COMBINING ATTITUDES AND TRAVEL BEHAVIOR - A COMPARISON OF URBAN MOBILITY TYPES IDENTIFIED IN SHANGHAI, BERLIN AND SAN FRANCISCO

Miriam Magdolen (corresponding author)
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Email: miriam.magdolen@kit.edu

Sascha von Behren
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Email: sascha.vonbehren@ kit.edu

Bastian Chlond
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Email: bastian.chlond@kit.edu

Marcel Hunecke
Ruhr University Bochum (RUB)
44780 Bochum, Germany
Email: marcel.hunecke@ruhr-uni-bochum.de

Peter Vortisch
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Email: peter.vortisch@kit.edu

Paper prepared for presentation and publication at the
98th Annual Meeting of the Transportation Research Board
6,172 words
4 tables (4 * 250 words = 1,000 words)
Total: 7,172 words (including abstract, text, references, tables)
Submission date: August 01, 2018

ABSTRACT

A detailed knowledge of potential travelers' behavior and underlying psychological factors is essential to estimate the potential of mobility-related services and improve transportation systems. The definition of mobility types allows the assignment of individuals to the respective groups of people with similar mobility needs. Previous research has mainly focused on one dimension only, either attitudes or travel behavior, for identifying distinct mobility types with cluster analysis. Considering both dimensions allows to uncover dissonances and consistencies between attitudes and behavior. Further, only a few studies compare mobility types in an international setting. In our study, we try to identify two-dimensional urban mobility types and compare them between cities in different cultural contexts. Therefore, we develop an integrated clustering approach and support it by machine learning algorithms in pre- and post-analysis. To combine attitudes and behavior in different urban mobility types, we use data from a standardized survey, conducted in Berlin, Shanghai and San Francisco. This survey is based on the concept of a travel skeleton that allows us to collect typical weekly travel behavior as well as psychological constructs. Based on the clustering processes, we identify 11 distinct urban mobility types. The results show clusters with dissonances between attitudes and behavior (e.g., Cluster 10 "Car-Enthusiasts with high Norms") and clusters with consistent characteristics (e.g., Cluster 4 "Convinced Bicycle and Public Transportation Users"). Further, the comparison between the cities highlights city specifics. Berlin and Shanghai are more similar in terms of occurring mobility types and thus mobility needs than San Francisco.

INTRODUCTION

Market analysis for any form of vehicles (e.g., cars, e-scooters) or service-related mobility products (e.g., sharing concepts) in urban contexts require a detailed understanding of the mobility demand. Therefore, it is important to segment the mobility market into different target groups with similar characteristics (mobility types). To capture mobility comprehensively, we must consider different aspects including realized travel behavior, norms and attitudes, sociodemographic as well as spatial environment. This is, for example, of particular interest for the adaptation of new mobility offers to the particular needs of the prospective target groups and the identification of market potential.

Previous research has mainly focused on either objective (i.e., travel behavior) or subjective (i.e., attitudes) dimensions to identify mobility types. A separate consideration of the dimensions do not reveal potential discrepancies between travel behavior and psychological factors. This phenomenon is often addressed in different contexts (e.g., cognitive dissonance, captive drivers). It is also not possible to show the consistency of attitudes and behavior. In order to define mobility types appropriately, we recommend a combined consideration of both dimensions. In addition, sociodemographic characteristics and spatial structures should not be ignored for subsequent analyses. This causes the following research questions: Which two-dimensional mobility types are prevalent in urban structures? Are there mobility types with dissonances between realized travel behavior and attitudes? Which differences and similarities of mobility types are observable between cities?

This paper aims to build an extension on existing research approaches and provides a meaningful contribution to the discussion on urban mobility types. To define differing mobility types, we used the data from an international standardized survey conducted in Berlin (Germany), San Francisco (USA) and Shanghai (China). The survey is based on the concept of a travel skeleton combining "objective" questions on travel behavior with "subjective" questions on individual attitudes and norms. This allows for a segmentation into urban mobility types with two dimensions. We calculated appropriate variables to define mobility types. As a segmentation method, we used a two-step clustering approach to form mobility types with distinct characteristics. Machine learning algorithms supported the explorative clustering approach in pre- and post-analysis. We provide a framework for determining mobility types in our surveyed cities. Therefore, we were able to research on the travel behavior and attitudes of people in three different cultures and their respective markets. Results showed similarities between Shanghai and Berlin. In San Francisco, we obtained specific car-oriented mobility types.

The following sections describe the outcome and conclusions of our analysis, and are structured as follows; after a literature review and outline of the survey approach, we describe our methodology to identify urban mobility types. The explanation of the cluster analysis and the application of the machine learning algorithms follows. In particular, we discuss the simultaneous inclusion of attitudes and behavior in the segmentation. After evaluating the cluster solution, the distinct mobility types and their characteristics are presented and interpreted. We then examine our mobility types from an international perspective. Finally, we discuss our approach, emphasize new insights from our study and refer to further research.

LITERATURE REVIEW

Travel behavior is a complex process influenced by various external and internal factors. In addition to spatial and sociodemographic influences, psychological characteristics are also considered as decisive factors (*1*). The effects of attitudes and norms on actual behavior are described in Ajzen's (*2*) Theory of Planned Behavior (TPB). Due to many determining factors, the investigation why people behave in a certain way is of particular interest in travel behavior research. Therefore it is essential to consider the combination of influences from the spatial, sociodemographic and psychological dimension, as this may influence different groups of people differently (*3*). To investigate this aspect, the literature presents segmentation approaches that allow grouping people with similar characteristics. The classification of people into specific mobility types with distinct characteristics serves to understand travel behavior and its determinants. Anable (*3*), Hunecke et al. (*4*) and Collum and Daigle (*5*) demonstrated the application of a cluster analysis based on attitudes as a suitable approach to identify distinct mobility types. Prillwitz and Barr (*6*) performed an attitude-based cluster analysis besides a cluster analysis based on aspects of travel behavior to compare the results of both approaches. Von Behren et al. (*7*) tested a two-dimensional clustering approach in a first attempt to identify urban mobility types. Results showed both similarities and contrasts between the behavioral and attitudinal characteristics. This promising method was used in a specific application with a homogenous group of people regarding their mobility needs. Only two similar districts in German cities (Hamburg and Berlin) were investigated. A consideration of whole cities and a comparison of mobility types between different cities require an enhancement of the approach.

