Predicting car use with Big Data

Alin Semenescu[a], Balazs Feher-Gavra[a], Denis Coca[b], Alin Gavreliuc[b] and Corina Ilin[b]

[a]Institutul pentru Cercetări Avansate de Mediu (ICAM), Universitatea de Vest din Timișoara

[b]Departamentul de Psihologie, Universitatea de Vest din Timișoara

Abstract

The present longitudinal study investigated participants’ car use behavior for a week, by using objective mobility data collected with Google Location through participants’ smartphones. 48 residents of a large Romanian town provided their mobility data, together with their answers on a questionnaire. Several psychological, structural and contextual predictors of car use were investigated in a multilevel model with two levels. Present findings indicate that instrumental motivations such as distance needed to travel, the need to pick-up/drop-off children to/from school and the need to carry heavy objects have the largest influence on car use behavior. Additionally, access to alternative transport infrastructure and the strength of people’s car use habits also significantly influence the number of kilometers travelled by car. Psychological variables such as those described by the theory of planned behavior or norm activation model, together with weather variables such as wind speed, temperature and precipitations, did not directly influence car use behavior. Results contribute to a more comprehensive understanding of car use behavior and are relevant for policymakers and city planners in their attempts to shape a sustainable urban environment that is adapted to one of the greatest challenges of the present.

Keywords: car use, structural predictors, psychological predictors, contextual predictors, objective measurements

1. Introduction

With few exceptions (e.g. Ma, Mulley, & Liu, 2017; Rodriguez & Rogers, 2014), the overwhelming majority of studies in transportation research used self-reported measures of car use. Respondents are generally asked how often they use their car for transportation purposes or asked to keep a daily mobility diary, in which they note all travels during a week. Even though this is a general practice, to my knowledge, there are no published studies that confirm the validity of such data by comparing it with objectively measured transportation behavior. This represents an important limitation in the field, as it allows the possibility for bias, which can significantly influence research conclusions.

The present study addresses this important limitation by investigating car use behavior with the help of Google Location data, which represents an objective measure of car use. Even though this is an innovative method of measurement in transportation research, a previously conducted study showed that Google Location data is as valid as data provided by traditional GPS devices (see Ruktanonchai, Ruktanonchai, Floyd, & Tatem, 2018). Aside from accurately measuring movement, Google Location data also provides a large amount of information that can be utilized for predicting car use. For example, participants' residential address, their work location, whether they were shopping in a particular day or even if they have children or not, can be inferred from the data.

The present study addresses several limitations of the field. First, it quantifies *car use* behavior with the help of objective measurements. Second, it distinguishes between multiple levels of variability in car use behavior, namely between intra-individual variation (i.e. day-to-day variation) and inter-individual variation (i.e. from individual to individual), by structuring and analyzing the data in a multilevel format. Third, it includes a series of objectively measured predictors of car use, such as *distance*, *wind speed*, *day of the week*, *picking-up or dropping-off children* to school/kindergarten or *doing shopping* and measures their longitudinal, day-to-day influence, on car use behavior.

1.1. Theoretical background

1.1.1. Contextual factors

Contextual variables such as the weather can significantly impact mobility decisions. However, there are only few studies focused on investigating the way individual travel patterns are influenced by the variability in weather conditions (Liu, Susilo, & Karlström, 2015). Two of the most investigated weather aspects are *temperature* and the level of *precipitations*. Studying the impact of temperature, Bergström and Magnussen (2003) found that car trips were more frequent in winter compared to summer, while for bicycle trips it was the opposite. Research in the UK also showed that increases in temperature resulted in higher levels of cycling (Parkin, Wardman, & Page, 2008), while in Germany, Müller, Tscharaktschiew and Haase (2008) found that cycling in summers was three times higher compared to winter.

On the other hand, the level of precipitation is frequently mentioned as the most negative weather aspect (Brandenburg, Matzarakis, & Arnberger, 2004) and a reason not to cycle. However, research findings concerning the impact of precipitations on travel mode choice have revealed conflicting results. While some authors found an increase in precipitation determines a switch from active transportation to public transport and private cars (Sabir, Koetse, & Rietveld, 2008; Sabir, van Ommeren, Koetse, & Rietveld, 2010; Saneinejad, Kennedy, & Roorda, 2012) others found that car traffic is reduced with rainfall (Hassan & Barker, 1999; Keay & Simmonds, 2005). However, according to a recent review investigating the impact of weather on individual travel behavior, warm and dry weather positively impact active transport modes, while rain, snow, wind, cold and hot weather determine a switch from active to other modes which offer shelter from the elements, such as the car (Böcker et al., 2013).

Another contextual factor that may influence transportation mode is related to the *distance* between the place of origin and the destination. While certain individuals live in remote places from daily necessities to be able to travel in a non-motorized manner, others live so close that walking or cycling are viable travel alternatives. Distance can therefore directly influence the viability of different transportation modes as well as travel decisions (e.g. Ding, Wang, Liu, Zhang, & Yang, 2017; Schlossberg, Greene, Phillips, Johnson, & Parker, 2006).

Aside from such factors, research shows that *wind speed* can significantly impact travel decisions. For example Flynn, Dana, Sears and Aultman-Hall (2012) showed that in a sample of 163 American commuters, the likelihood of cycling decreased significantly with increased wind speeds, bicycle trips being replaced by other modes such as the car. A similar result was described by Heinen, Maat and van Wee (2011), who investigated travel behavior in a sample of 633 part-time cycling commuters and found that, among other factors, wind speed significantly decreased the likelihood of cycling.

