

Classification of tours in the U.S. National Household Travel Survey through clustering techniques

Original

Classification of tours in the U.S. National Household Travel Survey through clustering techniques / Pirra, Miriam; Diana, Marco. - In: JOURNAL OF TRANSPORTATION ENGINEERING. - ISSN 0733-947X. - STAMPA. - 142:6(2016), p. 04016021. [10.1061/(ASCE)TE.1943-5436.0000845]

Availability:

This version is available at: 11583/2647388 since: 2016-09-05T15:12:26Z

Publisher:

American Society of Civil Engineers

Published

DOI:10.1061/(ASCE)TE.1943-5436.0000845

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

CLASSIFICATION OF TOURS IN THE U.S. NATIONAL HOUSEHOLD TRAVEL SURVEY THROUGH CLUSTERING TECHNIQUES

Miriam Pirra; Marco Diana

31st May 2016

This document is the post-print (i.e. final draft post-refereeing) version of an article published in the *Journal of Transportation Engineering*. Beyond the journal formatting, please note that there could be minor changes and edits from this document to the final published version. The final published version of this article is accessible from here:

[http://dx.doi.org/10.1061/\(ASCE\)TE.1943-5436.0000845](http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000845)

This document is made accessible through PORTO, the Open Access Repository of Politecnico di Torino (<http://porto.polito.it>), in compliance with the Publisher's copyright policy as reported in the SHERPA-ROMEO website:

<http://www.sherpa.ac.uk/romeo/search.php?issn=0733-947X>

Preferred citation: this document may be cited directly referring to the above mentioned final published version:

Pirra, M. & Diana, M. (2016) Classification of tours in the U.S. National Household Travel Survey through clustering techniques. *Journal of Transportation Engineering*, vol. 142(6), pp. 04016021-1-04016021-13.

CLASSIFICATION OF TOURS IN THE U.S. NATIONAL HOUSEHOLD TRAVEL SURVEY THROUGH CLUSTERING TECHNIQUES

Miriam Pirra, Ph.D¹; Marco Diana, Ph.D²

Abstract

Tours are increasingly being considered as an appropriate unit of observation of mobility behaviors and are one of the key ideas underpinning contemporary activity-based modeling approaches. Identifying typologies of tours would benefit both modelers and decision makers, striving to set up more tailored actions aimed at promoting environmentally benign travel choices. Different a-priori classifications based on activity kinds have been proposed, none of which seems clearly preferable on empirical grounds. This paper takes a complementary approach and defines a data-driven segmentation through a cluster analysis of tours that were derived from the trip records of the U.S. National Household Travel Survey of 2009. The socioeconomic characterization of each cluster is finally carried out to link travelers' profiles with specific kinds of tours. Four main tour clusters have thus been identified: non-work tours for compulsory activities done by young individuals, tours done by elder or retired, short and secondary tours within the travel day and tours dominated by the working activity. Their relevance on a modeling and policy viewpoint is discussed.

Keywords

tour-based analysis, cluster analysis, k-means, multimodality, market segmentation, classification

¹ Research Fellow, Dept. of Environmental, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino – ITALY, miriam.pirra@polito.it

² Associate Professor, Dept. of Environmental, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino – ITALY, marco.diana@polito.it (*corresponding author*)

Introduction

Contemporary research in travel demand models is increasingly pointing at the limitations of the traditional trip-based approach, in which single trip characteristics (e.g. destination, travel means) are directly inferred as a function of some explanatory variables. The development of tour-based and activity-based models is grounded on the explicit acknowledgement of the difficulty of studying travel demand without considering both activity patterns of individuals and the additional constraints in their choices that are induced by trip chains (for example, travelers do not have the choice of driving their cars for a return trip if they used transit in the outward journey). According to standard definitions, a *tour* is a sequence of trips starting and ending at the same location (Axhausen, 2007), where the usual reference location is the house, while a *trip chain* is more generally a sequence of two or more trips, irrespective of the location of trip ends.

The implementation of an activity-based model normally implies an a-priori definition of a fairly articulated and multi-level typology of activity patterns and resulting tours, following the scheme of nested logit models (Bowman and Ben-Akiva, 2001). The spatial and temporal distribution of individual activities and their relative importance is first considered, along with the constraints in choices given by household interactions and long term commitments (e.g. residential locations, car availability). A classification of trip chains or tours is then derived, on the basis of these activity patterns and of the interactions among tours occurring in the same day or even week. A variety of classifications has been proposed following these steps, sometimes dictated by data availability issues. Related to this, data sources from different kinds of surveys (e.g. revealed preferences, stated preferences, household-based or screen-line) are often jointly exploited (e.g. Shiftan et al., 2003) to meet the data requirements of such models, as clarified in Arentze and al. (2000).

Different activity classifications are proposed in a number of papers. Krizek (2003) and Primerano et al. (2008) provide good reviews of earlier research in this field and discuss the usefulness of the classical distinction into subsistence or mandatory, maintenance and discretionary activities, where mandatory activities include work and education and are planned first, followed by maintenance activities (such as shopping and health care) and, at the lowest priority level, discretionary activities (social and recreational activities, eating out etc.). Variants to such definitions have been proposed: for example, subsistence activities sometimes do not include attending school. Nishii et al. (1988) and Primerano et al. (2008)

consider primary and secondary activities and their relative order in the trip chain. Strathman and Dueker (1995), O'Fallon and Sullivan (2005) and Stopher et al. (2010) among others follow similar schemes, distinguishing between work, education and non work/non education activities. Even more articulated classifications are proposed for example by Valiquette and Morency (2010) and Nowrouzian and Srinivasan (2012), whereas the importance of considering the spatial localization of activities when classifying tours has recently been advocated (Ho and Mulley, 2013).

According to the state of the art in the research, a number of issues related to tour classification are still open. Concerning the hierarchical structure of models, it has been shown that conceptual issues related to prioritizing activities and, therefore, tours would deserve more research efforts (Buliung and Kanaroglou, 2007). Intuitively appealing assumptions, such as postulating that subsistence activities are planned before maintenance and discretionary ones come later, seem not supported by empirical investigation (Doherty 2006, 2011). On the other hand, characteristics that cannot be metrically expressed, such as activity types and travel modes, are difficult to aggregate at the tour level, since all the possible combinations of cases should be taken into consideration (Axhausen, 2007) and household interactions, such as the presence of "accompanying someone" among trip purposes, further jeopardize the whole process (Cyganski et al., 2013).

The above review shows the intricacies in defining typologies of tours on the basis of ex-ante judgments and assumptions related to the kind of activities, given the number of factors that need to be considered. This paper explores the potential of an alternative strategy to classify tours, by following an empirical and data-driven approach. An ex-post segmentation of tours is therefore proposed through a data clustering technique. These techniques have been largely applied in the past mainly to define homogeneous groups of travelers (e.g. see Anable, 2005; Diana and Mokhtarian, 2009; Elgar and Bekhor, 2004; Jensen, 1999 for some recent works), leading, for example, to more specific travel demand management strategies. However, to the best of the authors' knowledge, they have never been applied to classify tours. Previous descriptive studies looked at more general trip chaining behaviors in relation with socioeconomics, activity patterns or modal usages (Currie and Delbosc, 2011; McGuckin et al., 2005; Valiquette and Morency, 2010), whereas other data mining techniques making use of tour classifiers have been explored to study mode choice at the tour level (Biagioni et al., 2009). Similarities in activity patterns have been studied with more complex methods such as multidimensional sequence alignment (Joh et al., 2002); however, computational

requirements of such methods are likely to make them not easily applicable to classify tours in a very large dataset.

Customer segmentation approaches have already proven their usefulness, for example in order to define tailored and, therefore, more effective travel demand management strategies. In a similar vein, the definition of homogeneous groups of tours and the profiling of the respective groups of travelers would be of great interest for transport policy makers, beyond the above mentioned activity based framework. The final outcome could be the definition of more tailored actions aimed at achieving desirable transport policy goals, such as decreasing environmental impacts or improving congestion. The effectiveness of such policies could be improved, since they would jointly consider both individual characteristics (as in customer segmentation approaches) and practical constraints given by trip chains. From a more theoretical perspective, a growing body of research (that is not reviewed here since it is not the focus of this work) is studying trip chaining behaviors as a function of personal and household characteristics, activity and trip-related attributes and land use patterns. Other studies focus on the link between trip chaining and mode choice (Hensher and Reyes, 2000; Vande Walle and Steenberghen, 2006; Ye et al., 2007). Considering a data-driven classification of tours could contribute to an advancement of such efforts, since models could be developed within each segment. Models calibrated on a more homogeneous set of observations stemming from customer profiling analyses might in fact have a better explanatory power, as was already observed in studies dealing with stated preferences data (Fowkes, 2000, p.48; Louviere et al., 2000, pp. 372-375; Diana and Pronello, 2010).

The objective of this paper is therefore to build clusters of tours and check to which extent the resulting classification is useful both on a transport modeling and on a transport policy viewpoint by identifying tailored policy actions for each group that could be more effective in promoting a behavioral change towards more environmentally sustainable transport choices. Country-level data from the U.S. National Household Travel Survey (NHTS) of 2009 are used to this effect (NHTS, 2009). According to previous research (Timmermans et al., 2003), travel patterns are relatively stable across different spatial settings within an activity-based framework, so that the proposed classification should also be valid beyond the case study here considered, namely the entire U.S., and therefore useful also in different geographical contexts. Since observations related to entire tours are not readily available from the NHTS public use dataset, they are derived from individual trip records. Some initial descriptive

statistics are then presented, aimed at showing the complexity of a tour in terms of activities performed, tour length and joint use of different travel means.

