STUDYING TRAVEL-RELATED INDIVIDUAL ASSESSMENTS AND DESIRES BY COMBINING HIERARCHICALLY STRUCTURED ORDINAL VARIABLES

Marco Diana, Tingting Song, Knut M. Wittkowski

7th January 2009

This is the post-print (i.e. final draft post-refereeing) version of an article published in the journal *Transportation*. Beyond the journal formatting, please note that there could be minor changes from this document to the final published version. The final published version of this article is accessible from here: http://dx.doi.org/10.1007/s11116-009-9186-z

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=0049-4488

This document may be cited directly referring to the above mentioned final published version:

Diana, M., Song, T. and Wittkowski, K.M. (2009) Studying travel-related individual assessments and desires by combining hierarchically structured ordinal variables, *Transportation*, vol. 36(2), pp. 187-206.

STUDYING TRAVEL-RELATED INDIVIDUAL ASSESSMENTS AND DESIRES BY COMBINING HIERARCHICALLY STRUCTURED ORDINAL VARIABLES

Marco Diana*
Politecnico di Torino, Dipartimento di Idraulica, Trasporti e Infrastrutture Civili
Corso Duca degli Abruzzi, 24, 10129 Torino – ITALY
Phone: +39 011 564 5638 - Fax: +39 011 564 5699 - marco.diana@polito.it

Tingting Song
The Rockefeller University, Center for Clinical and Translational Science
1230 York Avenue, Box 322, New York, NY 10021 – U.S.A.
Phone: +1 212 327 8021 - Fax: +1 212 327 8450 - tsong01@mail.rockefeller.edu

Knut M. Wittkowski
The Rockefeller University, Center for Clinical and Translational Science
1230 York Avenue, Box 322, New York, NY 10021 – U.S.A.
Phone: +1 212 327 7175 - Fax: +1 212 327 8450 - kmw@rockefeller.edu

** Corresponding author*

ABSTRACT

Ordinal measures are frequently encountered in travel behavior research. This paper presents a new method for combining them when a hierarchical structure of the data can be presumed. This method is applied to study the subjective assessment of the amount of travel by different transportation modes among a group of French clerical workers, along with the desire to increase or decrease the use of such modes. Some advantages of this approach over traditional data reduction technique such as factor analysis when applied to ordinal data are then illustrated. In this study, combining evidence from several variables sheds light on the observed moderately negative relationship between the personal assessment of the amount of travel and the desire to increase or decrease it, thus integrating previous partial (univariate) results. We find a latent demand for travel, thus contributing to clarify the behavioral mechanisms behind the induced traffic phenomenon. Categorizing the above relationship by transportation mode shows a desire for a less environmental-friendly mix of modes (i.e. a greater desire to use heavy motorized modes and a lower desire to use two-wheeled modes), whenever the respondents do not feel to travel extensively. This result, combined with previous theoretical investigations concerning the determinants of the desire to alter trips consumption levels, shows the importance of making people aware of how much they travel.

KEYWORDS

Subjective mobility, Travel desires, Ordinal variables, Factor analysis, Multidimensional scaling

1 INTRODUCTION

Transportation engineering and planning processes traditionally make use of large amounts of information gathered through either mobility surveys (trip length and duration, cost etc.) or field measurements campaigns (traffic counts, odometer readings, air pollutants monitoring, etc.). These data essentially quantify one or more physical features of the transportation system, and are then used to feed a wide range of models postulated to support the decision making processes of public authorities. However, it now is widely acknowledged that mobility has deep roots also in personal attitudes, preferences and intentions, which are, thus, strong determinants of the observable behaviors and choices. Affective factors play in fact crucial role in the personal preferences for the alternatives (Steg *et al.*, 2001; Anable and Gatersleben, 2005) and in subsequent transportation choices. A deeper study of the evaluation processes at the individual level clearly is one of the most promising ways to improve our understanding of the mobility phenomenon and, thus, to set up planning tools that can more effectively pursue a stated objective.

Intensive research has been performed to identify some self-related factors that seem to be most predictive of travel behavior, such as individual differentiations or market segments (TRB, 1977; Dobson and Tischer, 1978; Tardiff, 1979; Gensch and Torres, 1980; Salomon and Mokhtarian, 1998; Outwater *et al.*, 2003; Anable, 2005) and habits (see for example Verplanken *et al.*, 1994; Verplanken *et al.*, 1997; Aarts and Dijksterhuis, 2000; Gärling *et al.*, 2001; Garvill *et. al.*, 2003; Bamberg *et al.*, 2003; Thøgersen, 2006). However there are other factors equally worth considering. In particular, individual representation and subjective assessment of personal mobility are important in determining individual choices and ultimately the characteristics of the demand for transportation. On the other hand, investigating the personal desire to change the amount of travel of a given kind can be of great help for example in projecting transportation trends. Mokhtarian and Salomon (2001) study these two concepts by introducing the factors Subjective Mobility (SM) and Relative Desired Mobility (RDM), whose main determinants are investigated in more details by Collantes and Mokhtarian (2007) and Choo *et al.* (2005).

Integrating SM and RDM is the main objective of the present paper. Understanding the relationship between SM and RDM can be in fact very important to better know the self-related mechanisms that underlie the travel behaviors and choices we observe. For example, it would be interesting to understand if being “fed up” with a particular transportation mode can induce a desire to decrease the total amount of transportation by decreasing the use of that particular

mode, or to maintain overall transportation levels by shifting to alternative modes. That the SM and RDM concepts are related is something intuitive, but analytically determining the functional relationship between these factors is not trivial, since they are highly sensitive to the particular mobility segment (i.e. the group of travelers) under consideration, but also to the way the mobility concept is specified, for example in terms of transportation modes. Some preliminary results regarding the SM-RDM relationship are already available in the published literature concerning these two main sources of variability. Diana and Mokhtarian (2007, 2009) study the effect of individual differences and define different user segments based on their objective, subjective, and relative desired travel amounts through different transport means. On the other hand, Choo *et al.* (2005) found empirical evidence of a negative relationship between RDM and SM within many categories of trips. It is however difficult to draw more general conclusions from these findings.

The dependency of the SM-RDM relationship on the experimental context is perhaps the main obstacle to generalizing the available partial results, i.e., to draw a more systematic picture. SM and RDM are in fact customarily measured through a survey asking something like “How much do you think you travel?” and “Would you like to travel more, the same or less?” to assess SM and RDM respectively (see the pioneering study of Mokhtarian and Salomon, 2001). However those questions are not to be directly posed in such general terms, since they could be rather baffling for the respondent. To have reliable responses it is instead necessary to formulate questions with reference to a more specific context, for example by categorizing them by transportation mode or purpose, hence asking the same questions for each category of trip we want to study. This fragmentation of the observations explains the tendency of having partially valid and context-specific results concerning the SM-RDM relationship. While investigating such relationship in specific contexts is something useful to formulate a theory that can explain the empirical observations, being able to aggregate results can be quite important, both from a practical and from a policy viewpoint. For instance, it could be important to assess the global desire to use transit and non motorized means, so that RDM across public transportation categories, bikes and feet needs to be accounted for. It could also be valuable to understand if making people more aware of the amount of their travel is *overall* limiting their desire to travel more, irrespective of the transport mean or trip category.

The precondition to study the SM-RDM relationships in such general terms is then to find a method that effectively combines several related, but distinct responses. This problem is exacerbated by the difficulty to express SM and RDM through metric (interval-scaled) variables,

since they are based on ordinal rating scales. For addressing the SM-RDM relationship in a specific context, where an individual observable variable suffices, non-parametric methods, such as Spearman rank-correlation and the Mann-Whitney u-test can be used, but, until recently, non-parametric methods to deal with multivariate data were lacking. Wittkowski *et al.* (2004) developed on a statistical technique for integrating several ordinal variables and applied it to the diagnosis and the treatment of psoriasis. This method is useful whenever some ordinal information has to be treated, and it is increasingly being used in different fields. Other published applications range from cancer treatment (Paczesny *et al.*, 2004) and animal studies on addiction (Spangler *et al.* 2005) to machine learning (Sapir *et al.*, 2005). In the field of economics, this technique was recently applied by Cherchye and Vermeulen (2006, 2007) to score life-time achievements of Tour-de-France cyclists. We argue that such method could be a valuable addition to the repertoire of analysis tools in the transport research field as well, where ordinal variables are routinely encountered in many travel surveys, e.g., when income brackets or educational levels are asked for or when subjective measures of involvement, desire, appreciation, etc. are to be dealt with. Even if measures are technically on an interval scale, such as frequency of using a given mode or going to a place, they often do not have a linear relationship with the underlying latent factor of interest, so that methods for ordinal data are often more appropriate.

