Analysing interpersonal variability for homogeneous groups of travellers

Working Paper

Author(s):
Schlich, Robert; Axhausen, Kay W.

Publication date:
2005-09

Permanent link:
https://doi.org/10.3929/ethz-b-000023522

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 296

Submitted to *Environment and Planning*

Analysing interpersonal variability for homogeneous groups of travellers

R Schlich

KW Axhausen

Arbeitsbericht Verkehrs- und Raumplanung 296 September 2005

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Arbeitsberichte Verkehrs- und Raumplanung 296

Vergleich interpersoneller Verhaltensvariabilität

R Schlich
SBB Fernverkehr
Wylerfeldstr. 125
CH – 3001 Bern

Telephon: +41-512-20 34 63
Telefax: +41-512-20 43 94
robert.schlich@sbb.ch

KW Axhausen
IVT
ETH Zürich
CH – 8093 Zürich

Telephon: +41-1-633 3943
Telefax: +41-1-633 1057
axhausen@ivt.baug.ethz.ch

September 2005

Kurzfassung

Die Einteilung der Bevölkerung in Gruppen mit ähnlichem Verkehrsverhalten ist in der Verkehrsverhaltensforschung bereits recht lange eine wichtige Frage. Das Ziel einer solchen Klassifizierung ist es, Gruppen zu identifizieren, der Mittglieder innerhalb einer Gruppe zueinander ähnlich sind, sich aber vom Verkehrsverhalten der Personen anderer Gruppen deutlich unterscheiden. Trotz der langen Tradition solche Gruppen zu konstruieren, hat es in den letzten 20 Jahren wenig Fortschritte auf diesem Gebiet gegeben. Einige ältere (Kutter, 1972, Pas, 1983; Schmiedel, 1984, Huff and Hanson, 1986, 1988b) stellen immer noch den "state of the art" dar.

Dies ist umso erstaunlicher wenn man sich vor Augen führt, dass diese Einteilungen alles andere als zufriedenstellend sind, da innerhalb der Gruppen eine sehr grosse Variabilität verbleibt. Dies liegt in erster Linie an zwei Hindernissen: Erstens mangelt es an geeigneten Langzeitdaten und zweitens an der Frage, wie man Ähnlichkeiten adäquat messen soll. Beiden Fragen wird in diesem Aufsatz nachgegangen. Der Aufsatz untersucht das Ausmass intrapersoneller Variabilität mit der Methode der Sequenz analyse, die nicht nur die Art, sondern auch Dauer du Reihenfolge von Aktivitäten berücksichtigt. Die Ergebnisse – basierend auf der Langzeitstudie Mobidrive zeigen, dass das Ausmass intrapersoneller Variabilität recht hoch ist. Aus diesem Grund wurden pro Person drei verschiedene typische Tage identifiziert und basierend auf diesen Tagen die (ebenfalls mit Sequenzanalyse berechnete) interpersonelle Variabilität zwischen den Personen als Ausgangspunkt für eine Clusteranalyse herangezogen wurde.

Die Ergebnisse zeigen, dass bezüglich der täglichen Aktivitätenprogramme sehr homogene Cluster gebildet werden können – hinsichtlich der herkömmlichen betrachteten Verhaltensindikatoren und sozidemographischen Merkmalen ist sie jedoch gering.

Schlagworte

Verkehrsverhalten, Variabilität, Messung, Clusteranalyse, Mobidrive

Zitierungsvorschlag

Schlich, R. und K.W. Axhausen (2004) Analysing interpersonal variability for homogeneous groups of travellers, *Arbeitsbericht Verkehrs- und Raumplanung*, **296**, IVT, ETH Zürich, Zürich.

Working Paper

Analysing interpersonal variability for homogeneous groups of travellers

R Schlich	KW Axhausen
SBB Fernverkehr	IVT
Wylerfeldstr. 125	ETH Zürich
CH – 3001 Bern	CH – 8093 Zürich
Telephone: +41-512-20 34 63	Telephone: +41-1-633 3943
Telefax: +41-512-20 43 94	Telefax: +41-1-633 1057
robert.schlich@sbb.ch	axhausen@ivt.baug.ethz.ch

September 2005

Abstract

The segmentation of the population into groups of people with homogenous travel behaviour has been an important issue to travel behaviour analysis for a long time.. Aim of this classification is to identify groups of people who are very similar to each other concerning their travel behaviour but clearly distinct from the members of other groups. Despite the long tradition to construct behavioural homogenous groups, there has not been much progress in the last 20 years. Some older classifications (Kutter, 1972, Pas, 1983; Schmiedel, 1984, Huff and Hanson, 1986, 1988b) are still the state of the art.

This is even more surprising as those classifications are far from satisfactory because they only explain a small amount of variability within the groups. This is mainly due to two different obstacles: The first obstacle is the lack of suitable longitudinal data, the second is the gap how similarity is measured and how the order of activities is considered in the measurement. Both obstacles shall be addressed in this paper.

This paper examines the amount of intrapersonal variability with the method of multidimensional sequence alignment which does not only consider the type of the performed activities but also their order and timing. The results based on the longitudinal Mobidrive data show that the intrapersonal variability is quite high. For each person three typical days were calculated and based on theses days the similarity of each person to the others was used as a basis for cluster analyses.

The results show that the members of each cluster are very similar in terms of daily activity programms, but not similar in terms of sociodemograhics and traditional behavioural indicators.

Keywords

Travel behaviour, travel diary, similarity, cluster analysis, Mobidrive

Preferred citation style

Schlich, R. and K.W. Axhausen (2004) Analysing interpersonal variability for homogeneous groups of travellers, *Arbeitsbericht Verkehrs- und Raumplanung*, **296**, IVT. ETH Zürich, Zürich..

1 Introduction: Homogeounus groups

The segmentation of the population into groups of people with homogenous travel behaviour has been an important issue to travel behaviour analysis for a long time.. Aim of this classification is to identify groups of people who are very similar to each other concerning their travel behaviour but clearly distinct from the members of other groups. Despite the long tradition to construct behavioural homogenous groups, there has not been much progress in the last 20 years. Some older classifications (Kutter, 1972, Pas, 1983; Schmiedel, 1984, Huff and Hanson, 1986, 1988b) are still the state of the art. This is even more surprising as those classifications are far from satisfactory because they only explain a small amount of variability within the groups.

Obviously there still is a need for market segmentation and homogeneous groups, as a proper classification helps in the estimation of behavioural models as they are required for the description of simulated persons/agents in micro-simulation or aggregate models.

This paper wants to explore the potential of a new measurement method of travel behaviour similarity based on multi-day diaries to be the basis for the identification of homogenous groups of travellers and thus to overcome two major obstacles with existing classifications:

The first obstacle is the absence of suitable data. The data is insufficient insofar as earlier classifications are based on cross sectional data. This means that some aspects of behaviour (e.g. the question of intrapersonal variability) could not be addressed by these classifications at all. Moreover, Huff and Hanson (1988b) showed, that the chance of misclassification of a person is much higher if cross sectional data is used – and thus the source of variability within the clusters. The second problem is the gap how similarity is measured and how the order of activities is considered in the measurement. Both obstacles (insufficient data and measurements that cannot consider the order of activities) shall be addressed in this paper.

In the following section, the data obtained from the longitudinal study Mobidrive is briefly presented which allows for the measurement of intrapersonal variability. In the third section the sequence alignment method as a major improvement to measure behavioural similarity is introduced. The next section deals with the question how variable the behaviour of one person actually is and how many days should be required for the following clustering process.

The fifthth section describes the methodology used for the clustering The resulting groups of similar behaviour are then described in terms of their behavioural variability as well as their sociodemographic structure and analyses. Finally the clustering results are discussed and further methodological gaps are discussed.

This paper is based on Schlich (2004) and expands the earlier results of Schlich and Axhausen (2003), as well as of Schlich (2003, 2001).

2 The Mobidrive six-week diary

The main cause of the lack of suitable data for the analysis of intrapersonal variability is the difficulty of obtaining the necessary data about travel behaviour from respondents over long periods (Pas, 1987; Zumkeller and Chlond, 1995). The specific difficulties of long-duration surveys result from the high response burden of the participants. In addition, it must be ensured that no self-selection of respondents with special interests in particular topics or with particular travel patterns takes place. Furthermore, people may neglect or forget to report trips, especially short trips, in the later parts of a multi-day or multiple-week survey (Golob and Meurs, 1986). For these reasons, longitudinal travel behaviour data has rarely been collected until the arrival of GPS-based tracking in the mid-1990s, which by-design has to omit important items common in travel diaries (See for example Axhausen, Schönfelder, Wolf, Oliveira and Samaga, 2004).

This gap was addressed by collecting a data set that is to a large extent unique[1,2]. It is the result of a six-week travel diary conducted for the research project Mobi*drive*. Funded by the German Federal Ministry for Education and Research, 361 persons reported 52,273 trips on 14,360 person days in spring and autumn of 1999 in Karlsruhe and Halle/Salle, Germany. The project consortium consisted of the PTV AG (Karlsruhe), the Institut für Stadtbauwesen (RWTH Aachen) and the Institute of Transport Planning and Systems (IVT) at ETH Zurich[3]. Sampling procedures, the survey instruments and data administration are presented in Axhausen, Zimmermann, Schönfelder, Rindsfüser and Haupt (2002). The data set is

[1] The survey has since been replicated in 2003 in Switzerland. This data set has only become available during late 2004 and was not included in this analysis (Axhausen, Löchl, Schlich, Buhl and Widmer, 2005).