At the international level, segmentation is often used to classify cities. Spatial-structural characteristics serve as differentiating variables whereby similar structures of cities are identified. Wulfhorst et al. (*8*) studied different types of megacities with regard to mobility cultures. For this purpose, spatial structures such as land use and built environment, mode-related transport qualities as well as aspects of travel behavior were taken into account. They determined a cluster solution into which megacities worldwide can be grouped. The examined cities from different continents and thus different cultures show common characteristics and therefore influence residents' urban mobility similarly. Timmermans et al. (*9*) present a study that focused on the comparison of travel behavior in an international context. To investigate the influence of spatial structures on travel behavior, travel data from surveys in the USA, UK, Japan, Canada and the Netherlands were evaluated. The results indicate a slightly greater influence of psychological principles on activity patterns than the characteristics of the city. In their study, Timmermans et al. also point to a lack of international comparisons of travel patterns in the literature and suggest further research.

Hence, our paper address two main aspects: First, by applying a cluster analysis with the simultaneous consideration of attitudinal and behavioral elements, we identify more information on the factors influencing travel behavior. Second, the analysis of people from Berlin, Shanghai, and San Francisco aims to contribute to international comparisons of people and their travel behavior. This is of particular interest to understand the influence of urban structures in different cultural environments.

DATA COLLECTION AND SURVEY DESIGN

Our analysis is based on a unique data collection approach, especially in terms of capturing comprehensive information about many travel related aspects. These aspects consist of daily and occasional travel behavior (including longitudinal aspects such as variability, multimodality, and long-distance traveling) and attitudes towards different modes (including social and individual norms). After explaining the data collection, we give more information on how we captured psychological factors and travel behavior using a travel skeleton.

Data collection

The research presented in this paper is based on data collected through three similar surveys, conducted in Germany (Berlin), China (Shanghai) and the U.S. (San Francisco) between October-2016 and January-2017. The three surveyed cities are well-developed and offer good public transport systems. Each city has specific innovative transport services such as ODM (e.g., Uber, Didi or DriveNow). Berlin and San Francisco are “hybrid cities”, which exhibit dense public-transit-oriented urban cores, surrounded by low-density car-oriented suburban areas. Shanghai is considered more of a “non-motorized” city, with a high population density which supports the use of non-motorized transport (*10*). Furthermore, Shanghai has a comparably low car ownership rate resulting from restrictive transport policies. To generate comparable datasets from each city, we used a standardized survey approach, which has already been carried out in a previous study in Germany (*7*), based on a computer-assisted personal interview (CAPI). The total sample size was 1,800 individuals with 600 respondents from each city. We conducted quota sampling regarding age, gender, household size, and net income to develop a representative survey group for each captured city. A professional market research firm carried out the survey by using a slightly different recruitment in each city, based on local cultural norms. In all cities, an access-panel with telephone screening was used. On-street recruitment was applied in Berlin and San Francisco. The survey aimed to capture behavior and psychological factors for individuals above the age of 17 and, as far as possible, for the whole household.

Concept of a travel skeleton

Travel behavior is highly variable, affects many aspects of life, and cannot be measured intra-individually by considering only short periods, and as such, a collection of longitudinal data is required. However, common surveys, which are based on trip diaries, are expensive and increase the respondent burden of the participants. This high respondent burden also limits the inclusion of additional supplementary questions. To create a cost-effective survey alternative, we developed a "travel skeleton", which focuses on typical elements of everyday travel as well as long-distance travel. The skeleton provides a reasonable compromise between the level of detail needed and the required effort to survey travel behavior. The idea of using a skeleton to identify routines and typical behavior is common in travel behavior research, an overview is given by von Behren et al. (*7*). "Typical" behavior refers in our research to the frequent, daily repetition of activities across many weeks in different areas of life. Similar to trip diary surveys, in this survey, the respondents had to report their behavior in a typical week in order to capture their usual mobility pattern and its determinants (e.g., chauffeuring of children). Thus, the skeleton approach reduces the impact of intrapersonal variance and has the advantage of requiring a smaller sample size to achieve similar research outcomes.

Psychological factors

Significant elements of travel behavior cannot be explained using the "objective" dimension only. We assume, based on existing research, that knowing more about people's attitudes towards different transport modes helps us to understand their travel behavior. To survey this psychological dimension, we used a standardized item set that is based on a Likert scale. This item set was developed by Hunecke (*4*) and has been applied in previous studies (*7; 11*). Table 1 shows the used psychological questions.

Table 1 Standardized psychological item set

Category / variable	***Items***	***Questions***
Public transportation privacy (PrivacyPT)	1	In public transportation people sometimes come too close to me in an unpleasant manner.
	2	In public transportation my privacy is restricted in an unpleasant manner.
Public transportation autonomy (AutonomyPT)	1	I can structure my everyday life very well without a car.
	2	I can take care of what I want to with public transportation.
	3	It is difficult for me to travel the ways I need to go in everyday life with public transportation instead of by car.
	4	If I want, it is easy for me to use public transportation instead of a car to do my things in everyday life.
Public transportation experience (ExperiencePT)	1	I appreciate public transportation, because there is usually something interesting to see there.
	2	I can easily use the traveling time on the bus or train for other things.
	3	I like to ride buses and trains, because I don't have to concentrate on traffic while doing so.
	4	I can relax well in public transportation.
Public transportation intention (IntentionPT)	1	It is my intention to use public transportation instead of a car for the things I do in everyday life.
	2	I have resolved to travel the ways I need to go in everyday life using buses and trains.

Subjective norm (SubjectiveNorm)	1	People who are important to me think it is good if I would use public transportation instead of a car for things I do in everyday life.
	2	People who are important to me think that I should use public transportation instead of a car.
Personal norm (PersonalNorm)	1	Due to my principles, I feel personally obligated to use eco-friendly means of transportation for the things I do in everyday life.
	2	I feel obligated to make a contribution to climate protection via my choice of transportation.
Car orientation (CarOrientation)	1	When I sit in the car I feel safe and protected.
	2	Driving a car means freedom to me.
Perceived mobility necessities (ForcedMobility)	1	My everyday organization requires a high degree of mobility.
	2	I constantly have to be mobile in order to comply with my everyday obligations.
Bicycle orientation (BicycleOrientation)	1	I like to be out and about by bike.
	2	I can relax well when riding a bike.
	3	I ride a bicycle because I enjoy the exercise.
Weather resistance (WeatherResistance)	1	I don't like to ride my bike when the weather is cool.
	2	I also ride my bike when the weather is bad.

Data Preparation

In most statistical analyses, observations with missing data have to be excluded. As we had to handle various missing data in the attitudinal item set, we chose to run an imputation process. Aware of the disadvantages and negative influences imputation may show on the data, it helped us to include 1,662 respondents instead of 1,213 people for further analyses. We tested two different imputation methods to minimize the resulting bias. First, we ran a *Multiple Imputation* (MI) with logistic regression as proposed by Rodriguez de Gil and Kromrey (*12*). Second, we conducted the package *missForest* by Stekhoven and Bühlmann (*13*) that relies on machine learning algorithms. To feed the imputation methods, not only the item responses were included but also information on sociodemographic characteristics such as age and gender as well as the share of car use and public transportation use of the individual. The solution with the least imputation error was the imputed data with *missForest*, measured by the index 'proportion falsely classified' (PFC). Furthermore, existing literature recommends this imputation method because of lower out-of-bag error estimates by mixed-type data in contrast to multivariate imputation methods (*13; 14*). To reduce the error of the imputed data, we excluded individuals with 14 or more missing values in the item set. Additionally, items with many missing values (>16% of all respondents) were not included for further analyses. In order to obtain representative results, we introduced a city-specific weighting to the surveyed data (based on spatial type, household size, age and gender).

