Translating node of Ranvier currents to extraneural electrical fields: a flexible FEM modeling approach

Evaluating car-sharing switching rates from traditional transport means through logit models and random forest classifiers

Riccardo Ceccato, Andrea Chicco and Marco Diana

14th April 2021

Riccardo Ceccato , Andrea Chicco & Marco Diana (2021) Evaluating car-sharing switching rates from traditional transport means through logit models and Random Forest classifiers, *Transportation Planning and Technology*, 44:2, 160-175, DOI: 10.1080/03081060.2020.1868084

Evaluating car-sharing switching rates from traditional transport means through logit models and random forest classifiers

Riccardo Ceccato[a], Andrea Chicco[a]* and Marco Diana[a]

[a]*Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Turin, Italy;*

*Department of Environment, Land and Infrastructure Engineering

Politecnico di Torino, 24 Corso Duca degli Abruzzi, 10129 Turin, Italy

Email: andrea.chicco@polito.it

Tel. +39 011 090 5602

Evaluating car-sharing switching rates from traditional transport means through logit models and random forest classifiers

Positive impacts of car-sharing, such as reduction in car ownership, congestion, vehicle-miles-travelled and greenhouse gas emissions, have been extensively analysed. However these benefits are not fully effective if car-sharing subtracts travel demand from existing sustainable modes. This paper aims to evaluate substitution rates of car-sharing against private car and public transport using Random Forest classifier and Binomial Logit Model. The models were calibrated and validated on a Stated-preferences travel survey and applied on a Revealed-preferences survey, both administered to a representative sample of the population living in Turin (Italy). Results of the two models were compared obtaining that the predictive power of both models is comparable, albeit logit model tends to estimate predictions with higher reliability. Random Forest produces higher positive switches towards car-sharing, however results from both models suggest that the substitution rate of private car is almost five times than the one of public transport, on average.

Keywords: car-sharing, data mining, mode choice, multimodality, sustainability, travel demand

Introduction

Car-sharing is an innovative alternative transport system where members pay to use a shared-vehicle (Martin and Shaheen 2011), which is provided and maintained by a company or private individual (Dill, McNeil, and Howland 2019). Through this service users can benefit from affordability (Zhou and Kockelman 2011), like public transport (PT), flexibility and privacy (D. Efthymiou and Antoniou 2016), like a private car, without bearing the costs and constraints of vehicle ownership (Kim, Rasouli, and Timmermans 2017; Shaheen, Cohen, and Martin 2011). In part because of these advantages, the use of car-sharing has spread throughout the world (Shaheen, Cohen, and Chung 2010). In particular, in Italy, from 2015 to 2018, the number of car-sharing rents per year increased by around 100% for free-floating and by about 50% for station-based (Ciuffini et al. 2019).

Furthermore several authors analysed car-sharing impacts on congestion and environment (Clewlow 2016). In particular, they reported that car-sharing reduces car ownership (Ko, Ki, and Lee 2019; Martin and Shaheen 2016; Becker, Ciari, and Axhausen 2018), parking space (Millard-Ball et al. 2005) and related congestion (Dowling and Kent 2015), and vehicle-miles travelled (Martin and Shaheen 2016); moreover, if shared-cars are electric vehicles (EVs), they contribute to decreasing emissions from fossil fuels (Hu et al. 2018), even fostering the acceptance of private EVs (Schlüter and Weyer 2019). However the effectiveness of these positive impacts depends on where car-sharing travel demand is originated (Chapleau, Gaudette, and Spurr 2019; Heilig et al. 2017). Specifically, in order to take advantage of these positive aspects, car-sharing should not subtract travel demand from existing sustainable modes (e.g. active means and public transport). Therefore, understanding how car-sharing might complement or substitute existing travel means is useful for both policy makers, whose target is to promote sustainable travels, and car-sharing providers, who aim to maximize the use of car-sharing.

Indeed, changes in travel demand after the introduction of car-sharing are often reported (Clewlow 2016), in particular for public transport, even if existing results are often contrasting (Ceccato and Diana 2019). Beyond previous works in which these variations are observed in before-after scenarios (Shaheen, Cohen, and Chung 2010; Ko, Ki, and Lee 2019; Martin and Shaheen 2016), some authors developed models to forecast car-sharing potential demand, predicting both membership (D. Efthymiou and

Antoniou 2016; Costain, Ardron, and Habib 2012) and future trips (Rotaris, Danielis, and Maltese 2019; Heilig et al. 2017). However many studies were focused on a specific sample of the population (e.g. students (Rotaris, Danielis, and Maltese 2019)), therefore results cannot be generalized. Moreover predicted demand were not analysed considering which travel modes were previously used (El Zarwi, Vij, and Walker 2017; Kim, Rasouli, and Timmermans 2017; Li et al. 2018), thus without identifying substitution relationships. Finally, modelling results sometimes are not used to quantify car-sharing impacts in a given scenario (Heilig et al. 2017; Costain, Ardron, and Habib 2012).

Unlike previous works, this paper aims to evaluate the car-sharing impacts on travel demand by analysing the switch from different existing modes, in order to identify and quantify the relationship of substitution between car-sharing and public and private means. In particular, two models were calibrated on a Stated-preferences travel survey administered to a representative sample of the population living in Turin (Italy). Results are generalized and then applied to another dataset with real trips performed in the same area, in order to quantify the number of trips that might be carried out on car-sharing.