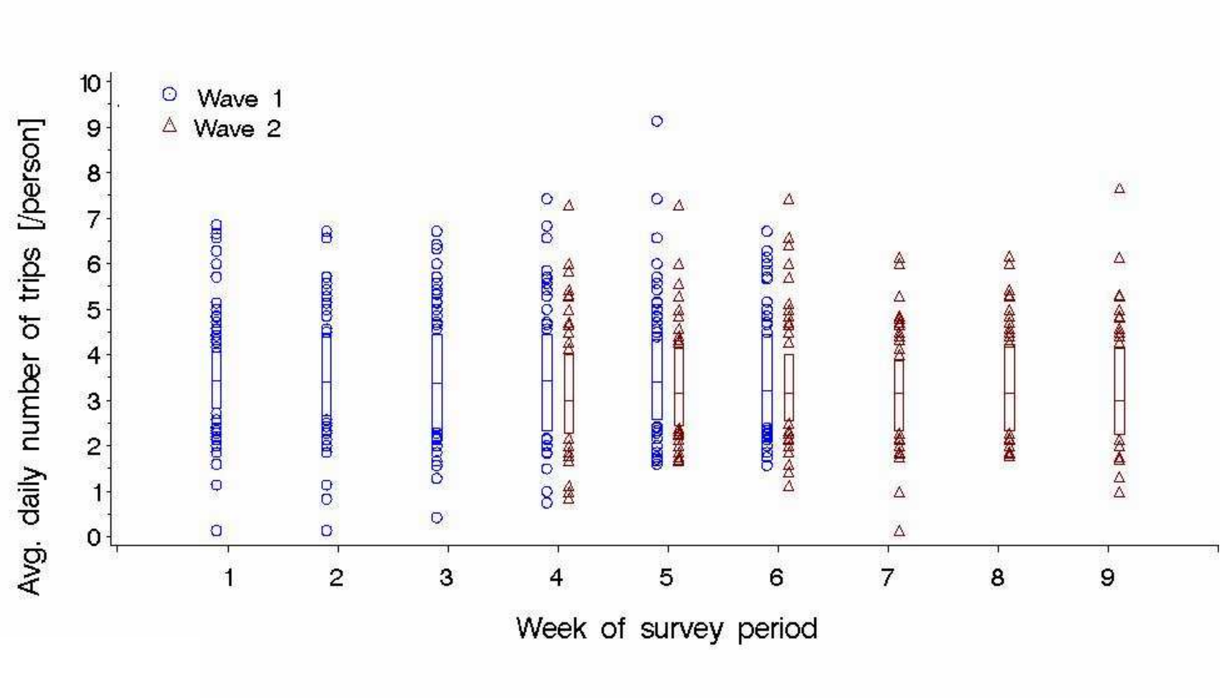
[2] There is one comparable survey, which covers a period of five weeks - the Uppsala survey. This survey was conducted 1971 and is the basis of a series of publications by Hanson and collaborators concerning the stability of travel behaviour (e.g. Hanson and Huff 1982, 1986 and 1988; Hanson and Burnett 1981 and 1982; Huff and Hanson 1986 and 1990). It was not used here, as certain desirable items are missing for the trips and as the set of socio-demographic variables is very limited (See Schlich, Schönfelder, Hanson and Axhausen, 2004)

[3] Further information is available at http://www.ptv.de/mobidrive/

documented in Schönfelder, Schlich, König, Horisberger and Axhausen (2000) and available for download at http://129.132.96.89/index.html.

Figure 1 shows for the city of Karlsruhe that the average number of reported trips on mobile days per persons and week exhibits no trend consistent with fatigue (These boxplots give the quartile values and the persons with values beyond the 25th and the 75th percentile by survey wave).

Figure 1 Average number of daily trips per wave on mobile days per persons and week



Source: Axhausen *et al.* (2002)

The in-depth analysis in Axhausen *et al.*, 2002 and its extension in Axhausen et al., 2005 and Madre, Axhausen and Brög, 2004 confirm the absence of fatigue in the data.

While the recruited households have a higher income, more cars and more working members in Halle, the general indicators of travel behaviour are remarkably similar to those obtained from one-day diaries in Karlsruhe and Halle– given the massive methodological differences between those studies (Axhausen *et al.*, 2002). The Mobi*drive* survey is therefore a unique opportunity to analyse the patterns of interpersonal and intrapersonal similarity.

3 Measuring similarity: Sequence alignment method

The major problem for previous work on the similarity of travel patterns was the lack of a generally accepted procedure to identify and measure it, which accounted properly for the multi-dimensional properties of a trip. Standard indicators such as the number of trips per day, mean trip distance, or mean trip duration neither consider the temporal dimension of the activity chains, nor the complexity of behaviour and are thus unsuitable. Various more complete measurement methods have been suggested in the past. Schlich and Axhausen (2003) provide a literature review and present selected comparative results for the Mobi*drive* data. It is particularly controversial in the literature which attributes to examine, how to classify and to weight them, and how the values of the attributes should be compared. Thus, different measures lead to different results for the same data (Burnett and Hanson, 1982). Hanson and Huff (1988) generally notice that the more detail the procedure considers, the smaller are the observed similarities.

Wilson (1998a, 1998b) recently introduced the sequence alignment method to travel behaviour research – a measurement approach that considers the order of actives properly. (See Bargeman, Joh and Timmermans, 2002; Berger 2000a, 2000b; Hertkorn and Kracht, 2002; Rindsfüser and Doherty, 2000; Schlich, 2001 and Wilson, 2001 for further applications to travel behaviour data). Since then Joh, 2004 and Joh, Arentze and Timmermans, 2001a, 2001b, 2001c, 2002 have extended the approach, so that it is now possible to measure the similarity of travel patterns considering their multi-dimensional nature and their sequence. In this form it fulfils the full requirements for the task at hand (See below for details).

The sequence alignment method was originally developed in molecular biology to compare DNS or protein strings (Sankoff and Kruskal, 1983). The idea of comparing strings consisting of a sequence of different elements was also adopted in other scientific fields, e.g. speech recognition. Wilson (1998a, 1998b) was the first to introduce sequence alignment to travel behaviour research, but the method had been adopted by social scientists for some time under the name *optimal matching* (e.g. Billari, 1999; Schaeper, 1999; Erzberger and Prein, 1997 or Abbot and Tsay, 2000 for an overview).

The one-dimensional measurement of similarity is based on attributes such as activity type, transport mode, starting time, trip or activity duration or trip destination. In general, one attribute is compared for each trip of a sequence – e.g. the second trip duration is compared to the second trip duration of another day. Previous approaches than calculated the similarity as, for example, the sum of the Euclidean distances between the attribute values.

These methods do not incorporate the sequential order of activities properly. $\boldsymbol{s}=\boldsymbol{s}[s_1.......s_m]$ and $\boldsymbol{g}=\boldsymbol{g}[g_1......g_m]$ are two trip sequences with *n* and *m* trips respectively. Imagine the sequences **s** (**s**ource) and **g** (tar**g**et) displayed in Figure 2, where each element represents activities in 15 minute intervals.

Figure 2: Pair wise comparison of two sequences representing activities

Example of two sequences

s: *WW* WW TS ST HH HH TL LL LT HH

g: WW TS ST HH HH TL LL LT HH *HH*

Calculation of similarity:

$$d(s,g)= \sum_{i=1}^{n} f(x) \qquad \text{and } f(x)= 1 \text{ if } s_i \neq g_i$$

$$f(x)= 0 \text{ if } s_i = g_i$$

Activities:

W: working; T: travel; S: shopping, L: leisure, H: home

Source: Schlich. (2004)

If the distance between both chains would be measured by pairwise comparisons with a score of one for a mismatch and zero for a match, the distance between both sequences would be 12 (for a string of 20 elements) although in both sequences the same activities are performed in the same order and for the same duration. The only difference is, that in the second sequence all activities after work start half an hour (or two intervals) earlier and that the first string possesses two intervals of working at the beginning instead of staying at home in the other one. Intuitively they are very similar. Improvements to this naïve way of calculating similarities were introduced by Pas (1983) and Hanson and Huff (1986) who added for example differential weights and considered the serial dependence of different attributes. Nonetheless their similarity functions ignore the sequential order of activities and their interdependencies.

The core idea of the sequence alignment method is to convert two sequences **s** and **g** by different, well defined operations in each other. The operations are substitution, insertion and deletion. Insertions and substitutions are sometimes called *indels*. The implied effort of each operation can be accounted for by different weights. Mostly, a weight of one is assigned to the operations deletion and insertion. The weight of a substitution can be understood as the sum of the consecutive operations of a deletion and an insertion and is thus given a value of two.

There is usually more than one way to change the sequence **s** into **g** by substituting, deleting and inserting characters into the strings. The way with the smallest sum of the weighted operations is selected and defines the Levenshtein distance (Levenshtein, 1968), as an equivalent to the shortest path distance in networks. Each way of equalising the sequences is called an alignment. An example is given in Figure 3.