DETERMINATION OF CLUSTER-FORMING VARIABLES

In cluster analysis, attitudes and norms are possible input variables to identify mobility types (see e.g., *3; 5*). However, including travel behavior aspects may further enhance the analysis. The combined approach leads to clusters that show specific characteristics in a behavioral and psychological dimension at the same time (*7*). Since the skeleton approach provides information on attitudes and behavior, we defined cluster-forming variables in both dimensions (see following two sections).

Attitudes towards modes

For the consideration of the psychological dimension in our analyses, a set of attitudinal items on a 5-point-Likert scale was available (see Table 1). We performed a common technique to densify the information of the surveyed items: a *Principal Component Analysis* (PCA) (see *4; 6; 15*).

Out of the 25 selected items, we obtained six consolidated components using *Kaiser's Criterion*, which requires an eigenvalue above one (*4; 6*). We also calculated *Cronbach's Alpha* which is often used as a criterion for the extraction of components (*5; 15*) and requires at least a value above 0.65 (*3*). Table 2 shows the result of the PCA. The highest loadings of each item on a component are indicated. The first component *PT Orientation* includes only items regarding public transportation. The second component *Bicycle Excitement* combines the items on Bicycle Orientation (1-3) with the item WeatherResistance1. *Norm* includes all items of the Personal Norm and Subjective Norm. The component *Adaptability* comprises the items on privacy when using public transportation as well as one item of weather resistance. We interpret this component as an expression of comfort aspects. The last two components *Forced Mobility* and *Car Excitement* include the items describing mobility necessities and the attitudes towards cars. With this solution, we obtained components that represent attitudes on the main modes (public transportation, bicycle, car) as well as three additional key elements of behavioral psychology: *Norm*, *Forced Mobility*, and *Adaptability*.

PT Orientation, *Bicycle Excitement*, *Norm* and *Forced Mobility* have sufficient values for *Cronbach's Alpha*. Only *Adaptability* and *Car Excitement* barely miss this criterion. Since not only the quality measures but also the interpretation is essential, we decided to continue with all six components. Especially with *Car Excitement,* we assume an essential element in the characterization of the psychology in the context of investigating travel behavior. The quality of the PCA was confirmed by *Kaiser´s Measure of Sampling Adequacy* (MSA) and *Bartlett's Test of Sphericity* (*7; 16*).

Table 2 Results and criteria of the Principal Component Analysis (PCA)

Principal Component Analysis (PCA) - Varimax Rotated Factor Pattern

	Components					
	PT Orientation	*Bicycle Excitement*	*Norm*	*Adaptability*	*Forced Mobility*	*Car Excitement*
Cronbach's Alpha	α = *0.93*	α = *0.92*	α = *0.81*	α = *0.62*	α = *0.80*	α = *0.60*
Items in PCA						
AutonomyPT2	0.834					
IntentionPT2	0.823					
ExperiencePT1	0.809					
ExperiencePT4	0.808					
IntentionPT1	0.791					
ExperiencePT2	0.778					
ExperiencePT3	0.765					
AutonomyPT4	0.687					
AutonomyPT1	0.668					
AutonomyPT3	0.526					
BicycleOrientation1		0.900				
BicycleOrientation2		0.892				
BicycleOrientation3		0.890				
WeatherResistance2		0.778				
PersonalNorm1			0.835			
PersonalNorm2			0.824			
SubjectiveNorm2			0.618			
SubjectiveNorm1			0.548			
PrivacyPT2				0.853		
PrivacyPT1				0.807		
WeatherResistance1				0.421		
ForcedMobility2					0.884	
ForcedMobility1					0.874	
CarOrientation1						0.835
CarOrientation2						0.764

Printed is the maximum loading of each item

Criteria of quality for PCA

	Value	*Pr > Chi-Square*
Kaiser´s Measure of Sampling Adequacy (MSA)	0.904>0.9 (marvelous)	
Bartlett's test of Sphericity	χ^2 (185) = 2039.28	p***

Travel behavior

Four travel indicators were calculated to represent different aspects of travel behavior in our analysis. The selection of these indicators is adapted from previous research (*7*) and describes important aspects of travel behavior: activities, mode choice, trip volume and long-distance travel.

First, we calculated the average *Trips per Day* based on the given information about trips in a typical week. The second indicator is *Share of Car Usage* and includes all trips done by car (driver, passenger and on-demand services by car). This indicator represents the proportion of car usage of the individual modal split in a typical week and shows the importance of the car. This indicator has a range between 0 (no car usage) and 1 (car is the only used mode). The third indicator *Share of Mandatory Trips* describes the proportion of trips to work or school of all trips in a typical week. As the last indicator, we used the number of *Long-Distance Trips* (overnight stays and day trips with distances > 100 km) during one year. Since all four indicators showed a certain interpersonal variation across all participants, they were considered as cluster-forming variables for the following segmentation.

CLUSTER ANALYSIS AND EVALUATION

In the following sections, we present our clustering approach and the used methods. This includes a two-step cluster methodology as well as machine learning algorithms for cluster evaluation. In addition, we give further insights into the clustering formation by illustrating the importance of each cluster-forming variable.

Clustering approach

For identifying mobility types, we decided to include all the people from the three different cities in the cluster analysis together, which guarantees the comparability of the obtained clusters. This may lead to a non-optimal solution to describe the mobility types occurring in each specific city but allows us to identify people with the same characteristics in all three cities. The cluster analysis was undertaken with the simultaneous inclusion of psychological variables and calculated travel indicators as described in the previous section: six attitudinal and four behavioral variables. A robust two-step cluster methodology was performed (see *3; 15*). First, we used the *Ward Method* (hierarchical method) to identify the structure in the data by merging those two observations respectively clusters that produce the lowest increase in variance. Based on the Cubic Cluster Criterion (CCC) and Pseudo t^2 we obtained an 11-cluster solution by using the software SAS. This solution from the hierarchical clustering served as input for the second part of our segmentation: a k-means clustering approach. This method, which is often used in segmentation (*3; 5; 17*), helps to stabilize the allocation of observations to a given number of clusters. Since the k-means algorithm allows a modified allocation or exchange of observations between the clusters, the overall solution is optimized. Because of performing the clustering procedure with different settings, we obtained two suitable solutions with 11 clusters. The different allocation of observations results in different cluster centers.

