Traditional „homogenous behavioural groups“: still valid analysing-tool or scientific anachronism ?

Torsten Luley (torsten.luley@dlr.de)
Julius Menge (julius.menge@dlr.de)
Barbara Lenz (barbara.lenz@dlr.de)
Institute of Transport Research
German Aerospace Center
Rutherfordstrasse 2
12489 Berlin-Adlershof

Abstract: Models of behaviour have the aim to build up reality in a simplified way. A good model is characterized by the possibility to picture human behaviour in the real world with some few determinants in the model. The concept of homogenous behavioural groups dating from the 1970ies offers such a reduction of complexity. The present article tries a first approach to answer the question if this concept is also valid in 2002 using data from the year 2002 survey “Mobilitaet in Deutschland” (“Mobility in Germany”). The intention is to identify general behaviour patterns in the transportation system and to evaluate these patterns with socio-demographic and socio-economic variables against the background of the traditional concept of homogenous behavioural groups.

Key Words: travel behaviour, homogenous behavioural groups, trip chains, Mobility in Germany 2002

1. Problem definition and research question

For long, major objectives of transport research have been to identify and forecast the use of infrastructures by the description of traffic flows in time and space. One widespread methodology to approach this problem has been to analyse the origin and flow of traffic empirically and to translate the results into models. Concerning passenger transport many models use the so-called “homogenous behavioural groups” which refer to the idea that particular socio-demographic and socio-economic attributes match particular ways of travel behaviour. This concept was brought up and tested empirically in the early 1970ies by Eckhard KUTTER who identified nine groups determined by only a few variables. Although since then KUTTER’S work has been discussed more or less controversially within the scientific community and has also found some suggestions for modification, it has still to be regarded as an almost sacrosanct quasi-standard for transportation research and modelling. However, having a look at the discussion on changing travel behaviour initiated in particular by social transportation research [1/2], the question is if the determining variables identified by KUTTER and later modified by others are still valid for a grouping of individuals as to their travel behaviour. So the present article tries a first approach to answer this question using data from the year 2002 survey “Mobilitaet in Deutschland” (“Mobility in Germany”) and analysing travel behaviour on the basis of trip chains. The intention is to identify general behaviour patterns in the transportation system and to evaluate these patterns with socio-demographic and socio-economic

variables against the background of the traditional concept of homogenous behavioural groups.

2. Modelling travel behaviour

Models of behaviour have the aim to build up reality in a simplified way. A good model is characterized by the possibility to picture human behaviour in the real world with some few determinants in the model. Such models also play an important role for description or prognosis in the transportation research area. In the early seventies of the last century KUTTER et al. [3] identified in empirical studies clusters of homogenous behaviour determined only by a few socio-demographic variables. Even though in such studies a causal connection could not be established between variables like “employment” or “car availability” and a specific behaviour in the transportation system, in an aggregated reflection there were significant correlations between social clusters and travel behaviour. Based on these perceptions it was possible to assume a specific allocation of related behaviour because of the knowledge of a certain social structure in a defined geographic area.

The analytical focussing on activities such as work, shopping or leisure and their intraday arrangements bases on the simple but fundamental statement made by HÄGERSTRAND and others [4/5] that activities out of home are the origin cause of trips and thus of physical mobility. The concept of homogenous behavioural groups that will be re-analysed in the present paper is such an individual behavioural model. In this model the holders of specific attributes are assigned to a cluster which at least at medium-term shows similar activity patterns and which differs significantly from other clusters. The „individual-factor-model” published by KUTTER [6] was the first prominent model based on the described reflections, and he deduced individual behaviour out of person- and household-specific factors. The factor analysis was accomplished with socio-demographic and geographic variables like employment, household size or place of residence. As a result of the identified factors and plausibility considerations groups were built which were in turn the basis of the identification of corresponding activity patterns. KUTTER described nine homogenous behavioural groups that could be defined with few variables like employment or car-availability. The concept of the homogenous behavioural groups has been established as a quasi-standard for transportation research and especially modelling for the last 30 years.