In the present paper we also propose an extension of the method for situations, where some variables are more interrelated than others. In our case, for example, buses and trams are “public transportation” modes. To take advantage of such study characteristics, we will extend the above works to situations where a hierarchical structure among the variables can be presumed, thereby improving the dispersion (information content) of the resulting measure. We will then use the information content of the scores resulting from various hierarchical models as a criterion to select a particular model. This new technique is then applied for the first time in the published literature to study the SM-RDM relationship through the combination of several SM and RDM observations within different trip categories (by mode). Finally, a brief discussion compares the outcome of our analysis with that of a traditional data reduction technique such as factor analysis.

2 CASE STUDY AND SM-RDM CORRELATIONS WITHIN CATEGORIES

A Web-based survey among staff working at INRETS (the French National Institute for Transport and Safety Research) was implemented in 2004 to investigate personal attitudes and feelings regarding mobility (see Diana, 2005 for details). A total of 164 subjects were questioned on their SM and RDM levels regarding 14 different transportation modes. A 10 point ordinal scale was used for SM, ranging from “I feel I do not travel at all” to “I feel I travel a lot” (by that particular mode). The RDM scale had 11 points, ranging from “I would like to travel much less” to “I would like to travel much more”, passing through the neutral point “I would like to travel the same amount as now” (with that particular mode).

Table 1 shows the Spearman rank-order correlations between SM and RDM by transportation mode in ascending order, with modes for which SM and RDM are almost unrelated in the center. Correlations vary widely; in particular, negative correlations are observed for many heavily used motorized modes and positive correlations are more typical of less space-consuming and more “leisurely” modes, which are also likely to be much less frequently used, at least for “compulsory” trips.

The SM-RDM correlations within categories may be useful in itself to address individual aspects. Several comments could be made on these results, that are however beyond the scope of the paper; their interpretation is thus left to the interested reader. In the following we propose a novel method to combine these ordinal measures to study a unique SM and RDM index and draw some general conclusions. The remainder of this paper is devoted to the presentation of the method and its application to this problem.

Table 1. Spearman correlations *r* and corresponding *p*-values between SM and RDM measures

	Car driver	Metro	Bus	Car pax.	Plane	Tram	Feet	Interc. train	Suburb train	Taxi	Boat	Moped motorc	Cycle	Roller
r	**-0.310**	**-0.295**	**-0.215**	**-0.162**	-0.154	-0.073	-0.054	-0.004	0.024	0.139	0.140	**0.182**	**0.197**	**0.378**
p	<0.001	<0.001	0.006	0.042	0.053	0.366	0.499	0.957	0.768	0.086	0.082	0.027	0.013	<0.001

Note: Correlations significant at the 5% level are reported in bold.

3 A METHOD TO COMBINE HIERARCHICAL ORDINAL DATA

The procedure that will be used to compute a global SM and a RDM index for each respondent is a generalization of the methodology described by Wittkowski *et al.* (2004), which we will briefly recall before showing how it can be improved by allowing for hierarchically structured data.

Let k index m subjects, each characterized by L variables. A partial order among the subjects $\boldsymbol{x_k} = (x_{k1}, \ldots, x_{kL})'$ can easily be defined. Subject k' is "superior" to (or better than) subject k if k' is higher in at least one variable and lower in none:

$$x_k < x_{k'} \Leftrightarrow \{\forall_{\ell=1,\ldots,L} x_{k\ell} \le x_{k'\ell} \wedge \exists_{\ell=1,\ldots,L} x_{k\ell} < x_{k'\ell}\}. \quad (1)$$

With untied univariate observations, i.e when all the observations are different and $L = 1$, the order (1) is complete, since all the $m(m - 1) / 2$ pairwise orderings can be decided. With several variables, a pairwise ordering can be ambiguous if subject k is higher than subject k' in some variable(s), but lower in others. Still, all subjects can be scored. Wittkowski *et al.* (2004) define the μ- (multivariate u-) score of a subject $u(x_k)$ as the number of subjects being worse minus the number of subjects being better (with respect to the partial ordering (1)). We refer the interested reader to their paper for the study of the statistical properties of this index.

A simple example computing the μ-score for 4 subjects A, B, C and D, characterized by 4 ordinal variables $x_1, \ldots, x_4$, is shown in Figure 1. The top table shows the initial ordinal measures. Pairwise orderings are shown in the four central matrices, each matrix comparing the four subjects among each other with respect to one variable. When the row i subject has a higher, lower, or equal value for a particular variable than the column j subject, the (i, j) element is "+1", "-1", or "0", respectively. If either subject is "NA" (missing), the (i, j) element reports a "?". Hence those four matrices are antisymmetric by construction. The bottom table shows the aggregated partial ordering and the computation of the resulting μ-scores. Cell (i, j) is set to

- "+1" if subject i is "superior" to subject j according to (1) (i.e., no (i, j) cell among the four central matrices contains a "-1" and at least one contains a "+1"),
- "-1" if it is inferior,
- "0" if the two vectors of variables are identical and
- "?" in case of ambiguity (that is, when at least one cell in the group reports a "-1" and also at least one reports a "+1").

The μ-scores of the four subjects are then simply the algebraic sum of the elements of each row in the bottom table.

The above example illustrates a limitation of this method. When the variables are less than perfectly correlated, which is often the case, e.g., with SM or RDM measures pertaining to different transportation means, the risk for pairwise orderings to become ambiguous increases with the number of variables included. The number of pairwise orderings that can be decided for a given subject defines this score's information content. In the extreme, all pairwise orderings for a given subject could become ambiguous, rendering the resulting score for this subject non-informative, i.e, be in effect "missing".

The situation can be improved when clusters of variables can be formed on the basis of empirical considerations or prior knowledge. Considering the set of 14 transportation modes upon which SM and RDM measures were taken, it is possible to assume that some perceptions and desires related to some modes are more related than others, as we discuss in the next section when we apply the methodology to our case study. Then, it is possible to extend the

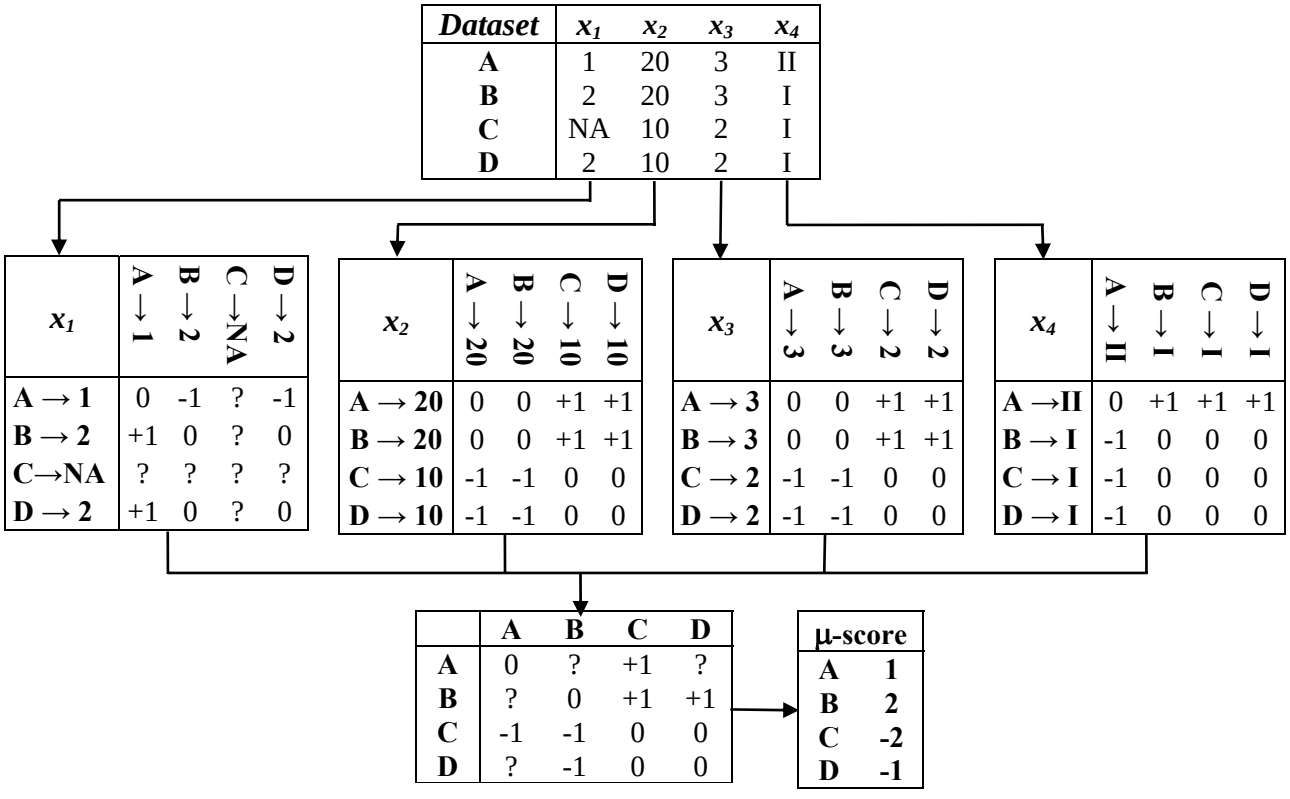

Figure 1. Example of μ-score computation according to Wittkowski et al. (2004): letters A to D represent the observations and x_1 to x_4 are the ordinal variables