Evaluating cluster solutions

To decide which cluster solution is appropriate, we evaluated both solutions by using the machine learning algorithm *Random Forest* (*18*). This algorithm is based on decision trees of various subsamples of the dataset. With the help of this classification tree algorithm, we tested the allocation of people to clusters. The evaluation of the clustering solution can be seen as a supervised learning process (*19*), as the *Random Forest* tries to learn what the reasons for allocation to certain clusters are. The algorithm divides the data into a training and test subset. With the training subset, the algorithm learns the influence (independent variables) and tries to predict the cluster allocation (dependent variable) in the test subset. If the predictive accuracy is good than the results are reliable.

To compare both solutions, we performed in both cases a *Random Forest* with 2,000 trees with the total sample size of 1,662 respondents. We used the cluster themselves as a dependent variable (*19*). As independent variables, we used our 10 cluster-forming variables from the clustering process. To evaluate how good the allocation works, we look at the predictive accuracy of the *Random Forest*: Out-of-Bag (OOB) prediction error and the confusion matrix. The OOB prediction error of solution 1 is lower (9.09%) than the error of solution 2 (11.25%). The confusion matrix, which illustrates the comparison of true cluster allocation with predicted cluster allocation, shows also better results for cluster solution 1. Based on these results, we decided to take cluster solution 1 as our final solution.

Analyzing clustering process

Applying a *Random Forest* for evaluating cluster solutions also enabled us to examine the clustering process in detail. By using the *Variable Importance Measure* (VIM), we identified the importance of the cluster-forming variables for the segmentation. Figure 1 shows the importance scores of the different cluster-forming variables. The score of each variables should be used only for comparison between variables.

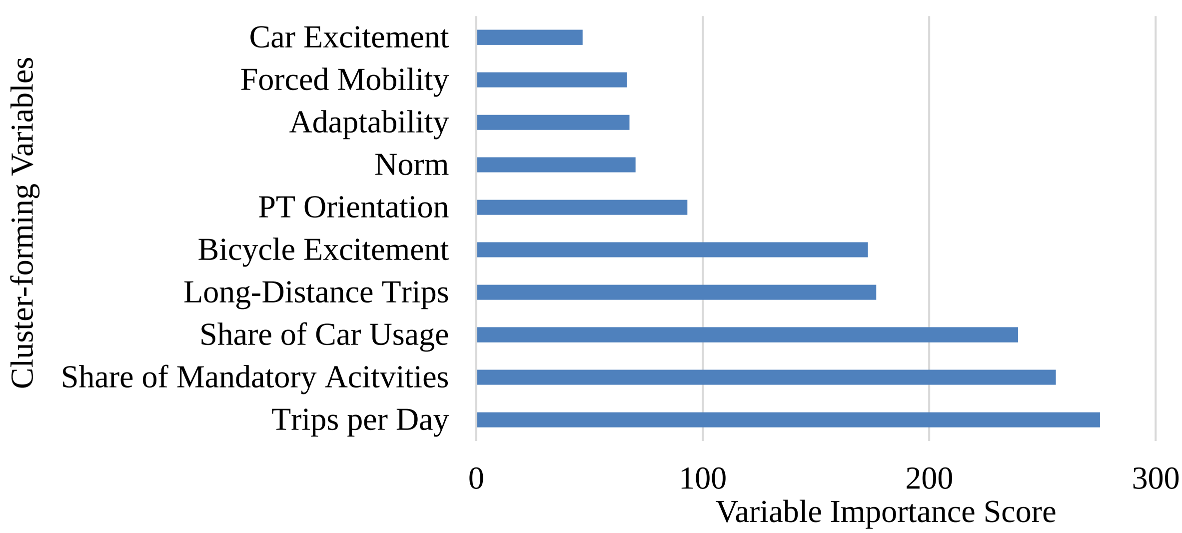

Figure 1 Variable Importance Measure of Random Forest

By looking at the VIM, we see the highest importance of the travel behavior variable *Trips per Day*. The lowest importance score for allocating people to clusters has the *Car Excitement*. In total, results show a higher average score of travel behavior variables in comparison with psychological variables. However, we used in the clustering process more psychological variables based on the PCA than behavior variables. Therefore, the impact of psychological variables is divided into more variables with lower importance. The results of the *Variable Importance Measure* confirm the idea of using attitudes and behavior for clustering at the same time because we can see a relatively important influence of psychological variables on the cluster allocation.

In addition to the VIM, we used *Partial Dependency Plots* (PDPs) of the *Random Forest* to illustrate the probability of cluster allocation depending on the cluster-forming variables. We can use it further as a supporting tool for the cluster interpretation (see Cluster description). Figure 2 and 3 visualize the PDPs of the cluster-forming variables. By looking on *Trips per Day*, we see a high probability for people with more than 6 trips per day to belong to Cluster 3. People with a trip rate between 5 and 6 are more likely to be in Cluster 9. The *Share of Mandatory Activities* over 0.4 increase the probability for Cluster 4. The plot of *Share of Car Usage* illustrates that monomodal car user a more likely to belong to Cluster 8. Considering Long-Distance Trips, we see a very high probability for Cluster 2, when people have more long-distance travels than 6.