Tour clusters will be obtained following a multi-level approach, that is, iteratively running a k-means algorithm to partition a cluster in two at each iteration. Clusters will then be characterized on a socioeconomic point of view, and the concluding section will discuss to which extent each cluster is representing an identifiable mobility market segment.

Building tours from the NHTS dataset

The main base in the creation of the dataset is the 2009 National Household Travel Survey (NHTS), which provides information on daily travel in the United States (NHTS, 2009). The analysis is related to the identification of all the tours that could be assigned in a day to a person.

Tours from the NHTS dataset are built by consolidating each sequence of observations in the “Day Trip” file that starts and ends at home and these sequences are named “Home-Based (HB) tours”. However, the travel day of the NHTS is a 24-hours period starting at 4am. Therefore, some tours might not completely be recorded: Home-Destination (HD) tours are therefore defined when the individual is on a tour at the beginning of the travel day and the tour ends within the travel day, Home-Origin (HO) tours when the individual is on a tour at the end of the travel day and the tour started within the travel day, and Not Home-Based (NH) tours when the tour started before the travel day and ended after it, or when the respondent is not dwelling at home. 416,180 tours of these four kinds are identified from the 1,167,321 trips in the dataset (about 1.28 tours per individual). The weight of each tour in the sample was computed as the sum of the weights of each trip composing it, the latter being provided in the NHTS dataset. After weighting, 92.7% of the tours are HB-type, the remainder being roughly equally split among the three other categories.

In the following, tours are characterized by deriving from the original dataset some measures of activity durations and travel times: these derived variables are presented in Table 1. To have a manageable number of cases, activities at trip ends are preliminarily consolidated in four classes: “Home” (which defines also the boundaries of a tour), “Work-related”, “Education-related” and “Other”. This is a partially different classification of activities compared to those discussed in the introduction. However, in the present data-driven approach, no hierarchy among activities is postulated, also because this was not empirically

supported by previous research (Doherty, 2006). On the other hand, since the amount of time spent in different activities is considered here, having a more detailed classification for mandatory activities (work and education) seems desirable, also considering that these typically are the most time consuming activities. A similar classification of activities has also been proposed by Yagi and Mohammadian (2008).

The travel means are also grouped in five classes: motorized individual, public transport, bike, walking and other, and their intensity of use in terms of travel times will be another input of the clustering algorithm. According to the structure of the questionnaire, the primary travel means for each trip is known: the sum of the corresponding travel times for all trips in a tour is reported in the “MODE...” variables that are listed in Table 1. Additionally, if the primary travel means was transit, the survey recorded all modes being used get to/from the terminal (including another transit service): these data are not considered here for simplicity.

Table 1. Derived Variables Related to Tours

Variable	Description	Weighted mean value or breakdown
TOURTYPE	Type of tour, one of the following four: Home-Based (HB), Home-Origin (HO), Home-Destination (HD), Not home-based (NH)	HB: 92.7%; HO: 2.3%; HD: 2.3%; NH: 2.7%
ACTIWORK	Total duration of “Work” activities (minutes)	117.2
ACTIOTHE	Total duration of “Other” activities (minutes)	92.5
ACTIEDU	Total duration of “Education” activities (minutes)	28.8
ACTITOT	Total activity time (sum of ACTIWORK, ACTIEDU and ACTIOTHE) (minutes)	238.5
MODEPERS	Total travel time by a motorized individual transport mode (minutes)	50.9
MODEPUBT	Total travel time by public transport (minutes)	4.4
MODEBIKE	Total travel time by bicycle (minutes)	0.5
MODEWALK	Total travel time by walk (minutes)	4.0
MODEOTHE	Total travel time by other modes (minutes)	0.5
MODETOT	Total travel time (sum of MODEPERS, MODEPUBT, MODEBIKE, MODEWALK and MODEOTHE) (minutes)	60.3
TOURTOT	Total tour duration (sum of ACTITOT and MODETOT) (minutes)	298.8
NUMTRIPS	Number of trips composing the tour (pure number)	3.4
TOURWGHT	Tour weight (sum of the trip weights) (pure number)	933,021

Several other variables could have been used to define clusters, according to the above literature review. From Table 1, it can be seen that the choice is to focus on the amount of time spent either in performing a given activity or in travelling, without distinguishing

between more important and less important activities or transport means. The underlying principle is to cluster tours taking the point of view of the traveler’s overall experience and perspective, that are mostly influenced by those activities (including travelling) that last longer.

It is worth noting in passing that the trip chaining dataset that is available in the NHTS website (NHTS, 2009) presents a similar aggregation of the trip-level observations; however, consistently with the definitions given in the introduction, trip chains need not necessarily start and end at the same location, as in this case they are simply bounded by any activity that lasts at least 30 minutes. Therefore, such dataset is not usable in the present analysis.

Descriptive statistics

Analysis of HB tours by activity type

A preliminary analysis is based on the activities that are performed during the tour. Table 2 lists eight main different types of tours that stem from the above introduced four classes of activities (home, work, education and other), along with some descriptive statistics related to tours that were entirely made during the survey reporting period (the so-called home-based, or HB tours). In particular, the second column reports both the absolute numbers of tours not considering sample weights and the weighted percentages over the total number of HB tours, while the last four columns report for each type of tour the average number of trips it contains (NUMTRIPS), the total time spent in performing activities during the tour (ACTITOT), the total travel time (MODETOT) and the sum of the latter two figures, that represents the total tour duration (TOURTOT). Since tour weights are the sum of the relative trip weights, tour types with higher NUMTRIPS mean values show relatively larger percent values in the second column of the table.

Table 2. Definition of the Activities for Home-Based (HB) Tours

Kind of tour	Number of observed tours (weighted %)	NUMTRIPS (mean)	ACTITOT (mean) [min]	MODETOT (mean) [min]	TOURTOT (mean) [min]
Home – Work – Home (HWH)	43487 (10.6%)	2.5	433.3	64.6	497.9
Home – Education – Home (HEH)	16581 (4.2%)	2.0	393.0	41.7	434.7
Home – Other – Home (HOH)	283852 (64.4%)	3.2	120.4	54.3	174.7

HOH_personal	101247 (31.9%)	2.7	70.2	42.7	112.9
HOH_social	105197 (30.2%)	2.3	131.8	42.0	173.8
HOH_transport	19745 (6.9%)	2.2	28.6	31.6	60.2
HOH_personal+social	38842 (18.9%)	4.4	181.6	79.2	260.8
HOH_other	18821 (12.1%)	5.1	180.9	89.7	270.6
Home – (Work and Other) – Home (HOWH)	36934 (17.1%)	5.0	477.1	90.3	567.4
Home – (Other and Education) – Home (HOEH)	6778 (3.1%)	4.1	472.0	67.9	539.9
Home – (Work and Education) – Home (HWEH)	302 (0.2%)	3.2	517.3	86.2	603.5
Home – (Work, Other and Education) – Home (HOWEH)	244 (0.2%)	5.8	605.0	114.7	719.7
Home – Home (HH, no activity)	1414 (0.2%)	1.0	0	24.8	24.8
All HB tours	389592 (100%)	3.4	238.4	61.6	300.0

The most frequent types of tours contain activities classified as “Other”, be these latter alone or in combination with work activities. Table 2 therefore reports some additional indented lines with a breakdown of Home-Other-Home (HOH) tours considering three main sub-categories of “Other” activities: “personal”, not necessarily involving a social dimension (e.g. shopping, doing errands, using medical, personal or professional services); “social”, such as having a meal, playing sports, going to religious activities, visiting someone or having vacations, and “transport”, dealing with accompanying someone. Two additional indented lines report descriptive statistics of more complex tours that combine either personal and social activities, or several activities not falling in the previous three groups.

It is quite surprisingly to see that only 10.6% of HB tours exclusively contains work-related activities (first line of the table). HB tours without intermediate activities (HH tours) are a minimal fraction. Such tours could, for example, be related to jogging, going for a simple walk or taking out the dog. Concerning the mean number of trips NUMTRIPS, it is obviously one for HH tours, equal to 2 for Home-Education-Home (HEH) tours and greater than 2 for the others. Among these latter, Home-Work-Home (HWH) tours are the simplest ones, the number of trips not being barely equal to 2 due to the existence of jobs such as salesmen. In the other cases and in particular for HOH and HOWH (Home-Other-Work-Home) tours, the average number of trips increases due to the greater complexity of related activity patterns.

The average duration of tours and activities in the classes of Table 2 is also an interesting aspect to be considered and analyzed. For the type HWH, the average daily work time ACTITOT is a little more than 7 hours, and the total work-related travel time including

commuting MODETOT is a little more than one hour. Similar figures are found for HEH trips, but with shorter travel times. Conversely, HOH tours have both shorter activity durations and travel times. This is also confirmed by looking at the indented lines: HOH_personal tour activities last about an hour, which could be the time necessary to shop, while half an hour could be sufficient to transport someone (MODETOT equal to 31.6 minutes in HOH_transport). A meeting with friends, instead, takes usually more time, as confirmed by the ACTITOT value for HOH_social that exceeds 2 hours.

When both work or education and some other activities are included in the tour, a longer time is taken for both activities and trips. Assuming that the average work time is the same for individuals performing both HWH and HOWH tours, the difference between ACTITOT values in those two kinds of tours can be interpreted as an estimation of the average time spent performing non-work activities during HOWH tours. This time is equal to about $477 - 433 = 44$ minutes and it can be seen as a kind of upper bound to the duration of the activity that could be added, in a day, to a work-related tour. Computing respectively HOWH/HWH and HOEH/HEH ratios on the number of trips allows concluding that the trip chain propensity is much higher when working activities are involved, rather than education activities.