On a methodological viewpoint, several models have already been adopted to evaluate mode choice of users in presence of car-sharing. Models based on Random Utility Maximization theory has been extensively used (Chapleau, Gaudette, and Spurr 2019; F. Wang and Ross 2018; Tang, Xiong, and Zhang 2015; Sekhar, Minal, and Madhu 2016), in particular multinomial logit (MNL) (Liang et al. 2018). However these models require statistical and mathematical assumptions (X. Wang and Kim 2019; Chen et al. 2018), such as independence of irrelevant alternatives (IIAs) for MNL (Lindner, Pitombo, and Cunha 2017; Chen et al. 2018; Liang et al. 2018); moreover correlations among explanatory variables are to be considered *a priori* (Chapleau, Gaudette, and Spurr 2019). Other models were developed to overcome these limitations (e.g. probit, nested logit, mixed logit) (Xie, Lu, and Parkany 2007), even considering non-linear relationships among variables (Lee, Derrible, and Pereira 2018). On the other hand several data mining techniques were adopted to evaluate travel mode choice of users (F. Wang and Ross 2018), such as Decision Tree (Lindner, Pitombo, and Cunha 2017; Xie, Lu, and Parkany 2007; Tang, Xiong, and Zhang 2015; Sekhar, Minal, and Madhu 2016), Artificial Neural Network (Lee, Derrible, and Pereira 2018; Xie, Lu, and Parkany 2007), Extreme Gradient Boosting (F. Wang and Ross 2018) and Support Vector Machines (Hagenauer and Helbich 2017). Unlike traditional logit models, these algorithms are more flexible (Chen et al. 2018; Tang, Xiong, and Zhang 2015; Xie, Lu, and Parkany 2007), since they do not require any particular statistical assumption (X. Wang and Kim 2019; F. Wang and Ross 2018), and in some cases were more accurate than econometric methods (Hagenauer and Helbich 2017). Moreover they extract significant patterns from input data (Hagenauer and Helbich 2017; Lindner, Pitombo, and Cunha 2017; Xie, Lu, and Parkany 2007) and they can easily manage large (Chapleau, Gaudette, and Spurr 2019) and unbalanced datasets (F. Wang and Ross 2018), like the one introduced in the following where only a tiny minority of interviewees declared a propensity to switch to car-sharing. Nevertheless, they often lack of interpretability (Waddell and Besharati-Zadeh 2019), without providing traditional parameters, such as demand elasticity and Value Of Time, which are useful for policy applications (Zhu et al. 2018).

Among different data mining techniques to model mode choice, Random Forest (RF) (Breiman 2001) often produces the highest classification accuracy (Chapleau, Gaudette, and Spurr 2019; Hagenauer and Helbich 2017). This algorithm has been mainly used for classification purposes (B. Wang, Gao, and Juan 2018; Ermagun, Rashidi, and Lari 2015; Liang et al. 2018; A. Efthymiou et al. 2019; Lee, Derrible, and Pereira 2018; Chen et al. 2018; Hagenauer and Helbich 2017; Sekhar, Minal, and Madhu 2016). On the other hand, few applications to forecast travel demand in presence of car-sharing are reported (Toque et al. 2018; Waddell and Besharati-Zadeh 2019; Chapleau, Gaudette, and Spurr 2019). Unlike these works, in this paper a classical logit and a RF models were calibrated and applied to estimate mode switch from existing modes to car-sharing. Results from the two models are compared and then applied to a real dataset of trips, in order to obtain the number of trips on car-sharing.

Field activities

The study area is the Turin Metropolitan Area, which includes the Turin Municipality, with about 900.000 inhabitants, and the municipalities surrounding the city, with about 544.000 inhabitants. Two free-floating car-sharing operators provide services within an operational area by using conventionally fuelled cars, whereas a third free floating operator has a fully electric fleet with charging stations. Car-sharing can thus be with a viable alternative to existing public and private modes, at least for trips within the operational area that is similar to the three services and roughly covers 45% of the surface of the Turin municipality (130 km^2), where about 750.000 inhabitants live.

Stated-preferences dataset

The survey was administered to a representative sample of the population living in Turin (4466 persons) and it collected socio-economic characteristics of the respondents, their travel diary and activities carried out up to 24 hours before the interview. In particular, locations were entered through Google Maps APIs to better estimate distances covered by the reported transport mode. In the last part of the survey, Stated-preferences experiments were carried out, where respondents had to express their shifting propensity for a randomly selected trip chain from the travel mode it was used to an alternative one (including car-sharing). Attributes of the current mode were estimated by directly considering the trip characteristics reported by the interviewee, whereas attributes of the alternative mode were evaluated using information on the trip chain through Google Maps APIs, along with data from public transit operators, car-sharing providers and fuel costs. In this way, questions were based on a real trip with realistic attributes, increasing the realism and reliability of the answers. Respondents had to answer using a 5-points scale ranging from “very unlikely” to “very likely” to switch. A detailed description of the survey can be found in (Ceccato & Diana, 2019).

