age	Women [%]	size
Retired (f)	68	100.0	928
Retired (m)	67	0.0	720
Job seeking	43	54.1	738
Students	16	48.5	2349
Employees (m)	39	0.0	4207
Empl. (f) with children	32	100.0	736
Empl. (f) below 42	32	100.0	1648
Empl. (f) above 42	51	100.0	1249
Housewives below 42	35	98.5	715
Housewives above 42	54	98.2	836
Free-lancers	45	28.4	1135
other Workers	37	91.2	102

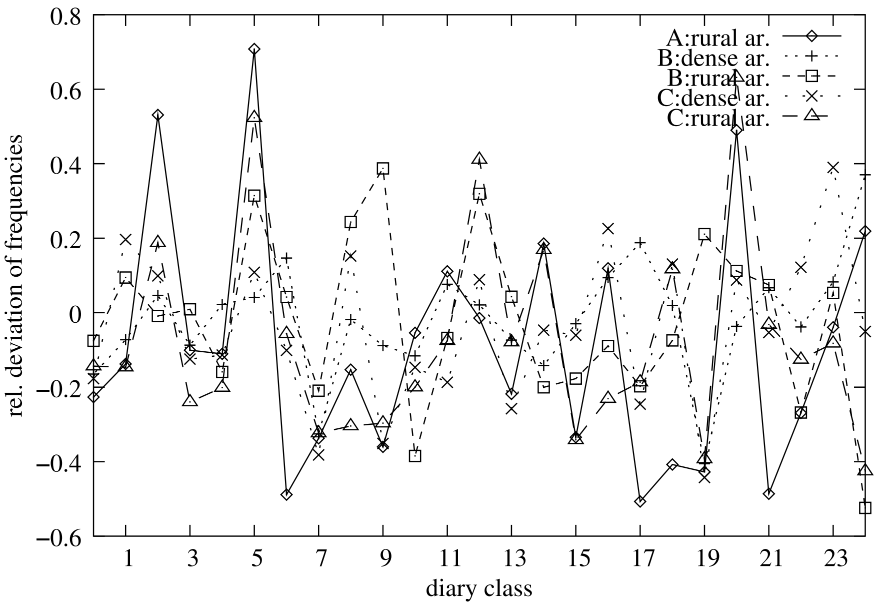
classes 8, 10 – 13, 19, and 21 exhibit more than three hours of *housekeeping*. But this is combined with *payed work* in class 8, with time for child care or care for elderly people in the classes 13 and 19 or with leisure times in the classes 10 – 12, and 21. Class 20 is the only one with several hours of *hobby/work at home* and similarly school days gather in one class, number 23.

It should be noted that the sequence analysis method which was used for the comparison of diaries is not only sensitive on budgets but also on the time of day when the activities are performed. This is illustrated by figure 2. Time of the day is on the abscissa, and the width of the lines corresponds to the share of diaries where an activity is marked for the corresponding point in time. Both, class 1 and class 4 contain full-time working days, but *payed work* starts usually at 6:00 in class 1 and nearly three hours later in class 4. Accordingly, trips are made earlier in class 1 than in class 4. In both classes the peak for travelling is more pronounced in the morning than in the afternoon.

Schmiedel (1984) linked socio-demographic variables directly to the different time use patterns that he obtained from his classification procedure. Here a more flexible approach is applied. It is not claimed that the diary classes correspond one to one to a group of people with more or less homogeneous behaviour. Instead, the correlation of socio-demographic variables with the frequency of diaries in the different classes is exploited to compute a probability for the assignment of time use patterns in the model. The set of the socio-demographic variables is again the result of a clustering procedure. This time, the elements of the set to be clustered are groups within the sample that are obtained by a very fine classification of the population in the sample. The comparison of the groups is based on the shares of diaries in the diary classes. The clusters are characterised by looking for common socio-demographic values among the groups of the original fine classification. Table 1 lists the person categories that were obtained.

The basic population of the time use survey was the entire population of Germany. To get an idea whether it can be justified to use the data for the simulation of travel demand in a certain area, it is important to know whether differences in time use patterns are likely to be influenced by the spatial characteristics of the area. The administrative districts in Germany are classified by the *Bundesamt für Bauordnung und Raumwesen* into 9 categories from core cities in dense regions to rural areas in rural regions. The cross-tabulation of the frequencies of the diaries with respect to diary classes and the categories of districts shows that the latter are only of minor influence compared to socio-demographic variables (Hertkorn and Kracht, 2002). Nevertheless, for most of the diary classes the rural districts show the same tendency when the relative deviations of the number of diaries are computed: $\nu_i = (n_i - e_i)/e_i$ with n_i being the observed number of diaries in class i, and e_i the expected number of diaries in i. (Figure 3).