Finally, HH tours are the shortest ones since they do not involve out of home activities, but the single trip which constitutes them is on average longer than the trips in most of the other kinds of tours, as apparent when comparing MODETOT/NUMTRIP ratios in all lines of the table.

Analysis by number of trips

In the remainder of the section all tours are considered, including the home-based ones that were previously analyzed. The mean number of trips in a tour is 3.4, whereas the modal value of the corresponding distribution is 2, since most tours are made of an outwards and a return trip. It is however interesting to analyze how such distribution changes according to the kind of tour (HB, HO, HD, NH), as shown respectively in Figure 1a, 1b, 1c and 1d. The plot “HB” reflects the behavior of the most general case, since it represents 92.7% of the total number of tours, i.e. 389592 trips, while we have 10,178 HO, 6,500 HD and 9,910 NH trips in the dataset. The frequency of tours having an increasing number of trips sharply decreases, yet more than 10% of HB tours are made of more than 5 trips. The distribution of the number of trips radically changes in the other three cases: modal values for HO and HD trips are respectively 1 and 3, and after such values the distribution is decreasing. All three

distributions are more dispersed, especially NH trips. These results are pointing at a greater complexity of tours that are not starting and ending at home: this could be due to the “boundary effects” related to incomplete tour recordings within the survey reporting period. Such kind of tours should be separately considered in a modeling effort to improve the results.

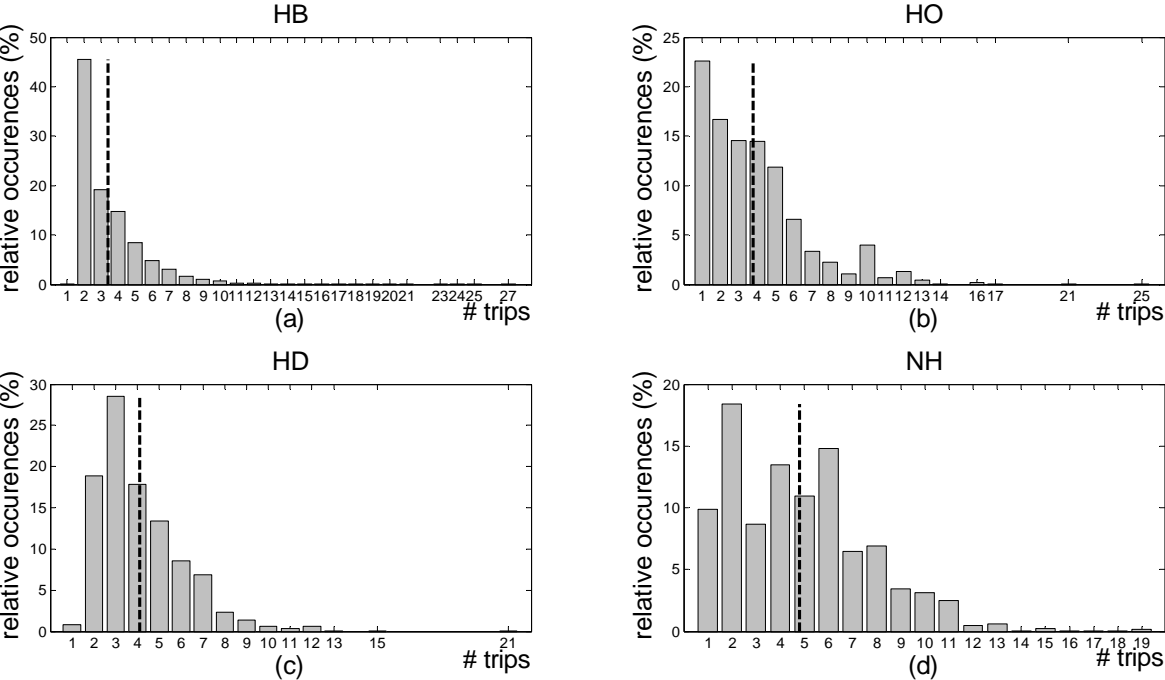


Fig. 1a-d. Relative distribution (%) of the number of trips in a tour by TOURTYPE category: (a) Home-based, (b) Home-Origin, (c) Home-Destination, and (d) Not Home-based. Dotted lines represent mean values.

Analysis by travel means

Considering mobility at the tour level gives a clearer view of the patterns of use of different travel means. The percentages presented in Table 3 can be seen as a representation of the multimodality behaviors of the U.S. population. As expected, people usually use only one transportation mode when they move, and in most cases this is a car, in fewer cases they walk and public transport plays a minor role in monomodal tours. However, when turning the attention to multimodal tours, every 1.4 tours completely done by public transport there are around 4.6 tours in which transit is used together with other modes (such figure can be computed by summing up the percentages of the last column of the table for the relevant rows).

Table 3. The Degree of Multimodality of the Tours

Number of modes in a tour	Mode combinations (*)	Number of occurrences	Within group weighted %	Column total weighted %
1	IM	344,913	88.4	79.0
	WA	36,886	8.6	7.7
	PT	5,436	1.5	1.4
	BI	3,887	1.0	0.9
	OT	1,576	0.5	0.4
	Total	392,699	100.0	89.4
2	IM, WA	11,924	54.4	5.1
	PT, WA	3,509	24.7	2.3
	IM, PT	4,116	13.1	1.2
	IM, OT	1,000	3.9	0.4
	IM, BI	451	2.1	0.2
	BI, WA	227	1.0	0.0
	WA, OT	144	0.5	0.1
	PT, OT	71	0.2	0.0
	PT, BI	35	0.1	0.0
	BI, OT	14	0.0	0.0
	Total	21,491	100.0	9.3
3	IM, PT, WA	1,582	85.5	1.1
	IM, WA, OT	92	4.2	0.1
	PT, BI, WA	42	3.5	0.1
	PT, WA, OT	83	3.1	0.0
	IM, BI, WA	52	1.6	0.0
	IM, PT, OT	60	1.1	0.0
	IM, PT, BI	15	0.8	0.0
	BI, WA, OT	3	0.2	0.0
	IM, BI, OT	2	0.1	0.0
	Total	1,931	100.0	1.3
4	IM, PT, WA, OT	46	68.7	0.0
	PT, BI, WA, OT	4	18.3	0.0
	IM, PT, BI, WA	8	13.0	0.0
	Total	58	100.0	0.0
5	all means	1	100.0	0.0

* IM = Individual motor means; PT = public transport; BI = bicycle; WA = walking; OT = other

Analyzing public transport usage at the tour rather than at the trip level allows better appreciating its importance for travelers, beyond its low market share. The flexibility of public transport that allows travelling without having to retrieve a personal vehicle in the same location where it was left at the end of a previous trip is in fact very useful in specific circumstances to efficiently complete a tour. Considering the values in the last column of Table 3, we see that every 79 monomodal tours by car (first row) we have only 8.1 tours where cars are combined with other means (sum of the percent values in rows where “IM” mode is present). Considering monomodal and multimodal tours involving bicycles, the ratio becomes 0.9 (fourth row) versus 0.3 (sum of the percent values in rows where “BI” mode is

present). Conversely, the same ratio for PT is 1.4 versus 3.5. Individual means are therefore more difficult to combine with others compared to public transport, probably given the constraints related to the availability of the private vehicle in the location where it is needed. These findings are also relevant to understand that the utility of using public transport can be better appreciated at the tour rather than at the trip level; therefore, standard trip-based models are not too fit to fully capture the relevance of public transport, especially in more complex travel patterns.

Cluster analysis

The above descriptive statistics have underlined the inherent complexity and variability of tours. This paper follows a cluster analysis approach aimed at their classification. At the outset, we make a distinction between clustering variables, upon which the definition of the clusters is based, and characterization variables, that are not considered in the cluster building phase but whose variation of the mean value among clusters is analyzed to learn more on the characteristics of the tours within each group.

The above reported analysis based on descriptive statistics put into evidence some interesting issues related to activity patterns on one side, and levels of use of different means on the other. Taking one step forward, we would like now to better characterize the relationships between mode choices and activity types at the tour level. However, the number of clustering variables needs to be sufficiently small to make the interpretation of clusters manageable, so that only a subset of the variables listed in Table 1 are used in the clustering algorithm. “Education” and “Other” activities are therefore jointly considered, so that a new variable ACTIEDOT is created including both ACTIEDU and ACTIOTHE. Thus, the following seven clustering variables are finally retained: ACTIWORK, ACTIEDOT, MODEPERS, MODEPUBT, MODEBIKE, MODEWALK and MODEOTHE.

Clusters are then assessed by looking at how some socioeconomic characteristics of the travelers vary among them: these characterization variables are listed in Table 4. Tour-related variables from Table 1 that were not considered among the seven above clustering variables, namely the number of trips composing the tour and the tour classification, are included as well. In doing so, the relationships between clustering variables (patterns of activities and modal usages) and the traveler’s characteristics can be investigated, taking a complementary approach compared to previous works that more specifically modeled single features of tours (number of stops, chaining sequence) as a function of selected socio-demographic factors.

The preliminary step preceding the clustering is the normalization of the values of the above listed seven variables through the computation of Z-scores. In this way, through a linear transformation, all the values are converted to a common scale, characterized by an average of zero and standard deviation of one (Han et al., 2011). This improves the consistency and stability of the dataset, making the comparison of variables and the computation of distance measures easier.