The advantage over conventional measurements becomes clear, if one imagines the sequences ***s****=**s**[ABCDEFGH]* and ***g1****=**g1**[ADEBFGCH]*, respectively ***g2****=**g2**[AFGBDECH]* (Joh *et al.*, 2002). A pairwise comparison gives a distances of 6 units – with the sequence alignment it is $d(s,g1) = 4$ respectively $d(s,g2) = 6$ units. The sequence alignment distinguishes correctly between the "wrong position but the same order" and "wrong positions and different orders" (Joh *et al*., 2002). This approach can be used both to calculate a similarity between two strings as well as a distance between them[4].

Figure 3: Two possible alignments for the sequences **s** and **g**

Sequences:

s=CAMBRIDGE
g=CAMPING

Distance sequence alignment :

1) substitute s_4(B:P), s_5(R:I), s_6(I:N), s_7(D:G) delete s_8(G), s_9(E) => d=10
2) substitute s_4(B:P), delete s_5(R), substitute s_6(D:N), delete s_8 (E) => d=6

Source: Schlich. (2004)

The adoption of a method from a completely different context creates problems. The major problem is in our case, that travel behaviour cannot be represented by a single attribute, as discussed above. Instead it has to be characterised by multiple attributes such as trip purpose, trip destination, travel mode or trip departure time. Unfortunately all these attributes have different measurement scales so that the methods for multidimensional alignment used in molecular biology (see McClure, Vasi and Fitch, 1994) cannot be adopted.

If all variables were independent from each other then the distances for k attributes could be calculated separately. The distances could then be weighted according to their importance with weighting factors β_k and summed. This method is called uni-dimensional sequence analysis (UDSAM):

[4] The same fact can be expressed by the term distance (Joh *et al*. 2002) or similarity (Wilson, 1998b). Wilson (1998b) states that none of the two expressions has clear advantages over the other. In this paper the terminology of Joh is used.

$$d(s,g) = \sum_{k=1}^{K} \beta_k (s_k, g_k)$$

In reality different attributes of an activity or trip depend on each other – for instance, the choice of a travel mode is influenced by the chosen location. If all attributes were connected to each other in the same way, it would be sufficient to calculate the distance as the distance of the attribute which is given the maximum weight. With this measurement the distance would be smaller than measured with UDSAM. However, treating the different attributes of an activities as either totally dependent or independent is not justified in most cases.

Generally, different attributes are partially dependent. If each attribute is represented by a single sequence, then for those attributes which are connected, the same operation has to be performed at the same position in the other sequences – for those which are independent from each other the operations will differ. Elements in the sequence which can be aligned simultaneously without calculating the cost twice because the same operations are performed across attributes are called segments (Joh *et al.*, 2002).

Identifying segments leads to a reduction in the total alignment costs and is thus a major task for the calculation of the Levenshtein distance for sequences with multiple attributes – this method is called multidimensional sequence alignment (MDSAM). The only way to obtain a correct optimal result is to calculate all possible alignments and to identify the minimal cost one. MDSAM calculations are at present not possible for large samples of trips due to their enormous computing costs. Joh (2004) developed a heuristic approach, which approximates the complete multidimensional approach. He proposed that not all alternative alignments have to be calculated across all attributes. The search for an approximation starts with those combinations, which scored the minimum distance for an attribute and tests further combinations with small deviations from this starting point. If this solution reduces the distance, then it chosen for further iterations. This approach is called the Hybrid MultiDimensional Sequence Alignment Method (Hy-MDSAM). It reduces the required computing time substantially. Although the resulting distances correlate very strongly with the unidimensional approach (UDSAM) this is an important advance and the method was adopted here (with a limit of 50 iteration without progress).

Further problems arise from the measurement scales of the attributes. While some attributes are categorical, like the trip purpose, other attributes are interval-scaled, such as trip duration or trip distance. The differences between the interval-scaled attributes of two sequences are lost, as these variables are currently transformed into discrete classes.

Furthermore, the choice of the attributes and of their weights lack a theoretical justification at present. As mentioned before, there is also little justification for the costs or penalties of the different operations (deletion, insertion and substitution). It is common practice in the social sciences (Schaeper, 1999), that the weight of an insertion or a deletion is fixed as half a substitution. For the costs of substitutions there is no unambiguous criteria. In an ideal case the costs should differ depending on the type of substitution – for example a substitution of a trip made by foot by a bicycle trip seems to be less expensive than a substitution by a car trip. Unfortunately there is currently no theoretical framework for the determination of these costs. In the application below, all attributes were weighted equally. Deletions and insertions had the weight one. Substitutions had uniformly the weight two.

Dollase, Hammerich and Tokarski (2000) point to the differences due to the different ways of incorporating activity duration. If travel behaviour is recorded with a travel diary the beginning and end time of an activity is known. It is then possible to divide a day into time intervals and to assign a main activity to each interval. Interval lengths used in previous applications were 5 to 15 minutes. The duration of the intervals will influence the results, because long intervals will neglect short activities. Furthermore the duration aspect dominates the resulting similarities. A related problem in this context is the question of how to deal with night hours. As they are normally equal in terms of all attributes and numerous they bias the results. In the application below, duration was neglected and only the sequence of activities considered.

4 Intrapersonal day-to-day variability

4.1 Literature

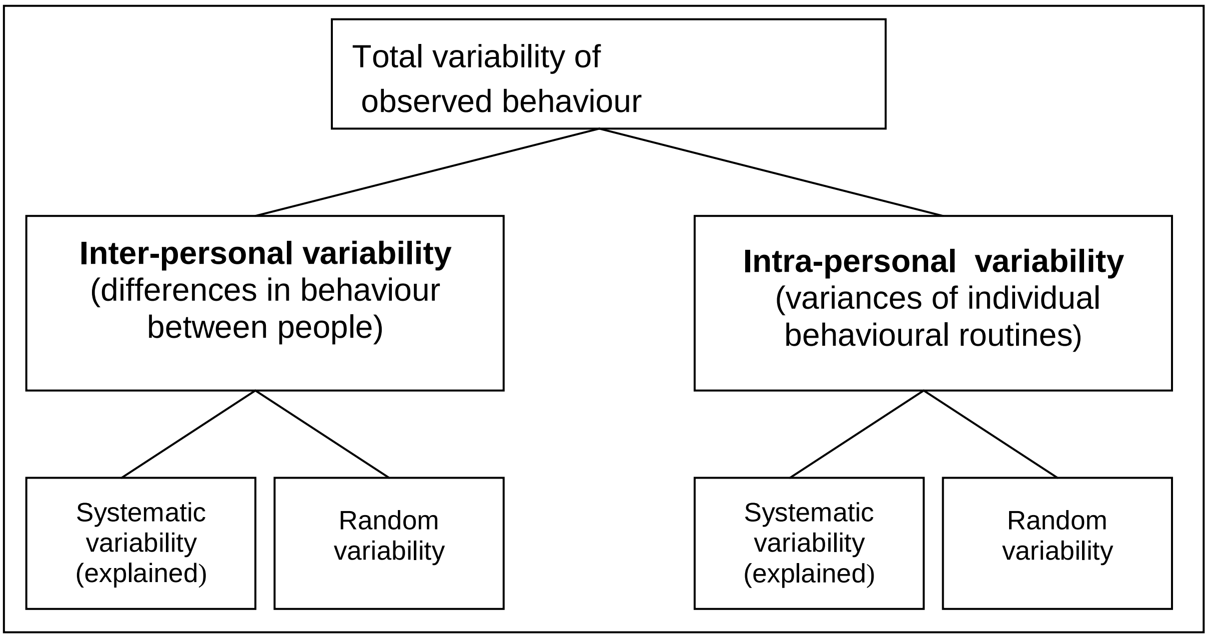
Travel behaviour analysis and transport models assume that routine behaviours dominate and that regular trips are performed (nearly) identically by travellers with respect to destination, mode, timing, duration and other relevant aspects of the movement. It is reasonable to assume that humans perform a high proportion of actions regularly because constraints and obligations do not change every day. For example, one can assume that commuting trips resemble each other concerning mode choice, route choice or departure times due to nearly stable constraints of these trips. Besides, it is unlikely that humans will judge their activities anew every time. They will rather repeat an activity pattern that offered them a satisfying experience without carefully judging the alternatives.

The constraints or obligations may be similar from day to day – but still the chosen activities are not necessarily identical. Differences occur because people do not have the same needs every day. For example, it is not necessary to go to a grocery store every day. In particular the motives for leisure are not identical from day to day, as are the social contexts, i.e. the persons present. A further cause of behavioural variability is environmental variation, such as the weather and other unexpected or irregular events, e.g. accidents, special sporting events etc.

The question of how repetitious travel actually is has been the subject of investigation for many years. However, intrapersonal variability (different behaviour of one person from day to day) has played a minor role in comparison with research on interpersonal variability (differences in the behaviour of different persons), thus there are few empirical results. This is surprising, since the question of intrapersonal variability is of great interest to transport planning.

A comprehensive review of the literature is available in Pendyala, Muthyalagari and Parashar (2000). Intrapersonal (within a person) and interpersonal (between person) variability can have systematic, as well as random causes (see. Figure 4) (Pas und Sundar, 1995). Kunert (1994) suggested that in previous studies much of the interpersonal variability is caused by intrapersonal variability. This is incidentally a reason for the well-known difficulties when defining market segments with regards to travel behaviour.