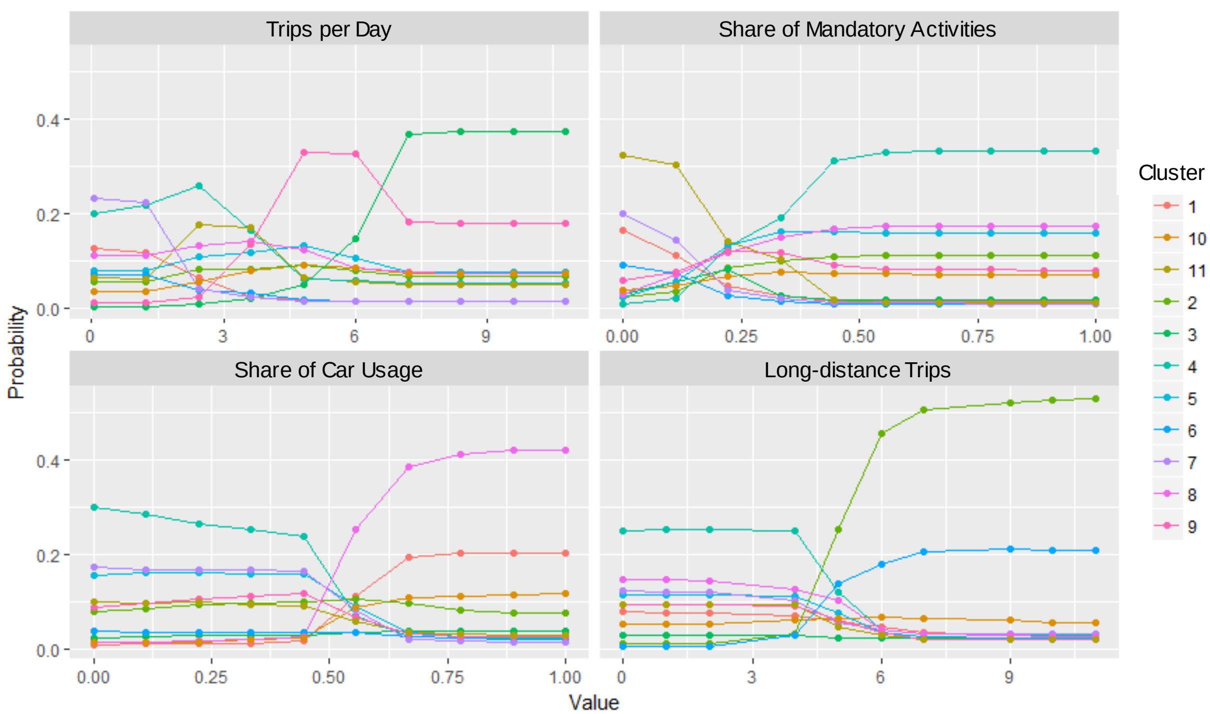

Figure 2 Partial Dependency Plots of Random Forest for Travel Behavior Indicators

By analyzing the psychological components of the clustering, we see a relevant effect of *Bicycle Excitement* on the cluster allocation (see Figure 3). People with a positive *Bicycle Excitement* are more likely to be in Cluster 4 and people with a low excitement a more likely to be in Cluster 5. A positive *PT Orientation* increases the probability for Cluster 5. Variables with lower importance are also relevant for the cluster allocation, as we can see by looking at *Car Excitement* in Figure 3. With an increase of the excitement, it gets more likely to belong to Cluster 9. Figure 3 shows a high probability for Cluster 4 (~20%). As we saw in Figure 1, the clustering process is more dominated by behavior than by attitudes. However, results of the PDPs show also an important influence of attitudes on the allocation to several clusters (e.g., Cluster 5).

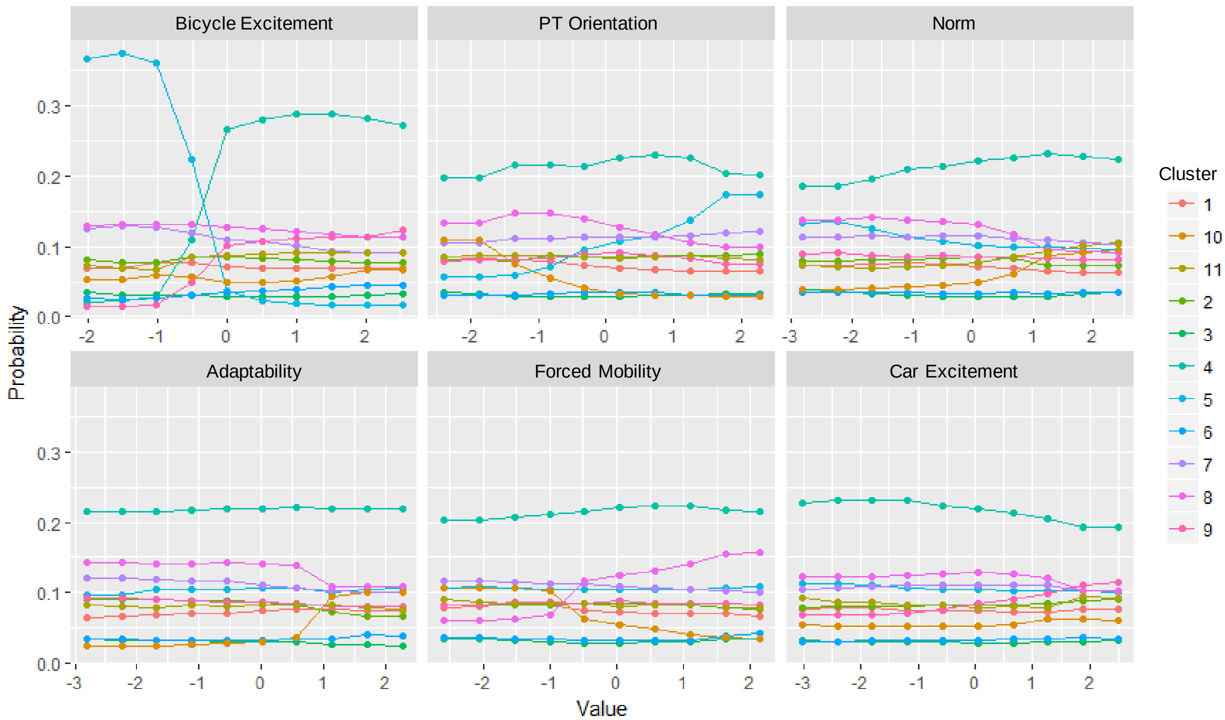

Figure 3 Partial Dependency Plots of Random Forest for Psychological Components

RESULTS

Based on the performed cluster analysis and the evaluation of the cluster solutions, we identified 11 clusters. These clusters represent distinct mobility types to which people from all three cities are assigned (1,662 observations in total). Besides differences between the clusters regarding the cluster-forming variables, we expect further variations in the sociodemographic attributes as well as in travel behavior. In the following, we illustrate the differences and similarities between the obtained mobility types. Therefore, we analyze and discuss the characteristics of the sociodemographic, psychological and behavioral dimension to differentiate the distinct clusters. In addition, an analysis of the international aspect of our study is carried out.

Cluster description

The evaluation and interpretation of the cluster characteristics are essential for the application of a cluster analysis. Each cluster shows unique attributes that differ from the others. Because of the combined consideration of attitudes and behavior in the analysis, it is possible to identify if both dimensions coincide or contradict each other within a mobility type. Table 3 provides an overview of the mean characteristics of our obtained clusters. We see different cluster sizes, which vary from 49 to 365 observations. The variables to describe the clusters are separated into three sections. The first section shows the cluster-forming variables. The four indicators of the travel behavior and the six components representing the attitudes are given. The six components are standardized to allow direct comparisons in the attitudinal dimension. The four behavioral indicators are shown in their original scale. Sections 2 includes more details on the sociodemographic characteristics of the mobility types. Beside to information on a personal level such as age and gender, we also evaluated details on the household level. These include, among others, the number of cars in the household and the income class, which is differentiated into five categories. The third section consists of variables regarding travel behavior. We evaluated the mean km per day and the share of the usage of public transportation as well as the share of walking and cycling.