The same survey was administered through both CATI (Computer Assisted Web Interviewing) and CAWI (Computer Assisted Telephone Interviewing) protocols 7 days a week in three different 4-weeks periods (in 2016 and 2017). Given the aim of this paper, only experiments in which the alternative mode is car-sharing are considered here. Moreover only motorized trips performed by non car-sharing users living in the Turin Municipality were considered, since the number of bike trips was low and it was assumed that car-sharing cannot substitute walking trips. Finally, neutral answers were discarded, thus obtaining 1050 observations, which were grouped in two alternative answers: “stay with the current means” and “switch to car-sharing”. The sample is quite unbalanced, since only 61 people took the switch option; the RF will therefore be used beyond more commonplace econometric methods.

Revealed-preferences dataset

The Revealed-preferences survey was run in the cities of Milan and Turin between the 13th and the 28th of May 2019, as one of the activities of the European project STARS (http://stars-h2020.eu/). The survey was aimed at understanding the differences between car-sharing users and non-users in terms of mobility habits (e.g. frequency of usage of different transport means, PT and bike sharing subscriptions), changing in car ownership and in the use of different modes to perform within-city trips. Additionally, interviewees were asked to provide information on the last trip performed with car-sharing (users) or with any other mode (non-users). Detailed trip characteristics related to the specific reported mode were derived from Google Maps API (e.g. distance, in-vehicle and walking time, waiting time at transit stop), transit agencies and car-sharing providers. Finally, socio-economic information at both the household (e.g. number of members, workers, cars, income) and the individual (education, occupational status) level were collected.

The survey was administered through both CAWI and CATI protocols to a representative sample of licensed drivers living in the Municipality of Turin; the weighted sample was stratified by gender and age to ensure representativeness. A total of 1070 complete questionnaires were collected, 734 (68.5%) in the city of Milan and 336 (31.5%) in Turin; given our study area only the latter observations are considered in the following. Additionally, since the aim of the paper is to investigate the switching intention from the mode used to car-sharing in performing different within-city motorized trips, only the 200 non-users interviews were considered.

Table 1 shows the main characteristics of the non-users sample both at household and individual level. About 57% of interviews were collected through CATI; the number of males and females is almost the same. The average age of the interviewees is 52.4 years; the majority of the non-users has a mid-level education (56% has a high school diploma) and works out of home (68.5%). The majority of households have two members owning a driving license; the average household income ranges between 2500€ and 3000€ per month. Most of these households have at their disposal one car (47.5%). About 4 out of 10 persons interviewed have a PT subscription and 9% of individuals are bike sharing members. Concerning the modal share, results showed that most of the recorded trips were performed by car (about 65%), one third by PT (35%).

[Table 1 near here]

Method

Two alternative models were developed in order to understand variables affecting the shift towards car-sharing and to predict the number of potential trips with this mode. The first one is a binomial logit and the second one is a RF model, which were estimated through Biogeme (Bierlaire, 2018) and RapidMiner (Hofmann & Klinkenberg, 2013), respectively. The endogenous variable of both models has two levels: "stay with the current means" and "switch to car-sharing" if the same trip had to be performed in the future. Since one of the aims of the paper is to compare performances and prediction results, both the models were calibrated, validated and applied on the same Stated-preferences dataset and had the same specification (with variables taken from the pool in Table 2), which considered trip attributes and socio-economic characteristics both at household and individual level. In order to evaluate models' predictive accuracies the entire dataset was divided into two stratified random samples: a training dataset with 70% of the observations and a testing dataset with the remaining 30%. Then, the two estimated models were applied to a Revealed-preference dataset carried out in the same study area.

[Table 2 near here]

Logit models are a standard method in travel demand studies and therefore they are not formally introduced here (Ben-Akiva, Lerman, and Lerman 1985). On the other hand, RF is a classifier that combines a set of many tree predictors ("forest") (Breiman

2001; X. Wang and Kim 2019). Each tree is generated according to the following procedure. From the training dataset, a sample is selected through the bootstrap method, which minimizes model variance and avoids overfitting (X. Wang and Kim 2019). Then the tree is built on this subsample. In particular, starting from the first node of the tree and for each of the consecutive ones, the splitting into leaf nodes is performed according to a defined criterion (e.g. Gini index) and considering only a random set of explanatory variables; this approach is called bagging features technique (Toque et al. 2018) and it reduces the correlation among all the trees in the “forest”, generated by the RF algorithm (X. Wang and Kim 2019). By applying this tree to the original training dataset, each observation is classified, storing the results. Following this framework, a specified number of trees are generated, obtaining many classification results for each observation. Finally, using the majority voting strategy (Chen et al. 2018), the classification outcome is obtained as the result that is predicted in most of the single trees (Sekhar, Minal, and Madhu 2016).

Results and discussion

Models calibration

Binomial logit

Estimation results of binomial logit model are reported in Table 3.

Like in previous studies (Clewlow, 2016; Mishra et al., 2015), car-sharing is adopted by households with high income (INCOME_AVG is positive) and with a low number of cars per members owing a driving license (CARPERLICENCE is negative). Furthermore coefficients related to monthly usage frequencies of other transport modes are all positive (F_CAR, F_PT and F_BIKE), confirming that people with multimodal travel habits are more willing to switch towards new transport solutions (Diana, 2010).

However, for the specific considered trip, car-sharing is less likely to substitute public transport rather than private car (PT is negative). People reporting positive switch are also bike sharing members (BIKE_SHARING), pointing out that the propensity to sharing positively affects the use of other shared means.