Figure 4 Components of total variability



Adapted from Pas (1987), S. 432

Using a multi-day data set Pas and Sundar (1995), and Pas (1987) showed with an analysis of variance that between 35 percent and 50 percent of the total variance is due to intrapersonal variance (with minor differences for different behavioural attributes). Pas and Koppelmann (1986) demonstrated, that the intrapersonal variance (with regards to the number of trips) varies for the different respondents. This results was confirmed by Jones und Clarke (1988) using their own measure of similarity between days for working and non-working persons. Lipps (2002) suggested lower amounts of intrapersonal variability (20%) and recommended to neglect it. This results is based on time-budget data, which may be less variable.

The most comprehensive analyses using the previously best dataset were performed by Huff and Hanson (1986) and Hanson and Huff (1988a und 1998b), who identified a core of regular and consistent activities. Still, these are recombined in multiple ways, so that neither a typical work day, nor typical weekend days could be derived.

Previous work has not been able to measure the level of day-to-day variability properly, nor has it been able to describe the range of behaviours fully. If this assessment is true, then a number of widely employed approaches and results become dubious. A prime example are the homogenous groups in their current form, as they are based on one and two-day surveys only (Pas, 1987; Hanson and Huff, 1988a, Herz, 1983).

4.2 Results

This section presents a first overview over the amount of intrapersonal variability calculate with the proposed method of sequence alignment for the Mobidrive data. For each day of one person the distance to all other days was calculated using distance, mode, purpose, and departure time as attributes. The attributes had up to twelve classes. The average intrapersonal variability by age, sex and car ownership is given in Table 1.

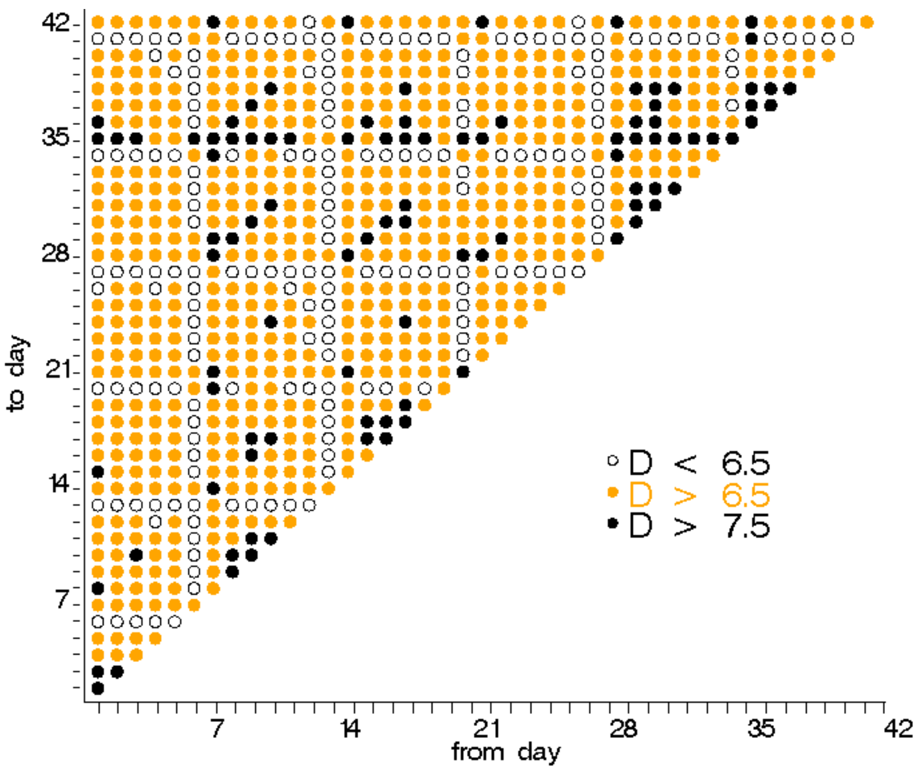
Table 1 Average intrapersoneal variability of the daily activity patterns (Hy-MDSAM) by sex, age and household car ownership [Levenshtein distance]

	Females						Males					
Number of cars in the household	Up to 30 years		30-60 Years		60 and more years		Up to 30 years		30-60 Years		60 and more years	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
0	8.6	2.3	6.6	1.1	5.3	3.1	5.7	1.6	7.3	2.9	9.6	2.9
1	7.4	3.7	7.0	2.6	5.5	2.3	6.5	3.2	7.2	2.5	6.4	2.6
2 and more	8.6	5.4	7.9	2.7	10.4	3.8	5.9	2.5	6.9	2.5	4.2	2.5
Total	7.8	4.1	7.2	2.5	5.8	2.8	6.3	2.9	7.1	2.5	6.5	2.6

Females, who are either younger or live in households with cars, have more variable patterns than males, who show more variability when older or living in households with fewer cars. A first interpretation suggests family responsibilities and work patterns for these differences. In any case, these results confirm Pas and Koppelman's 1996 conclusions, that there a substantial differences between person groups.

Next to the differences between person groups, the differences between types of days over time are of special interest. Figure 5 shows how similar the days are by displaying the average multidimensional Levenshtein distances resulting from the pair wise comparison of the 42 days of the Mobidrive reporting period for all respondents. The bulk of the days is not all that similar to each other. The Saturdays (6., 13., … day) are closer to many weekdays, then the weekdays are to each other. The days with the biggest average distance to each other are spread across the whole reporting period and the types of days.

Figure 5 Average day-to-day multi-dimensional Levenshtein distance (Mobidrive, all respondents) [Levenshtein distance]



Source: Schlich (2001)

4.3 Characterising persons and the necessary number of days

Ideally, one would describe the differences between persons by comparing their daily patterns for all available days. In the case of the Mobidrive data set this would require the calculation of the multidimensional distance measure for over 115 * 10^6 pairs of days. An estimated computing effort of 15 months (0.7 * 10^6 minutes of CPU time for a 2003 Pentium-based personal computer) plus data handling time were infeasible and are likely to be infeasible for similar projects. It is therefore necessary to capture day-to-day intrapersonal variability by a small set of typical days This would allow the analyst to characterise a person in a parsimonious way.

Lipps (2001) tried to identify such a typical day by focusing on the structure of the schedule. His work is typical of the previous work in this area. Using a-priori defined main activities he classified tours by main activity types and position during the day into four types. The most frequently observed tour types were reassembled into a reference schedule. The observed day which was closest to this reference schedule was selected as the reference day for the person.

This method systematically underrepresents leisure, shopping and service trips (Heller-Kemp and Lipps, 2000). Additionally Hanson and Huff's (1986) results show that even in the best case, a number of days are needed, if one can clearly identify types at all.

If one day is not sufficient to represent a person's behaviour then the question is how many days are ? The Mobidrive dataset makes it possible to explore this question through the analysis of the 360 41*41 matrices of Levenshtein distances calculated above. Using the Ward – cluster algorithm (Ward. 1963) the days were grouped for each person. The dendrogram in Figure 6 shows for example how the days were grouped for person 3. One can see for example, how the Saturdays (days 6, 13, ... 41) are increasingly brought together.

It is reasonable to assume, that the number of clusters which is required to explain a certain share of the behavioural variance will vary from person to person. The average variance explained across the respondents by number of clusters is shown by Figure 7. On average ten clusters are needed to obtain an R^2 of 0.75. Three are needed to account for half of the explained variance. It is obvious that a single representative day is insufficient to characterise a person's behaviour.

Figure 6 Cluster analysis of the reported days of person 3 (dendogram; 1 = Monday of the first reporting week)