Figure 3: Relative deviation of the frequencies of diaries for rural areas



Other indispensable sources of information for activity-based travel demand estimation are data about the locations for activities in the city to be studied and data about the travel infrastructure: the street network, public transport connections, and travel times for the different modes. These data were provided by the *Institut für Stadtbauwesen und Stadtverkehr* in Aachen and the Centre of Applied Computer Science (ZAIK) within the project SimVV (Beckmann *et al.*, 2002). The synthetic population for the City of Cologne was also adopted from this project. The population was generated by the authors of this article.

3. The structure of the model

The general idea of the model is to select a scheme for every person of the synthetic population and to complete the scheme with locations for the different activities and travel modes for the trips. We call the person whose schedule undergoes the different modelling steps P. The schemes are derived from the diaries, and to do this the information represented by the classification of the diaries is used. A continuous period of time in a diary which is filled with one activity is called an episode. In a scheme each episode has a starting time (the one that was reported in the original diary) and a duration, but also parameters that indicate the variance of the starting time and duration, respectively. For the calculation of the variance similar episodes in other diaries of the same class are used. The restriction to diaries of the same class enhances the probability that the episode is compared with episodes of the same kind not only with respect to the activity that is performed but also with respect to the structure of the schedule. A leisure activity in a working day schedule may be different with respect to its flexibility than a leisure activity in a non-working day.

The selection of locations is a crucial step in the modelling process, because it determines how the participation in activities is linked to travel demand. The same schedule may require trips of different lengths in a different environment. As a basic principle the model of intervening opportunities is applied (Ortúzar and Willumsen, 1999). If travel time was the only criterion for the choice of a location, the task would be to look for the closest location or to search the combination of locations that leads to the shortest tour in the case of a trip chain. However, the list of alternatives that is used in the model will usually not coincide with the choice set of the individual. Some of the alternatives may be unknown to P or they are not acceptable for reasons of personal taste. Therefore, it is assumed that an alternative is only accepted with probability $p < 1$. Then the probability for the ith alternative is given by the modified geometric distribution: $f(i; p) = pq^{i-1}$ with $q = 1 - p$. The parameter p depends on the mapping of activities to possible locations. If for instance *going to the theatre* and *visiting a organ concert* are both classified as *visiting a cultural event*, then p has to be smaller compared to the situation where these activities are treated separately with two distinct sets of possible locations.

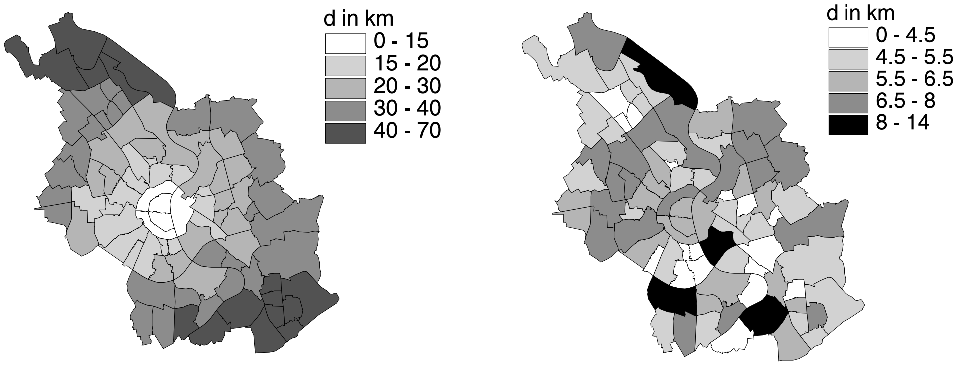
To estimate the travel times that are needed for the ordering of the locations, a mode has to be selected for the corresponding trip. For most of the activities there is an overwhelming number of locations that have to be considered. Therefore, the locations are aggregated to destination zones. The algorithm is applied to zones instead of single locations, the selection is modified to take into account the capacity of the zones, and the location is chosen at random within the selected zone. If several locations are visited on a tour, the evaluation of all possible combinations of locations would be very time consuming. To avoid this, the choices are made successively. The episodes are sorted according to their duration, and the location for the longest episode is selected first, then the locations for the other episodes are chosen. The priority of activities is used for the choice of the travelling mode as well. If an individual vehicle is chosen for one of the prioritised episodes of a tour, then this vehicle must be used for the other trips as well. On the other hand, if no vehicle was taken along to reach the location of a prioritised activity, then it is not available for the other trips, either.