All the coefficients of variables related to the occupational status (WOOH, WAH and RET) are negative, indicating that workers and retired people are less likely to shift than students. Comparing the absolute values of these coefficients, as expected, the willingness to switch is less negative for people working out of home, since they need to perform more trips. Concerning the household composition, negative coefficients of both HH_FAMNC and HH_COUPLE suggest that households with more than one member have a lower propensity to switch toward car-sharing compared to people living alone or without relatives (such as students).

Positive switch intentions are reported for trips performed during non working days (NO_WORK_DAY is positive). Nevertheless there is no evidence that car-sharing might be adopted for systematic or discretionary trips (HBO, HBW and NH are all negatives). Finally mid-level of education (high school diploma) has a negative effect (HS) and men tend to have a lower propensity to switch rather than women (GENDER is negative).

[Table 3 near here]

Random Forest

Unlike MNL, RF does not provide a readable output of model estimation parameters, but only a measure of relative importance of variables. Figure 1 shows relative importance measures of each explanatory variable, which was calculated as the total amount of improvements that a variable provided at a node according to the splitting criterion. As such, RF nicely complements the information that is provided by MNL and helps in giving a more complete picture of the intertwined relationships between predictors, both at the individual and at the trip level, and outcome. Since RF does not consider different attributes for different alternative choices, differences of trip cost, distance and time between car-sharing and the revealed current mode were computed.

Observing Figure 1 one can note that the most important variables are socio-economic characteristics of household and individual, travel habits and trip attributes. In particular, age, income and education level of the respondent are in line with previous works (Rotaris, Danielis, and Maltese 2019). Moreover monthly use frequency of PT and private car, as well as differences in cost and duration of trips have a great impact on car-sharing adoption.

Comparing the variable significance values of binomial logit in Table 3 and the RF importance ranking in Figure 1, frequency of private car, income and cost are among the most important parameters in both models. However, other variables, like age and occupational status, play different roles on the outcomes of the two models, which might be due to the linear formulation of the MNL systematic utility or to the unbalanced sample.

[Figure 1 near here]

Models validation

In order to have a comparison between the two calibrated models, performance measures were estimated and summarized in Table 4. Overall binomial logit and RF have similar performances in terms of model accuracy and classification error. However, due to the unbalanced prediction classes, recall and precision were analysed. Considering these indicators, binomial logit still has the highest performances except for the recall of “SWITCH” class.

[Table 4 near here]

Trip-level switching probabilities

By applying both models to the Revealed-preferences dataset, switching probabilities towards car-sharing were obtained for each reported trip. In order to estimate the substitution rate of car-sharing respect to existing travel modes, trips were grouped according to the current reported mode. However only motorized means were considered (private car, public transport), since analysing data it was observed that nobody selected car-sharing as an alternative to walking. In order to evaluate the substitution ratio of car-sharing respect to existing modes, percentage of positive switches among trips performed with a specific mode, considering different switching probability lower thresholds obtained with binomial logit and RF, are respectively reported in Figure 2 and Figure 3. For the sake of visualization only results with a threshold greater than 50% (thus switching is more likely than staying with the current mode) are shown.

As expected, both the figures below show that generally if the threshold increases, the switching rate decrease for all the modes and is never above 7%, confirming the general inertia to keep on using their current mode. However in Figure 2 there is a slight increase in the switching rate of car trips between the threshold of 75% and 85%, since in this range the number of positive switches is constant while the number of negative switches decreases. Furthermore in both the outputs, switching rates for car trips are greater than the ones for public transport trips, confirming that car-sharing is more likely to substitute private car rather than transit. However, binomial logit produces positive switches up to 95% of probability, whereas RF up to 60%, suggesting that logit model tends to estimate predictions with higher reliability.

In addition, comparing the number of trips with positive switch for car and transit, it was estimated that the substitution rate of private car is almost five times than the one of public transport, on average.

[Figure 2 near here]

[Figure 3 near here]

In order to estimate the total number of trips which might be performed on car-sharing and to compare them with the actual number of trips carried out by inhabitants in the study area, estimated results of both the models were expanded to the trip universe of Turin. These results are summarized in Table 5 considering different used modes. Observing this table one can note that the two models provided different outcomes, in particular RF produces a number of trips on car-sharing higher than the one of binomial logit, even if both models predict a higher number of car trips substituted by car-sharing rather than public transport trips.

[Table 5 near here]

Conclusions

In this paper two models were developed to predict the number of potential trips on car-sharing considering the travel mode that this new service is likely to substitute.

In particular, a classical binomial logit model and a Random Forest (RF) classifier were calibrated and validated on a Stated-preferences travel survey, and then applied to a Revealed-preferences travel survey. Both surveys were administered in Turin, Italy. In this way, the switching propensity towards car-sharing was estimated and then expanded to the whole universe of population and trips.

Models performances and results were compared. Performance indicators show that binomial logit and RF have similar predictive power. However, only the former provides a deep understanding of the effect of explanatory variables. In particular, this model confirms that car-sharing is adopted by households with high income and with a low number of cars per members owning a driving license. Furthermore, results point out that people with multimodal travel habits are more willing to switch towards new transport solutions, even if for the analysed trips, car-sharing is less likely to substitute public transport rather than private car. The average effect of exogenous variables in the RF model is simply ranked considering their importance. The analysis shows that the some of the most important variables are those with high significance values in the binomial logit model. Moreover, binomial logit produces positive switches up to 95% of switching probability, whereas RF up to 60%, suggesting that logit model tends to estimate predictions with higher reliability.