The need to transport goods has been identified as another important factor that weights on people's decision about which travel mode to choose for a particular journey. When people do their *shopping* and need to carry heavy loads, they are less likely to choose non-motorized means of transportation such as the bicycle (Heinen, Maat, & van Wee, 2011; Heinen, Maat, & van Wee, 2013) and may prefer to use the car instead. A similar situation arises for parents who need to *pick-up or drop-off their children* to school or kindergarten. Research shows that a large proportion of trips to school are carried out by car (e.g. DiGuiseppi, Roberts, Li, & Allen, 1998; Owen et al., 2012). Additionally, longer distances from home to school determine parents to switch from active transportation in favor of car transport (DiGuiseppi et al., 1998; Ferdinand, Sen, Rahurkar, Engler, & Menachemi, 2012) and this is especially true when children are small and not allowed to travel independently (Westman, Friman, & Olsson, 2017).

Several studies have found that the *day of the week* has a specific influence on travel patterns. Even though research in this area is scarce, studies conducted in New York area and San Francisco Bay area, found that the number of person trips per household at the weekend was somewhat lower than during weekdays, yet the average distance travelled per trip was larger on weekends compared to weekdays. The net result was that kilometers travelled were about the same during weekdays and weekend (Bhat & Lockwood, 2004). Other research, conducted on 10650 residents in the Beijing area, found that holiday trips showed a stronger dependency on cars than weekday trips (Yang, Shen, & Li, 2016), resulting in more kilometers travelled.

1.1.2. Structural factors

Structural measures, colloquially known also as *hard* measures (Bamberg, et al., 2011), are focused on changing travel behavior by modifying the physical environment (e.g. improving infrastructure) or by changing legal or economic policies (e.g. forbidding cars in city centers, control of parking, higher taxes on fuel, etc.). Evaluation studies conducted to examine the behavioral responses to such measures provided evidence that structural measures were often effective in decreasing car use. For example, improvements in *travel infrastructure* such as upgrades in public transport service (e.g. Cairns, Sloman, Newson, Anable, Kirkbride, & Goodwin, 2004; Kristensen & Marshall, 1999) or building new infrastructure (e.g. Arentze, Borgers, Ponje, Stams, & Timmermans, 2001), have revealed positive effects in reducing car use as well as in promoting more sustainable travel alternatives (see Pucher, Dill, & Handy, 2010). For example, Dill and Carr (2003) found that for every additional mile of cycling lane per square mile there was an increase of about 1% in the proportion of commuters using the bicycle as transportation mode.

Even though the idea that an improved public transport infrastructure plays an important role in reducing car transport is widely accepted (Potoglou & Kanaroglou, 2008), a different viewpoint is advanced by scientists who found that PT access does not necessarily result in less car use. For example Chatman (2013) found that in New Jersey, easy rail access did not predict the amount of car use by nearby residents, while a research conducted in Beijing found no association between *proximity* to a metro station and work-related vehicle kilometers travelled (Li & Zhao, 2017). Such results are supported by Ewing and Cervero's (2010) meta-analysis, who found that the average elasticity of car use in relation to the distance to nearest transit stop was very small and practically non-significant.

1.1.3. Psychological factors

Structural measures are typically costly to implement and sometimes may even be politically unfeasible, as some measures (especially restrictive policies and disincentives) can be strongly opposed by the public. They have also been considered as insufficient for reducing car use (Stopher, 2004). Consequently, interest in *psychological* (or the so-called *"soft"*) measures has increased. Psychological measures are defined by Steg (2003, p. 190) as "strategies aimed at influencing people's perceptions, beliefs, attitudes, values, and norms", which are focused on changing travel behavior through voluntary, instead of coercive means. Allowing people the freedom to choose represents the primary reason why such measures are also better received by the public (Taylor, & Ampt, 2003).

Within the psychological paradigm, two theories were predominantly employed in transportation research to explain car use behavior, namely *the theory of planned behavior* (TPB;

Ajzen, 1991) and the *norm activation model* (NAM; Schwartz, 1977). TPB is a rational choice theory which stipulates that behavior is determined by behavioral *intentions* which, in turn, are influenced by *attitudes* (the degree of favorable or unfavorable rational evaluations of the behavior), *subjective norms* (perceived pressures from the social environment or significant others to behave in certain ways) and *perceived behavioral control* (the perceived easiness of performing the behavior). TPB model was useful in predicting travel mode choice in a considerable number of empirical studies (e.g. Bamberg, Ajzen, & Schmidt, 2003; Bamberg & Schmidt, 2003; Heath & Gifford, 2002; Klöckner & Matthies, 2009), while two meta-analyses found that all TPB constructs significantly correlated with car use behavior (see Gardner & Abraham, 2008; Lanzini & Khan, 2017).

NAM, on the other hand, assumes that individual behavior is shaped by moral considerations. The model assumes that behavior is directly predicted by *personal norms*, which are defined as moral standards that people hold for themselves. According to the model, personal norms are activated only if one is aware of the consequences of her/his behavior (*awareness of consequences*) and feels a sense of responsibility for such consequences (*ascription of responsibility*). In transportation research, NAM was extensively used to predict travel mode choice (e.g. Bamberg & Schmidt, 2003; Nordlund & Garvill, 2003), while meta-analytical evidence shows personal norms are a consistent and significant predictor of car use (Gardner & Abraham, 2008; Lanzini & Khan, 2017).