All obtained clusters represent mobility clusters with distinct characteristics. For example, Cluster 2 and Cluster 6 both show a high number of *Long-Distance Trips*. This goes along with the *Partial Dependency Plots* of *Long-Distance Trips* (see Figure 2). However, these two clusters differ regarding their *Bicycle Excitement*. Cluster 2 shows neutral attitudes towards bicycles. Cluster 6, on the other hand, has a high positive attitude. We also see differences in the sociodemographic characteristics: More than 80% of the allocated people in Cluster 2 are employed part- or full-time. In Cluster 6, this proportion is only 34%. To support the interpretation of all clusters, we provide a short description in Table 4. We also named the distinct mobility types. These names should be seen as a suggested term and do not represent a fixed definition.

As we included both psychological and behavioral variables, the evaluation of these dimensions within the mobility types were of particular interest to us. CL 8 "Car Users with the Need to Be Mobile" shows a high *Share of Car Usage* (0.855) and a relatively high *Car Excitement* (0.273). Also, the attitudes towards public transportation and towards bicycles are negative. As a result, the psychology and behavior of the people in this cluster match. The opposite applies to CL 10 "Car Enthusiasts with High Norms". This mobility type has the highest value for the *Norm* (0.903), which in general implies an eco-friendly behavior. However, the highest value for *Car Excitement* (0.437) can also be found in CL 10. A high *Share of Car Usage* (0.820) reflects this. Due to the high value for *Norm*, one would rather expect the people to be more public transportation or bicycle oriented. Looking at the other characteristics of this cluster, we see high income and a relatively high number of cars in the household. Additionally, people allocated to CL 10 tend to live in multi-person households. The high average *km per Day* indicate relatively long distances in daily travel. This aspect may be an indication of the high car orientation because the car provides flexibility in their everyday life. We see a lack of realization of the perceived norms. Reasons for the discrepancies of the high *Norm* and the high *Share of Car Usage* may be external influences, such as a gap in public transportation supply or obligations within the household that force the use of the car or social desirability bias.

1 **Table 3 Cluster characteristics**

	Cluster	1	2	3	4	5	6	7	8	9	10	11
	Number of Observations	124	137	49	365	172	56	183	210	141	91	134
	in %	7%	8%	3%	22%	10%	3%	11%	13%	8%	5%	8%
Cluster-forming variables	Long-Distance Trips	0.914	8.043	1.496	1.418	1.555	7.271	1.010	1.282	1.971	2.570	1.594
	Share of Car Usage	0.859	0.362	0.483	0.114	0.171	0.281	0.190	0.855	0.191	0.820	0.136
	Share of Mandatory Acitvities	0.004	0.454	0.153	0.578	0.453	0.017	0.001	0.518	0.279	0.432	0.062
	Trips per Day	1.119	3.083	7.371	2.514	3.129	2.403	1.136	2.662	4.557	3.380	2.796
	PT Orientation	-0.982	0.084	-0.185	0.475	0.892	-0.060	0.252	-0.738	0.220	-1.498	0.340
	Bicycle Excitement	-0.544	0.079	-0.095	0.723	-1.120	0.477	-0.452	-0.362	0.796	-0.072	0.078
	Norm	-0.738	0.156	-0.448	0.159	-0.322	-0.038	-0.047	-0.359	-0.026	0.903	0.483
	Adaptability	0.575	-0.274	0.000	0.050	-0.036	0.082	-0.175	-0.009	-0.298	1.275	-0.049
	Forced Mobility	-0.757	0.157	0.294	0.020	0.152	-0.691	-0.780	0.732	0.159	-0.543	-0.209
	Car Excitement	-0.365	0.062	0.289	-0.083	-0.451	0.007	-0.085	0.273	0.275	0.437	-0.047
Sociodemographic Characteristics	*Individual Level*											
	Driving License (yes in %)	42.05	68.17	92.09	56.84	65.50	60.12	48.54	95.28	56.54	84.71	51.60
	Gender (male in %)	23.90	53.40	18.47	54.70	33.51	34.87	35.25	54.02	48.48	48.10	40.35
	Age	66.22	41.82	41.85	33.57	36.42	53.40	58.07	42.99	37.48	33.93	53.34
	Employed Full + Part-time (in %)	7.98	83.51	66.13	65.17	60.96	34.18	16.75	84.70	51.60	61.82	18.58
	College Student (in %)	0.00	3.58	9.67	25.13	22.43	0.46	0.29	4.23	19.52	18.05	8.74
	Retired (in %)	73.35	8.05	17.89	0.14	4.97	49.23	61.66	6.90	16.60	6.12	56.38
	Household Level											
	Income class (1=low to 5=high)	2.92	2.45	2.17	2.07	2.21	2.14	2.08	2.96	2.14	3.49	1.83
	Number of Household Members	3.28	2.16	1.72	2.12	2.28	1.97	2.31	2.64	2.41	3.50	2.25
	Number of Children	0.20	0.27	0.36	0.22	0.30	0.14	0.21	0.34	0.46	0.31	0.13
	Number of Cars	0.96	0.76	0.80	0.46	0.61	0.43	0.45	1.46	0.55	1.56	0.38
Travel Behavior	km per Day	10.20	23.10	72.05	21.68	19.47	13.98	4.74	32.19	25.37	49.29	10.11
	Share of PT Usage	0.04	0.22	0.18	0.36	0.49	0.16	0.33	0.05	0.29	0.01	0.30
	Share of Walking and Cycling	0.10	0.38	0.33	0.49	0.30	0.54	0.46	0.09	0.51	0.17	0.55

Given is the weighted mean value or proportion of the characteristics within the clusters

Table 4 Cluster profiles

CL	Cluster Name	%	Cluster Description
1	Low-Mobile Car Users	7.46	■ Low daily trip rates with low kilometers per day ■ Lowest share of long-distance travel ■ Negative attitudes to all means of transport
2	Multimodals with Affinity to Long-Distance Travel	8.24	■ Highest number of long-distance trips ■ Mostly employed people ■ Usage of all means of transport
3	Car-Affine High Mobiles	2.95	■ Highest daily trip rates and most km per day ■ High perceived mobility necessities and high car orientation ■ Mostly women, high rate of children in household
4	Convinced Bicycle and Public Transportation Users	21.96	■ Largest cluster ■ Highest share of mandatory activities ■ High affinity towards bicycle and public transportation
5	Public Transportation Enthusiasts	10.35	■ Highest affinity to public transportation, highest share of public transportation usage ■ Lowest affinity towards bicycles and cars ■ Mainly employed and students
6	Multi-Locals without Obligations	3.37	■ Few mandatory activities and a low forced mobility ■ High number of long-distance trips ■ High share of trips by foot and by bicycle
7	Low-Mobiles within Small Distances	11.01	■ Lowest km per Day, low number of trips per day ■ Mainly retired people ■ Mostly negative attitudes to the means of transport
8	Car Users with the Need to Be Mobile	12.64	■ High share of car usage and high share of mandatory activities ■ Highest value for forced mobility ■ Mostly men, highest employment rate
9	Open-Minded Multimodals	8.48	■ High number of trips per day ■ Highest bicycle affinity ■ High rate of children in household
10	Car Enthusiasts with High Norms	5.48	■ Highest affinity to cars, lowest to public transportation ■ High ecological norm at the same time ■ Household with high income, multi-person households
11	Non-Motorists within a Close Range	8.06	■ Lowest share of car usage, lowest motorization rate ■ High value for the norm ■ Highest share of walking and cycling