There have been many modifications and further developments, in particular for the calculation and evaluation of homogenous behavioural groups, introducing criteria like number of trips per day, distances, trip duration or sequence of activities. This explains the diversity of classifications that differ by number of groups or clusters as well as by analytical discrimination between groups or clusters [7/8]. It was especially SCHMIEDEL [9] who dealt with the empirical identification of a manageable number of so called “homogenous personal subgroups” in his dissertation. With a cluster analysis of activity time budgets he examined behaviour data of the KONTIV 1976 (“Kontinuierliche Erhebung zum Verkehrsverhalten”) after he grouped the individuals with selected sociodemographic and geographic attributes. The resulting seven “homogenous behavioural personal subgroups” are similar to the “homogenous behavioural groups” of KUTTER. SCHMIEDEL could demonstrate persuasively that at time of data collection there was an interrelation between individual attributes and behaviour in the transportation system. He also proved that the inclusion of more subgroups with homogenous behaviour does not lead to a significant enhancement of the explained variance.

3. Reference day data as analytical basis

To evaluate the validity of the traditional concept of “homogenous behavioural groups” the present contribution uses reference day data – in this case data from the study of the year 2002 “Mobility in Germany” (MIG).

A fundamental problem at reference day data is the fact that intra-individual variations of behaviour cannot be reproduced. Individuals normally have multiple typical behavioural daily and weekly patterns with a high intrapersonal variance and a relatively small share of completely repeated daily patterns [10]. To bypass this problem there was a first diary survey in 1972 in Uppsala (Sweden) with 296 households. The collected time use data were analysed particularly by HUFF and HANSON [11/12] to identify behavioural routines and behaviour patterns. Time use and diary data were also used in different other studies. MAHMASSANI, HATCHER and CAPLICE [13] exploited mobility data of the year 1989 from Austin (Texas, USA) from commuters over a time period of 10 working days. AXHAUSEN and others [14] and SCHLICH [15] worked with six week diary data that were collected in line with a German project called MOBIDRIVE. LIPPS [16] analysed activity patterns with one week diary data. HERTKORN and KRACHT [17] in turn used two reference days of the German time-use-study and analysed the data with a relatively new method called „sequence alignment“ [18]. With this method HERTKORN and KRACHT [19] were able to do sequence analysis and SCHLICH [20] could optimise consistent concepts of homogenous behavioural groups in reference to the intrapersonal variance. Alongside there are much more current scientific activity based approaches utilizing time use data [21/22/23].

The problem of all approaches that use reference day data is the fact that getting valid empirical material is very difficult. On the one side there are methodological problems like effects of habituation, breakup behaviour or social acceptability (for example information about the socially not accepted intensive usage of a car for short trips). On the other hand time use surveys are very expensive in proportion to the complexity of the object of investigation.

4. Data analysis

In the following we will analyse with data of the MIG 2002 whether today there is also an interrelation between similar behaviour and socio-demographic and socio-economic indicators. The survey MIG 2002 contains 62,729 person and 167,851 trip data sets. For our analysis we define the individual as the smallest indivisible cluster of behaviour. Similarity of behaviour is understood as behaviour that is documented in the same succession at a reference day.