In addition to TPB and NAM, a third line of research focused on the habitual nature of driving, which was neglected by both theories. *Habits* are defined as “relatively stable behavioral patterns, which have been reinforced in the past […] and are executed without deliberate consideration” (Verplanken, Aarts, van Knippenberg, & van Knippenberg, 1994). As mobility decisions are repeatedly taken under stable context situations, habits can become particularly important in explaining such decisions (Verplanken et al., 1994). For example, when choosing between various alternatives for a particular journey, people rarely consciously deliberate which travel mode to use. They frequently undertake the same journeys under the same situational conditions, over and over, which can result in automatic behavioral patterns. Empirical studies found that car use habits were a significant predictor of car use behavior (e.g. Bamberg & Schmidt, 2003; Eriksson, Garvill, & Nordlund, 2008) and that they significantly increased the explained variance in travel mode choice, over the explanatory power of rational decision processes (e.g. Verplanken, Aarts, van Knippenberg, & Moonen, 1998; Verplanken et al., 1994).

1.2. The present study

Unlike other studies conducted on the topic that used self-reported measures, the present study uses an innovative method for objectively measuring car use, based on Google Location data, which is analyzed using a multilevel modeling approach. At level 1 (intra-individual level), contextual variables such as weather variables (i.e. *wind speed, temperature, precipitation*), *distance* travelled, *day of the week* and whether participants went *shopping* or *picked up children from school/kindergarten* will be used to predict the daily variation in car use behavior. At level 2 (inter-individual level), psychological variables such as *attitudes* towards car use reduction, *subjective norms* about car use, *perceived behavioral control* to reduce car use, *awareness of negative consequences* derived from car use, *ascription of responsibility* for these consequences, *personal norms* related to sustainable transportation, car use *habits* and *intentions* to reduce car use, together with two structural variables (i.e. *proximity to public transport* and access to alternative travel *infrastructure*) will be used to predict individual differences in car use (see Figure 1).

At level 1, according to the previously reviewed literature, it is expected that *wind speed*, level of *precipitation*, *distance*, *shopping* and *picking- up children* will positively predict car use, while *temperature* will negatively predict it. Because the influence of the *day of the week* on car travel is ambiguous, no hypothesis was formulated and instead this relationship will be exploratively investigated. At level 2 it is expected that *attitudes* towards car use reduction, *subjective norms* about car use, *perceived behavioral control* to reduce car use, *awareness of negative consequences*, *ascription of responsibility*, *personal norms* about sustainable transportation, *intentions* to reduce care use and alternative travel *infrastructure* will negatively predict car use, while car use *habits* will positively predict it. The relationship between *proximity to PT* and car use behavior is ambiguous and will therefore be investigated in an explorative manner, together with demographic variables such as *gender* and *age*.

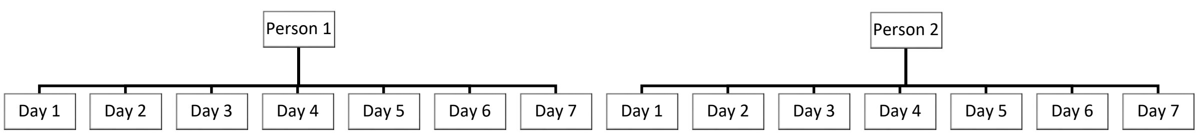

Figure 1. A graphical representation of the nested model

2. Method

2.1.Participants and procedure

Before the data collection process started, an ethics approval was obtained from the Ethical Committee of the West University of Timisoara (RCE2019-26). Participants were recruited in two ways: First, announcements on Facebook groups were posed and second, seminar credits were offered to psychology students enrolled at the West University of Timisoara who helped the research by bringing in participants for the study. As an incentive for participation, a bicycle (worth about 250 euros) was offered through a random draw. Participants were required to complete a questionnaire measuring all the psychological constructs and to upload their Google Location data into a private data storage account. Anonymity was guaranteed by asking participants to create a private code for both the questionnaire and the name of the Google Location data file, which allowed us to link the two pieces of data. Data was collected during late spring period (May and June), when the weather in Timisoara generally permits walking and cycling. Participants were required to own a driver’s license and to be residents in the city of Timisoara, which is the third largest city in Romania, with a population of about 330000. A total of 50 participants uploaded their data and completed the questionnaire, yet two participants were removed from analyses because their Google Location data did not contain enough information to accurately estimate movement. Therefore, the final sample consisted of 48 individuals, whose travel behavior was followed for a full week. They were 43.8% female and ranged in age from 18 to 53 years (M = 29.042, SD = 8.691). Of them, 87.5% also owned a car or had unrestricted access to one and 91.7% owned a bicycle or had unrestricted access to one. Regarding their highest educational achievement, 37% finished high-school, 4.3% finished vocational training, 32.6% university, 17.4% master and 8.6% finished a PhD, indicating a sample skewed towards the more educated.

2.2.Measures

Car use (measured in meters) was the dependent variable and was calculated for each individual, for each particular day, by summing up all distances travelled by car during the day. For each participant, seven days of travelling within a whole week (Monday to Sunday) were considered, which was chosen to be the closest possible to the date the person completed the questionnaire, resulting in a total of 336 days analyzed. Car use was manually coded with the help of Location History Visualizer Pro v. 1.6.1 software, which displays GPS points on the map, connected by a line that indicates movement. The software also estimates the distance between two GPS points and the speed of movement, making it possible to distinguish between different modes of transportation (see Image 1).