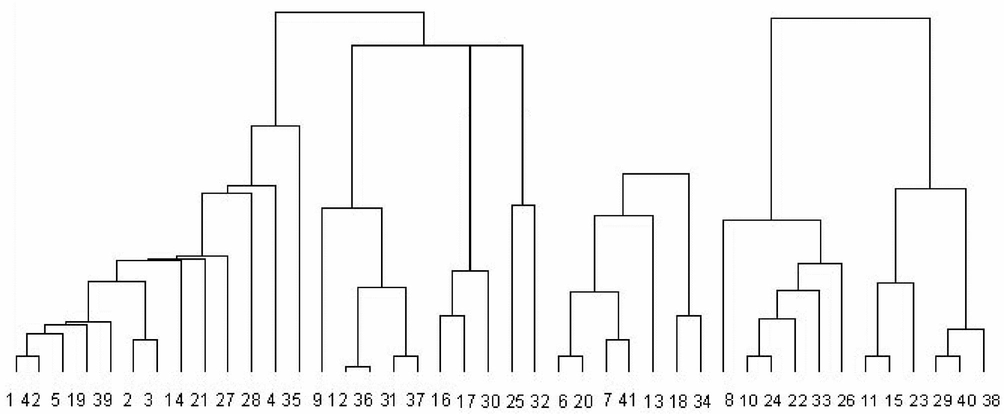
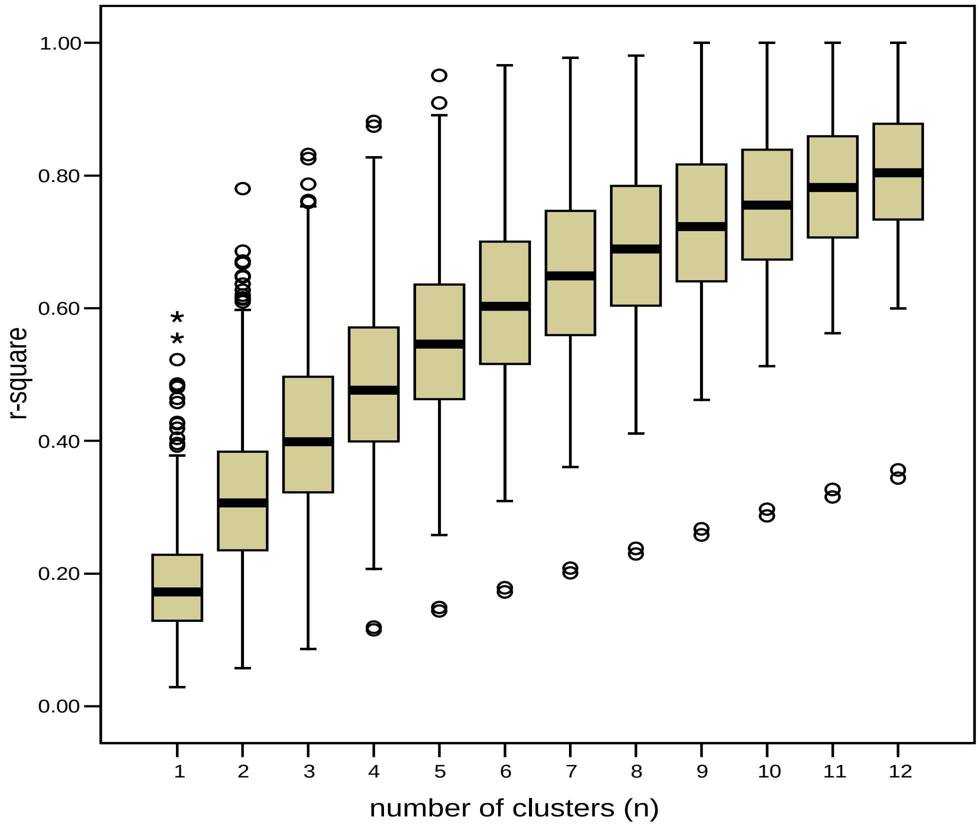


Figure 7 Average R^2 by number of clusters across 360 cluster analyses of the Mobidrive respondents



The pattern of the increase in explanatory power does not have a kink, which would allow the identification of an optimal number of clusters. Next to the explanatory power it would be useful, if the solution has an internal coherence and does not require too much computing time in the next step, the identification of the homogeneous groups. The three cluster solutions have on average an R^2 of 0.45 and would allow to describe the respondents in a parsimonious way. In the next step, the internal coherence of the clusters of the three-cluster solution was analysed. The three clusters of each person were sorted by the number of days belonging to them. This makes the clusters more comparable, as a possible cluster of Sundays will always be smaller than a possible cluster of weekdays.

The largest clusters cover nearly ¾ of the reported days and the different days are represented roughly equally (Table 2). In the second, much smaller cluster the weekend days are overrepresented, while Fridays and days with below average number of trips mark the last and smallest cluster (less then 10% of the reported days). As the three-cluster solution provides a coherent picture of the days and fulfils the requirement of parsimony with a reasonable to good explanatory power, it was adopted as the basis for the next step of the analysis.

Table 2 Properties of the clusters of the three-cluster solutions

	Largest cluster	Second largest cluster 2	Smallest cluster
Average number of days [n]	30.1	8.6	3.3
Average number of trips per day [n]	3.7	3.7	3.5
Share of Mondays [%]	14.9	12.0	14.2
Share of Tuesdays [%]	15.0	11.7	13.5
Share of Wednesdays [%]	14.8	13.1	12.8
Share of Thursdays [%]	15.0	11.7	14.1
Share of Fridays [%]	14.3	13.5	16.4
Share of Saturdays [%]	12.7	19.9	14.7
Share of Sundays [%]	13.4	18.1	14.3

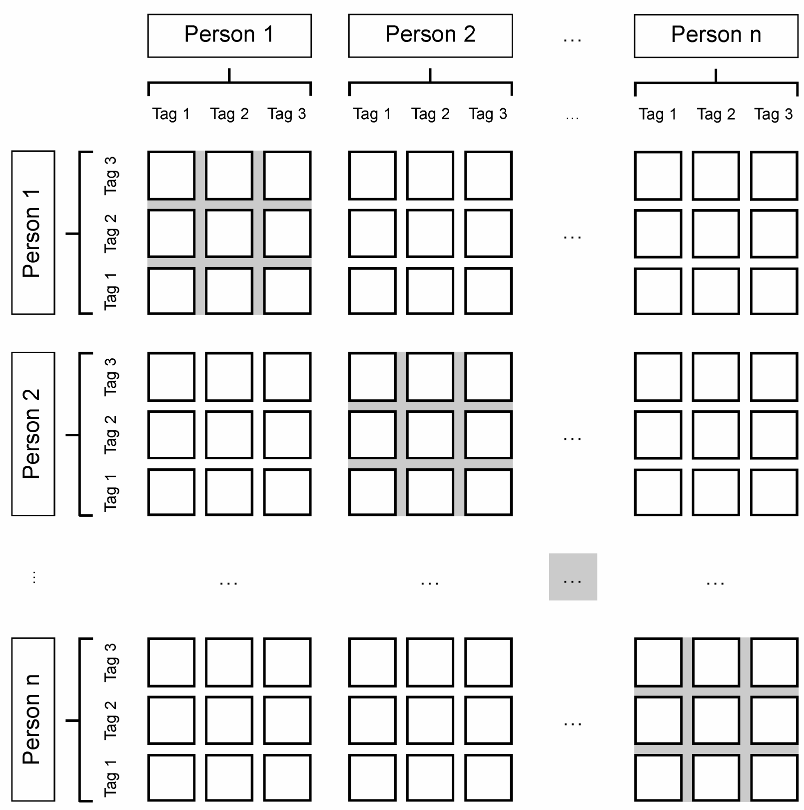
The typical or representative day of each cluster was selected as that day, which had the minimum average Levensthein distance to all other members of that cluster, i.e. the median day.

5 Homogeneous groups and interpersonal variabiltiy

5.1 Identifying person groups based on interpersonal variabilty

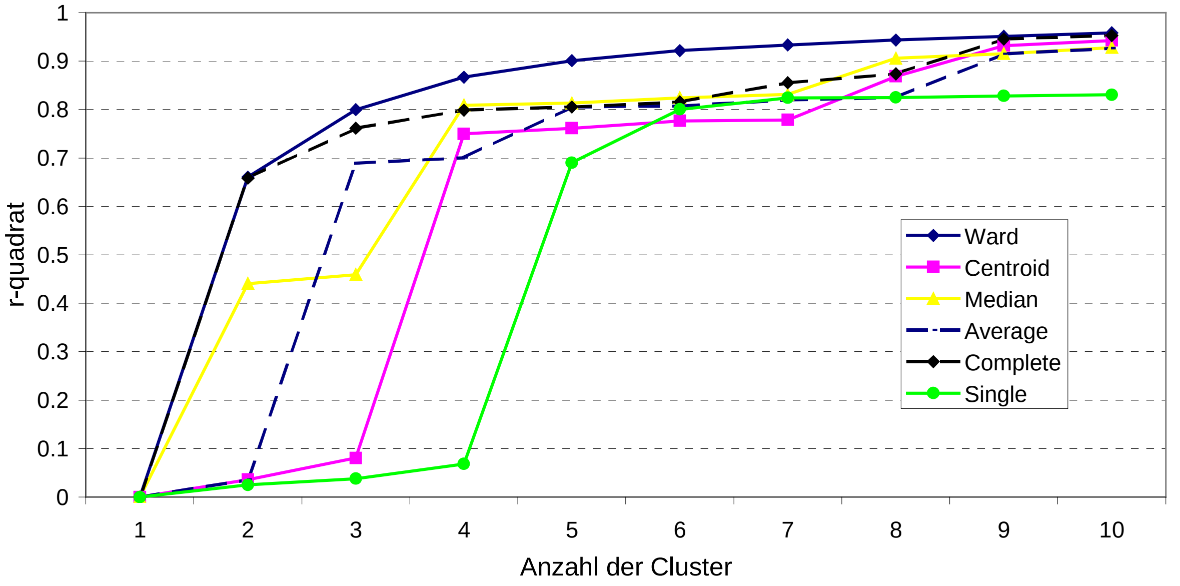
The availability of three representative days for each person makes it possible to construct a distance matrix between all persons with reasonable effort. It is clear, that both clustering and the selection of the median day in those clusters reduces the total variability, but in contrast to earlier work this approach maintains the complexity of the daily patterns to some extent, while ealier work worked with averages, such as average number of trips, median distance per trips etc. For each person pair nine Levenshtein distances were calculated (See Figure 8)

Figure 8 Comparison of the three median days per person



The resulting matrix of 360 * 360 average multi-dimensional Levenshtein distances was analysed using different clustering algorithms. The choice of the preferred number of clusters was based on the explanatory power (R^2) of the solution and if the contribution of the next higher number of clusters to the explanatory power was strongly decreasing. Figure 9 indicates that a solution employing five clusters is sensible. First, the further increase in explained variance is small for solutions beyond 5 clusters. Second, all algorythms need at least five to capture the bulk of the variance. For the preferred five cluster solution the explained variance is between 70 and 90 perscent of the total variance. The solution of the Ward algorythm has the highest power and is preferred in the further analysis.