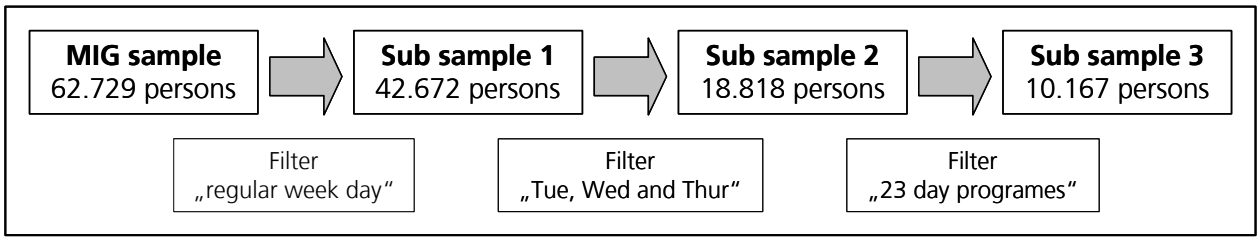

Figure 1: Steps to the researched sub sample

Table 1: Day programs

day program	frequency	percent
Home-Work-Home	**2512**	**24,7**
Immobile	**1772**	**17,4**
Home-Shopping-Home	**855**	**8,4**
Home-Leisure-Home	**709**	**7,0**
Home-Work-Home-Leisure-Home	**579**	**5,7**
Home-Work-Shopping-Home	**546**	**5,4**
Home-Apprenticeship-Home	**397**	**3,9**
Home-Private Errands-Home	**342**	**3,4**
Home-School-Home	**338**	**3,3**
Home-Shopping-Home-Leisure-Home	**328**	**3,2**
Home-School-Home-Leisure-Home	**273**	**2,7**
Home-Apprenticeship-Home-Leisure-Home	**267**	**2,6**
Home-Nursery School-Home	**160**	**1,6**
Home-Shopping-Home-Shopping-Home	**128**	**1,3**
Home-Private Errands-Shopping-Home	**126**	**1,2**
Home-Work-Home-Private Errands-Home	**120**	**1,2**
Home-Work-Leisure-Home	**111**	**1,1**
Home-Shopping-Home-Private Errands-Home	**109**	**1,1**
Home-Private Errands-Home-Shopping-Home	**106**	**1,0**
Home-Work-Home-Work-Home	**103**	**1,0**
Home-Leisure-Home-Leisure-Home	**102**	**1,0**
Home-Private Errands-Home-Leisure-Home	**95**	**0,9**
Home-Shopping-Home-Work-Home	**89**	**0,9**
Total	**10167**	**100**

The data were prepared in several steps (Figure 1). The described reference day problem was considered in a way that only reference days were taken for analysis which were termed explicitly as “regular” by the respondents. Furthermore only Thursdays, Tuesdays and Wednesdays were analysed to avoid the well-known contortions on the weekdays Monday and Friday. Saturday and Sunday were also excluded from analysis. Because of the fact that immobile persons (means “persons not leaving home at the survey day”) represent a relevant share of the sample, the notion “trip chain” could not be used, instead we use the notion “day program”. All respondents that have the same day program are represented in specific day program clusters. For statistical-methodical reasons we analysed only day program clusters whose frequency in the overall sample is above 80. In total we analysed 10,167 day personal data sets (which corresponds to 16% of the total sample) with 23 day programs (which corresponds to 47% of all 3,560 identified day programs in MIG 2002). In the following all percent indications are referred to those 23 day programs.

Each of the identified day program cluster shows a typical day program within a specific variance. For illustration clusters are shown in the following six day programs. Thereby the day time data represent the arithmetic mean of the respective day time cluster.