Rules used to code car use

i. Only travels made within 15 km in a straight line from the city center were considered
ii. Days that did not contain data, in which the person did not travel at all or in which the person travelled to a destination outside the city that was further than 15 km in a straight line from the city center, were not considered. For these dates, data from the closest similar day in which the person travelled within 15 km of the city center was used instead. For example, if on Saturday 25th May, the person stayed at home, then Saturday 18th May was the next date considered. If on this date the person travelled outside the city, then Saturday 11th May was considered, and so on.
iii. A trip was considered as ended if 1) there was a change of transportation mode (e.g. from car to bicycle) or 2) the person remained in the same place for at least 10 minutes.
iv. Travel distances shorter than 100 meters were not considered
v. Transportation mode was distinguished by movement speed and pattern of movement. Movement speeds between 2 – 7 km/h was generally coded as walking, between 8 – 25 km/h as cycling and higher than 26 km/h as car use or PT use. In addition to speed, the pattern of movement was also considered. For example, movement on a pedestrian road or in the opposite way of a one-way road was only coded as walking or cycling. Car use was distinguished from PT based on speed and time needed to complete the route (PT is generally slower, due to stops) and based on the specific pattern of GPS data. For

example, if a person's GPS location was registered as stationary for a period of time near a bus or tram station, followed by a period of fast movement that ended in another bus or tram station, this was coded as a public transportation ride. In case of doubt, the same route was checked on Google Maps by introducing the starting point and the destination. If the route corresponded to an existing PT route and was accomplished in approximately the same time, this was coded as a PT ride. If fast movement was registered on routes where there is no PT, this was coded as car transportation.

vi. If a stop was made at a market or supermarket, we considered that the person was shopping.

vii. If a stop was made at a school or kindergarten, we considered that the person was dropping-off or picking-up his/her children to school or kindergarten.

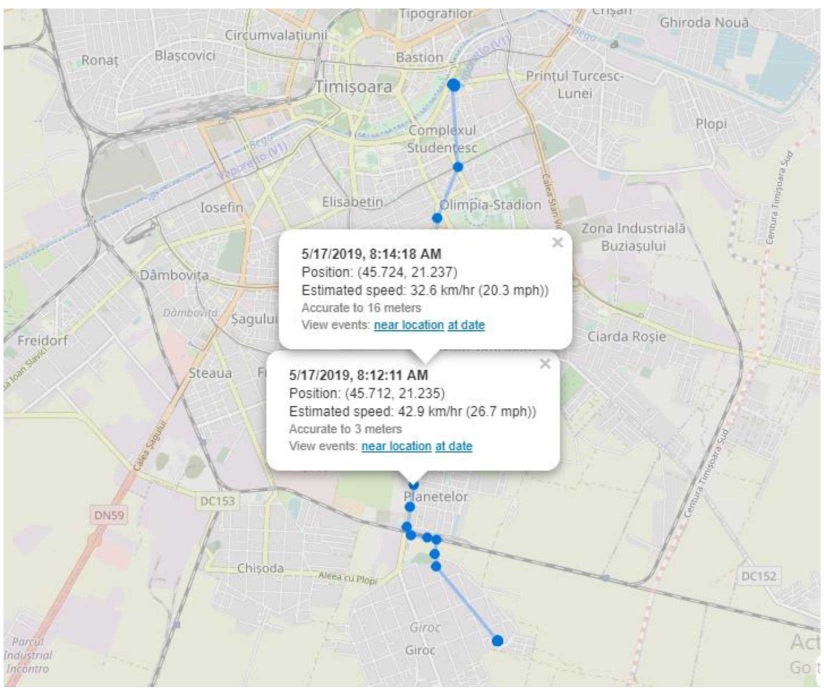

Image 1. Example of a route in which the car was used

The reliability of this measurement was verified by comparing, for 25 randomly chosen dates, the distance coded through this method against the data provided by Google Maps Timeline, which is a service that quantifies movement based on Google Location data. The correlation coefficient for car use estimated with the two measurements was $r = .996$, $p < .001$, while the mean estimated kilometers per day were $M = 24.340$, $SD = 33.088$, for the manual coding and $M = 23.040$, $SD = 31.990$, for Google Maps Timeline. The absolute difference between the two methods in each of the 25 days was also calculated and a one-sample-t-test was performed on the difference. The absolute difference was significantly different from 0, $t(24) = 2.757$, $p = .011$, indicating significant discrepancies between the two methods. The average difference was 1.580 km per day higher, when the manual method was used. However, this represents a discrepancy of only 6.8% (39.5 km discrepancy for a total of 576 km calculated by Google Maps Timeline during the 25 days). We also calculated the proportion of car use (kilometers travelled by car divided by the total number of kilometers travelled) and compared the proportions obtained with the two methods. The difference in the proportion of car use coded manually (p =.8096) was no different than the proportion calculated by Google Maps Timeline (p = .7937), $Z = .141$, $p = .889$. This suggests that the manual method slightly overestimated travelled distances for all modes (i.e. car, bicycle, walking and public transport), yet the car modal split share (proportion of car kilometers to the total kilometers travelled) remained relatively the same. This may be the case because GPS locations estimated with low accuracy may be eliminated during data processing by the Google algorithm, while they were kept in the current analysis, increasing thus estimated distances.

2.2.1. Level 1 variables

Wind speed (WS), *temperature (TEMP)* and the level of *precipitation (PRECIP)* were coded for each individual, for each particular date, by retrieving historical weather data from VisualCrossing.com database. *WS* was coded by considering the maximum wind speed of the day (in km/h), *TEMP* was coded by considering the maximum temperature of the day (in degrees Celsius), while *PRECIP* was coded by summing up all precipitation (in mm) that fell during the day.

Distance that the person needed to cover was coded by measuring the distance (in meters), in a straight line, between the furthest two GPS points recorded during the day (see Image 2).