Figure 9 Explained variance (R^2) by number of clusters and algorythm (average multi-dimensional Levenshtein distance between three typical days for each person)

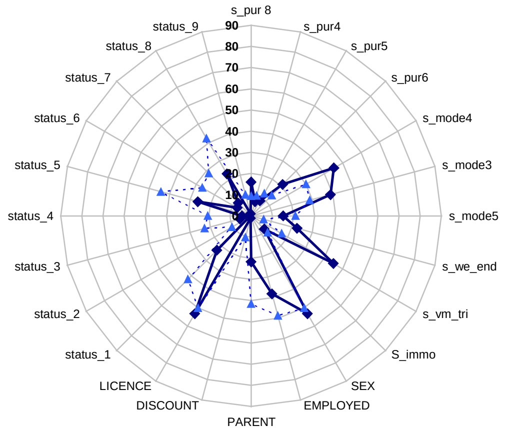


Analysing the size of the clusters for the different algorithms and number of clusters reveals that the five cluster solution of the Ward algorithm avoids very small and very large clusters, which is desirable. The others produce residual clusters with only a few members. The step from the four to the five cluster solution of the Ward algorithm breaks up the largest previous cluster, while the next step only identifies a tiny subgroup of the second largest cluster in the four cluster solution.

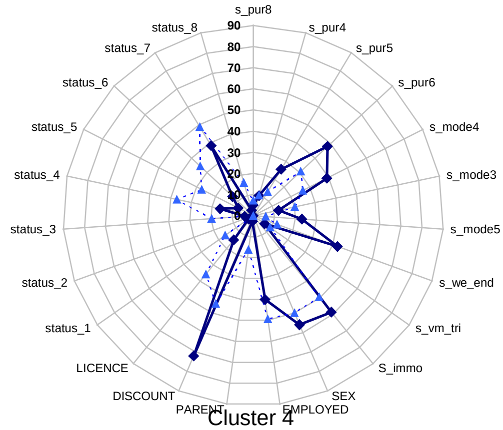
Figure 10 visualises the results of Table 3 and Table 4, which describe both the sociodemographic and travel behavioural characteristics of the cluster members. It is noticeable that the clusters do not distinguish themselves from each other along these dimensions, which is a surprising result, given that the previous literature had stipulated the power of the socio-demographic variables to explain travel behaviour, in particular the volume of trips.

Figure 10 Sociodemographics and travel characteristics of the cluster members [in %]

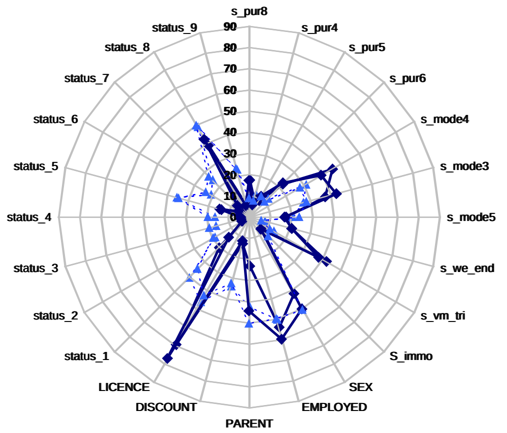
Cluster 1



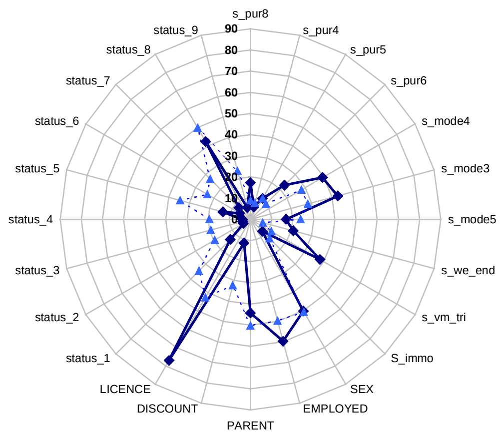
Cluster 2



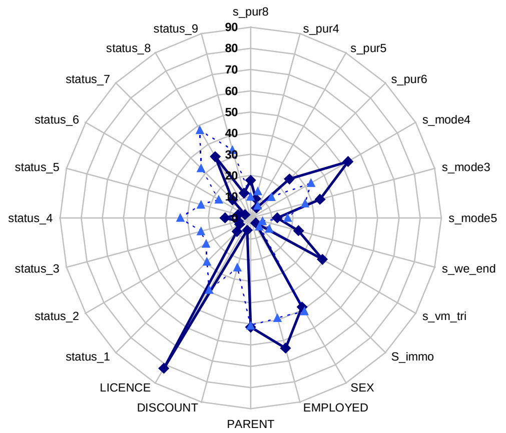
Cluster 3



Cluster 4



Cluster 5



s_pur8: % of leisure trips	Parent: % partens
s_pur4: % of school trips	Discount: % discount pt
s_pur5: % of working trips	Licence: % driving licence
s_pur6: % of shopping trip	status_1: % pupil
s_mode4: % of car trips	status_2: % student
s_mode3: % of unmotorised trips	status_3: % Apprentence
s_mode5: % of pt trips	status_4: % Housemaker
s_we_end: % of weekend trips	status_5: % Retiree
s_vm_tri % share of am trips	status_6: % Unemployed
s_immo: % of immobile dayd	status_7: % Parttime
Sex: share females	status_8: % Fulltime
Employed: % employed	status_9':% Self-employed

Mean: Standard deviation:

Table 3 Travel behaviour characteristics of the 5-cluster solution (Ward based on average multi-dimensional Levenshtein distance between three typical days for each person)

	Cluster 1 (n=92)		Cluster 2 (n=75)		Cluster 3 (n=108)		Cluster 4 (n=52)		Cluster 5 (n=33)	
Variable	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Leisure trips [%]	16.0	9.5	15.2	7.2	17.4	9.0	17.3	9.0	17.7	10.1
Education trips [%]	7.0	9.9	5.0	7.2	8.3	11.1	6.0	8.1	9.3	13.1
Work trips [%]	8.2	12.3	9.9	10.0	8.9	11.1	11.4	11.4	5.4	6.6
Shopping trips [%]	21.1	13.8	25.5	13.2	22.0	12.3	22.7	10.3	25.8	13.9
Car based trips [%]	45.2	29.9	48.0	30.7	45.1	30.7	39.4	27.8	53.1	32.9
Non-motorised trips [%]	38.8	28.6	39.0	26.4	36.8	26.1	42.8	28.1	33.8	26.6
Public transport trips [%]	15.1	21.0	12.3	20.1	17.2	20.2	16.8	23.7	12.6	17.5
Mean trip distance [km]	8.1	9.1	7.5	7.2	9.9	6.5	9.3	7.6	8.2	5.4
Weekend trips [%]	22.5	6.1	23.1	5.9	20.4	7.4	20.9	6.1	23.5	5.8
A.M. trips [%]	44.9	16.6	42.4	12.1	41.8	13.2	38.0	11.4	39.1	10.2
Number of persons/trip [n]	1.5	0.5	1.6	0.4	1.7	0.6	1.6	0.6	1.6	0.3
Days without trips [%]	8.7	11.0	6.6	9.7	7.5	8.5	8.0	12.7	3.4	5.8
Number of trips/day [n]	2.9	1.0	3.9	1.1	3.6	1.2	3.8	1.5	4.8	1.2
Mean Levenshteindistanz	4.9	1.7	6.4	1.9	8.7	2.9	7.0	3.7	8.3	1.8

Table 4 Sociodemographic characteristics of the 5-cluster solution (Ward based on average multi-dimensional Levenshtein distance between three typical days for each person)

	Cluster 1 (n=92)		Cluster 2 (n=75)		Cluster 3 (n=108)		Cluster 4 (n=52)		Cluster 5 (n=33)	
Variable	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
Age [Years]	41.3	21.7	41.2	18.4	41.7	49.5	39.9	17.1	48.5	50.8
Household cars [n]	1.2	0.5	1.0	0.6	23.1	42.4	1.2	0.5	63.6	48.9
Males [%]	53.3	50.2	58.7	49.6	13.0	33.7	50.0	50.5	51.5	50.8
Being parent [%]	21.7	41.5	40.0	49.3	69.4	46.3	44.2	50.2	6.1	24.2
Rail discount card [%]	1.1	10.4	2.7	16.2	20.4	40.5	11.5	32.3	81.8	39.2
Driving licence [%]	53.3	50.2	72.0	45.2	3.7	19.0	76.9	42.5	9.1	29.2
Pupil [%]	22.8	42.2	14.7	35.6	2.8	16.5	13.5	34.5	6.1	24.2
University student [%]	1.1	10.4	2.7	16.2	2.8	16.5	3.8	19.4	6.1	24.2
In education [%]	5.4	22.8	0.0	0.0	14.8	35.7	3.8	19.4	12.1	33.1
Housewife [%]	4.3	20.5	4.0	19.7	4.6	21.1	3.8	19.4	6.1	24.2
Retired [%]	26.1	44.2	16.0	36.9	6.5	24.7	13.5	34.5	3.0	17.4
Unemployed [%]	7.6	26.7	8.0	27.3	38.9	49.0	5.8	23.5	12.1	33.1
Part time [%]	8.7	28.3	13.3	34.2	5.6	23.0	7.7	26.9	33.3	47.9
Full time [%]	22.8	42.2	38.7	49.0			42.3	49.9	31	46
Selfemployed [%]	1.1	10.4	2.7	16.2	3	17	5.8	23.5	4	19

5.2 Analysing person groups

The next steps were an ANOVA and a discriminate analysis to check the quality of the classification and analyse the impact of various variables on cluster membership. The ANOVA method was used to analyse, whether the cluster membership has a significant influence on the value of the tested variable. This is indicated by two asterisks in Table 6. The amount of explained variability for this variable is given in the column R^2.