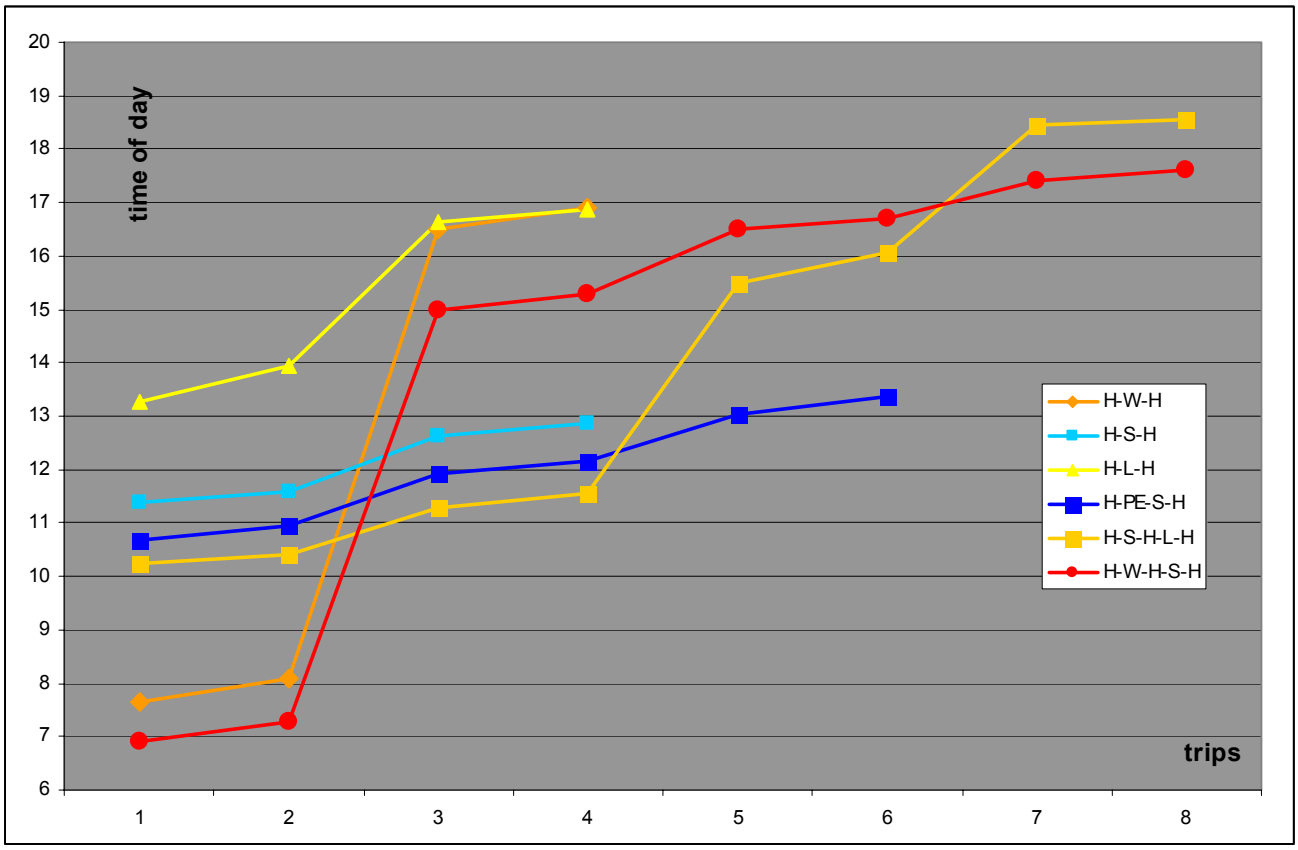

Figure 2: Day time means of selected day program clusters

(1) H-W-H:
Little surprising is that members of the day program cluster “Home-Work-Home” move to work from 7:40 to 8:00 a.m. Then they work until 16:30 and move back home from 16:30 to 16:50.

(2) H-S-H:
Persons from the cluster „Home-Shopping-Home“ move in the mean from 11:20 to 11:35 to the shopping place, do their shopping for an hour and move back home from 12:40 to 12:55.

(3) H-L-H:
Members of the cluster „Home-Leisure-Home“ go out of their home shortly after 1 p.m., move to the place of their leisure activity in about 35 minutes, do their activities until 4:40 p.m. and then return home at about 5 p.m.

(4) H-PE-S-H:
Persons of the cluster „Home-Private Errands-Shopping-Home“ leave home at 10:40 a.m., move about 15 minutes to the place of their activity, do their private errands for about one hour, then move 10 minutes to a shopping place, do their shopping for one hour and then move back home for about 20 minutes.

(5) H-S-H-L-H:
Members of the cluster „Home-Shopping-Home-Leisure-Home“ leave home shortly after 10 a.m., then move for 10 minutes to their place of shopping, do their activities for about 40 minutes, go back home, then do activities at home until 3.30 p.m., move for about 30 minutes to their place of leisure activity, do their activities for 2 hours and are back home at 6:30 p.m.

(6) H-W-H-S-H:
Persons of the cluster „Home-Work-Home-Shopping-Home“ leave home at 7 a.m., then move for 20 minutes to their place of work, work until 3 p.m., then move home for about 20 minutes, stay there for one hour, move to the place of shopping for 10 minutes, do their shopping activities for about 30 minutes and are back home at 5:45 p.m.

In regard to the modal split it is striking that there are big differences between the day program clusters if there is a sequence of work in it. If the respondent is an employee the share of the car is much higher than in the other clusters. In the following figure the six selected day program clusters from above are built up again.