International Comparison

On the basis of the special dataset, we were able to include people from Berlin, Shanghai and San Francisco in the clustering. Through the simultaneous inclusion of observations from all three cities, the cluster solution gives us an insight into the mobility types occurring in urban environments worldwide. However, it is also of particular interest to investigate if people only from one of the cities primarily characterize a cluster. This would allow us to identify a link between city-specific characteristics and mobility types. Spatial, infrastructural and cultural aspects can serve this purpose. For each mobility type, we examined the proportion from Shanghai, San Francisco, and Berlin (see Figure 4). A comparison of the proportions "Within the Clusters" highlighted city-specific clusters: In CL 1 "Low-Mobile Car Users", 82.3% of the allocated people are from San Francisco. In CL 10 "Car-Enthusiasts with High Norms", the value is even 96.7%. Both clusters have a high *Share of Car Usage* in common. CL 10 is of particular interest as mentioned above: People of this cluster have the highest orientation towards the car (*Car Excitement*) and at the same time the highest value on *Norm*. This contrast confirms the high standing of the car as a mode in the USA. In CL 1, we see a negative attitude towards the car. Since the *Share of Car Usage* is still high, we conclude a high impact of city-specific conditions on the car usage. The characteristics of CL 1, as well as CL 10, may reflect car-friendly urban planning and poor or unattractive transport alternatives. To make reliable statements about this, a more detailed examination the spatial structure and the topography would be useful for more detailed analyses.

Figure 4. Representation of Each City Within the Cluster

Only CL 1 and CL 10 can be directly assigned to a city (i.e., San Francisco). No other cluster is formed almost exclusively by people from one city only. However, an unequal number of people of the three cities should be noted (Shanghai = 502, San Francisco = 570, Berlin = 590), which may lead to a slight bias in this interpretation. For this reason, we also examined to what extent the 11 identified mobility types occur within each of the three cities (see Figure 5). According to the results above, we find only a low occurrence of CL 1 and CL 10 in Berlin and Shanghai. We even see further similarities between these two cities: CL 2, 4, 5, 6, 9 and CL 11 occur to almost the same extent in Berlin and Shanghai. For example, 11.4% of respondents from Shanghai and 11.0% of respondents from Berlin are assigned to CL 9 "Open-minded Multimodals". All clusters mentioned above characterize mobility types in Berlin and Shanghai with multimodal travel behavior or with the affinity to the use of public transportation or non-motorized means of transport. This indicates a comparable supply of alternatives to cars as means of transport in both cities. In Berlin and Shanghai, the rail-bound public transportation is well developed and additionally the motorization rate is low, compared to other cities in the respective countries. By segmenting all three cities together, we see commonalities in the travel behavior and the psychological characteristics of the people in the mobility types. Furthermore, the cultural and spatial differences are partly manifested in the formation of the obtained clusters. The results of our clustering should be interpreted with consideration of city-specific differences.

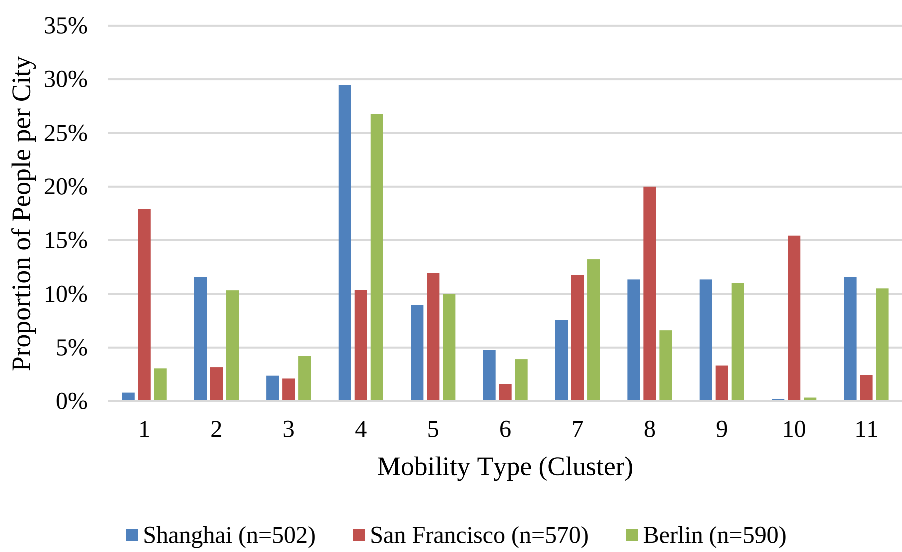

Figure 5. Visualization of the Cluster Distribution for Each City

CONCLUSIONS

In our paper, we segmented people to urban mobility types, analyzed and compared the received mobility types by using data from an international survey. The data is based on a travel skeleton approach that provides a reasonable compromise between the level of detail needed and the required effort to survey travel behavior and psychological factors at the same time. To improve the clustering, we extended the process by data preparation and evaluation of potential clustering solution. Therefore, we combine classic clustering approaches with suitable machine learning algorithms (i.e., *Random Forest*). The evaluation provides further insights by identifying the importance of variables on the cluster formation. In addition, we were able to investigate the influence of variables on the allocation of people to clusters with the help of *Partial Dependency Plots.*

Based on the final 11-cluster solution, we analyzed clusters regarding their specifics. Therefore, we considered cluster-forming variables as well as cluster-describing variables (e.g., sociodemographic characteristics). Some clusters show a dissonance between attitudes and behavior (e.g., Cluster 10 "Car-Enthusiasts with high Norms"). We also obtained clusters with conformity between both dimensions (e.g., Cluster 8 "Car Users with the Need to Be Mobile"). The largest cluster is Cluster 4 "Convinced Bicycle and Public Transportation Users" and is dominated by people from Berlin and Shanghai. For the consideration of the international setting, we additionally analyzed the distribution of the clusters in the three cities. On the one hand, the results show the occurrence of certain mobility types in all three cities. Hence, those clusters represent urban mobility types of the same kind in different cultural settings regarding travel behavior and attitudes. Especially between Berlin and Shanghai, we see parallels in the distribution of mobility types. Multimodality, as well as the affinity to use public transportation and non-motorized means of transport, are the common attributes. On the other hand, some clusters represent city-specific characteristics. "Low-Mobile Car Users" (CL 1) and "Car-Enthusiasts with High Norms" (CL 10), which both show a high level of car usage, are almost exclusively represented by people from San Francisco. These findings offer the linkage between a mobility type and its characteristics with the distinguishing spatial and infrastructural as well as the cultural framework of each city. Our analysis shows that the data from a survey approach, which combines a survey on typical travel behavior with attitudinal questions, is qualified to find and to analyze mobility types in an intercultural setting. Questioning participants in each of the cities of Berlin, San Francisco, and Shanghai allowed for the comparison of people's travel behavior and the determinants in the three different cultures.