The main goal of cluster analysis is the partitioning of objects into clusters, each of them being characterized by homogeneity of the elements (in the present case, tours) within the clusters and heterogeneity between the clusters. In the first case, data belonging to the same group should be as similar as possible, while, in the latter, elements from different clusters should be as dissimilar as possible. The applied clustering algorithm is k-means, which performs a crisp clustering assigning a data vector to a cluster on the basis of the Euclidean distance of the vector from the cluster mean (Tan et al., 2006). The “crisp” attribute indicates that each object is assigned exactly to one specific cluster. A different approach would be the “fuzzy” one, where a gradual membership allows the attribution of each element to different clusters. Finally, we do not consider observation weights in this process.

Table 4. Tour-Related and Socioeconomic Variables of the Traveler which Are Used to Characterize Clusters

Variable	Description	Categories/Range	Number of tours (weighted %)
R_AGE	Traveler age	5-88	414172 (99.7%)
		92 = 89+	2008 (0.3%)
R_SEX	Traveler gender	Male	197020 (48.5%)
		Female	219160 (51.5%)
TRAVDAY	Travel day - day of week	Su = Sunday	55910 (12.2%)
		Mo = Monday	58839 (14.2%)
		Tu = Tuesday	61039 (14.8%)
		We = Wednesday	61202 (14.7%)
		Th = Thursday	59343 (14.4%)
		Fr = Friday	62790 (15.4%)
		Sa = Saturday	57057 (14.3%)
LIF_CYC	Life cycle classification for the traveler household	1 = one adult, no children	20157 (7.4%)
		2 = 2+ adults, no children	82841 (19.0%)
		3 = one adult, youngest child 0-5	1697 (1.2%)
		4 = 2+ adults, youngest child 0-5	51785 (18.6%)
		5 = one adult, youngest child 6-15	7379 (2.9%)
		6 = 2+ adults, youngest child 6-15	90060 (25.2%)
		7 = one adult, youngest child 16-21	3001 (1.1%)
		8 = 2+ adults, youngest child 16-21	29086 (9.0%)
		9 = one adult, retired, no children	23450 (3.0%)
		10 = 2+ adults, retired, no children	106724 (12.6%)

PRMACT	Primary activity of the traveler in the previous week	WRK = Working	176860 (51.6%)
		TMA = Temporarily absent from a job or business	10694 (2.9%)
		LFW = Looking for work	8062 (2.8%)
		HMK = A homemaker	36158 (8.0%)
		SCH = Going to school	14573 (5.6%)
		RET = Retired	104340 (11.4%)
		OTH = Doing something else	17027 (4.4%)
		UNC = Unclassified (not used)	48466 (13.3%)
TOURACTI	Tour classification compounding TOURTYPE categories and, for HB tours, activity types from Table 2	HOH = Home – Other – Home	307211 (66.5%)
		HWH = Home – Work – Home	43487 (9.8%)
		HOWH = Home – Other – Work – Home	37480 (16.2%)
		HH = Home – Home	1414 (0.2%)
		HO = Home – Origin	10178 (2.3%)
		HD = Home – Destination	6500 (2.3%)
		NH = Not home-Based	9910 (2.7%)
NUMTRIPS	Number of trips composing the tour	1-27	Mean value: 3.4 Std deviation: 2.0

A bisecting k-means algorithm is used in the following analysis. At the first level, the set of tours is split in two clusters. On the basis of the analysis results, one of the two clusters is in turn split in two and the process is iterated once more, leading to a third-level clustering. Figure 2 visually shows the partition of the entire dataset in the different clusters. An alternative and more commonplace approach would be to directly split the set into $n > 2$ clusters. However, research has shown that the bisecting k-means technique can outperform standard k-means (Steinbach et al., 2000) by bringing in some of the advantages of hierarchical clustering approaches, given the fact that these latter are not applicable for such large datasets. Bisecting k-means are also less susceptible to initialization problems, that is, the problem of the choice of the cluster seeds from which distances are computed (Tan et al., 2006, p. 509). Finally, we chose not to apply more than three times the bisecting algorithm to have a manageable number of clusters to analyze.

Cluster analysis completely follows a data-driven and a-theoretical approach, for this it is often criticized for the general lack of transferability of its results compared to statistical models. However, the NHTS dataset is a nationally representative sample of individuals (and of related tours), unlike many studies in the transport sector that apply clustering techniques on more or less limited samples. Therefore, we are confident that the typologies of tours that we describe in the following pages can have a more general validity. As a preliminary form of validity assessment, we randomly split the sample in two parts P1 and P2 respectively containing 75% and 25% of all observations and we repeated the above analysis (Figure 2). Resulting clusters largely replicated those that are described in the following.

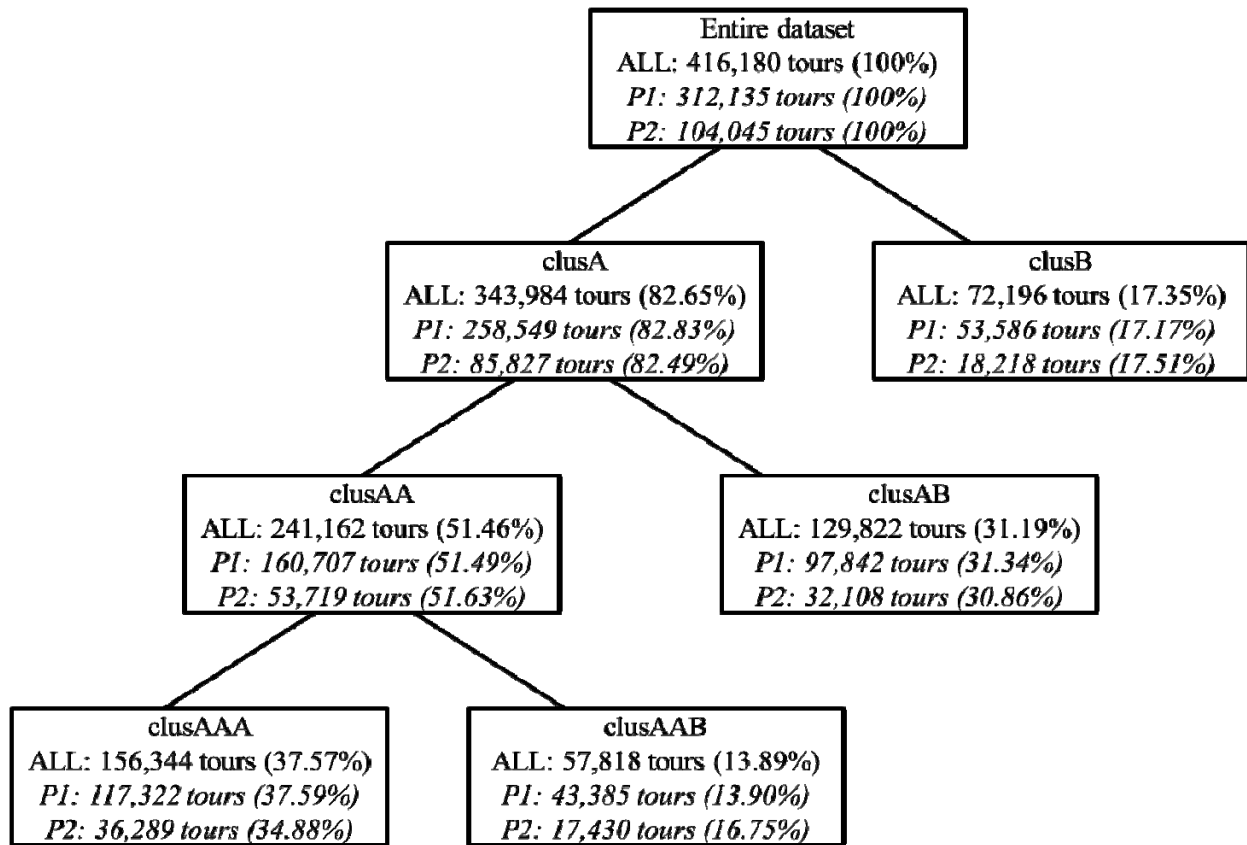


Fig. 2. Hierarchy of the clusters according to bisecting k-means algorithm, for the entire dataset (ALL) and for partitions P1 and P2

First-level solution with two clusters

The easiest possible solution is to simply generate two clusters (clusA and clusB in the following): the resulting dimensions are of 343,984 and 72,196 elements, respectively. A first general analysis could be led looking at the mean values for those features and for each obtained cluster (Table 5, rows two and three). These values were computed not considering observations weights, consistently with the aforementioned process to derive clusters.