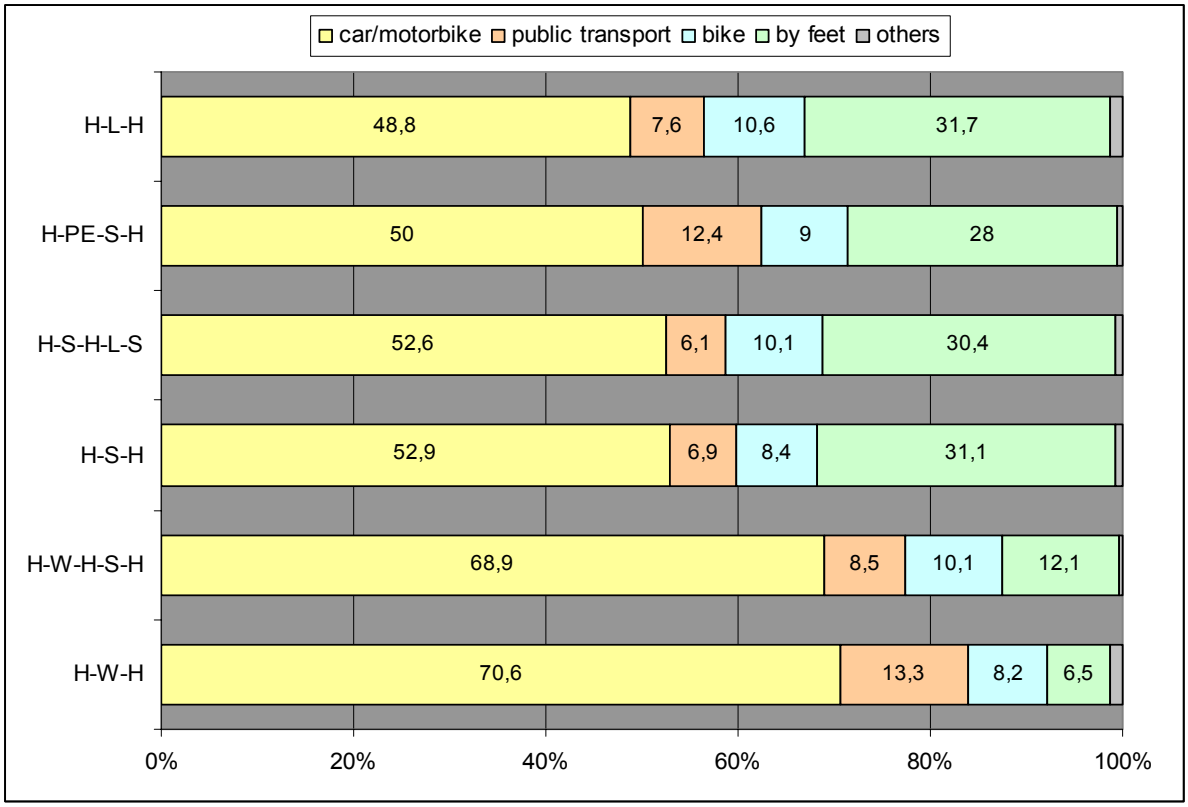

Figure 3: Modal Split of selected day program clusters

5. Empirical verification of the homogenous day program clusters

In the following potential statistical correlations between the 23 day program clusters and different socio-demographic and geographic variables are verified per cross tabulation. Because of the fact that nearly all analysed variables are scaled nominally the chi square test to determine significance (α) is being used. The value of chi square (χ^2) and the variability (df) through the chi square test are used to calculate the probability of an existing variance between observed and expected values even if there is no correlation between the variables in the parent population. It is examined whether an observed variance from expected values leads to a conclusion about a correlation. The strength of a correlation is calculated with the coefficient Cramer V (*V*) because the value of this coefficient lies always between 0 and 1 whereas the value 1 could be reached in all tables independently from the number of cells. The Cramer V is calculated with the following term:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Here *N* terms the range of the sample. If the table contains fewer rows than columns *k* terms the number of rows, otherwise the number of columns. We tested the correlations between the 23 day program clusters and the variables employment, age, education, car availability, gender, household size, income and place of residence (following the typology of the Bundesamt für Bauwesen und Raumordnung (BBR)) for each case.