Car-sharing potential trips were analysed considering switching probabilities provided by both models. Results of both models indicate that car-sharing substitution rates of private car are greater than the ones of public transport. In particular, the substitution rate of private car is almost five times than the one of public transport, on average. Nevertheless, binomial logit predicts lower shifting values with higher probabilities rather than RF. Overall, the former model estimates that car-sharing might substitute about 20% of car trips and 12% of public transport trips, whereas the latter predicts 36% and 33%, respectively.

Results are helpful for policy makers to understand how car-sharing might substitute existing motorized travel means.

Acknowledgements

This study was partly sponsored by the European project "Shared mobility opporTunities And challenges foR European citieS" (STARS), which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 769513. Financial support to collect the SP dataset also came through a "Ricerca dei Talenti" grant from Fondazione CRT.

Disclosure statement

No potential conflict of interest is reported by the authors.

References

Becker, Henrik, Francesco Ciari, and Kay W. Axhausen. 2018. “Measuring the Car Ownership Impact of Free-Floating Car-Sharing – A Case Study in Basel, Switzerland.” *Transportation Research Part D: Transport and Environment* 65 (August): 51–62. https://doi.org/10.1016/j.trd.2018.08.003.

Ben-Akiva, Moshe E, Steven R Lerman, and Steven R Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Vol. 9. MIT press.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45: 5–32. https://doi.org/10.1007/9781441993267_5.

Ceccato, Riccardo, and Marco Diana. 2019. “Substitution and Complementarity Patterns between Traditional Transport Means and Car Sharing: A Person and Trip Level Analysis.” *Transportation*. https://doi.org/10.1007/s11116-018-9901-8.

Chapleau, Robert, Philippe Gaudette, and Tim Spurr. 2019. “Application of Machine Learning to Two Large-Sample Household Travel Surveys: A Characterization of Travel Modes.” *Transportation Research Record*. https://doi.org/10.1177/0361198119839339.

Chen, Xuewu, Frank Witlox, Long Cheng, Jonas De Vos, and Xinjun Lai. 2018. “Applying a Random Forest Method Approach to Model Travel Mode Choice Behavior.” *Travel Behaviour and Society* 14 (May 2018): 1–10. https://doi.org/10.1016/j.tbs.2018.09.002.

Ciuffini, Massimo, Raimondo Orsini, Sofia Asperti, Valeria Gentili, Davide Grossi, Delia Milioni, Luca Refrigeri, et al. 2019. *3° Rapporto Nazionale Sulla Sharing Mobility*. http://osservatoriosharingmobility.it/wp-content/uploads/2019/07/comesta-la-sharing-mobility_III-Rapporto-SM_13-e-FRONT.pdf.

Clewlow, Regina R. 2016. "Carsharing and Sustainable Travel Behavior: Results from the San Francisco Bay Area." *Transport Policy* 51: 158–64. https://doi.org/10.1016/j.tranpol.2016.01.013.

Costain, Cindy, Carolyn Ardron, and Khandker Nurul Habib. 2012. "Synopsis of Users' Behaviour of a Carsharing Program: A Case Study in Toronto." *Transportation Research Part A: Policy and Practice* 46 (3): 421–34. https://doi.org/10.1016/j.tra.2011.11.005.

Dill, Jennifer, Nathan McNeil, and Steven Howland. 2019. "Effects of Peer-to-Peer Carsharing on Vehicle Owners' Travel Behavior." *Transportation Research Part C: Emerging Technologies* 101 (February): 70–78. https://doi.org/10.1016/j.trc.2019.02.007.

Dowling, Robyn, and Jennifer Kent. 2015. "Practice and Public-Private Partnerships in Sustainable Transport Governance: The Case of Car Sharing in Sydney, Australia." *Transport Policy* 40: 58–64. https://doi.org/10.1016/j.tranpol.2015.02.007.

Efthymiou, Alexandros, Emmanouil N. Barmpounakis, Dimitrios Efthymiou, and Eleni I. Vlahogianni. 2019. "Transportation Mode Detection from Low-Power Smartphone Sensors Using Tree-Based Ensembles." *Journal of Big Data Analytics in Transportation* 1 (1): 57–69. https://doi.org/10.1007/s42421-019-00004-w.

Efthymiou, Dimitrios, and Constantinos Antoniou. 2016. "Modeling the Propensity to Join Carsharing Using Hybrid Choice Models and Mixed Survey Data." *Transport Policy* 51: 143–49. https://doi.org/10.1016/j.tranpol.2016.07.001.

Ermagun, Alireza, Taha Hossein Rashidi, and Zahra Ansari Lari. 2015. "Mode Choice for School Trips." *Transportation Research Record: Journal of the Transportation Research Board* 2513 (1): 97–105. https://doi.org/10.3141/2513-12.

Hagenauer, Julian, and Marco Helbich. 2017. "A Comparative Study of Machine Learning Classifiers for Modeling Travel Mode Choice." *Expert Systems with Applications* 78: 273–82. https://doi.org/10.1016/j.eswa.2017.01.057.

Heilig, Michael;, Nicolai; Mallig, Tim; Hilgert, Martin; Kagerbauer, and Peter Vortisch. 2017. "Large-Scale Application of a Combined Destination and Mode Choice Model Estimated with Mixed Stated and Revealed Preference Data." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2669: pp 31--40. https://doi.org/10.3141/2669-04.