Table 5. Unweighted Mean Durations (min) for the Variables Characterizing Each Cluster and, in Brackets, Row % of Total Activity and Travel Times

	Number of tours	ACTIW ORK	ACTIED OT	ACTIT OT	MODEP ERS	MODE PUBT	MODE BIKE	MODEW ALK	MODE OTHE	MODE TOT	TOU RTOT
Entire dataset	416,180	82.0 (43.8%)	105.4 (56.2%)	187.4 (100%)	46.5 (86.4%)	2.8 (5.3%)	0.5 (0.9%)	3.6 (6.6%)	0.4 (0.8%)	53.8 (100%)	241.2
clusA	343,984	97.0 (64.6%)	53.1 (35.4%)	150.1 (100%)	37.4 (87.9%)	0.6 (1.5%)	0.5 (1.2%)	3.7 (8.7%)	0.3 (0.7%)	42.5 (100%)	192.6

clusB	72,196	10.9 (3.0%)	354.4 (97.0%)	365.3 (100%)	90.0 (83.5%)	13.3 (12.4%)	0.2 (0.3%)	3.0 (2.8%)	1.1 (1.0%)	107.6 (100%)	472.9
clusAA	214,162	142.8 (89.7%)	16.4 (10.3%)	159.2 (100%)	30.7 (80.7%)	0.9 (2.5%)	0.7 (1.9%)	5.4 (14.0%)	0.4 (0.9%)	38.1 (100%)	197.3
clusAB	129,822	21.3 (15.8%)	113.8 (84.2%)	135.1 (100%)	48.4 (97.1%)	0.2 (0.3%)	0.2 (0.3%)	0.9 (1.9%)	0.2 (0.4%)	49.9 (100%)	185.0
clusAAA	156,344	16.4 (46.9%)	18.6 (53.1%)	35.0 (100%)	19.1 (68.9%)	0.5 (1.7%)	0.9 (3.3%)	7.0 (25.1%)	0.3 (1.0%)	27.8 (100%)	62.8
clusAAB	57,818	484.7 (97.9%)	10.5 (2.1%)	495.2 (100%)	62.0 (94.0%)	2.2 (3.3%)	0.2 (0.3%)	1.0 (1.5%)	0.6 (0.9%)	66.0 (100%)	561.2

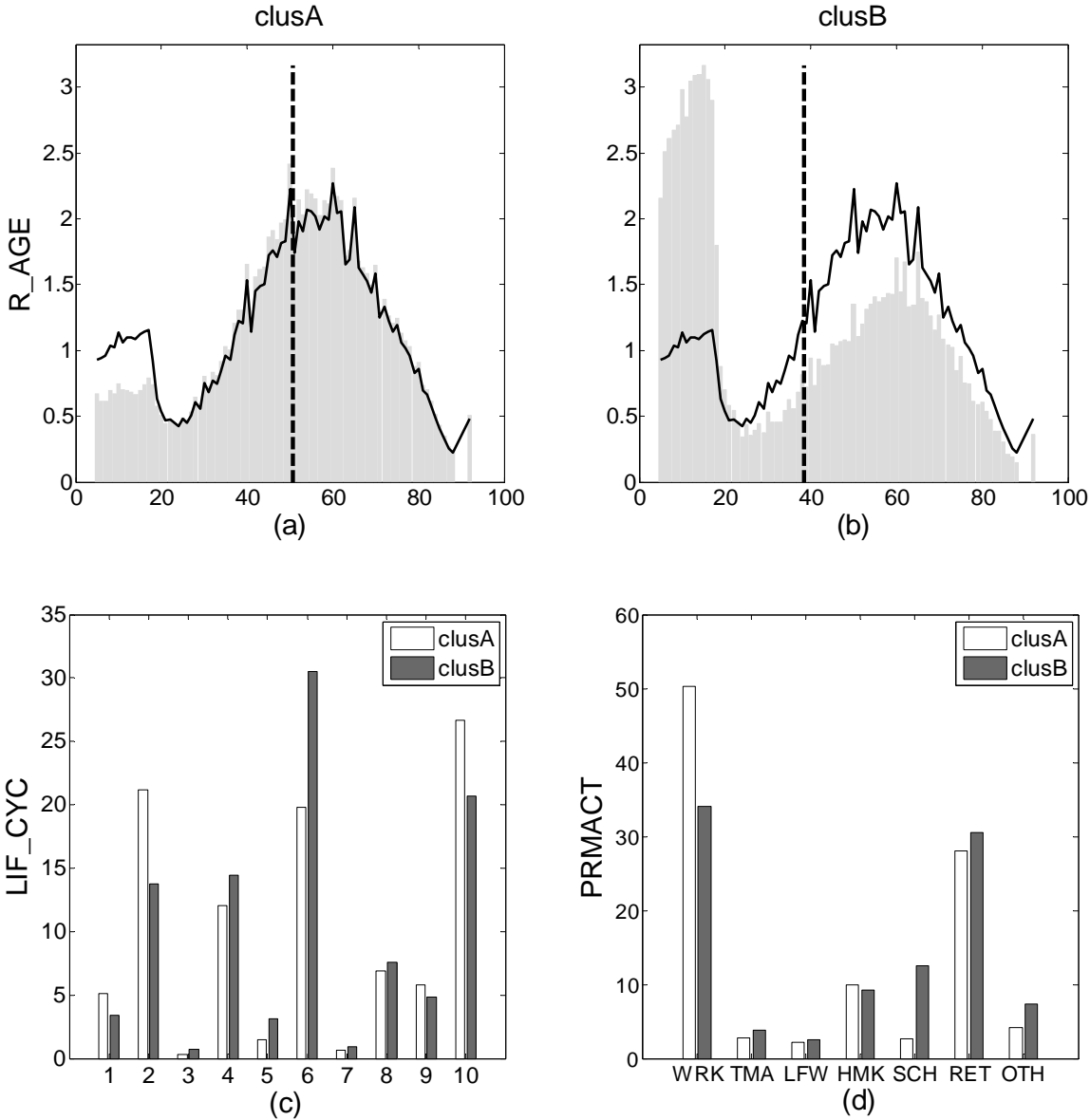
The first observation could be done by analyzing the parameters ACTIWORK and ACTIEDOT: the first one has a greater value in clusA (around 10 times more), while the latter has a bigger value in clusB (about 7 times more). Thus, clusA contains more work-related tours. However, this does not mean that many of the tours in clusA are not containing any work-related activity, as it is clear by comparing the cardinality of this set with that of the HWH and the HOWH tours that is reported in the second column of Table 2. The overall interpretation of the clusters must be done by jointly considering the mean values of all clustering variables.

Observing therefore the percentages of travel time through various transportation modes, it can be seen that the most used are always the motorized individual ones (cars, motorbikes etc.). Beyond such means, the second most frequently used mode is walking for clusA and public transport for clusB. This is also reflected by the overall higher travel time, due to the lower commercial speed of transit. This result is rather intriguing, since transit usage is usually higher in systematic mobility patterns, including work commute. Since “Other” activities include going to school, it is therefore likely that clusB contains many tours of students.

By looking at some of the socioeconomic characteristics of the clusters represented in Figure 3 this assumption can be checked, considering that the white bars show distributions of those characteristics marked on the y-axis for clusA and the gray bars show the distribution of the same characteristics for clusB. Results for R_AGE are instead presented in two distinct charts (Figure 3a for clusA and in Figure 3b for clusB) to improve readability. In this socioeconomic characterization, clusB reveals a pick of people less than 20 years old, while in clusA this group is underrepresented. Concerning PRMACT, 50% of people in clusA are workers, while, in clusB, students are many more albeit not the majority.

The comparison of the variable LIF_CYC shows that in clusA households are mostly composed by two or more adults, frequently retired (code 10 as from Table 4) and without children (code 2), while in clusB the most consistent are the cases of two or more adults with youngest child in 6-15 years old range (code 6). The last observation resoundingly confirms the great number of young people observed in R_AGE plot.

The remaining variables from Table 4 do not present any consistent variation between the two clusters. To sum up, clusB seems to contain tours with longer travel times and relatively intensive use of public transport means, predominantly made by young people and without work activities: this can be seen as a first market segment, albeit roughly identified. It is also noteworthy that the mean of NUMTRIPS is 3.5, a relatively higher value since this is always lower than 3 for all the defined clusters.