Table 5 ANOVA for cluster membership

Behavioural variables			Sociodemographic variables		
Variable	P > F	R^2	Variable	P > F	R^2
Leisure trips [%]	0.406	0.01	Age [Years]	0.857	0.00
Education trips [%]	0.133	0.02	Household cars [n]	0.015**	0.03
Work trips [%]	0.139	0.02	Males [%]	0.226	0.02
Shopping trips [%]	0.120	0.02	Being parent [%]	<0.001**	0.06
Car based trips [%]	0.610	0.00	Rail discount card [%]	<0.001**	0.04
Non-motorised trips [%]	0.319	0.01	Driving licence [%]	<0.001**	0.04
Public transport trips [%]	0.502	0.00	Pupil [%]	0.293	0.01
Mean trip distance [km]	0.199	0.02	University student [%]	0.646	0.01
Weekend trips [%]	0.020**	0.03	In education [%]	0.316	0.01
A.M. trips [%]	0.032**	0.03	Housewife [%]	0.254	0.01
Number of persons/trip [n]	0.615	0.01	Retired [%]	0.059*	0.03
Days without trips [%]	0.101	0.02	Unemployed [%]	0.776	0.00
Number of trips/day [n]	<.001**	0.17	Part time [%]	0.561	0.01
Mean Levenshteindistanz	<.001**	0.26	Full time [%]	0.077*	0.02
			Self employed [%]	0.088*	0.02

* significant (10% - level)

** significant (5% - level)

It is shown, that for the majority of variables no significant connection between the cluster membership and the characteristics of the variables exists. The number of trips in the morning and at weekends, the number of daily trips and the level of intrapersonal variability are exceptions among those variables which describe travel behaviour. These findings are not surprising, as the temporal order of activities is a major task of analysis in the sequence analyses.

Among the socio-demographic variables the number of pvs, driving licences and tickets for the public transport per household are as well significantly influenced by cluster membership as are the share of parents, retirees, fulltime- and self-employed people.

Yet, the amount of explained variability by cluster membership is below 7 percent for each sociodemographic variable. The highest share of explained variability is reached for the behavioural variables, particularly for the number of daily trips and the amount of intrapersonal variability with 17, respectively 26 percent.

To integrate the various variables a discriminant analysis was performed to determine the their power to predict cluster membership jointly and individually. The discriminant function, also called a *canonical root*, is estimated as a linear combination of discriminating (independent) variables, such that $L = b_1x_1 + b_2x_2 + ... + b_nx_n + c$, where the b's are discriminant coefficients, the x's are discriminating variables, and c is a constant. This is analogous to multiple regression, but the b's are discriminant coefficients which maximize the distance between the means of the criterion (dependent) variable. The maximum number of discriminant functions to be estimated is the number of independent groups minus one. The analysis here was peforemd with for different number of discriminant functions.

The functions will be evaluated by their eigenvalue and wilk’s lamda. The eigenvalue of each discriminant function reflects the ratio of importance of the dimensions which classify cases of the dependent variable. If there is more than one discriminant function, the first will be the largest and most important, the second next most important in explanatory power, and so on. The eigenvalues assess relative importance because they reflect the percents of variance explained in the dependent variable summing to 100% for all functions.

The following tables show, that the eigenvalue is low for all four functions, which means that the devation of the values between the clusters is low compared to the deviation within the clusters. The column percentage of variance shows which amount of deviation of all values can be explained by each discriminant function with the usual pattern of declining contributions.

Table 6 Explanatory power and significance of the discriminate functions

Discriminate function	Wilks' Lambda**	Significance	Eigenvalue	% of variance
1	0.81	0.00	0.12	51.36
2	0.90	0.00	0.07	32.52
3	0.96	0.24	0.03	15.01
4	1.00	0.93	0.00	1.11

** combination of functions (row 1: function 1-4, row 2: function 2-4, etc.)

Wilks's lambda is used to test mean differences, such that the smaller the lambda for a function, the more it contributes to the discrimination. Lambda varies from 0 to 1, with 0 meaning group means differ and 1 meaning all group means are the same.

For the last two functions the hypotheses of identical mean values cannot be rejected. For this reason – and because of the smaller eigenvalues of these functions – only the coefficients of first two discrimant functions will be considered. The standardised and non standardises values of the coefficients of those functions are shown the following table– to evaluate the impact of a variable the standardises coefficients have to be compared.

In the first discriminant function the possession of a driving licence or a rail discount card as well as gender have the highest impact on cluster membership. Being parent is the strongest variable in the second discriminant functions, followed by the possession of a a rail discount card and gender.

Table 7 Coefficients of the first two discriminant functions

	Function 1 Coefficient	standardised-Coefficient	Function 2 Coefficient	standardised Coefficient
Gender	-0.65	-0.32	0.90	0.45
Age	-0.03	-0.53	-0.01	-0.19
Driving licence [y/n]	1.48	0.68	0.29	0.13
Employed [y/n]	0.40	0.20	-0.32	-0.16
Rail discount card [y/n]	2.06	0.52	-1.96	-0.49
No. of pvs	0.03	0.02	-0.66	-0.38
Being parent [y/n]	0.71	0.32	1.75	0.80

6 Discussion and conclusions

In contrast to the earlier literature on the interpersonal comparison of travel behaviour and on homogeneous groups of travellers, which ignored the information contained in the sequence of activities, this paper is based on a comprehensive analysis of just this aspect of daily behaviour. The multidimensional sequence analysis of Joh (2004) is found to be an appropriate way to capture this sequence information and to derive suitable measure of similarity between days of the same or different persons. The substantial computing requirements limit its use to some extent today, but not to extent to exclude its use.

The use of this ignored aspect of travel behaviour leads to the conclusion, that the persons with similar behavioural sequences do not form groups which can be easily identified by their sociodemographics or their average travel behaviour. Licence ownership, rail discount card ownership and sex were able to predict the cluster membership to a some extent in ANOVA and discriminant analyses. This can be seen to be disappointing, but on the other hand it indicates new dimensions, which will need to be explored, especially as the groups show a high degree of internal behavioural coherence, as indicated by the large explanatory power of the five cluster solution analysed above.

This impression of the independence of this aspect from the aspects traditionally analysed was confirmed, when the results of this clustering were compared to a clustering based on average behavioural characteristics (Schlich, 2004). The membership in the cluster of one solution had no power to predict membership in the other.

The results presented here are unique both due to their methodology and the richness of the behavioural data (six week diary). There is a clear need for their replication with other multiple-day data sets. There is also a need to develop cruder measures of the structure of the sequences, which would allow us to characterise them and to understand their differences more cheaply. This is true in particular, if we want to use this information in agent-based simulation approaches to activity scheduling, where the maintenance and replication of the true level of behavioural variability is crucial.

7 Acknowledgements

We would like to thank the Mobi*drive* project colleagues T Haupt and A Zimmermann (PTV AG, Karlsruhe), KJ Beckmann and G Rindsfüser (Institut für Stadtbauwesen, RWTH Aachen) as well as our IVT colleagues A König and S Schoenfelder for the collaboration. We would also like to thank Chang-Hyeon Joh (Imperial College, London) and Prof. Harry Timmermans (University Eindhoven) for their support.

8 Reference

Abbott, A. and A. Tsay (2000) Sequence analysis and optimal matching methods in sociology: Review and prospect, *Social Methods & Research*, **29** (1) 3-33.

Axhausen, K.W. , M. Löchl, R. Schlich, T. Buhl and P. Widmer (2004) Fatigue in long duration surveys, *Arbeitsbericht Verkehrs- und Raumplanung*, **262,** IVT, ETH Zürich, Zürich.

Axhausen, K. W., S. Schönfelder, J. Wolf, M. Oliveira and U. Samaga (2004) 80 weeks of GPS-traces: Approaches to enriching the trip information, *Transportation Research Record*, **1870**, 46-54.

Axhausen, K.W., A. Zimmermann, S. Schönfelder, G. Rindsfüser and T. Haupt (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **29** (2) 95-124.

Bargeman, B., C.H. Joh and H.J.P. Timmermans (2002) Vacation behaviour using a sequence alignment method, *Annals of Tourism Research*, **29** (2) 320-337.

Berger, M. (2000a) Formation of typologies of similar changes and differences in activity behaviour – a multi-method-approach und application of the optimal-matching-technique, Vortrag bei der 9th International Association of Travel Behaviour Conference, Gold Coast/Queensland, Juli 2000.