Hu, Songhua, Hangfei Lin, Kun Xie, Xiaohong Chen, and Hongjie Shi. 2018. "Modeling Users' Vehicles Selection Behavior in the Urban Carsharing Program." *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2018-Novem: 1546–51. https://doi.org/10.1109/ITSC.2018.8569386.

Kim, Jinhee, Soora Rasouli, and Harry J.P. Timmermans. 2017. "The Effects of Activity-Travel Context and Individual Attitudes on Car-Sharing Decisions under Travel Time Uncertainty: A Hybrid Choice Modeling Approach." *Transportation Research Part D: Transport and Environment* 56 (August): 189–202. https://doi.org/10.1016/j.trd.2017.07.022.

Ko, Joonho, Hyeongyun Ki, and Soojin Lee. 2019. "Factors Affecting Carsharing Program Participants' Car Ownership Changes." *Transportation Letters* 11 (4): 208–18. https://doi.org/10.1080/19427867.2017.1329891.

Lee, Dongwoo, Sybil Derrible, and Francisco Camara Pereira. 2018. "Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling." *Transportation Research Record*. https://doi.org/10.1177/0361198118796971.

Li, Qing, Feixiong Liao, Harry J.P. Timmermans, Haijun Huang, and Jing Zhou. 2018.

"Incorporating Free-Floating Car-Sharing into an Activity-Based Dynamic User Equilibrium Model: A Demand-Side Model." *Transportation Research Part B: Methodological* 107: 102–23. https://doi.org/10.1016/j.trb.2017.11.011.

Liang, LeiLei, Meng Xu, Susan Grant-Muller, and Lorenzo Mussone. 2018. "Travel Mode Choice Analysis Based on Household Mobility Survey Data in Milan: Comparison of the Multinomial Logit Model and Random Forest Approach." In *Transportation Research Board 97th Annual Meeting*. Washington DC, United States.

Lindner, Anabele, Cira Souza Pitombo, and André Luiz Cunha. 2017. "Estimating Motorized Travel Mode Choice Using Classifiers: An Application for High-Dimensional Multicollinear Data." *Travel Behaviour and Society* 6: 100–109. https://doi.org/10.1016/j.tbs.2016.08.003.

Martin, Elliot, and Susan Shaheen. 2011. "The Impact of Carsharing on Public Transit and Non-Motorized Travel: An Exploration of North American Carsharing Survey Data." *Energies* 4 (11): 2094–2114. https://doi.org/10.3390/en4112094.

Martin, Elliot, and Susan Shaheen. 2016. "Impacts of Car2go on Vehicle Ownership, Modal Shift, Vehicle Miles Travelled, and Greenhouse Gas Emissions: An Analysis of Five North American Cities." Berkeley. http://innovativemobility.org/wp-content/uploads/2016/07/Impactsofcar2go_FiveCities_2016.pdf.

Millard-Ball, Adam;, Gail; Murray, Jessica; ter Schure, Christine; Fox, and Jon Burkhardt. 2005. "TCRP Report 108. Car-Sharing: Where and How It Succeeds." Washington D.C.

Rotaris, Lucia, Romeo Danielis, and Ila Maltese. 2019. "Carsharing Use by College Students: The Case of Milan and Rome." *Transportation Research Part A: Policy* *and Practice* 120 (January): 239–51. https://doi.org/10.1016/j.tra.2018.12.017.

Schlüter, Jan, and Johannes Weyer. 2019. “Car Sharing as a Means to Raise Acceptance of Electric Vehicles: An Empirical Study on Regime Change in Automobility.” *Transportation Research Part F: Traffic Psychology and Behaviour* 60: 185–201. https://doi.org/10.1016/j.trf.2018.09.005.

Sekhar, Ch Ravi, Minal, and E. Madhu. 2016. “Mode Choice Analysis Using Random Forrest Decision Trees.” *Transportation Research Procedia* 17 (December 2014): 644–52. https://doi.org/10.1016/j.trpro.2016.11.119.

Shaheen, Susan A., Adam P. Cohen, and Melissa S. Chung. 2010. “North American Carsharing.” *Transportation Research Record: Journal of the Transportation Research Board* 2110 (1): 35–44. https://doi.org/10.3141/2110-05.

Shaheen, Susan A., Adam P. Cohen, and Elliot Martin. 2011. “Carsharing Parking Policy.” *Transportation Research Record: Journal of the Transportation Research Board* 2187 (1): 146–56. https://doi.org/10.3141/2187-19.

Tang, Liang, Chenfeng Xiong, and Lei Zhang. 2015. “Decision Tree Method for Modeling Travel Mode Switching in a Dynamic Behavioral Process.” *Transportation Planning and Technology* 38 (8): 833–50. https://doi.org/10.1080/03081060.2015.1079385.

Toque, Florian, Mostepha Khouadjia, Etienne Come, Martin Trepanier, and Latifa Oukhellou. 2018. “Short & Long Term Forecasting of Multimodal Transport Passenger Flows with Machine Learning Methods.” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2018-March: 560–66. https://doi.org/10.1109/ITSC.2017.8317939.

Waddell, Paul, and Arezoo Besharati-Zadeh. 2019. “A Comparison of Statistical and Machine Learning Algorithms for Predicting Rents in the San Francisco Bay

Area." In *Transportation Research Board 98th Annual Meeting*. Location: Washington DC, United States.