Among other things, the results improve our understanding of how people use and evaluate different transport systems. The distinct characteristics of the clusters allow us to investigate which types might show an open mind, for example, towards ODM and thus gain a brief overview of the potential markets. Another application could be targeted policies and mobility offers as cluster-specific mobility solutions to increase the acceptance of the people. For example, people from CL 10 "Car Enthusiasts with High Norms" may be particularly interested in environmentally friendly technologies such as battery electric vehicles.

Overall, the application of an integrated clustering approach appeared to be a suitable method to define distinct mobility types in such an international and intercultural setting. In our study, we did not include spatial structures because the complexity of the clustering process would increase with an extra dimension. However, we still see spatial differences in our obtained urban mobility types by comparing Berlin and Shanghai with San Francisco. Further research could implement more detailed analyses on spatial structures. It would be also of interest to integrate cities with significantly different characteristics regarding the quality of transport systems.

ACKNOWLEDGMENTS

This paper presents analyses based on the Urban Travel Monitor (UTM) funded by the BMW Group, Munich. The authors confirm contribution to the paper as follows: study conception and design: Sascha von Behren, Bastian Chlond, Miriam Magdolen; data collection: Sascha von Behren; analysis of results: Miriam Magdolen, Sascha von Behren; interpretation of results: Miriam Magdolen, Sascha von Behren, Bastian Chlond, Marcel Hunecke, Peter Vortisch; draft manuscript preparation: Miriam Magdolen, Sascha von Behren. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Hunecke, M., S. Haustein, S. Grischkat, and S. Böhler. *Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. Journal of Environmental Psychology*, Vol. 27, 2007, pp. 277–292, http://dx.doi.org/10.1016/j.jenvp.2007.08.001.
2. Ajzen, I. *The theory of planned behavior. Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, 1991, pp. 179–211, http://dx.doi.org/10.1016/0749-5978(91)90020-T.
3. Anable, J. *'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory. Transport Policy*, Vol. 12, 2005, pp. 65–78, http://dx.doi.org/10.1016/j.tranpol.2004.11.004.
4. Hunecke, M., S. Haustein, S. Böhler, and S. Grischkat. *Attitude-Based Target Groups to Reduce the Ecological Impact of Daily Mobility Behavior. Environment and Behavior*, Vol. 42, No. 1, 2010, pp. 3–43, http://dx.doi.org/10.1177/0013916508319587.
5. Collum, K. K., and J. J. Daigle. *Combining attitude theory and segmentation analysis to understand travel mode choice at a national park. Journal of Outdoor Recreation and Tourism*, Vol. 9, 2015, pp. 17–25, http://dx.doi.org/10.1016/j.jort.2015.03.003.
6. Prillwitz, J., and S. Barr. *Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour. Journal of Transport Geography*, Vol. 19, 2011, pp. 1590–1600, http://dx.doi.org/10.1016/j.jtrangeo.2011.06.011.
7. von Behren, S., C. Minster, M. Magdolen, B. Chlond, M. Hunecke, and P. Vortisch. *Bringing travel behavior and attitudes together: An integrated survey approach for clustering urban mobility types. TRB 97th Annual Meeting Compendium of Papers,* Washington, D.C., 2018.
8. Wulfhorst, G., J. Kenworthy, S. Kesselring, T. Kuhnimhof, M. Lanzendorf, and R. Priester. *Mobility Cultures in Megacities: Results from a Global Study. 13th World Conference on Transport Research,* Rio de Janeiro, Brazil, 15-18 July 2013, 2013.

9. Timmermans, H., P. van der Waerden, M. Alves, J. Polak, S. Ellis, A. S. Harvey, S. Kurose, and R. Zandee. *Spatial context and the complexity of daily travel patterns: an international comparison. Journal of Transport Geography*, Vol. 11, No. 1, 2003, pp. 37–46, http://dx.doi.org/10.1016/S0966-6923(02)00050-9.
10. Institute for Mobility Research (ifmo) (Hrsg.), ed. *Megacity Mobility Culture - How Cities Move on in a Diverse World.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, http://dx.doi.org/10.1007/978-3-642-34735-1.
11. von Behren, S., C. Minster, J. Esch, M. Hunecke, P. Vortisch, and B. Chlond. *Assessing car dependence: Development of a comprehensive survey approach based on the concept of a travel skeleton. Transportation Research Procedia*, Vol. 32, 2018, pp. 607–616, http://dx.doi.org/10.1016/j.trpro.2018.10.015.
12. Rodriguez de Gil, P., and J. D. Kromrey. *What Score Should Johnny Get? Missing_Items SAS Macro for Analyzing Missing Item Responses on Summative Scales. Proceedings of the SAS® Global Forum 2013 Conference*, 2013.
13. Stekhoven, D. J., and P. Bühlmann. *MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics (Oxford, England)*, Vol. 28, No. 1, 2012, pp. 112–118, http://dx.doi.org/10.1093/bioinformatics/btr597.
14. Doove, L. L., S. van Buuren, and E. Dusseldorp. *Recursive partitioning for missing data imputation in the presence of interaction effects. Computational Statistics & Data Analysis*, Vol. 72, 2014, pp. 92–104, http://dx.doi.org/10.1016/J.CSDA.2013.10.025.
15. Rode, P., C. Hoffmann, J. Kandt, D. Smith, and A. Graff. *Towards New Urban Mobility: The case of London and Berlin*, LSE Cities/InnoZ. London School of Economics and Political Science.
16. Bühner, M. *Einführung in die Test- und Fragebogenkonstruktion.* Pearson Studium, München, 2006.
17. Haustein, S., and T. A. S. Nielsen. *European mobility cultures: A survey-based cluster analysis across 28 European countries. Journal of Transport Geography*, Vol. 54, 2016, pp. 173–180, http://dx.doi.org/10.1016/j.jtrangeo.2016.05.014.
18. Breiman, L. *Random Forests. Machine Learning*, Vol. 45, No. 1, 2001, pp. 5–32, http://dx.doi.org/10.1023/A:1010933404324.
19. Cardoso, M., I. Themido, and F. Pires. *Evaluating a clustering solution: An application in the tourism market. Intelligent Data Analysis*, Vol. 3, No. 6, 1999, pp. 491–510, http://dx.doi.org/10.1016/S1088-467X(99)00035-9.