Berger, M. (2000b) Abbildung und Erklärung von Unterschieden zwischen Aktivitätenmustern – ein Multimethodenansatz unter Verwendung der Optimal-Matching-Technik, *Stadt Region Land*, **69**, 145-155.

Billari, F.C. (1999) Sequence analysis in demography: changes and wishes, Vortrag beim Workshop on Longitudinal Research in Social Science: A Canadian Focus, London, Oktober 1999.

Dollase, R., K. Hammerich and W. Tokarski (2000*) Temporale Muster – die ideale Reihenfolge der Tätigkeiten*, Leske und Buderich Verlag, Opladen.

Erzberger, C. and G. Prein (1997) Optimal Matching Technik: Ein Analyseverfahren zur Vergleichbarkeit und Ordnung individuell differenter Lebensverläufe, *ZUMA-Nachrichten*, **21**, 52-81.

Golob, T.F. and H. Meurs (1986) Biases in response over time in a seven-day travel diary, *Transportation,* **13** (2) 163-181.

Hanson, S. (1980) Spatial diversification and multipurpose travel: Implications for choice theory, *Geographical Analysis,* **12,** 245-257.

Hanson, S. and K.O. Burnett (1981) Understanding complex travel behaviour: Measurement issues, in P.R. Stopher, A.H. Meyburg and W. Brög (Eds.) *New Horizons in Travel-Behaviour Research*, 207-230, Lexington.

Hanson, S. and K.O. Burnett (1982) The analysis of travel as an example of complex human behaviour in spatially-constraint situation: Definition and measurement issues, *Transportation Research A*, **16** (2) 87-102.

Hanson, S. and J.O. Huff (1982) Assessing day-to-day variability in complex travel patterns, *Transportation Research Record*, **891,** 18-24.

Hanson, S. and J.O. Huff (1986) Classification issues in the analysis of complex travel behaviour, *Transportation*, **13** (4) 271-293.

Hanson, S. and J.O. Huff (1988a) Systematic variability in repetitious travel*, Transportation*, **15** (2) 111-135.

Hanson, S. and J.O. Huff (1988b) Repitition and day-to-day variability in individual travel patterns: Implication for classification, in R.G. Golledge and H. Timmermans (Eds.) Behavioural modelling in geography and planning, Croom Helm . London.

Hertkorn,G. and M. Kracht (2002) Analysis of large scale time use surveys with respect to travel demand an regional aspects, Vortag bei der International Association of Time Use Research (IATUR), Annual Conference 2002, Oktober, 2002, Lissabon.

Herz , R. (1983) Stability, variability and flexibility in everyday behaviour, in S. Carpenter und P.M. Jones (Hrsg.) *Recent Advances in Travel Demand Analysis,* 385-400, Gower, Aldershot.

Huff, J.O. and S. Hanson (1986) Repetition and variability in urban travel, *Geographical Analysis*, **18** (2) 97-114.

Huff, J.O. and S. Hanson (1990) Measurement of habitual behaviour: Examining systematic variability in repetitive travel, in P.M. Jones (Ed.) *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, 229-249, Gower, Aldershot.

Joh, C.H. (2004) Measuring and predicting adaptation in multidimensional activity-travel patterns, *bouwstenen faculteit bouwkunde,* **79**, Dissertation an der Technischen Universität Eindhoven. Eindhoven.

Joh, C.H., T.A. Arentze, F. Hofman and H.J.P. Timmermans (2002) Activity pattern similarity: A multidimensional alignment method, *Transportation Research B*, **36** (5) 385-403.

Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2001a) A position-sensitive sequence alignment method illustrated for space-time activity-diary data, *Environment and Planning A*, **33** (2) 313-338.

Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2001b) Multidimensional sequence alignment methods for activity-travel pattern analysis – A comparison of dynamic programming and genetic algorithms, *Geographical Analysis,* **33** (3) 247-270.

Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2001c) Pattern recognition in complex activity-travel patterns: A comparison of Euclidean distance, signal processing theoretical, and multidimensional sequence alignment methods, *Transportation Research Record,* **1752,** 16-22.

Jones, P. and M. Clarke (1988) The significance and measurement of variability in travel behaviour, *Transportation*, **15** (1) 65-87.

Kitamura. R. and T. van der Hoorn (1987) Regularity and irreversibility of weekly travel behaviour, *Transportation*, **14** (2) 227-251.

Kunert, U. (1994) Weekly mobility of life cycle groups, *Transportation*, **21** (3) 271-288.

Kutter, E. (1972) Demographische Determinanten des städtischen Personenverkehrs, *Veröffentlichungen des Instituts für Stadtbauwesen der TU Braunschweig,* **9**, TU Braunschweig, Braunschweig.

Lipps, O. (2002) Inter- und intrapersonelle Variation der Aktivitätsrhythmen, in *Heureka 02 – Optimierung in Verkehr und Transport,* Forschungsgesellschaft für Strassen- und Verkehrswesen, Köln, 469-479.

Lipps, O. (2001) Modellierung der individuellen Verhaltensvariation bei der Verkehrsentstehung, *Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe*, **58**, Dissertation an der Universität Karlsruhe, Karlsruhe.

Madre, J.-L., K.W. Axhausen and W. Brög (2004) Immobility in travel diary surveys: An overview, *Arbeitsbericht Verkehrs- und Raumplanung*, **207**, IVT, ETH Zürich, Zürich.

McClure, M.A., T.K. Vasi and W.M. Fitch (1994) Comparative analysis of multiple protein sequence alignment methods, *Molecular Biology and Evolution*, **11** (4) 571-592.

Pas, E.I. (1983) A flexible and integrated methodology for analytical classification of daily travel-activity behavior, *Transportation Science,* **17** (4) 405-429.

Pas, E.I. (1987) Intrapersonal variability and model goodness-of-fit, *Transportation Research A,* **21** (6) 431-348.

Pas, E.I. (1988) Weekly travel-activity behaviour, *Transportation,* **15** (1) 89-109.

Pas, E.I and F.S. Koppelmann (1986) An examination of the determinants of day-to-day variability in individuals' urban travel behaviour, *Transportation*, **13** (2) 183-200.

Pas, E.I. and S. Sundar (1995) Intrapersonal variability in daily urban travel behaviour: Some additional evidence, *Transportation*, **22** (2) 135-150.

Pendyala, R.M., A. Parashar and G.R. Muthyalagari (2000) Measuring day-to day variability in travel characteristics using GPS data, paper presented at the 79th Annual Meeting of the Transportation Research Board, January 9-14, 2001 Washington.

Recker, W.W., M.G. McNally and G.S. Root (1985) Travel/Activity anaysis: Pattern recognition, classification and interpretation, *Transportation Research A*, **19** (4) 279-296.

Rindsfüser, G. and S. Doherty (2000) Konzepte, Module und Datenerfordernisse für SMART – Simulationsmodell des Aktivitäten-(Re)Planungsprozesses, *Stadt Region Land*, **69,** 109-131.

Sankoff, D. and J.B. Kruskal (1983) *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*, Addison-Wesley, Reading.

Schaeper, H. (1999) Erwerbsverläufe von Ausbildungsabsolventinnen und Absolventen – eine Anwendung der Optimal Matching Technik, *Arbeitspapier des Sonderforschungs-bereichs 186 der Universität Bremen*, **57**, Universität Bremen.

Schlich, R. (2004) Verhaltenshomogene Gruppen in Längsschnitterhebungen, Dissertation an der ETH Zürich, Zürich.

Schlich, R. (2003) Homogenous groups of travellers, Vortrag bei der 10th International Conference on Travel Behaviour Research, Luzern, August 2003.

Schlich, R. (2001) Analysing intrapersonal variability of travel behaviour using the sequence alignment method, Vortrag bei der European Transport Research Conference (ETC), Cambridge, September 2001.

Schlich, R. and K.W. Axhausen (2003) Habitual travel behaviour - evidence from a six week travel diary, *Transportation,* **30** (1) 13-36.

Schlich, R., S. Schönfelder, S. Hanson and K.W. Axhausen (2004) The structures of leisure travel: Temporal and spatial variability, *Transport Reviews,* **24** (4) 219-228.

Schmiedel, R. (1984) Bestimmung verhaltensähnlicher Personenkreise für die Verkehrsplanung, Dissertation an der Universität Karlsruhe, Karlsruhe.

Schönfelder, S. and K.W. Axhausen (2001) Mobi*drive* - Längsschnitterhebungen zum individuellen Verkehrsverhalten: Perspektiven für raum-zeitliche Analysen, paper presented at CORP 2001, February 2001, Wien.

Schönfelder, S., R. Schlich, A. König, A. Aschwanden, A. Kaufmann, D. Horisberger und K.W. Axhausen (2002) Mobi*drive*: Data format guide, *Arbeitsberichte Verkehr- und Raumplanung,* **116**, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.

Ward, J.H. (1963) Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, **58** , 236-244, 1963.

Wilson, W.C. (1998a) Activity pattern analysis by means of sequence alignment methods, *Environment and Planning A*, **30** (6) 1017-1038.

Wilson, W.C. (1998b) Analysis of travel behavior using sequence alignment methods, Transportation Research Record, **1645**, 52-59.

Zumkeller, D. and B. Chlond (1995) Nutzen and Realisierungsprobleme einer bundesweiten Paneluntersuchung zum Verkehrsverhalten, *Internationales Verkehrswesen*, **47** (1-2) 20-25.