Wang, Bao, Linjie Gao, and Zhicai Juan. 2018. "Travel Mode Detection Using GPS Data and Socioeconomic Attributes Based on a Random Forest Classifier." *IEEE Transactions on Intelligent Transportation Systems* 19 (5): 1547–58. https://doi.org/10.1109/TITS.2017.2723523.

Wang, Fangru, and Catherine L. Ross. 2018. "Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model." *Transportation Research Record*. https://doi.org/10.1177/0361198118773556.

Wang, Xinyi, and Sung Hoo Kim. 2019. "Prediction and Factor Identification for Crash Severity: Comparison of Discrete Choice and Tree-Based Models." *Transportation Research Record*. https://doi.org/10.1177/0361198119844456.

Xie, Chi, Jinyang Lu, and Emily Parkany. 2007. "Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks." *Transportation Research Record: Journal of the Transportation Research Board* 1854 (1): 50–61. https://doi.org/10.3141/1854-06.

Zarwi, Feras El, Akshay Vij, and Joan L. Walker. 2017. "A Discrete Choice Framework for Modeling and Forecasting the Adoption and Diffusion of New Transportation Services." *Transportation Research Part C: Emerging Technologies* 79: 207–23. https://doi.org/10.1016/j.trc.2017.03.004.

Zhou, Bin, and Kara M. Kockelman. 2011. "Opportunities for and Impacts of Carsharing: A Survey of the Austin, Texas Market." *International Journal of Sustainable Transportation* 5 (3): 135–52. https://doi.org/10.1080/15568311003717181.

Zhu, Zheng, Xiqun Chen, Chenfeng Xiong, and Lei Zhang. 2018. “A Mixed Bayesian Network for Two-Dimensional Decision Modeling of Departure Time and Mode Choice.” *Transportation* 45 (5): 1499–1522. https://doi.org/10.1007/s11116-017-9770-6.

Tables

Table 1. Demographic characteristics of the sample

Individual characteristics	**N**	**(%)**	**Household characteristics**	**N**	**(%)**
Type of interview			Household members		
CAWI	114	57.0%	1	29	14.5%
CATI	86	43.0%	2	70	35.0%
Gender			3	54	27.0%
Male	97	48.5%	4	44	22.0%
Female	103	51.5%	More than 4	3	1.5%
Age			Household children		
18-24	10	5.0%	0	117	58.5%
25-29	9	4.5%	1	38	19.0%
30-34	11	5.5%	2	39	19.5%
35-44	25	12.5%	More than 2	6	3.0%
45-54	41	20.5%	Driving licences		
55-64	43	21.5%	1	50	25.0%
65-74	55	27.5%	2	112	56.0%
More than 75	6	3.0%	3	29	14.5%
Education level			More than 3	9	4.5%
Not high school graduate	23	11.5%	Household cars		
High school graduate	112	56.0%	0	8	4.0%
Degree or Ph.D.	65	32.5%	1	95	47.5%
Occupational status			2	81	40.5%
Work out of home	137	68.5%	3 or more	16	8.0%
Work at home	16	8.0%	Household income [€/month]		
Student	9	4.5%	Less than 1000	17	8.5%
Retired	32	16.0%	1000-1500	28	14.0%
Unemployed	6	3.0%	1500-2000	28	14.0%
PT subscription			2000-2500	33	16.5%

Yes	80	40.0%	2500-3000	39	19.5%
No	120	60.0%	3000-4000	26	13.0%
Bike sharing subscription			4000-5000	12	6.0%
Yes	18	9.0%	5000-10,000	9	4.5%
No	182	91.0%	More than 10,000	8	4.0%

Table 2. Candidate exogenous variables

Variable	**Description**	**Type**	**Level**
AGE	Age	Metric	Individual
BASE_COST	Current mode trip cost [€]	Metric	Trip
BASE_DIST	Current mode trip distance [m]	Metric	Trip
BASE_DUR	Current mode trip duration [min]	Metric	Trip
BASE_LEG	Current mode trip legs	Metric	Trip
BASE_WAIT	Current mode waiting time [min]	Metric	Trip
BASE_WALK_DIST	Current mode walking distance [m]	Metric	Trip
BASE_WALK_DUR	Current mode walking duration [min]	Metric	Trip
BIKE_SHARING	Bike sharing subscription (Y: yes, N: no)	Categorical	Individual
CARPERLICENCE	Number of cars per driving licences	Metric	Household
CS_COST	Car-sharing trip cost [€]	Metric	Trip
CS_DIST	Car-sharing trip distance [m]	Metric	Trip
CS_DUR	Car-sharing trip duration [min]	Metric	Trip
CS_LEG	Car-sharing trip legs	Metric	Trip
CS_WAIT	Car-sharing waiting time [min]	Metric	Trip
CS_WALK_DIST	Car-sharing walking distance [m]	Metric	Trip
CS_WALK_DUR	Car-sharing walking duration [min]	Metric	Trip
DELTA_COST	Difference between CS_COST and BASE_COST [€]	Metric	Trip
DELTA_DIST	Difference between CS_DIST and BASE_DIST [m]	Metric	Trip
DELTA_DUR	Difference between CS_DUR and BASE_DUR [min]	Metric	Trip
DELTA_LEG	Difference between CS_LEG and BASE_LEG	Metric	Trip
DELTA_WAIT	Difference between CS_WAIT and BASE_WAIT [min]	Metric	Trip