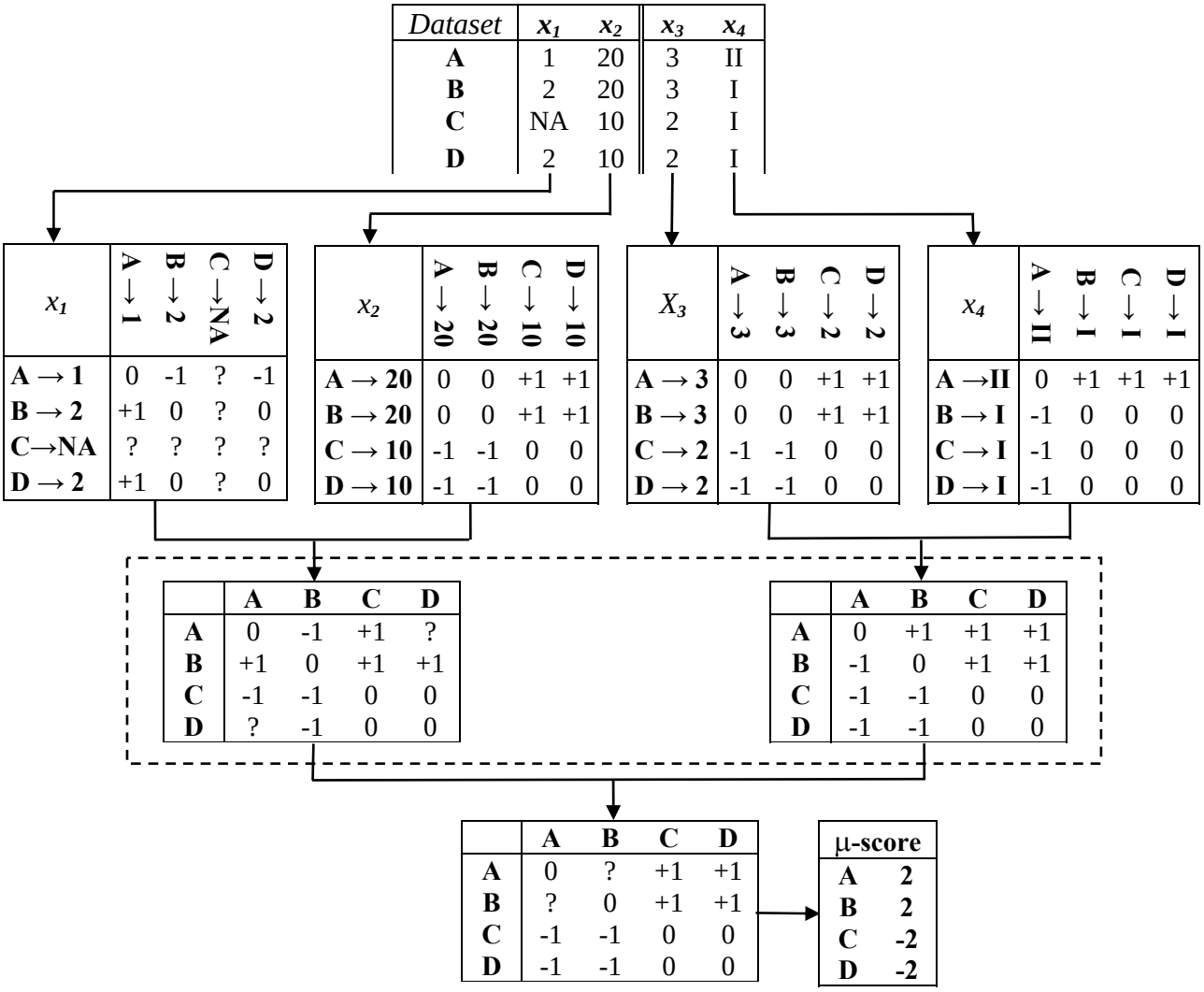

Figure 2. Example of μ-score computation in presence of a hierarchical data structure: the additional step in computing μ-scores is shown in the dotted rectangle

method of μ-scores to utilize information about a hierarchical structure among the variables to avoid ambiguities and, thus, to increase information content of the scores by separately applying the second step of the above described method within each cluster, and then to apply it again to combine different clusters, iteratively if necessary.

Recalling the previous example of Figure 1, we now assume that the variables x_1 and x_2 are related, as are x_3 and x_4. We separately apply the method for those two groups, and then repeat the analysis to combine the two groups themselves (see Figure 2). As a result, the number of ambiguities in the final table has been halved compared to Figure 1 (from 4 to 2), so that the new scores are more dispersed and, thus, more informative. The gain can be even more

substantial when analyzing datasets of realistic size and more complex patterns among the variables.

With programming languages like S or R, which are particularly suited for statistical algorithms, this method is easily implemented. The open-source package *muStat* of R (http://cran.r-project.org) has been used in the application described in section 5.

4 HIERARCHICAL CLASSIFICATIONS OF THE TRANSPORT MODES

Before we can apply the hierarchical method to our problem, we need to define the classification of the ordinal measures. We preliminarily performed two multidimensional scaling analyses from the correlation matrices of the SM and the RDM measures, respectively. The resulting two-

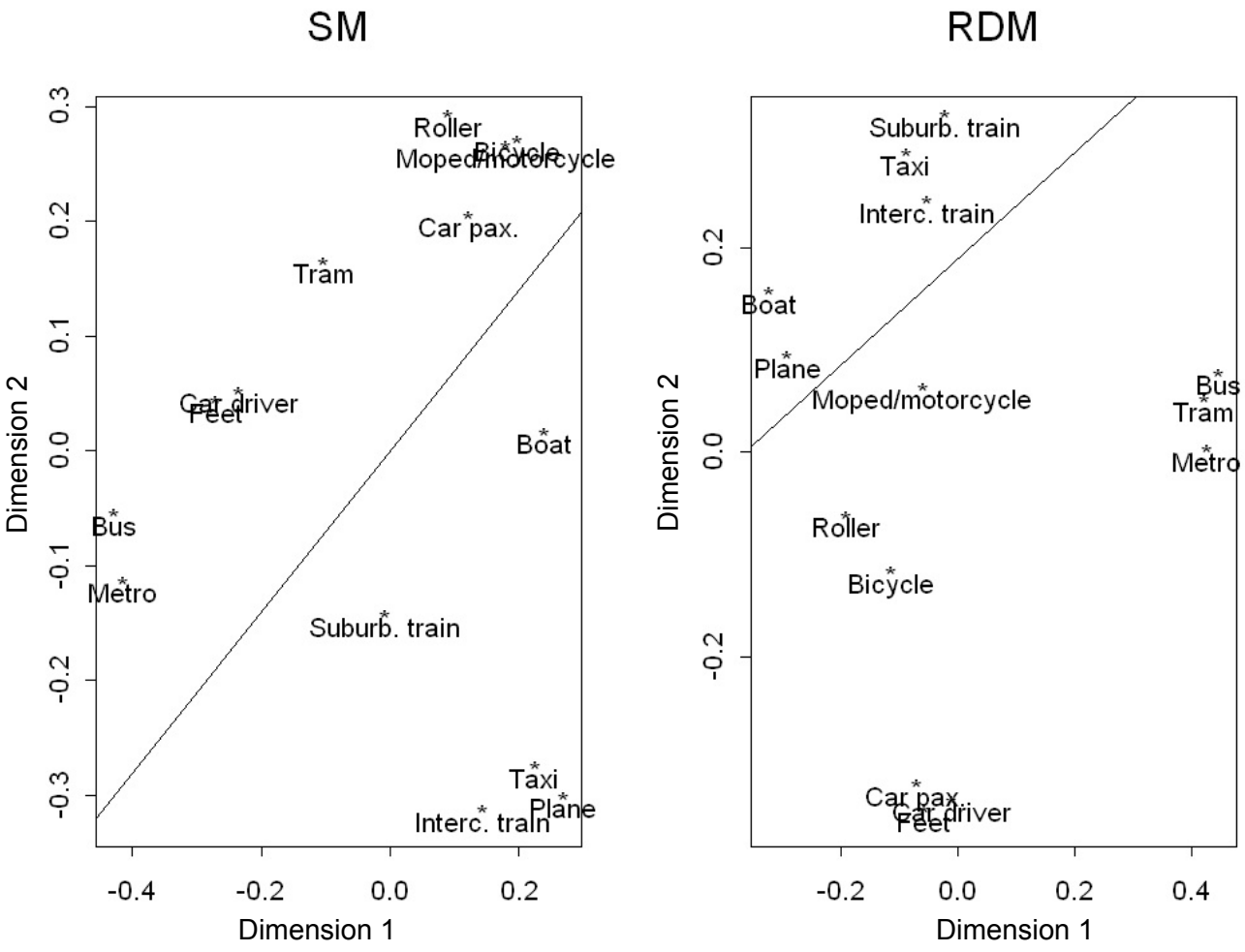

Figure 3. Multidimensional scaling for SM and RDM measures

dimensional perceptual maps are reported in Figure 3. The transportation modes are arranged so that the distance between two items fits as closely as possible their dissimilarity, measured as the "inverse" ($\sqrt{1-r^2}$) of their correlation *r*. Although different clusters of modes could be defined on the basis of these plots, the only possible cluster definition common to both SM and RDM measures consists of splitting the 14 modes into long-distance (including suburban and intercity train, boat, plane and taxi) and short-distance or urban modes (the remaining nine modes), as indicated by the two diagonal lines. We see here a confirmation of the importance of the territorial factors for the study of travel behaviors, that can also affect attitudes and desires related to different modes.