Day of the week (DOTW) was coded with 0 if weekday and with 1 if weekend, *shopping (SHOP)* was coded with 0 if no market or supermarket was visited during the day and with 1 if a stop to a supermarket or market was made, while *picking-up children (CHILD)* was coded with 0 if no stop to a school or kindergarten was made during the day, and with 1 if otherwise.

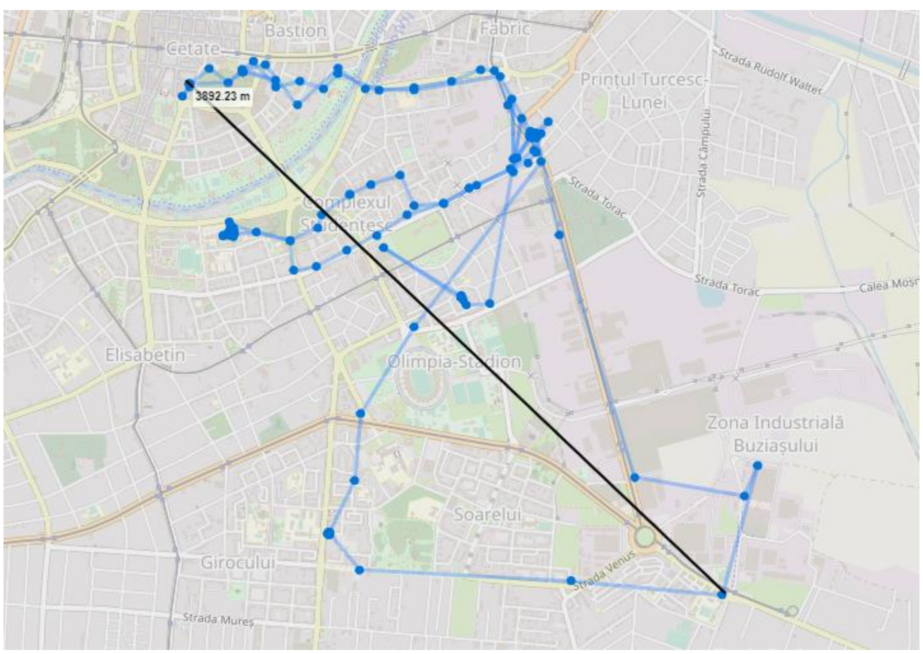

Image 2. Example of coding for distance

2.2.2. Level 2 variables

Proximity to PT (PPT) was coded by measuring the walking distance (in meters) between participant's residential location and the nearest PT station. Participant's residential location was inferred by investigating the GPS coordinates recorded during the night (02:00 – 06.00 AM), during multiple days.

Attitudes (ATT) towards car use reduction were measured with five semantic differentials, each on a 7-point scale. Participants had to rate to what extent reducing their car use is unattractive/attractive, bad/good, harmful/beneficial, unpleasant/pleasant and unworthy/valuable.

Subjective norms (SN) related to car use reduction were assessed with four items (e.g. "Most people who are important to me would support me in using the car less", "Most people who are important to me think that I should reduce car transport", etc.), adapted from Bamberg et al. (2003) and measured on a 7-point scale (1 = totally disagree, 7 = totally agree).

Perceived behavioral control (PBC) to reduce car use was assessed with a 7-point scale used by Bamberg et al. (2003). Participants had to respond to the following two items: "For me to reduce my car use in the future would be" (1 = difficult, 7 = easy) and "My freedom to reduce my car use in the future is" (1 = low, 7 = high).

Awareness of consequences (AC) of car use was assessed with a scale used by Ünal, Steg and Gorsira (2018). Participants rated their agreement (1 = totally disagree, 7 = totally agree) with the seven items of the scale (e.g. "The greenhouse effect resulting from road traffic is a serious problem", "Air pollution resulting from car traffic is a serious problem", "I am concerned about CO2 emissions resulting from road traffic", etc.).

Ascription of responsibility (AR) for negative consequences resulting from car use was measured with a scale composed of three items adapted from Jakovcevic and Steg (2013).

Participants rated their agreement or disagreement (1 = totally disagree, 7 = totally agree) with the following items: “I am jointly responsible for the problems caused by car use”, “Not just others, like the government, are responsible for heavy traffic, but me too” and “I feel joint responsibility for the contribution of car traffic to global warming”.

Personal norms (PN) related to car use reduction were assessed with a scale used by Jakovcevic and Steg (2013). Participants rated their agreement (1 = totally disagree, 7 = totally agree) with the eight items of the scale (e.g. “I feel personally obliged to travel in an environmentally sound way, such as by using a bicycle or public transport”, “I feel obliged to take the environmental consequences of car use into account when making travel choices”, etc.).

Car use *habits (HAB)* were assessed with the self-report index of habit strength (SRHI, Verplanken & Orbell, 2003). Respondents rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree) their agreement with the 12 items of the scale (e.g. “Using the car is something I do automatically”, “Using the car is something that belongs to my everyday routine “, etc.).

Infrastructure (INFR) for transportation alternatives was measured with three items. Participants rated how much they agree (1 = totally disagree, 7 = totally agree) with the following statements: “The transport infrastructure in the place where I live allows me to travel with other means than the car”, “Where I live there are other viable travel alternatives besides the car” and “If I wanted to, I could travel with other means of transportation besides the car”.

Participants’ *intention* to reduce their car use was measured with two items on a 7-point scale: “My intention to reduce my car use in the near future is” (1 = weak, 7 = strong) and “I intend to reduce my car use in the near future” (1 = unlikely, 7 = likely). See Table 1 for reliability statistics of each scale used. An average score was calculated from all the items of each scale, which was used in the analyses.