Table 2: Significance and correlation between day programmes and selected variables

cross tables	χ^2	df	α	*V*
employment/day program	**4107,597**	**60**	**0,000**	**0,434**
age/day program	**7637,715**	**132**	**0,000**	**0,384**
education/day program	**18111,970**	**374**	**0,000**	**0,351**
car availability/day program	**11141,310**	**242**	**0,000**	**0,316**
gender/day program	**318,983**	**22**	**0,000**	**0,192**
household size/day program	**2044,500**	**154**	**0,000**	**0,169**
income/day program	**995,128**	**242**	**0,000**	**0,096**
bbr-typology/day program	**257,374**	**176**	**0,000**	**0,056**

At each cross tabulation the significance is 0,000. This calculated value is clearly below the standard level of 5% and thus so low that an independency of each variable combination is implausible. The null hypothesis is thus declined for all variable combinations.

The strongest correlation exists between the day program clusters and the variable ***employment***. This means that a central statement of the homogenous behavioural groups approach is confirmed. Employment seems to have an essential influence on individual day programs.

Little surprising is the strength of the correlation between the variable ***age*** and the day program clusters if one considers the age dependent activities, in particular work, school and apprenticeship. Here the homogenous behavioural groups are also verified.

The correlation between the variable ***education*** and the day program clusters is also relatively strong. This indicator is normally not explicitly considered in the homogenous behavioural groups. This variable stronger should much more be considered for the building of behavioural clusters, particularly as this indicator is an element of the official statistics and can be used without any problem.
Also of substantial importance for the description of the day program clusters is the ***car availability***. This indicator also seems quite influential on the day programs of the employed part of the population in the year 2002.
Surprisingly, the correlation between day program clusters and ***gender and household size*** is rather moderate [24]. However, it has to be considered that members of bigger households have a more fragmented day program and so they are represented above average in those 50% of day programs that are not shown here.
The ***household income and the place of residence*** are also correlating only weakly with the day program clusters. In respect to income it has to be considered that younger people with higher incomes are represented above average in those 50% of day programs that are not shown here.

6. Conclusions

The scientific motivation for the review of the concept of homogenous behavioural groups was the hypothesis that mobility patterns have been significantly modified during the last decades in particular due to the change in individual lifestyles. This means that they cannot any longer be explained adequately by using only "traditional" socio-demographic variables. This hypothesis was checked with data from the survey "Mobility in Germany 2002". For an accurate evaluation of the results presented here, it must be kept in mind that for methodological reasons these results are not directly comparable to KUTTER'S analysis. But there is no doubt that because of these results the concept of homogenous behavioural groups has to be critically checked. The result of our analysis is twofold. It cannot be denied that there is still a strong interaction between socio-demographic and socio-economic variables and activity (and thus travel) patterns. So we have seen that one half of the trip chains analysed here – which is equivalent to 50% of the observed behaviour on a „regular" working day from Tuesday to Thursday – can be attributed to distinct socio-demographic and socio-economic variables. On the other hand this means that 50% of the trip chains can *not* be explained by the traditional variables, apparently because the behaviour of 50% of the persons surveyed does not allow the conventional assignments. The consequence is that their behaviour can not be deduced from socio-demographic and socio-economic variables offered by official statistics.
As a consequence for future research we suggest to advance at multiple steps. The first step would be to incorporate available data on individuals – in particular "education" – in further models of passenger behaviour. It has to be tested then, if this enlargement of data helps to "explain" larger shares of trip chaining, not only for "regular" days but also for other days like week-ends. In a second step existing data sets, e.g. from lifestyle and mobility style research should be used to identify new potential variables that could further improve the description and modelling of travel behaviour. Empirical testing – as a final step – would allow extracting the key variables that are able to elucidate travel behaviour of today's people.

References

[1] HUNECKE, M. (2000). Lebensstile, Mobilitätsstile und mobilitätsbezogene Handlungsmodelle – Forschungsstrategien zur Analyse personenbezogener Mobilität. In Institut für Landes- und Stadtentwicklung (Hrsg.), *U.MOVE – Jugend und Mobilität* (S. 30-39). Dortmund: ILS-Schrift Nr. 150.

[2] HUNECKE, M. (1997). Lebensstile als Determinanten umweltverantwortlichen Handelns - Typenbildung für die umweltpsychologische Interventionsforschung. In: Kaufmann-Hayoz, R. (Hrsg.), Bedingungen umweltverantwortlichen Handelns von Individuen. Bern: 75-80.