In the following we focus our attention on the above defined nine "urban transportation modes", for which complementarity, i.e., the possibility of using one instead of another, is much higher than for intercity travel, where the choice of the mode is often dictated by "hard" external constraints. Mode choice for long-distance travel is also more grounded on objective information, as a subject is more likely to objectively compare the performances of the available means and to make a rational choice, consistent with the random utility theory. We also exclude feet and rollers since they are usually not at the core of transportation policy plans and can less easily substitute the other modes.

Several ways to hierarchically classify the remaining seven modes can be envisaged. A technology-driven approach could suggest to group the modes primarily on the basis of the infrastructure they use (e.g. road versus rail). However, we should keep in mind that the variables considered (SM and RDM) are linked to individual evaluations and desires, so that a criterion that better suits the personal representation of different transportation modes should be more appropriate. In particular, previous research (Flamm, 2003, pp. 42-44) has shown that it is possible to define a classification of the transportation modes that is relevant from the travelers' point of view by considering the "implicit constraints" in their use (for example, need to respect schedules for transit services or availability of parking spaces), beyond their performances in terms of travel times and costs, as it is done in transportation models.

Our "Classification 1" (C1, see Figure 4) ranks transportation modes by increasing use constraints. Bicycles are the least constrained transport mode (no need of driver's licence or dedicated parking space, loosened circulation rules etc.). Moped and motorcycles share some constraints with cars, but consume less space, which can be extremely important in congested urban areas. Thus, the C1 top level classifies transportation modes as "bicycle",

“moped/motorcycle”, “car” and “transit”. Below that, the mode “car” is subdivided into “car driver” and “car passenger” (pax.), while mode “transit” is subdivided into “metro”, “bus,” and “tram”.

Alternatively, we could derive the modes “bicycle” and “moped/motorcycle” from an intermediate level mode “two-wheeled”, giving the typological classification “C2” (Figure 5). This is consistent with previous research that underlined the common characteristics on a subjective point of view between these two modes (Flamm, 2003, p. 95), although on a policy viewpoint it is surely more appropriate to keep them separate, for example because of their different environmental performances.

Finally, we could take an entirely different approach, starting by classifying the modes with respect to autonomy, ie, splitting the modes first based on whether the traveller is a driver or passenger. The active modes are then subdivided into “bicycle”, “moped/motorcycle”, and “car driver”, the passive modes into “car pax.”, “bus”, “tram”, and “metro, giving “C3” (Figure 6).

By performing the same analyses with all these classifications of the modes we can better assess the stability of the results concerning the SM-RDM relationship. In the following, these same classifications will be retained for both SM and RDM measures.

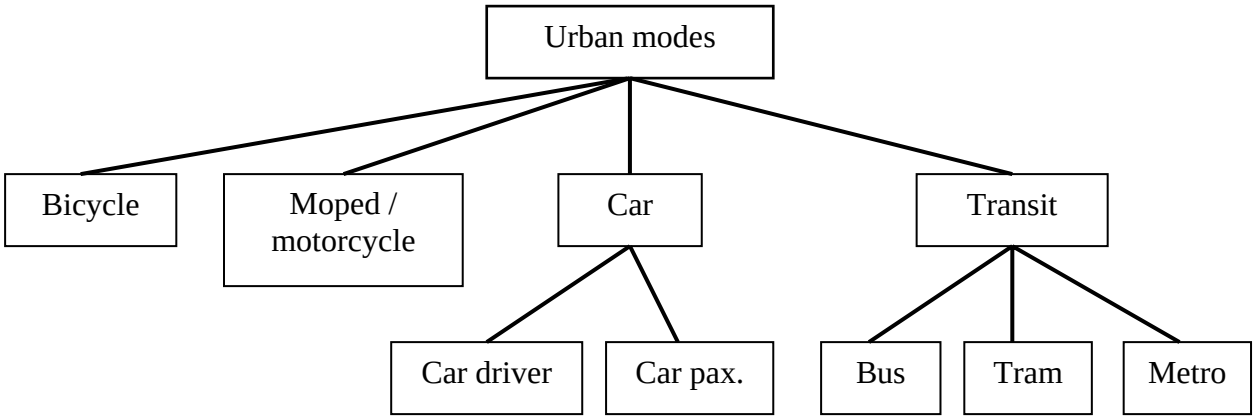

Figure 4. “C1” constraint driven hierarchical structure of transportation modes

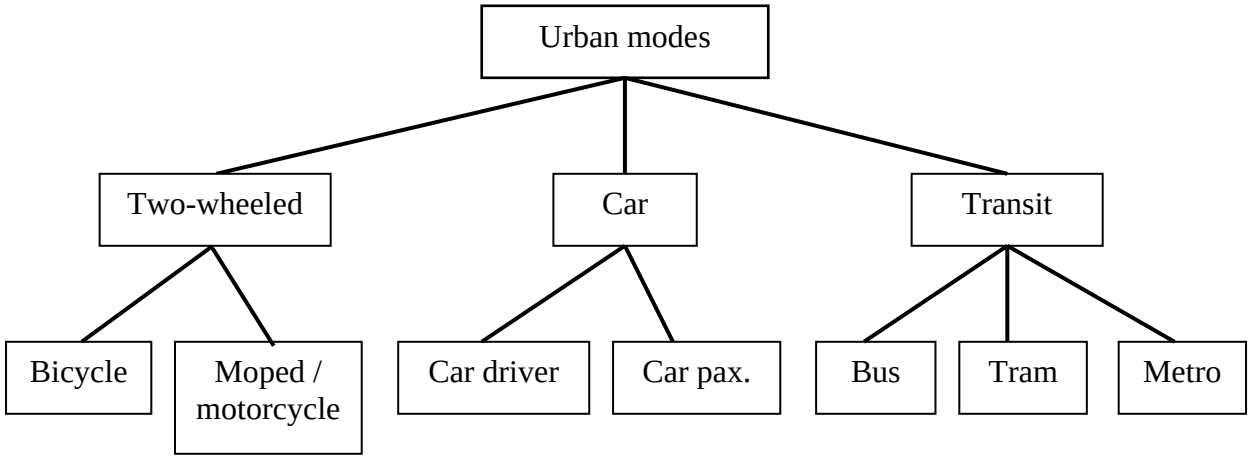

Figure 5. “C2” typology driven hierarchical structure of transportation modes

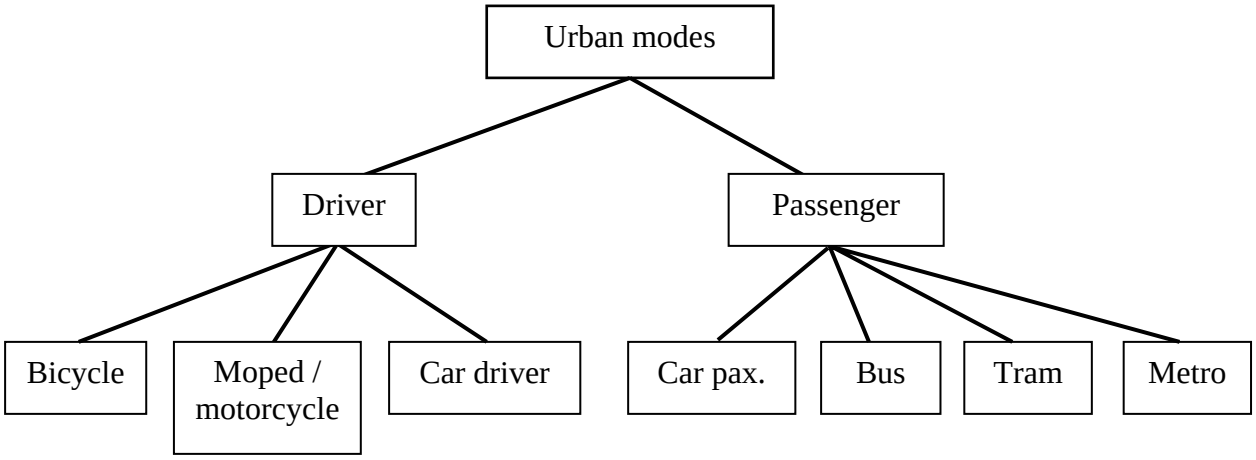

Figure 6. “C3” autonomy driven hierarchical structure of transportation modes

5 COMPUTATION AND ANALYSIS OF THE OVERALL SM AND RDM INDICES

We now study the relationship between subjective and desired mobility (SM and RDM), as measured by interviewing the INRETS staff, through the above defined C1 classification of the transportation modes, and compare the results to those using C2 and C3. We start by analysing the correlations between SM and RDM μ-scores of the above four C1 "intermediate level" modes (table 2). Pearson correlations are used, because it is essential to account for the distance among the different μ-scores. Obviously, the first two diagonal elements report the same values as table 1. The two lower diagonal elements confirm the negative relationship between SM and RDM both for cars and for public transport.