Table 1. *Reliability statistics of the scales used*

Variable	Number of items	Cronbach's alpha
ATT	5	.867
SN	4	.704
PBC	2	.773
AC	7	.941
AR	3	.912
PN	8	.823
HAB	12	.938
INFR	3	.907
INT	2	.918

Note: ATT = attitudes towards car use reduction; SN = subjective norms for car use reduction; PBC = perceived behavioral control to reduce car use; AC = awareness of consequences; AR = ascription of responsibility; PN = personal norms for car use reduction; HAB = car use habits; INFR = infrastructure for transportation alternatives; INT = car use reduction intentions

3. Results

A total of 1555 car trips took place in the 336 days analyzed. Participants travelled, on average, 7.84 km per day by car, 1.406 km by bicycle, 2.297 km on foot and 0.819 km by public transport. The descriptives for all continuous variables analyzed in the present study are presented in Table 2.

A multilevel model was created, with level 1 units being the 336 days in which data was collected and level 2 units, the 48 different individuals. All analyses were conducted with the help of HLM v 8.1.4.14 software. Preliminary analyses revealed no missing data and no extreme observation for any of the analyzed variables. An intercept only model that allowed the partitioning of the total variance between the two levels of the model, revealed that 59% of the variance was at intra-individual level (level 1) and 41% was at inter-individual level (level 2; the intraclass correlation). A chi-square test showed that the variance at the second level was significant, $\chi^2(47) = 275.401$, $p < .001$, indicating that a 2-level MLM is appropriate for the data.

Table 2. *Descriptive statistics for all continuous*

Variables	**Mean**	**SD**
Level 1		
Car use	7841.00	10798.67
WS	20.76	7.05
TEMP	19.62	6.33
PRECIP	3.79	6.83
Distance	3837.69	2954.47
Level 2		
PPT	383.08	379.15
ATT	4.05	1.52
SN	3.88	1.30
PBC	4.92	1.67
AC	6.11	1.02
AR	5.64	1.36
PN	5.16	1.16
HAB	2.35	1.05
INFR	5.23	1.56
INT	5.00	1.88
Age	29.04	8.69

Note: PPT = proximity to public transportation; WS = maximum wind speed; TEMP = maximum temperature of the day; PRCIP = level of precipitation; ATT = attitudes towards car use reduction; SN = subjective norms for car use reduction; PBC = perceived behavioral control to reduce car use; AC = awareness of consequences; AR = ascription of responsibility; PN = personal norms for car use reduction; HAB = car use habits; INFR = infrastructure for transportation alternatives; INT = car use reduction intentions

In the next step, all continuous predictors were grand centered and a saturated model with all level 1 predictors was created. Non-significant predictors were eliminated, one by one, following a backward procedure, starting at each step with the predictor that had the largest p-value. Following this procedure, PRECIP, $b = -22.865$, $t(281) = -.429$, $p = .661$, WS, $b = 13.407$, $t(282) = .290$, $p = .772$, DOTW, $b = -634.809$, $t(283) = -.925$, $p = .356$ and TEMP, $b = 41.956$,

$t(284) = .734$, $p = .464$ were removed from the model, because they did not significantly predict car use. Distance, $b = 2.801$, $t(285) = 22.657$, $p < .001$, CHILD, $b = 10835.026$, $t(285) = 4.032$, $p < .001$ and SHOP, $b = 2072.926$, $t(285) = 3.045$, $p = .003$ were significant and remained in the model, explaining together 58.3% of the day-to-day variance (variance at level 1) and 81.6% of the inter-individual variance (variance at level 2) in car use.

At the next step, all level 2 predictors were introduced in a saturated model and non-significant predictors were eliminated, one by one, in a similar manner. AR, $b = -35.103$, $t(35) = -.071$, $p = .944$, Age, $b = 12.893$, $t(36) = .196$, $p = .846$, INT, $b = -97.554$, $t(37) = -.253$, $p = .802$, PBC, $b = 103.773$, $t(38) = .231$, $p = .818$, PA, $b = -291.008$, $t(283) = -.468$, $p = .642$, Gender, $b = -681.808$, $t(40) = -.700$, $p = .488$, ATT, $b = 238.915$, $t(41) = .729$, $p = .470$, PN, $b = 262.438$, $t(42) = .590$, $p = .558$, SN, $b = 640.894$, $t(43) = 1.797$, $p = .079$, PPT, $b = 1.985$, $t(44) = 1.531$, $p = .133$ were non-significant and were removed from the model. HAB, $b = 1216.805$, $t(45) = 2.702$, $p = .005$ and INFR, $b = -732.147$, $t(45) = -2.348$, $p = .011$ significantly predicted car use and were retained in the final model, explaining an additional 5.5% of the level 2 variance (see Table 3). The total variance in car use explained by the five predictors (level 1 + level 2) was $R^2 = 70.1\%$.

The assumption of normally distributed errors was checked by investigating residuals at both levels of analysis. A visual inspection of Q-Q plots revealed no obvious deviations from normality for both plots. Normality of residuals was also checked by inspecting skewness and kurtosis. Skewness values were .089 for level 1 and -.672 for level 2 residuals, while kurtosis values were 5.187 for level 1 and 2.292 for level 2 residuals. These values are within the criterion proposed by Hair et al. (2010), who argued that skewness values between -3 and 3 and kurtosis values between -7 and 7 are still consistent with the normality assumption. We therefore conclude that the assumptions of the model were satisfactorily met.