[3] KUTTER, E. / HOLZAPFEL, H. / MARTENS, G. (1981): Ermittlung von Variablen und Parametern möglicher Gesamtmodelle für Verkehrsanalyse und Verkehrsprognose auf der Grundlage amtlicher Statistik. In: Forschung Straßenbau und Straßenverkehrstechnik, Heft 339.

[4] HÄGERSTRAND, T. (1970): What about people in Regional Science? Papers of the Regional Science Association 24, 7-21.

[5] CHAPIN, F.S. (1974): Human activity patterns in the city. New York.

[6] Kutter, E. (1972): Demographische Determinanten städtischen Personenverkehrs. Braunschweig.

[7] WEICHBRODT, C. (1977): Entwicklung und Anwendung integrierter Personenverkehrsmodelle auf der Basis individueller Verhaltensmuster. Braunschweig.

[8] HAUTZINGER, H. / KESSEL, P. / BAUR, R. (1980): Mobilitätschancen unterschiedlicher Bevölkerungsgruppen im Personenverkehr. In: Forschung Straßenbau und Straßenverkehrstechnik, Heft 310.

[9] SCHMIEDEL, R. (1984): Bestimmung verhaltensähnlicher Personenkreise für die Verkehrsplanung. Karlsruhe.

[10] LIPPS, O. (2001): Modellierung der individuellen Verhaltensvariationen bei der Verkehrsentstehung. Karlsruhe.

[11] HUFF, J.O. / HANSON, S. (1986): Repetition and variability in urban travel. Geographical Analysis. 18(2): 97-114.

[12] HUFF, J.O. / HANSON, S. (1988): Systematic variability in repetitious travel. Transportation 15(2): 111-135.

[13] MAHMASSANI, H.S. / HATCHER, S.G. / CAPLICE C.G. (1997): Daily variation of trip chaining, scheduling, and path selection behaviour of work commuters. In: Stopher, P. / Lee-Gosselin, M.E.H. (eds.) Understanding Travel Behaviour in an Era of Change. Oxford: 351-379.

[14] AXHAUSEN, K.W. / ZIMMERMANN, A. / SCHÖNFELDER, S. / RINDSFÜSER, G. / HAUPT, T. (2002): Observing the rhythms of daily life: A six-week travel diary. Transportation 29, 95-124.

[15] SCHLICH, R. (2003): Homogenous groups of travellers. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

[16] LIPPS, O. (2001): Modellierung der individuellen Verhaltensvariationen bei der Verkehrsentstehung. Karlsruhe.

[17] HERTKORN, G. / KRACHT, M. (2002): Analysis of a large scale time use survey with respect to travel demand and regional aspects, In: Fontainha, E. / Perista, H. (eds.): Work Time and Leisure Time - Dynamics and Convergence in Changing Contexts. International Association for time use research (IATUR)-Conference 2002, October 2002, Lisbon.

[18] JOH, C.-H. / ARENTZE, T. / HOFMAN, F. / TIMMERMANS, H. (2002): Activity pattern similarity: a multidimensional sequence alignment method. Transportation Research Part B(36): 385–403.

[19] HERTKORN, G. / KRACHT, M. (2002): Analysis of a large scale time use survey with respect to travel demand and regional aspects, In: Fontainha, E. / Perista, H. (eds.): Work Time and Leisure Time - Dynamics and Convergence in Changing Contexts. International Association for time use research (IATUR)-Conference 2002, October 2002, Lisbon.
[20] SCHLICH, R. (2003): Homogenous groups of travellers. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.
[21] KITAMURA, R. / PAS, E.I. / LULA, C.V. / LAWTON, T.K. / BENSON, P.E. (1996): The sequenced activity mobility simulator (SAMS): an integrated approach to modeling transportation, land use and air quality. Transportation 23, 267-291.
[22] BEN-AKIVA, M.E. / BOWMAN, J.L. / GOPINATH, D. (1996): Travel demand model system for the information era. Transportation 23, 241-266.
[23] GÄRLING, T. / KALÉN, J. / ROMANUS, J. / SELART, M. / VILHELMSON, B. (1998): Computer simulation of household activity scheduling. Environment and Planning A 30, 665-679.
[24] KWAN, Mei-Po (2000): Gender differences in time-space constraints. Area 2000 32(2): 145-156.