Table 2. Correlations matrix among the four "top-level" modes

		RDM			
		Bicycle	Moped/ motorcycle	Car	Transit
SM	Bicycle	0.197	0.006	-0.157	-0.086
	Moped/ motorcycle	0.109	0.182	0.024	0.047
	Car	-0.006	-0.152	-0.192	0.006
	Transit	-0.035	-0.063	0.150	-0.240

Off-diagonal elements show some interesting patterns. For example, those who feel to use bicycles a lot tend to have a lower desire to use public transport and cars in the future. Neither frequent car users[1] nor frequent transit users are satisfied with their mode usage. In addition, frequent car users also do not desire to use moped or motorcycle, whereas transit users find car use attractive. These different SM-RDM relationships across the modes considered may indicate deeper differences at a more personal level, so that a market segmentation technique based on the perceived and desired use of different transportation modes could be an effective way to set up transportation management policies targeted to the different groups of travellers, thus increasing the effectiveness of such policies (Diana and Mokhtarian, 2007).

In the next step, we compute both hierarchical and non-hierarchical global scores for SM and RDM for all the respondents. Figures 7 and 8 show the cumulative distribution functions for the SM and RDM indices respectively, i.e., for each possible score between -163 and 163, the value

[1] For simplicity, "frequent users (of a given mode)" from here on reflects Subjective Mobility, i.e., "those who think to use (a given mode) a lot", rather than the amount of travel that is actually done.

on the y-axis indicates the proportion of scores that are smaller or equal. Only the hierarchical scores C1 and C3 are reported, since C2 turned out to be an intermediate case. The inserts in these two figures represent the scatter plots of the hierarchical against the non hierarchical (NH) scores for each observation in the dataset.

As expected, the hierarchical scores are more dispersed and, thus, more informative. For the non-hierarchical SM score, for instance, all except two subjects are cramped into the range (-80,50) with the median 50% of all score values falling into the narrow range (-5,10). The hierarchical SM indices differentiate better, because their values are more evenly distributed across the range (-160,100), with the center 50% spread across the range (-30,40). As a result, they are less subject to random errors (see inserts of Figures 7 and 8). The total number of pairs of individuals with an ambiguous pairwise ordering (as among the $n(n-1)/2 = 13366$ pairs) was decreased from 325 for NH to 62 for C1, 63 for C2 and 70 for C3.

Similar results were found for RDM, although the gains are less impressive, as can be seen by comparing Figure 8 to Figure 7. The reduction of the number of ambiguously ordered pairs of individuals is in fact less marked (from 200 pairs to 143, 138 and 134 respectively for C1, C2 and C3). This is partly due to the fact that 12 individuals indicated they would like to travel exactly as they actually do by all the seven considered transportation modes, so that they receive the same score, irrespective of the method used (they are represented by the long vertical segments in the left part of Figure 8). These 12 individuals thus represent the neutral point in the ideal RDM scale encompassing the seven transportation modes. Individuals whose score is below this threshold (thus lying at the left of the vertical traits of Figure 8) express a global desire to travel less, whereas at the right end of the figure we have those who would like to overall travel more. Although some individuals may fall above or below the neutral point according to the classification of the transport means being considered, it is rather evident from Figure 8 than the majority of the respondents would globally like to travel more, and only few would like to travel less. Our results then constitute a quantitative confirmation of the existence of a latent travel demand, an assumption that is at the base of the abundant literature concerning the induced traffic (see for example Denvil, 1996; Abelson and Hensher, 2001; Noland and Lem, 2002).

C3 outperforms C1 and C2 since it differentiates best. This points at the importance of the fact of driving the vehicle or simply being carried by other and its relevance for the individual representation of different transport means, consistently with previous studies that put into evidence the importance of feeling in control of the travel experience (see for example Stradling

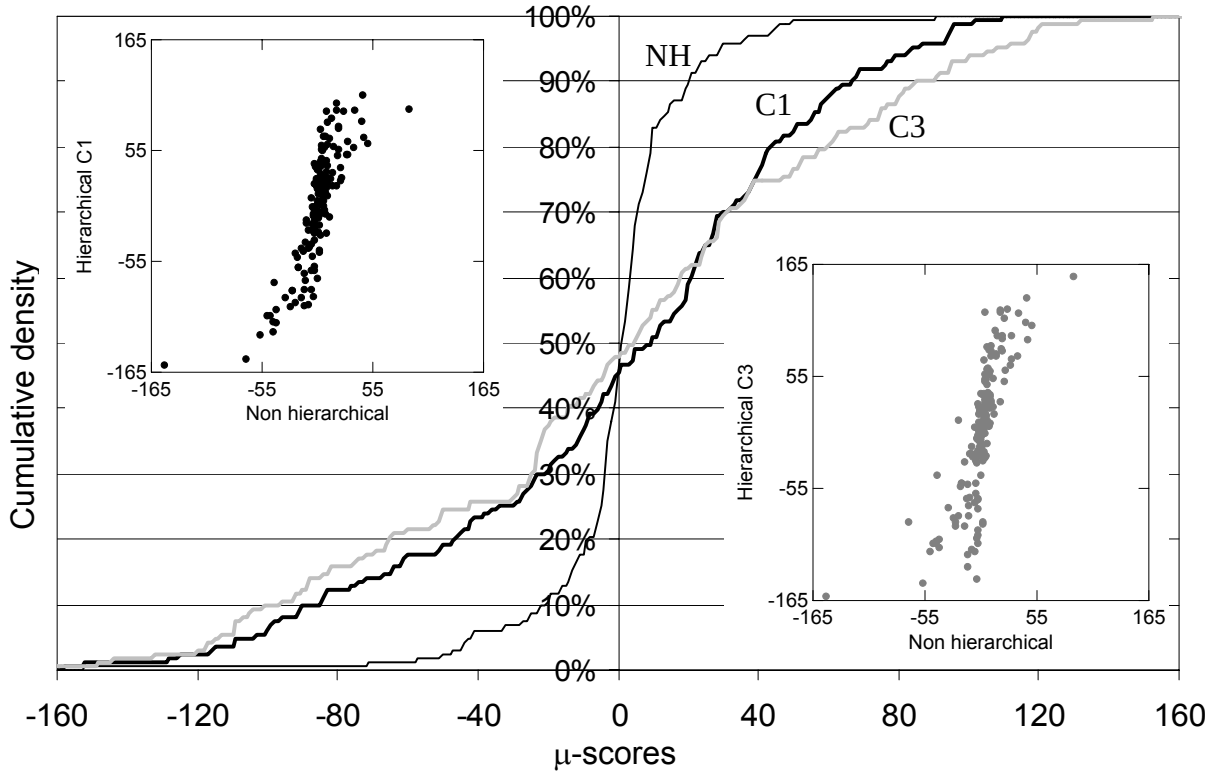

Figure 7. Variation of hierarchical (C1, C3) and non hierarchical (NH) SM indices