Table 3. *Regression coefficients in the three models used*

	M1 - Intercept only model			M2 – Model with first level variables			M3 – Model with first and second level variables			
Predictor	**Coefficient**	***SE***	***p***	**Coefficient**	***SE***	***p***	**Coefficient**	***SE***	***β***	***p***
Fixed part										
Intercept	7840.997	1099.634	< .001	6907.183	590.055	< .001	6947.926	524.070		< .001
Distance				2.801	0.123	< .001	2.729	0.122	.747	< .001
CHILD				10835.026	2687.465	< .001	9723.517	2658.005	.117	< .001
SHOP				2072.926	680.747	.003	2014.421	671.891	.089	.003
HAB							1216.805	450.265	.118	. 005
INFR							-732.147	311.774	-.106	.011
Random part										
σ^2_e	69337351.324	8326.905		28896206.556	5375.519		28853779.450	5371.571		
σ^2_{r0}	48136054.459	6938.015	< .001	8848898.515	2974.709	< .001	6169489.242	2483.845		< .001
Deviance	7086.023			6732.217			6691.664			

Note: CHILD = pick-up/drop-off children to school/kindergarten; SHOP = do shopping; HAB = car use habits; INFR = infrastructure for transportation alternatives.

Therefore, the final, mixed regression equation was:

$$Car\ use = 6947.926 + 1216.805*HAB - 732.147*INFR + 2.729*Distance + 9723.517*CHILD + 2014.421*SHOP + r_0 + e$$

Our analysis suggests that the further people needed to travel, the more they used the car. *Distance* to travel destination had the largest impact on car use (β = .747) which, according to Cohen's guidelines (1988), represents a large effect. For every kilometer increase in distance, there was a 2.729 km increase in car use, when controlling for the other variables in the model. The need to pick-up/drop-off children to school/kindergarten, the need to do shopping and the available infrastructure for transportation alternatives had all small effects, β = .117, β = .089 and β = -.106, respectively. People traveled by car, on average, with 9.723 km more when they needed to pick-up or drop-off their children, with 2.014 km more when they needed to do shopping and with 0.732 km less for every point increase in access to alternative transportation infrastructure, when all the other variables in the model were controlled. Car use habits was the only psychological variable that significantly predicted car use (β = .118), indicating that for every point increase in car use habits people traveled 1.216 km more by car, when all the other predictors in the model are controlled.

4. Discussion

The present study investigated multiple contextual, psychological and structural predictors by using a different statistical procedure, namely a multilevel modeling approach. Unlike previously conducted studies on the topic, the DV (i.e. car use) was objectively measured, which is a rare case in the literature on travel mode choice. Results suggest that, even though, from a moral perspective, people were highly aware of the negative consequence of car use (M_{AC} = 6.11), they assumed responsibility for such consequences (M_{AR} = 5.64) and had strong personal norms about car use reduction (M_{PN} = 5.16), these variables seem to play no role in people's actual behavior, as none of these variables significantly predicted the amount of kilometers travelled. Such results are contrary to previous findings in the travel mode choice literature (e.g. Bamberg & Schmidt, 2003; Hunecke, Blöbaum, Matthies, & Höger, 2001; Nordlund & Garvill, 2003), which generally found a significant relation between the variables described by the norm activation model and environmentally-relevant travel behaviors. However, such studies relied entirely on self-reported measures of travel behavior. A possible answer to this discrepancy might lie in people's motivation to pay attention more and to report behaviors that are in line with their moral convictions, at the same time neglecting or justifying those that are not. Therefore, the correlation between self-reported behaviors and self-reported attitudes and convictions could be artificially inflated in self-reported measures, biasing in this way research results.

Present results suggest that objective constraints such as travelled distance, available infrastructure and the need to transport children or goods play the highest role in explaining actual car use behavior. The strongest predictor was distance (β = .747), indicating that when people need to travel longer distances, they are more likely to do so by car. This may be due to instrumental motivations such as increased comfort, flexibility, speed and safety that the car provides over other modes of transportation (Gardner & Abraham, 2007; Steg, Vlek, & Slotegraaf, 2001). Also, having to transport children (β = .117) or goods (β = .089) significantly increased kilometers travelled by car, most probably because car transport can satisfy such goals with the least amount of effort in the shortest amount of time, when compared to all other transportation modes. On the other hand, accessibility to infrastructure for alternative means of transportation (β = -.106), significantly decreased kilometers travelled by car, a finding that is in line with other studies that show that the build environment has a significant influence on transportation choices (see Ewing & Cervero, 2010). Even though the effect is small, the difference in travelled kilometers between the two extremes of the scale is significant.

Individuals who have easy access to alternative travel infrastructure traveled by car with 4.392 km less each week than those who have the hardest access to such infrastructure. For a car that runs on petrol with an average consumption of 7l/100km, this is equivalent to an annual reduction of 37.683 kg CO_2 per person.

Concerning psychological predictors, car use habits (β = .118) was the only psychological variable measured in the present study that significantly influenced car use, suggesting that, over and above what instrumental reasons can account for, individuals for whom car use has become habitual are more likely to travel by car, compared to those with weaker car use habits. This is because individuals with strong car use habits are less likely to process new travel-relevant information in their environments, such as improvements in travel infrastructure (e.g. new cycling lanes, new PT lines, etc.), even if this might offer them a better alternative than the one they are used to and, consequently, become less adapted to the travel condition around them. By making the same travel choices over and over again, they form a default travel mode which they activate in most travel situations, irrespective of the change in travel conditions. Individuals with the strongest car use habits travelled by car with 4.867 km more each week than those with weak habits, which corresponds to an annual difference of 41.758 kg CO_2 per person, for a car that runs on petrol with an average consumption of 7l/100km.