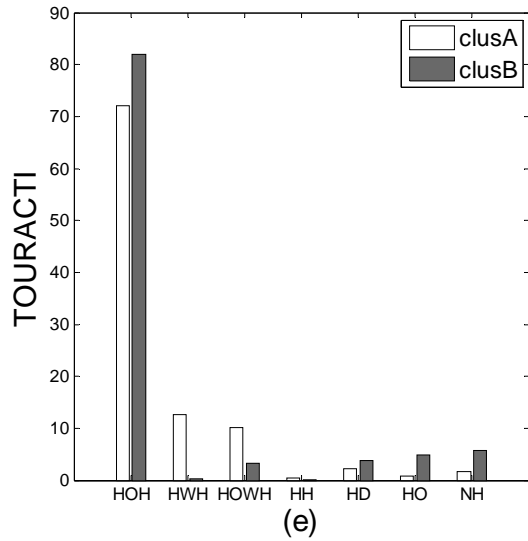


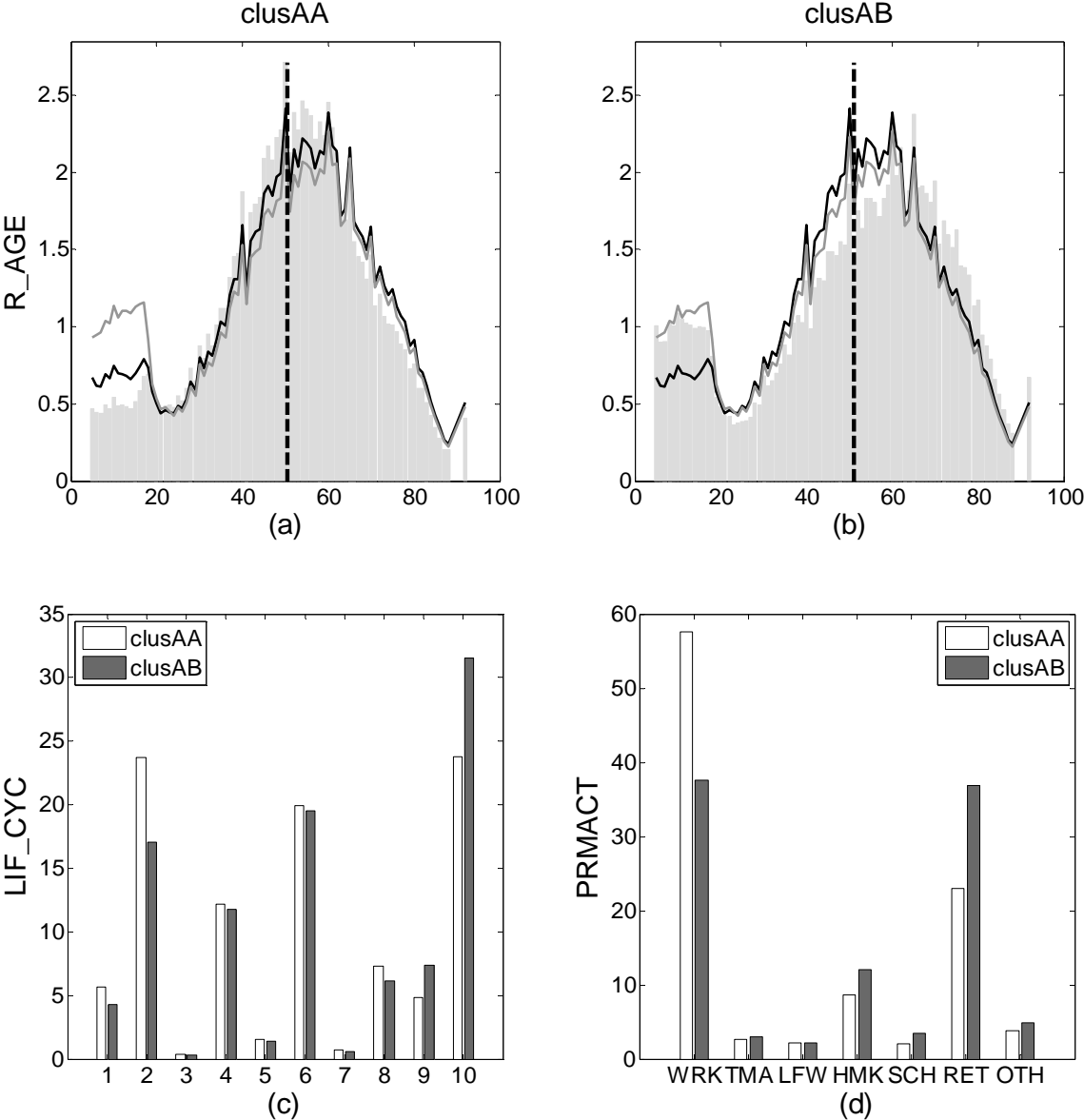
Fig. 3a-e. Histograms with relative frequencies for R_AGE for clusA (a) and clusB (b) and for (c) LIF_CYC, (d) PRMACT, and (e) TOURACTI for tours in clusA (white) and clusB (gray). The black line in R_AGE represents the distribution for the whole dataset and the dotted line the mean value. NB: people aged 89 or more are coded as “92” in the dataset.

Second-level solution with two clusters

ClusA was by far the most numerous cluster and also the less well-defined, concerning both the on-tour activity patterns (mix of work and other) and its socioeconomic characterization. Therefore, the clustering algorithm is again applied on clusA to generate two additional clusters (clusAA and clusAB) which are made up of 214,162 and 129,822 elements respectively. Rows four to five of Table 5 show the corresponding mean values. The division between people whose on-tour main activity is work (clusAA) and people doing mainly other things during the tour (clusAB) is again neat. Regarding modal usages, clusAB almost exclusively captures those using private means, whereas in clusAA 14.0% of travel time is by walk, a quite relevant percentage if related to the overall market shares of the different means. As a consequence, the total time spent travelling is longer for clusAB than for clusAA, since the use of active travel means and transit is in general associated with shorter trip chains.

The most relevant differences in terms of socioeconomic characteristics for these new clusters are displayed in Figure 4. ClusAB shows a lower number of individuals aged between 20 and 60, and in fact retired people are the relative majority, even if workers are well represented. By looking at TRAVDAY distributions, clusAA tours are equally split in workdays and less present in weekends, whereas the opposite happens for clusAB which also contains very few

tours involving working activities and relatively more complex tour structures (see the TOURACTI histogram of Figure 4). Concerning the travelers' gender, clusAB mainly collects women (56.6%) while the other one has a similar partition between the two genders (50.3% of women).



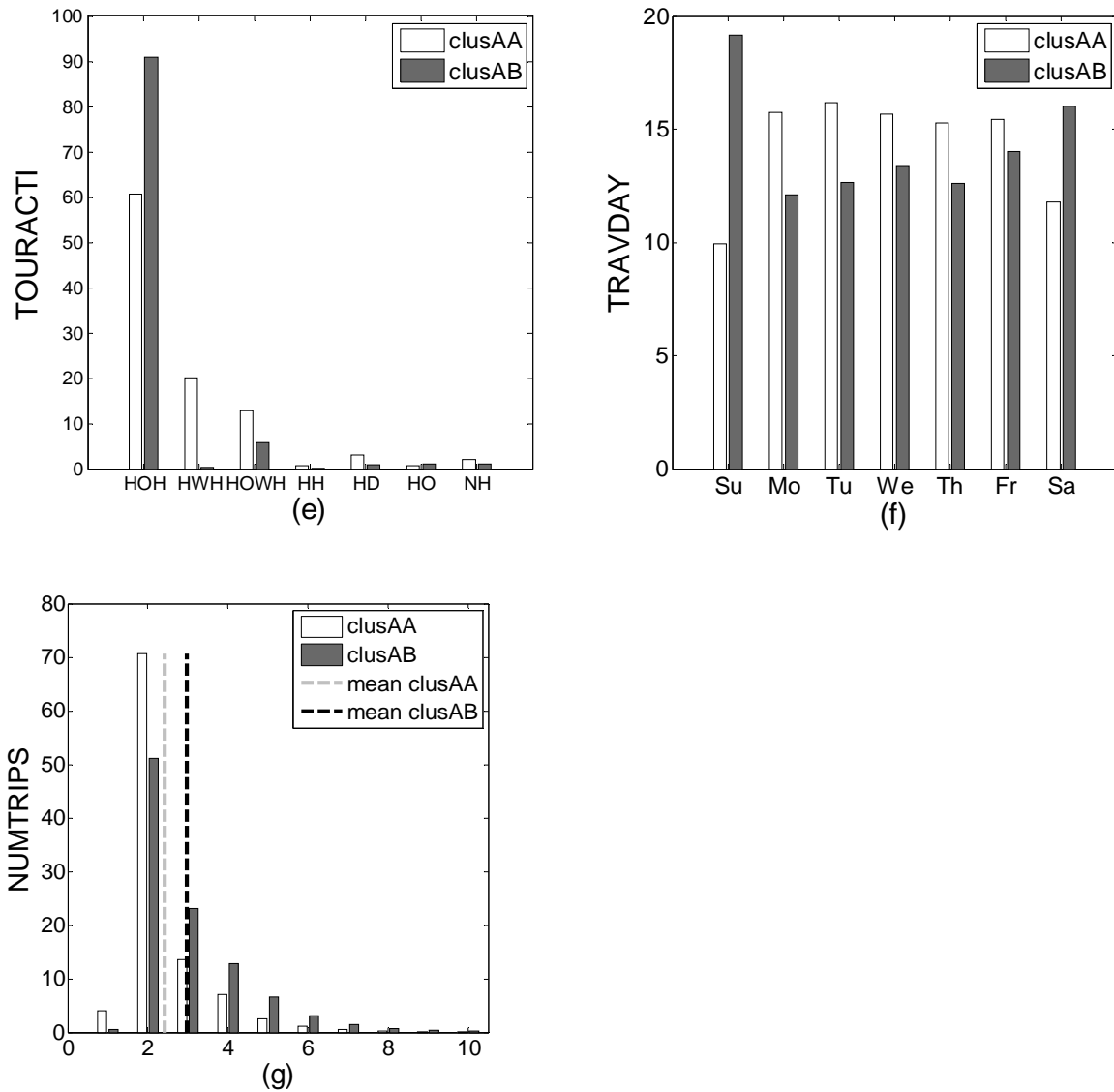


Fig. 4a-g. Histograms with relative frequencies for R_AGE for clusAA (a) and clusAB (b) and for (c) LIF_CYC, (d) PRMACT, (e) TOURACTI, (f) TRAVDAY and (g) NUMTRIPS for tours in clusAA (white) and clusAB (gray). The gray line in R_AGE represents the distribution for the whole dataset, the black one that for clusA and the dotted line the mean value. NB: people aged 89 or more are coded as “92” in the dataset.

To sum up, clusAB represents older individuals and retired persons who almost exclusively travel by car and, in general, those travelling in non-work days and mostly not going to work. This is reflected by the bigger size of the tour in terms of number of trips. ClusAA presents a more mixed profile in terms of life cycle classification of the household, activities performed during the tour and mixed used of different travel means, and it is therefore the candidate for the next iteration of the clustering algorithm.

Third-level solution with two clusters

When operating a k-means clustering, two groups originate from clusAA, namely clusAAA and clusAAB, of 156,344 and 57,818 elements respectively. The last two rows of Table 5 report the related mean values. Since this decomposition derives from clusAA, which could be seen as the group of mainly “workers” (high mean value of ACTIWORK), the mean values for ACTIEDOT for clusAAA and clusAAB are similar and not so high. ClusAAB however has a work time of almost 7 hours, much higher than that of any other cluster, therefore clearly identifying tours that contain an entire work day. For these, there is a relatively appreciable use of transit although the car is always clearly dominating. On the contrary, clusAAA shows the most intensive use of walk since the total travel time is quite short. This group is likely to represent short-distance tours (or even secondary tours in the activity-based models terminology) in which individuals are successfully combining work and other activities. ClusAAA has, in fact, the lowest value of MODETOT among all the clusters (27.8 minutes), while clusAAB reaches the second highest total travel time value after clusB (66.0 minutes). The low duration of work times is probably pointing at more flexible schemes (e.g. self-employed rather than salaried) that allow the traveler to more effectively combine different activities in a tour.