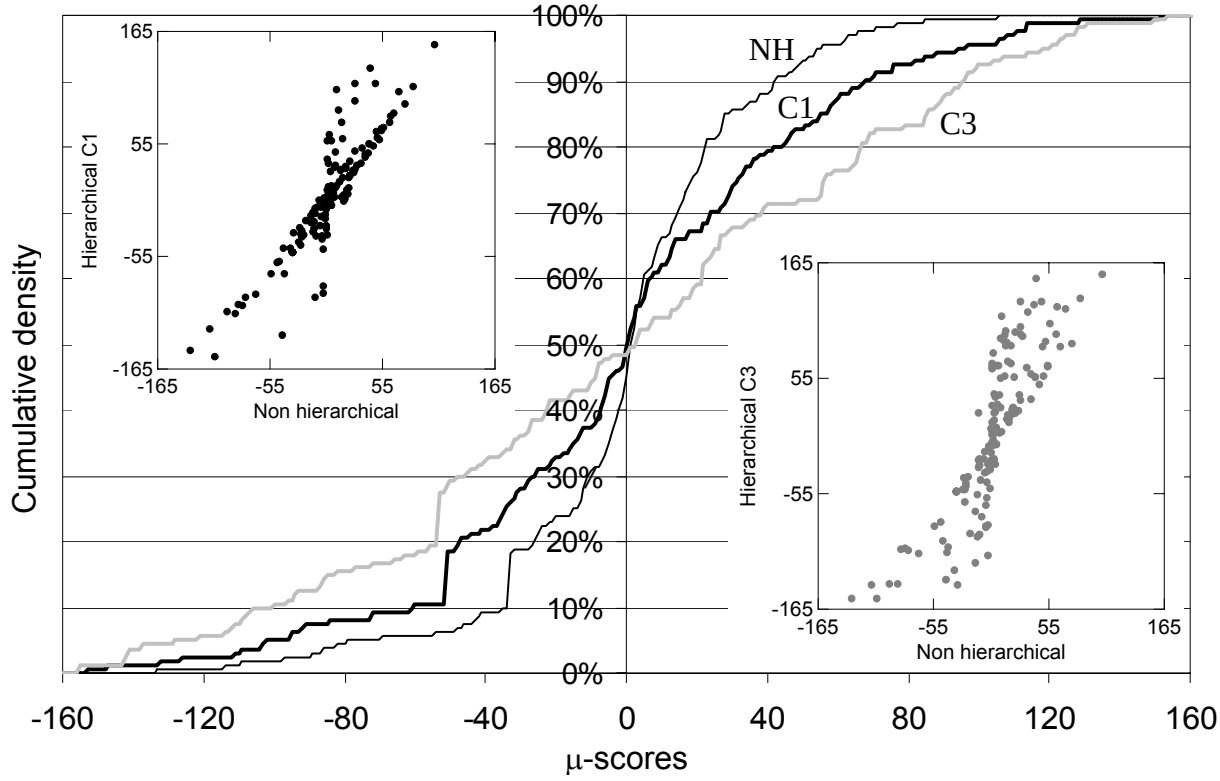

Figure 8. Variation of hierarchical (C1, C3) and non hierarchical (NH) RDM indices

et al., 2000; Anable and Gatersleben, 2005).

It is also interesting to note that SM and moreover RDM scatter plots of Figures 7 and 8 show two distinct clusters of respondents. One cluster lies along a straight line and the other forms an S-shaped curve. A closer inspection of the data indicates that this has to do with the way in which the respondents rated their SM and RDM for the different transport means. Some tend to give the same rating to every transportation mode. In this case, the score that is obtained by combining the different ordinal measures is insensitive to the aggregation method (NH, C1, C2, C3). Hence these users will form a diagonal line across the scatter plot. This straight line is less evident in Figure 7 than in Figure 8, as fewer subjects gave the same response to all SM questions than to all RDM questions.

On the other hand, a different group of respondents gives ratings that are quite differentiated across the different modes. In this case, the number of ambiguous pairwise orderings increases and the combined score tends to be flattened around the zero value when the NH method is used. Here the information gain in using a hierarchical procedure is maximised, so that C1 and C3 scores will be rather different than the NH one. Hence those subjects will align along an S-shaped curve. Comparing the two scatter plots of Figure 8, we can have a confirmation that C3 differentiates better than C1, since more subjects “belong” to the C3 versus NH S-shaped curve.

We now turn our attention to the global SM-RDM relationships. The correlation between the global SM and RDM μ-scores was -0.091 when considering C1, and similar for C2 (-0.160) and C3 (-0.115). The correlation between the non-hierarchical SM and RDM scores was -0.072. Combining the ordinal measures by considering a classification of the modes consistent with the subjective point of view of the travellers allows then to draw more general conclusions. Our results could possibly indicate some dissatisfaction tampered by fatalism, and are consistent with previous research reporting a negative relationship between SM and RDM for many categories of trips (Choo *et al.*, 2005). This rather weak correlation is the outcome of contradictory trends at the single mode level (table 1). We offer an interpretation of such results and we prospect the related policy implications in the concluding section.

6 DISCUSSION: μ-SCORES VERSUS FACTOR ANALYSIS

In the preceding sections we combined several mobility-related ordinal measures by using multidimensional scaling first to separate a suitable group of “urban” transportation modes. Then we computed μ-scores for several potential factor structures (“constraint”, “typology”, and “autonomy”) to select the factor structure resulting in the most informative SM and RDM indices. We would like now to compare this non-parametric methodology to traditional factor analyses, which have been widely used in travel and other behaviour research.

Factor analysis uses pairwise correlations among a set of n variables to postulate some underlying factors that could adequately describe the observed data and should be easily interpreted. The number of factors to be extracted k represents a compromise between the contrasting goals of reproducing the observed correlation patterns as close as possible with a saturated model ($k = n$) and of having a maximally parsimonious model ($k = 1$). Despite some widely used guidelines and rules of thumb, the researcher must basically rely on his/her own judgement in some key passages such as the selection of the number of factors to consider, the definition of a significance threshold for the factor loadings, or the choice of a rotation method to aid with interpretation of the results. Moreover, as a parametric method based on the linear model, factor analysis relies heavily on the assumption that the importance (e.g., in terms of policy implications) of an increase by one unit is the same across the whole scale of this variable. When this assumption of linearity can be reasonably justified, factor analyses can provide useful results. Concerning SM, however, the importance of “not at all” vs “rarely” or “often” vs “very often” cannot reasonably be assumed to be the same, even though the order “not at all” < “rarely” < “often” < “very often” can be assumed to be valid.

μ-scores can be seen as a data reduction technique, in that a single measure is provided to comprehensively assess overall preference. In our study, however, the aim was not the traditional aim of (exploratory) factor analysis, namely to reduce the number of variables to be included, but to select a model yielding optimal SM and RDM indices, among several models theoretically postulated *a priori*. As a consequence, we explicitly wanted to include all of the observed SM and RDM measures, rather than checking to which extent the dimensionality of the SM and RDM constructs can be reduced. One possibility would be to run a confirmatory rather than an exploratory factor analysis, postulating a measurement model where the mode-specific measures load on a single factor and then estimating it through a structural equation modelling approach (see Bollen, 1989 for a theoretical discussion on the exploratory and confirmatory

factor analysis concepts and Diana, 2008 for a recent joint application of both techniques in the travel behavior research field). However this seems not appropriate in our case: a simple inspection of Figure 3 shows the existence of “cluster of modes” (i.e. some modes are plotted near each other to form distinct groups) that point to the multidimensional nature of the SM and RDM by-mode measures. On the other hand, the fact of having combined these measures into a single index does not mean that we are trying to force-fit a single factor model, since the method we used does not postulate any particular correlation pattern across the measures.

Beyond the above “dimensionality issue”, a more specific concern is related to ordinal data. There are several examples of factor analysis performed on ordinal variables in travel behavior research (for example Mokhtarian *et al.*, 2001; Johansson, 2004; Anable, 2005). Some authors suggest applying factor analysis to ordinal data by replacing the data of each variable by its ranks (Bagge *et al.*, 2005). While this approach avoids some problems, such as sensitivity to outliers, it does not overcome the fundamental problem that the latent factors are combined as a linear combination with fixed “factor loadings” as coefficients (Brookins, 1970). The required assumptions (“multivariate normal”, “bivariate normal”, “normal ogive”, and “proportional odds”, see Jöreskog and Moustaki, 2001, for details) are often, as in our case, difficult to justify. Thus, the “rank transformation approach” (Lemmer *et al.*, 1968; Conover and Iman, 1981), when applied to situations with several factors, suffers from similar problems in factor analysis as in analysis of variance (Fligner, 1981; Blair et al, 1987; Akritas *et al.*, 1997; Haas, 1999; Shah and Madden, 2004).

To better illustrate the two above points, we run three different series of factor analyses with 2, 3 and 4 factors for the 7 SM and 7 RDM mode-specific measures that we considered. In the first series we directly used the respondent’s scores, in the second one we log-transformed them, and, in the third, we took their ranks. Compared to principal components analysis, factor analyses allow for “rotations” of the solution to yield factor structures that might ease the interpretation. In all the three cases we performed a varimax rotation of the solution. The resulting factor loadings are reported in Table 3. Loadings with an absolute value <0.10 are omitted, and the largest loading for each variable is shown in bold. The variance explained by each model is also reported.