The travel times of the original schemes reflect the situation of the person who recorded the underlying diary. The selection of locations in the area where the model is applied will usually lead to different travel times. The question is, how this affects the starting times and the duration of the episodes. As mentioned above, for each starting time and for each duration of an episode a variance has been computed. Shifts in the starting times or changes in a duration are used to calculate costs for the deviation from the original scheme. The starting times are altered in a way that the sum of the costs reaches a minimum. Thus, a trip in the evening that needs more time than was reserved can lead to an earlier starting time of the episodes that are scheduled before the trip. There is no unidirectional interaction of the episodes. The costs also allow for a feedback in the selection process of the schemes. This reflects the fact that people usually are aware of typical travel times when they allocate their time. If the costs exceed a threshold the scheme is discarded and other schemes are evaluated until an acceptable schedule could be built.

4. The influence of accessibility

In the following the influence of the parameter p on the distance that is travelled during a day is illustrated. The test bed is the City of Cologne. The travelling velocity is held constant, and the same time use patterns are assigned in both runs. In the first run the probability to reject a location for reasons of personal taste is set to the maximum value that still ensures that every person is satisfied with at least one location within the city. The left hand side of figure 4 shows that the distribution of locations is nearly irrelevant under these conditions. The geometric relations lead to a favorable situation for the city centre. When p is set to a value where people find an acceptable location among the first hundred alternatives, the peculiarities of the correspondence of the populations activities in a quarter with the locations there become visible. For a more elaborate interpretation the interplay of the composition of the population and the supply with locations has to be studied, which was not done yet.

Figure 4: Average travel distance per person and day by districts



5. Conclusion

It has been shown how travel demand can be derived from time use date by deriving activity patterns directly from diaries. It is proposed to classify the data set and to derive the variability of time use patterns on the basis of the classes. In a microscopic model the patterns are assigned to individuals of a synthetic population of the City of Cologne, and the patterns are complemented with locations where activities are performed, and with travel modes. In order to reflect the different conditions in different quarters of the city, the starting times of activity episodes can be shifted if travel times deviate from the travel times in the original diary. Schedules are rejected if these deviations exceed an acceptable range.

By focussing on the link between time use and travel demand, it is possible to develop a model with a set of parameters of moderate size. This make it easier to understand the interplay of the different causes for the predicted changes in travel behaviour. In addition, computation is fast. This is important if the model is combined with traffic flow simulation to capture the feed back of travel times on the choice of locations and travel modes.

The schedules that are generated by this model are directly derived from observed diary data. As a consequence, the behaviour that can be exhibited by the individuals in the simulation is confined to the behaviour that was reported by the population of the survey. This constraint is somewhat alleviated by the fact that temporal flexibility is introduced by allowing episodes to be shifted within the schedule. If one judges the data set to cover the variety in time use behaviour well, it is a strength of the model that the same variety can be found in the behaviour of the simulated population, because there are no restrictions due to a schematic description of the schedules. However, constraints due to regulations concerning working hours and opening hours are not represented explicitly in the model. Accordingly, the model can only be applied to scenarios where such regulations remain unchanged and where no drastic changes in time-use behaviour are expected. Therefore, for long term forecasts the model needs to be complemented with a module that tells if people adopt time use patterns from the preceding generation, if they stick to the patterns they are used to, or if there are circumstances where they invent new patterns.

References

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

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

Bowman, J. L. and M. E. Ben-Akiva (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A*, **35** 1–28.

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

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

Joh, C.-H., T. Arentze, F. Hofman and H. Timmermans (2002) Activity pattern similarity: a multidimensional sequence alignement method, *Transportation Research Part B*, **36** 385–403.

Lipps, O. (2001) Modellierung der individuellen Verhaltensvariationen bei der Verkehrsentstehung, Dissertation, Institut für Verkehrswesen, Universität Karlsruhe, Karlsruhe.

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

Schmiedel, R. (1984) Bestimmung verhaltensähnlicher Personenkreise für die Verkehrsplanung, Dissertation, Fakultät für Bauingenieur- und Vermessungswesen der Universität Karlsruhe (TH), Karlsruhe.