Figure 5 shows the socioeconomic characterization of those two latter clusters. Cluster clusAAA shows lower occurrences in the age interval 20-60 compared to clusAA (black line), while clusAAB concentrates more on that interval that represents the working age, thus confirming the previous interpretation. Such distinction of the two clusters is sharply put into evidence also by the other variables whose distributions are reported in the figure: clusAAB is characterized by a dominance of work-related tours made on weekday by workers without children, which can presumably stay longer at the workplace. Many tours in clusAAA are also likely to be completed by workers, but the working activity inside such tours is much less relevant.

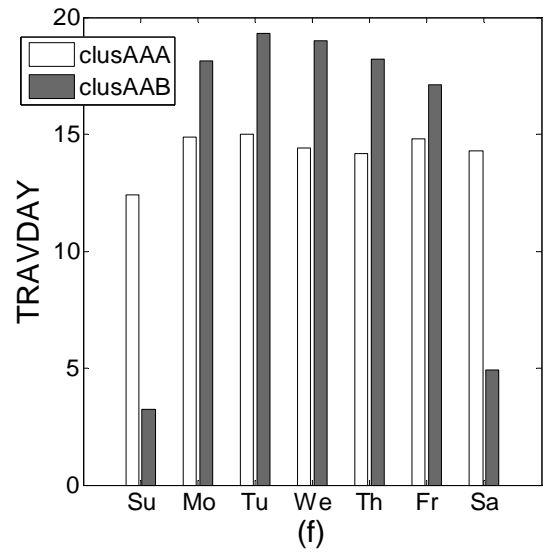
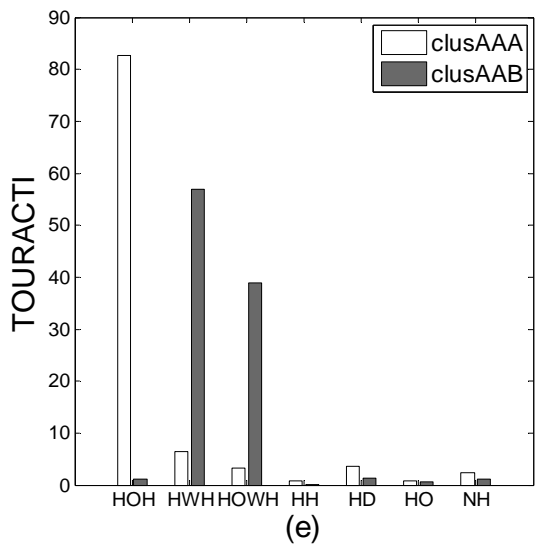
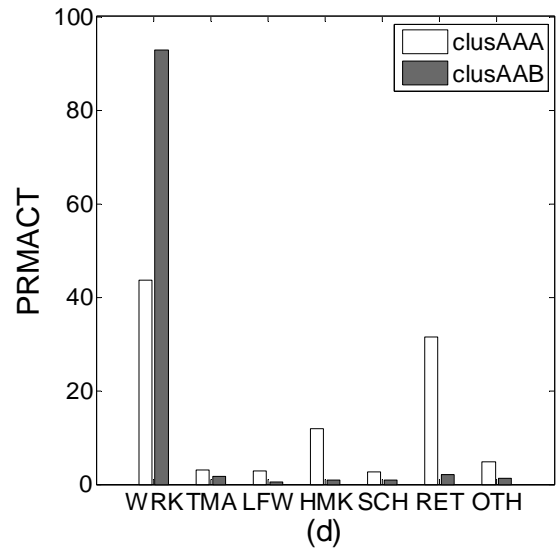
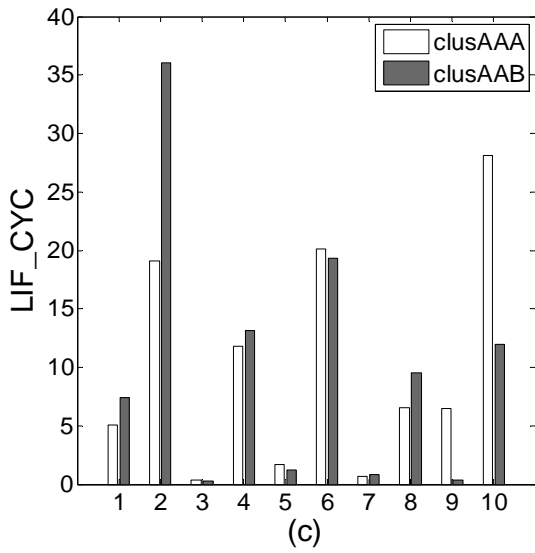
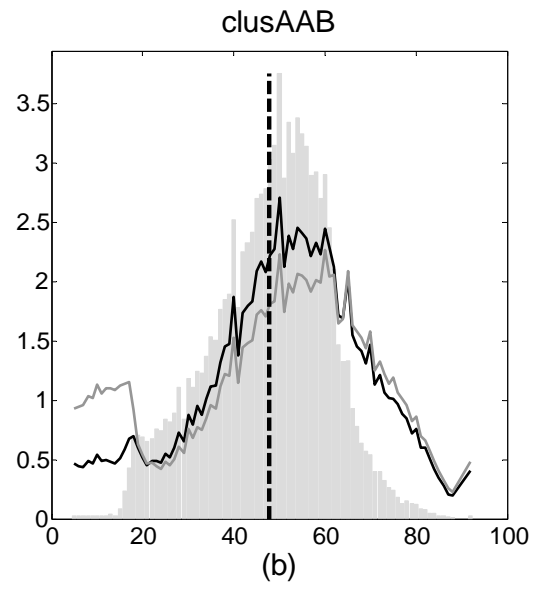
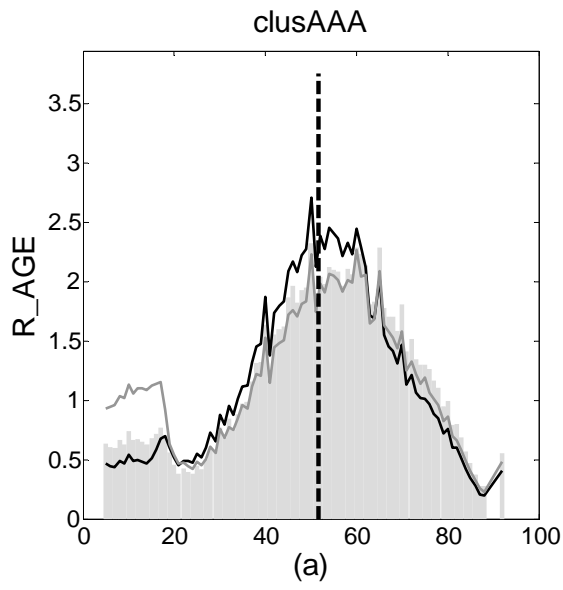


Fig. 5a-f. Histograms with relative frequencies for R_AGE for clusAAA (a) and clusAAB (b) and for (c) LIF_CYC, (d) PRMACT, (e) TOURACTI and (f) TRAVDAY for tours in clusAAA (white) and clusAAB (gray). The gray line in R_AGE represents the distribution for the whole dataset, the black one that for clusAA and the dotted line the mean value. NB: people aged 89 or more are coded as “92” in the dataset.

Discussion and conclusions

This paper presented some analyses aimed at uncovering the structure of tours within the sample of the U.S. National Household Travel Survey administered in 2009. By applying a multi-level clustering analysis, some groups are defined that are hopefully lowering the inherent complexity of tours within each group in terms of mix of mode uses and of activity types. Developing tour-based travel demand models within each cluster could hopefully help improving their fit and predictive power. Tour classifications based on our cluster analysis could also be useful to define the nested choice structure of the person-day activity scheduler within activity-based models. Additionally, the socioeconomic profile of the traveler inside each group has been checked, leading the way to the definition of customized marketing campaigns to more effectively promote “better” mobility behaviors, targeting specific groups of individuals and taking into consideration complete trip chains rather than individual trips.

The clusters that were identified are briefly recalled below. Jointly considering activity types, socio-demographic profile and travel means allowed identifying as well some interesting policy and transport modeling implications:

- ClusB: Tours predominantly made by young individuals, possibly going to school and in most cases not on their work trip. For these tours the use of transit is relatively higher. Given the traveler’s profile and the on-tour activity type, a strong influence of travel cost on mobility behaviors is to be expected. Concerning the performance of the means, the influence of factors such as on-board entertainment, smart information services and careful branding would probably be non-negligible compared to wait and travel times. Trip liking and travel utilities are in fact increasingly being affected by the range of activities that can be performed while travelling, enabled by new technology implementations (Lyons et al., 2013; Rasouli and Timmermans, 2014; Mokhtarian et al., 2015). However, embedding such factors in models is less straightforward compared to other instrumental variables such as travel times, costs or socioeconomic characteristics. On a policy viewpoint, measures focusing on prices on one side and on innovating services on the other, rather than aimed at improving the conventional level of service of a transport system in terms of

waiting times and commercial speed, could effectively promote the behavioral change towards more environmentally sustainable travel choices at a smaller cost.