For ordinal data, i.e., if the functional relationship between the observed measures and their meaning in terms of the underlying concept is not known, monotone transformations such as taking logarithms or computing ranks should ideally provide the same results. Models selected on the basis of the information content of the μ-scores are in fact unaffected by these (and other)

Table 3. Factor analyses results by assumed number of factors and transformation

	2 factors				3 factors						4 factors							
	SM		RDM		SM			RDM			SM				RDM			
No transform.																		
Bicycle		**-.28**	.11	**.40**		**-.40**			**.33**	-.20		-.13		**.42**		.14	-.13	**.48**
Motorcycle			.25	**.31**				.19	**.52**	.13			**.11**		.18	**.55**		.16
Car driver	-.50	**.54**	**-.16**	-.11	**-.61**	.31	.30	-.14		**.41**	**-.60**	.16	.33	-.23	-.13		**.48**	
Car pax.	.21	**.39**			.12	**.40**	.16			**.21**		**.49**		.19			**.18**	
Bus	**.93**		**.80**	-.17	**.92**	.16		**.77**	.26		**.87**	.34	-.11		**.76**	.31	-.15	
Tram	**.43**	.14	**.77**	.46	.41		**.52**	**.71**	.46	-.26	.38	.28	**.42**	.20	**.72**	.32	-.23	.29
Metro	**.81**		**.78**		**.80**			**.78**			**.86**		.13		**.81**			
var. explained	*37%*		*35%*		*41%*			*40%*			*45%*				*43%*			
Log transform.																		
Bicycle	**.10**		.27	**.47**			**-.32**	.25	.27	**-.45**			**-37**	.12	.18	**.59**	.15	
Motorcycle		**.15**	.34	**.59**		**.17**		.36	**.52**	.28		**.17**			.25	.37	**.55**	.15
Car driver	**-.56**	.39		**.20**	-.39	**.53**	.14		**.37**		-.36	**.55**	.15	-.11			.10	**.42**
Car pax.		**.20**			.11		**.44**			**-19**		.10	**-.42**	.15		-.11		**.15**
Bus	**.87**	.20	**.77**		**.89**	-.14		**.81**			**.86**	-.16		.24	**.78**		.26	
Tram	.35	**.49**	**.83**	.31	**.50**	.37		**.81**		.39	**.50**	.34			**.74**	.39	.26	-.17
Metro	**.76**	.21	**.81**		**.78**			**.80**			**.82**				**.83**	.14		
var. explained	*33%*		*41%*		*38%*			*45%*			*38%*				*48%*			
Rank transf.																		
Bicycle		**-.27**		**.28**		**-.41**	.25			**.30**	.14		-.20	**.42**			**.35**	-.12
Motorcycle			**.21**				**.29**		**.51**				.14	**.29**		**.52**		
Car driver	-.34	**.62**	-.20	**-.47**	**-.56**	.41	.13	-.25		**-.45**	**-.66**		.18	-.14	-.26		-.21	**.42**
Car pax.	.23	**.30**		**-.16**		**.32**	.11			**-.14**			**.52**					**.17**
Bus	**.91**	-.22	**.84**	.17	**.91**		.14	**.77**	.30	.16	**.79**	.42	.25		**.78**	.30		-.13
Tram	**.40**		**.75**	.26	.33	.11	**.34**	**.72**	.24	.24	.13	**.54**		.12	**.72**	.19	.30	
Metro	**.76**	-.23	**.67**		**.79**		.12	**.77**			**.67**	.45			**.77**			
var. explained	*34%*		*32%*		*38%*			*37%*			*42%*				*39%*			

monotonic transformations of the data. In contrast, as Table 3 shows, this is often not the case when using factor analysis. The three transit variables (Bus, Tram, Metro) heavily load on the same factor in most of the 18 analyses, but loadings of other variables vary widely. Some variables, especially “SM driver” and “RDM motorcycle”, have quite significant loadings across different factors in many models, thus jeopardizing a clear interpretation of the results. No single model stands out as the “best”.

Of note, merely using any “non-parametric” approach is not a panacea. The results based on the rank-transformed data consistently explain the smallest amount of the variance, even though our data clearly does not follow a Gaussian distribution.

While the results based on factor analyses do not allow for selecting a particular model, they are largely consistent with both the multidimensional scaling findings (figure 3) and the hierarchical structures of the transport modes that we assumed (figures 4 to 6). For example, the strong

cluster of the three transit variables is apparent in the RDM perceptual map, whereas “SM tram” is also nearer other variables such as “SM driver” and “SM passenger”, thus reflecting respectively the strong loading of “SM driver” on the first factor and the loading of the “SM tram” variable on factors other than the first in table 3. We can also observe that the three 4-factor-RDM models match our classifications C1, the 3-factor-RDM model without transformation matches C2 and the 2-factor-RDM model with log transformation matches C3, thus providing empirical evidence to the three postulated hierarchical structures.

Still, no single method can be a panacea. The number of possible hierarchical structures increases so fast, that an exhaustive search is impossible even for moderately sized studies. Thus, the proposed non-parametric approach should not be seen as an alternative to traditional analyses based on the linear model, but merely as an addition. In a first step, one might use factor analyses and multidimensional scaling, to limit the number of structures to be considered for analysis. The proposed method could then be used to resolve the ambiguities among models suggested under the arbitrary assumptions regarding transformations, number of factors, cut-off points, and rotations.

7 POLICY IMPLICATIONS AND CONCLUSIONS

This paper has investigated the relationship between subjective mobility (SM) and relative desired mobility (RDM) in a sample of clerical workers. To obtain actionable hypotheses, measures pertaining to several different transportation modes need to be combined into a single index. This was achieved through a modification of an innovative methodology. Our method utilizes μ-scores for multivariate data (Wittkowski *et al.*, 2004), yet allows for a hierarchical structure of the data to avoid ambiguities when ordering the respondents based on a set of ordinal measures. This modification not only improves the information content of the index and the significance level of the observed correlations, but also provides novel insight into the relationship among the transportation modes and into personal characteristics of their users.

The correlations between “intermediate” SM and RDM indices unveil some interesting features. In particular, travellers using “leisure” modes (bi- or motorcycle) seem to enjoy the use of their “toys”. Bicycle users, in particular, are adverse to cars or public transportation, while motorcycle users are neutral. In contrast, users of cars or public transportation are dissatisfied. Still, they would rather switch to the other non-leisure category, cars, than to either of the leisure modes. The combination of RDM measures into a single index shows that the majority of respondents

globally desires to travel more, and only few would like to travel less. The suggestion of a latent demand for travel contributes a novel argument based on quantitative results to the debate on the phenomenon of induced traffic.

Looking at the top-level indices that combine all seven urban transportation modes considered, some interesting policy implications can be inferred. As mentioned in the introduction, Choo *et al.* (2005) postulated that SM might be one of the determinants of RDM, and found a negative relationship for many categories of trips. We provide empirical evidence for a negative SM-RDM correlation overall, and we quantitatively uncover the existence of a latent demand for travel among a majority of respondents. Combining theoretical consideration from previous research and empirical evidence from our study shows that to lessen the desire for more travelling, people need to be aware how much they already travel, so that the latent demand decreases as well, a desirable goal of almost any modern transportation policy. Underestimating one's current travel can conversely be associated with a greater desire for more mobility.

More disaggregate results are also relevant on a policy viewpoint. SM underestimation can contemporarily prompt for a greater desire to use more energy-consuming modes (those reported on the left side of table 1 where SM-RDM correlations are negative) and a lesser desire to use more environmentally friendly modes (the right side of table 1), hence leading to a much worse mix of use of different modes from an environmental perspective. Most importantly, this doubly negative effect could be one of the mechanisms underlying the widely observed difficulties that modal diversion strategies encounter systematically when trying to substitute the use of cars. Not taking into account the role of the self-consciousness of the amount of trip that one consumes is then probably one of the reasons of the ineffectiveness of many modal diversion campaigns, particularly when the use of bicycles is promoted.

The potentials of combining ordinal measures by means of μ-scores for hierarchically structured ordinal variables could be further exploited in future research. In particular, it would be of interest to analyze SM and RDM data categorized by trip purpose, rather than by transport mode, to check if the general findings concerning the SM-RDM relationships are the same. This method could also have several other applications in transport research, given the quantity of ordinal data that are available through mobility surveys that are often difficult to properly analyze.

ACKNOWLEDGEMENTS

The survey originating the dataset that has been used in the present research was implemented thanks to the support of INRETS, the French National Institute for Transport and Safety Research. The project was also supported in part by Grant Number UL1RR024143 from the U.S. National Center for Research Resources (NCRR). Comments from three anonymous referees helped to improve a previous version of the paper.