Surprisingly, none of the constructs described by the theory of planned behavior significantly predicted car use. While in the case of subjective norms and attitudes towards car use reduction such results are consistent with others from the literature (e.g. Heath & Gifford, 2002; Noblet et al., 2014), a non-significant result for perceived behavioral control is contrary to the general findings in the literature. However, this discrepancy can be partly explained by its conceptual overlap with the construct of habits. This overlap was also reflected statistically, by a very high correlation between the two constructs, *r* = -.673. Indeed, PBC alone significantly predicted inter-individual variance in car use behavior, which disappeared as soon as HAB was also introduced in the model.

Another surprising result is that none of the weather variables investigated (i.e. temperature, precipitation and wind speed) significantly predicted daily variation in car use. However, as most dates investigated were in May and June, there was little variability in temperature and precipitation during the investigated days. An investigation conducted on more days, sampled from each month of the year, should provide a more nuanced picture of how (and if) such variables have an impact on car use.

From an interventionist perspective the results of the present study portray a grey picture. As the largest amount of variance in car use (58.3% in day-to-day and 81.6% between individuals) was predicted by instrumentalist motives such as distance needed to cover, the need to pick-up/drop-off children and the need to do shopping, there seems to be little room to decrease car use through targeted soft interventions. Results suggest that a more adequate perspective would be to structure the built environment in such a way that individuals have easy access to all daily necessities. Creating neighborhoods with greater densities and increased access to services and facilities will decrease the need to travel for longer distances, making it more likely that individuals will switch to active means of transportation such as cycling, or walking, instead of using the car (Kamruzzaman, Washington, Baker, Brown, Giles-Corti, & Turrell, 2016). Improvements in alternative travel infrastructure would also be needed for recreational purposes or when individuals need to travel for longer distances, which have generally been shown to be effective in promoting active transportation (e.g. Aziz, Nagle, Morton, Hilliard, White, & Stewart, 2018; Handy, Cao, & Mokhtarian, 2006). However, such “hard” measures should always be complemented by soft interventions, to promote their use by informing, educating or changing people’s attitudes, which will impact also people’s perceptions about the physical reality with which they interact (Noblet et al., 2014). One final possibility for changing car use behavior suggested by our study’s results is to address people’s habits. Even though habits might seem particularly resistant to change, soft measures in which techniques such as implementation intentions were used have shown promising results (see Armitage et al., 2011) and need to be investigated further.

4.1. Limitations and future studies

The present study suffers from a few limitations that are worth noting. First of all, the sample size used is small and non-representative of the population, making it likely that the relations identified between level 2 variables and car use to look somewhat different in a larger sample, which is also representative of the population of car drivers in Timisoara. Even though the sample size was large enough to detect small effects such as those of HAB or INFR, a low number of participants generally results in imprecise parameter estimates, with large errors. It is therefore recommended for future studies to include more participants, which will increase both internal validity of the findings as well as its external validity.

A second important limitation is related to the coding of some of the variables used in the present study (i.e. car use, CHILD and SHOP), which was done by a single person. Even though the coding was carefully done and was cross-validated against other data (i.e. Google Timeline for car use), there is a possibility of error that future studies need to address by including more than one coder.

As previously mentioned, weather variables TEMP and PRECIP suffered from low variability, as they were generally coded for the same week, for most participants. Therefore, to address this limitation, it is recommended that more diverse days, from each month of the year, are considered. This will likely produce more diverse data in temperature and precipitation, which will result in a clearer understanding of their relation with car use behavior.

An important limitation that needs to be explicitly addressed resides in the type of model used. The present study used a two-level model in which all level 2 predictors were assessed at the same level, which resulted in a large number of non-significant effects. However, this type of analysis did not account for other relations between different predictors or their possible indirect (i.e. mediated) effects on car use, which would be both more in line with the theoretical assumptions as well as more realistic. Future studies should address this limitation by incorporating such predictors in a more complex statistical model that also accounts for the various relations between predictors, which will greatly enhance our understanding about the relations between each link of the long chain of predictors of car use.

Even though the present study included a large number of predictors, which explained most of the variability in car use (58.3% in day-to-day variability, 87.1% in inter-individual variability and 70.1% overall), there is significant variability still unexplained. Future studies could depart from the present findings and investigate a more comprehensive set of predictors that could increase this percentage and enhance our understanding. For example, a future, more complex study, could investigate the same behavior using a multilevel perspective with more than two levels, like in the present study. Car use behavior could be investigated in multiple cities, within the same or between different cultures, in which case cultural variables might come into play. This approach would more realistically model real world phenomena, as significant differences between cultures in travel behavior have been observed (see Klinger & Lanzendorf, 2016).

4.2. Conclusions

The present study investigated car use behavior using objective mobility data, analyzed in a multilevel format. Present findings indicate that instrumental motivations such as distance needed to cover, the need to pick-up/drop-off children to/from school and the need to carry heavy objects have the largest influence on car use behavior. Additionally, access to alternative transport infrastructure and the strength of car use habits also significantly influence the number of kilometers travelled by car. Psychological variables such as those described by the theory of planned behavior and norm activation model, together with weather variables such as wind speed, temperature and precipitation, did not directly influence car use behavior. Results are relevant for policymakers and city planners in their attempts to shape a sustainable and modern urban environment that is adapted to the greatest challenges of the moment.