- **ClusAB:** Tours mainly done by elders not going to work or retired people, often completed during weekends and predominantly by car. Travel modes performances such as comfort and security would probably be the key element to influence such segment. The tour duration and the average length of individual trips are shorter, so targeting this group with travel demand management actions would probably have a smaller aggregate impact compared to other segments, even if these travelers are almost exclusively using private means. Since individuals performing such tours have probably a more flexible schedule, behavioral change can be achieved only acting on motivations rather than through external constraints, for example considering travel awareness raising actions or personalized travel planners.
- **ClusAAA:** short tours (probably secondary tours within the daily schedule) involving short activities. The use of active modes (walk and bike) is the highest of all clusters, even if cars still cover the majority of travel time. Mobility patterns of this segment are often overlooked in more naïve transport planning approaches, either deliberately (wrongfully thinking that active travel modes should not be considered since they do not contribute to congestion nor have environmental impacts) or due to the well-known problem of active modes trip underreporting in travel surveys (Wolf et al., 2003; Jin et al., 2013). Appropriate action should be taken to correct this, as discussed in Bayart and Bonnel (2012). On a policy viewpoint, multimodal practices could be further consolidated through some “customer loyalty marketing” strategies, since these travelers are already familiar with environmentally benign travel means.
- **ClusAAB:** tours involving a long-lasting working activity, so that the related trips are mainly for commuting purposes. These are the tours that are perhaps best addressed by classical travel demand models based on econometric theories, with a relatively non-negligible use of public transport and where probably a combination of travel times and costs can best explain behavioral patterns. Policy actions aimed at promoting more sustainable travel choices related to such tours should probably act on the costs, performances and levels of service of the competing means.

Given the nationally representative dataset that was used and on the basis of split-sample replications results, we believe that such clusters have sufficiently general validity. On the

other hand, repeating the analysis in more local or foreign context where travel and activity patterns are widely different from those throughout the United States would probably lead to different findings. Extending validation analyses would add insights on the transferability of these results.

Future research efforts will be aimed at better characterizing some tour structures that are relatively less frequent and that therefore are not easily identifiable with such a general analysis. It could be particularly interesting to focus on multimodal tours, that are “hidden behind” the massive use of cars, and on tour made by particular groups of individuals such as mobility impaired or transit captives.

References

- Anable, J. (2005) “Complacent car addicts” or “aspiring environmentalists”? Identifying travel behaviour segments using attitude theory. *Transport Policy* 12(1), 65-78.
- Arentze, T., Timmermans, H., Hofman, F., Kalfs, N. (2000) Data needs, data collection, and data quality requirements of activity-based transport demand models. *Transportation Research E-Circular*, No. E-C008, II-J/1-II-J/30.
- Axhausen, K.W. (2007) Definition of movement and activity for transport modelling. In: Hensher, D. A., Button, K. J., Handbook of Transport Modelling (Handbooks in Transport), 2nd Edition. Emerald Publishing Group.
- Bayart, C., Bonnel, P. (2012) Combining web and face-to-face in travel surveys: comparability challenges? *Transportation* 39(6), 1147-1171.
- Biagioni, J. P., Szczurek, P. M., Nelson, P. C., Mohammadian, A. (2009) Tour-based mode choice modeling: using an ensemble of conditional and unconditional data mining classifiers. In *Proceedings of Transportation Research Board 88th Annual Meeting*, Washington, D.C.
- Bowman, J. L., Ben-Akiva, M. E. (2001) Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice* 35(1), 1-28.
- Buliung, R. N., Kanaroglou, P. S. (2007) Activity–travel behaviour research: conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. *Transport Reviews* 27(2), 151-187.
- Currie, G., Delbosch, A. (2011) Exploring the trip chaining behaviour of public transport users in Melbourne. *Transport Policy* 18(1), 204-210.

- Cyganski, R., Justen, A., Schulz, A., Köhler, K. (2013) Generation of coherent trip chains for travel behavior analysis and modeling. In *Proceedings of the European Transport Conference*, Frankfurt, Germany.
- Diana, M., Mokhtarian, P. L. (2009) Grouping travelers on the basis of their different car and transit levels of use. *Transportation* 36(4), 455-467.
- Diana, M., Pronello, C. (2010) Traveler segmentation strategy with nominal variables through correspondence analysis. *Transport Policy* 17(3), 183-190.
- Doherty, S. T. (2006) Should we abandon activity type analysis? Redefining activities by their salient attributes. *Transportation* 33(6), 517-536.
- Doherty, S. T., Mohammadian, A. (2011) The validity of using activity type to structure tour-based scheduling models. *Transportation* 38(1), 45-63.
- Elgar, A., Bekhor, S. (2004) Car-rider segmentation according to riding status and investment in car mobility. *Transportation Research Record* 1894, 109-118.
- Fowkes, A.S. (2000) Recent developments in stated preference techniques in transport research. In: Ortuzar, J. de D. (Eds.), *Stated Preference Modelling Techniques*, PRTC, pp. 37-52.
- Han, J., Kamber, M., Pei, J. (2011) *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers.
- Hensher, D. A., Reyes, A. J. (2000) Trip chaining as a barrier to the propensity to use public transport. *Transportation* 27(4), 341-361.
- Ho, C. Q., Mulley, C. (2013) Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice. *Transportation Research Part A: Policy and Practice* 49, 206-219.
- Jensen, M. (1999) Passion and heart in transport – A sociological analysis on transport behavior. *Transport Policy* 6(1), 19-33.
- Jin, X., Wu, J., Asgari, H., Argote, J. A. (2013) Examining trip misreporting behavior using gps-assisted household travel surveys. In *Proceedings of Transportation Research Board 92th Annual Meeting*, Washington, D.C..
- Joh, C. H., Arentze, T., Hofman, F., Timmermans, H. (2002) Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological* 36(5), 385-403.
- Krizek, K. J. (2003) Neighborhood services, trip purpose, and tour-based travel. *Transportation* 30(4), 387-410.

- Louviere, J.J., Hensher, D.A., Swait, J.D. (2000) Cross validity and external validity of SP models. In: *Stated Choice Methods — Analysis and Application*. Cambridge University Press, Cambridge, pp. 354–381.
- Lyons, G., Jain, J., Susilo, Y., Atkins, S. (2013) Comparing rail passengers’ travel time use in Great Britain between 2004 and 2010. *Mobilities* 8(4), 560-579.
- McGuckin, N., Zmud, J., Nakamoto, Y. (2005) Trip-chaining trends in the United States: understanding travel behavior for policy making. *Transportation Research Record: Journal of the Transportation Research Board* 1917, 199-204.
- Mokhtarian, P. L., Papon, F., Goulard, M., Diana, M. (2015) What makes travel pleasant and/or tiring? An investigation based on the French National Travel Survey. *Transportation* 42(6), 1103-1128.
- NHTS (2009) U.S. Department of Transportation, Federal Highway Administration. 2009 *National Household Travel Survey*. URL: <http://nhts.ornl.gov>.
- Nishii, K., Kondo, K., Kitamura, R. (1988) Empirical analysis of trip chaining behavior. *Transportation Research Record: Journal of the Transportation Research Board* 1203, 48-59.
- Nowrouzian, R., Srinivasan, S. (2012) Empirical analysis of spatial transferability of tour-generation models. *Transportation Research Record: Journal of the Transportation Research Board* 2302, 14-22.
- O’Fallon, C., Sullivan, C. (2005) Trip chains and tours: definitional issues associated with household travel surveys. In *Proceedings of 28th Australasian Transport Research Forum*, Sydney, Australia.
- Primerano, F., Taylor, M. A., Pitaksringkarn, L., Tisato, P. (2008) Defining and understanding trip chaining behaviour. *Transportation* 35(1), 55-72.
- Rasouli, S., Timmermans, H. (2014) Judgments of travel experiences, activity envelopes, trip features and multi-tasking: A panel effects regression model specification. *Transportation Research Part A: Policy and Practice* 63, 67-75.
- Shiftan, Y., Ben-Akiva, M., Prousaloglou, K., de Jong, G., Popuri, Y., Kasturirangan, K., Bekhor, S. (2003) Activity-based modeling as a tool for better understanding travel behaviour. In *Proceedings of 10th Conference of the International Association of Travel Behavior Research*, Lucerne, Switzerland.
- Steinbach, M., Karypis, G., Kumar, V. (2000) *A comparison of document clustering techniques*. University of Minnesota Technical Report #00-034, 20pp.

- Stopher, P., Jiang, Q., Zhang, Y. (2010) Tour-based analysis of multi-day GPS data. In *Proceedings of 12th World Conference for Transport Research*, Lisbon, Portugal.
- Strathman, J.G., Dueker, K.J. (1995) Understanding Trip Chaining. Special Reports on Trip and Vehicle Attributes, 1990 NPTS Report Series. Washington D.C., U.S. Department of Transportation.
- Tan, P.-T., Steinbach, M., Kumar, V. (2006) Introduction to Data Mining. Pearson Addison Wesley, Boston.
- Timmermans, H., Van der Waerden, P., Alves, M., Polak, J., Ellis, S., Harvey, A. S., Kurose, S., Zandee, R. (2003) Spatial context and the complexity of daily travel patterns: an international comparison. *Journal of Transport Geography* 11(1), 37-46.
- Valiquette, F., Morency, C. (2010) Trip chaining and its impact on travel behaviour. In *Proceeding of 12th World Conference on Transport Research*, Lisbon, Portugal.
- Vande Walle, S., Steenberghen, T. (2006) Space and time related determinants of public transport use in trip chains. *Transportation Research Part A: Policy and Practice* 40(2), 151-162.
- Wolf, J., Lechl, M., Thompson, M., Arce, C. (2003) Trip rate analysis in GPS-enhanced personal travel surveys. In: Stopher, P. R., Jones, P. M. (eds.) *Transport survey quality and innovation*, Elsevier, London, 483-498.
- Yagi, S., Mohammadian, A. K. (2008) Modeling daily activity-travel tour patterns incorporating activity scheduling decision rules. *Transportation Research Record: Journal of the Transportation Research Board* 2076(1), 123-131.
- Ye, X., Pendyala, R. M., Gottardi, G. (2007) An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B: Methodological* 41(1), 96-113.