REFERENCES

Aarts, H., Dijksterhuis, A.: The automatic activation of goal-directed behaviour: The case of travel habit, *Journal of Environmental Psychology* 20(1), 75-82 (2000)

Abelson, P.W., Hensher, D.: Induced travel and user benefits: clarifying definitions and measurement for urban road infrastructure, in K.J. Button and D.A. Hensher (eds.), *Handbook of Transport Systems and Traffic Control*, Pergamon, Amsterdam, 121-141 (2001)

Akritas, M.G., Arnold, S.F., Brunner, E.: Nonparametric hypotheses and rank statistics for unbalanced factorial designs – Part I, *Journal of the American Statistical Association* 92(437), 258-265 (1997)

Anable, J.: 'Complacent car addicts' or 'aspiring environmentalists'? Identifying travel behaviour segments using attitude theory, *Transport Policy* 12(1), 65-78 (2005)

Anable, J., Gatersleben, B.: All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes, *Transportation Research Part A* 39(2-3), 163-181 (2005)

Bagge, C.L., Stepp, S.D., Trull, T.J.: Borderline personality disorder features and utilization of treatment over two years, *Journal of Personality Disorders* 19(4), 420-439 (2005)

Bamberg, S., Ajzen, I., Schmidt, P.: Choice of travel mode in the theory of planned behavior: the roles of past behavior, habit, and reasoned action, *Basic and Applied Social Psychology* 25(3), 175-187 (2003)

Blair, R.C., Sawilowsky, S.S., Higgins, J.J.: Limitations of the rank transform statistic in tests for interactions, *Communications in Statistics* 16(4), 1133-1145 (1987)

Bollen, K.A.: Exploratory and confirmatory factor analysis, in: K.A. Bollen, *Structural Equations with Latent Variables*, Wiley, New York, 226-232 (1989)

Brookins, O.T.: Factor analysis and gross national product: a comment, *The Quarterly Journal of Economics* 84(4), 648-650 (1970)

Cherchye, L., Vermeulen, F.: Robust rankings of multidimensional performances: An application to Tour de France racing cyclists, *Journal of Sports Economics* 7(4), 359-373 (2006)

Cherchye, L., Vermeulen, F.: Acknowledgement of priority, *Journal of Sports Economics* 8(5), 557 (2007)

Choo, S., Collantes, G.O., Mokhtarian, P.L.: Wanting to travel, more or less: exploring the determinants of the deficit and surfeit of personal travel, *Transportation* 32(2), 135-164 (2005)

Collantes, G.O., Mokhtarian, P.L.: Subjective assessments of personal mobility: What makes the difference between a little and a lot?, *Transport Policy* 14(3), 181-192 (2007)

Conover, W.J., Iman, R.L.: Rank transformations as a bridge between parametric and nonparametric statistics, *The American Statistician* 35, 124-134 (1981)

Denvil, C., ed.: Special issue on induced traffic, *Transportation* 23(1), 1-119 (1996)

Diana, M.: *An Exploratory Web-Based Attitudinal Travel Survey Administered to INRETS Staff*. Les Collections de l'INRETS – "Outils et méthodes" n°12, Arcueil, France, 187 pp, ISBN 2-85782-623-0 (2005)

Diana, M.: Making the "primary utility of travel" concept operational: a measurement model for the assessment of the intrinsic utility of reported trips, *Transportation Research Part A* 42(3), 455-474 (2008)

Diana, M., Mokhtarian, P.L.: The individual desire to change a "modal basket" and its relationship with the actual and perceived use of different transport means, *11th World Congress on Transport Research*, Berkeley, USA, 41 pp. (2007)

Diana, M., Mokhtarian, P.L.: Desire to change one's multimodality and its relationship to the use of different transport means, *Transportation Research Part F*, in press, doi:10.1016/j.trf.2008.09.001 (2009)

Dobson, R., Tischer, M.L.: Perceptual market segmentation technique for transportation analysis, *Transportation Research Record 673*, 145-152 (1978)

Flamm, M. : *Comprendre le choix modal – Les déterminants des pratiques modales et des représentations individuelles des moyens de transport*. Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 307 pp. (2003)

Fligner, M.A.: Rank transformations as a bridge between parametric and nonparametric statistics – Comment, *The American Statistician* 35, 131-132 (1981)

Gärling, T., Fujii, S., Boe, O.: Empirical tests of a model of determinants of script-based driving choice, *Transportation Research Part F* 4(2), 89-102 (2001)

Garvill, J., Marell, A., Nordlund, A.: Effects of increased awareness on choice of travel mode, *Transportation* 30(1), 63-79 (2003)

Gensch, D.H., Torres, P.T.: Perceived-difference segmentation model for mass transit marketing, *Transportation Research Record 765*, 16-22 (1980)

Haas, C.N.: On modeling correlated random variables in risk assessment, *Risk Analysis* 19(6) 1205-1214 (1999)

Johansson, M.: Psychological motivation of pro-environmental travel behaviour in an urban area, in T. Rothengatter and R. D. Huguenin (Eds.), *Traffic and Transport Psychology – Theory and Application – Proceedings of the ICTTP 2000*, Elsevier, Amsterdam, Netherlands, 435-442 (2004)

Jöreskog, K.G., Moustaki, I.: Factor analysis of ordinal variables: A comparison of three approaches, *Multivariate Behavioral Research* 36(3) 347-387 (2001)

Lemmer, H.H., Stoker, D.J., Reinach, S.G.: A distribution-free analysis of variance technique for block designs, *South African Statistical Journal* 2, 9-32 (1968)

Mokhtarian, P.L., Salomon, I.: How derived is the demand for travel? Some conceptual and measurement considerations, *Transportation Research Part A* 35(8), 695-719 (2001)

Mokhtarian, P.L, Salomon, I., Redmond, L.: Understanding the demand for travel: it's not purely "derived", *Innovation: the European Journal of Social Science Research* 14(4), 355–380 (2001)

Noland, R.B., Lem, L.L.: A review for the evidence of induced travel and changes in transportation and environmental policy in the US and the UK, *Transportation Research Part D* 7(1), 1-26 (2002)

Outwater, M.L., Castleberry, S., Shiftan, Y., Ben-Akiva, M., Zhou, Y.S., Kuppam, A.: Attitudinal market segmentation approach to mode choice and ridership forecasting – Structural Equation Modeling, *Transportation Research Record 1854*, 32-42 (2003)

Paczesny, S., Banchereau, J., Wittkowski, K.M., Saracino, G., Fay, J., Palucka, A.K.: Expansion of melanoma-specific cytolytic CD8+ T cell precursors in patients with metastatic melanoma vaccinated with CD34+ progenitor-derived dendritic cells, *Journal of Experimental Medicine* 199(11), 1503-1511 (2004)

Salomon, I., Mokhtarian, P.L.: What happens when mobility-inclined market segments face accessibility-enhancing policies? *Transportation Research Part D* 3(3), 129-140 (1998)

Sapir, M., Verbel, D., Kotsianti, A., Saidi, O.: Live logic^{TM}: Method for approximate knowledge discovery and decision making, in D. Ślęzak, G. Wang, M. Szczuka, I. Düntsch and Y. Yao (eds.), *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing – 10th International*

Conference, RSFDGrC 2005 – Proceedings, Part 1, Lecture Notes in Artificial Intelligence 3641, Springer, Berlin, 532-540 (2005)

Shah, D.A., Madden, L.V.: Nonparametric analysis of ordinal data in designed factorial experiments, *Phytopathology* 94(1) 33-43 (2004)

Spangler, R., Wittkowski, K.M., Goddard, N.L., Avena, N.M., Hoebel, B.G., Leibowitz, S.F.: Opiate-like effects of sugar on gene expression in reward areas of the rat brain, *Molecular Brain Research* 124(2), 134-142 (2004)

Steg, L., Vlek, C., Slotegraaf, G.: Instrumental-reasoned and symbolic-affective motives for using a motor car, *Transportation Research Part F* 4(3), 151-169 (2001)

Stradling, S.G., Meadows, M.L., Beatty, S.: Identity and independence: two dimension of driver autonomy, *International conference on traffic and transport psychology – ICTTP 2000*, Bern, 12pp (2000)

Tardiff, T.J.: Attitudinal market segmentation for transit design, marketing, and policy analysis, *Transportation Research Record 735*, 1-7 (1979)

Thøgersen, J.: Understanding repetitive travel mode choices in a stable context: A panel study approach, *Transportation Research Part A* 40(8), 621-638 (2006)

TRB – Transportation Research Board: Preferences, perceptions, and market segments in travel behavior, *Transportation Research Record 649*, 48 pp. (1977)

Verplanken, B., Aarts, H., van Knippenberg, A., van Knippenberg, C.: Attitude versus general habit: Antecedents of travel mode choice, *Journal of Applied Social Psychology* 24(4), 285-300 (1994)

Verplanken, B., Aarts, H., van Knippenberg, A.: Habit, information acquisition, and the process of making travel mode choices, *European Journal of Social Psychology* 27(5), 539-560 (1997)

Wittkowski, K.M., Lee, E., Nussbaum, M., Chamian, F.N., Krueger, J.G.: Combining several ordinal measures in clinical studies, *Statistics in Medicine* 23(10), 1579-1592 (2004)