DELTA_WALK_DIST	Difference between CS_WALK_DIST and BASE_WALK_DIST [m]	Metric	Trip
DELTA_WALK_DUR	Difference between CS_WALK_DUR and BASE_WALK_DUR [min]	Metric	Trip
EDUCATION	Level of education (NE: no education, PRIM: primary school diploma, SEC: secondary school diploma, HS: high school diploma, UNI: university degree, master or Ph.D.)	Categorical	Individual
EMPLOYEMENT_AGGR	Job status (RET: retired, STN: student, UNE: unemployed, WAH: work at home, WOOH: work out of home)	Categorical	Household
F_BIKE	Bike monthly use frequency	Metric	Individual
F_BS	Bike sharing monthly use frequency	Metric	Individual
F_CAR	Car monthly use frequency	Metric	Individual
F_PT	Public transit monthly use frequency	Metric	Individual
GENDER	Gender (M: male, F: female)	Categorical	Individual
HH_CAR	Number of cars	Metric	Household
HH_CHILDREN_U	Number of children (<18 years)	Metric	Household
HH_DRIVLICENCE	Number of driving licences	Metric	Household
HH_SIZE	Number of members (HH_SINGLE: one member, HH_COUPLE: two members, HH_FAM: household with less than six members and with underaged children, HH_FAMNC: household with less than six members and no	Categorical	Household

	children, HH_BIG: household with more than five members)		
HH_WORKERS	Number of workers	Metric	Household
INCOME_AVG	Average monthly income [1000€]	Metric	Household
MODE_USED	Current mode used (CAR: private car as a driver/passenger, PT: public transit)	Categorical	Trip
NO_WORK_DAY	Holiday (Y: yes, N: no)	Categorical	Trip
PT_SEASON_TICKET	Public transit pass (Y: yes, N: no)	Categorical	Individual
TRIP_PURP	Trip purpose (HBW: home-based work, HBEd: home-based education, HBO: home-based other, NH: non home-based)	Categorical	Trip
ZTL_TO_AP	Destination within a limited traffic zone (Y: yes, N: no)	Categorical	Trip

Table 3. Binomial logit estimation for switching intention

Name	Value	Std err	t-test	p-value	
F_CAR	0.038	0.0102	3.77	0.000	***
RET *(ref. STN)*	-0.994	0.2920	-3.41	0.001	***
CS_COST	-0.352	0.1130	-3.13	0.002	**
INCOME_AVG	0.238	0.0782	3.04	0.002	**
HBO *(ref. HBEd)*	-0.927	0.3540	-2.62	0.009	**
PT *(ref. CAR)*	-0.569	0.2180	-2.61	0.009	**
WAH *(ref. STN)*	-0.799	0.3060	-2.61	0.009	**
HH_FAMNC *(ref. HH_SINGLE)*	-0.504	0.2090	-2.41	0.016	*
NO_WORK_DAY *(ref. N)*	0.584	0.2470	2.37	0.018	*
F_PT	0.015	0.0067	2.32	0.021	*
HS *(ref. NE)*	-0.328	0.1440	-2.27	0.023	*
BASE_COST	-0.244	0.1100	-2.23	0.026	*
WOOH *(ref. STN)*	-0.583	0.2660	-2.19	0.029	*
HBW *(ref. HBEd)*	-0.788	0.3610	-2.18	0.029	*
F_BIKE	0.033	0.0154	2.13	0.033	*
CARPERLICENCE	-0.493	0.2780	-1.78	0.076	†
NH *(ref. HBEd)*	-0.783	0.4420	-1.77	0.076	†
BIKE_SHARING *(ref. N)*	0.975	0.5640	1.73	0.084	†
GENDER *(ref. F)*	-0.246	0.1430	-1.72	0.086	†
HH_COUPLE *(ref. HH_SINGLE)*	-0.304	0.1780	-1.71	0.087	†

*Significance codes: *** $p< 0.001$; ** $p< 0.01$; * $p< 0.05$; † $p<0.10$*

Statistics

Sample size:	1050
Init log likelihood:	-898.32
Final log likelihood:	-646.30
Likelihood ratio test for the init. model:	504.04
Rho-square for the init. model:	0.281

Rho-square-bar for the init. model:	0.258
Akaike Information Criterion:	1332.60
Bayesian Information Criterion:	1435.94

Table 4. Validation performance measures of binomial logit and RF

Performance measure	**Class**	**Binomial Logit**	**Random Forest**
Cohen's kappa	-	0.124	0.146
Accuracy	-	75.1%	75.0%
Classification error	-	24.9%	25.0%
Recall	Switch	12.2%	15.7%
	NoSwitch	96.9%	95.5%
Precision	Switch	58.1%	54.8%
	NoSwitch	76.1%	76.6%

Table 5. Number of trips switched and not switched (row percentage in brackets)

	Binomial logit		**Random Forest**		
	Switch	**NoSwitch**	**Switch**	**NoSwitch**	**Total**
CAR	88020 (20%)	355384 (80%)	161699 (36%)	281706 (64%)	443405 (100%)
PT	27230 (12%)	208925 (88%)	77686 (33%)	158469 (67%)	236155 (100%)
Total	115250 (17%)	564309 (83%)	239385 (35%)	440175 (65%)	679559 (100%)

Figures

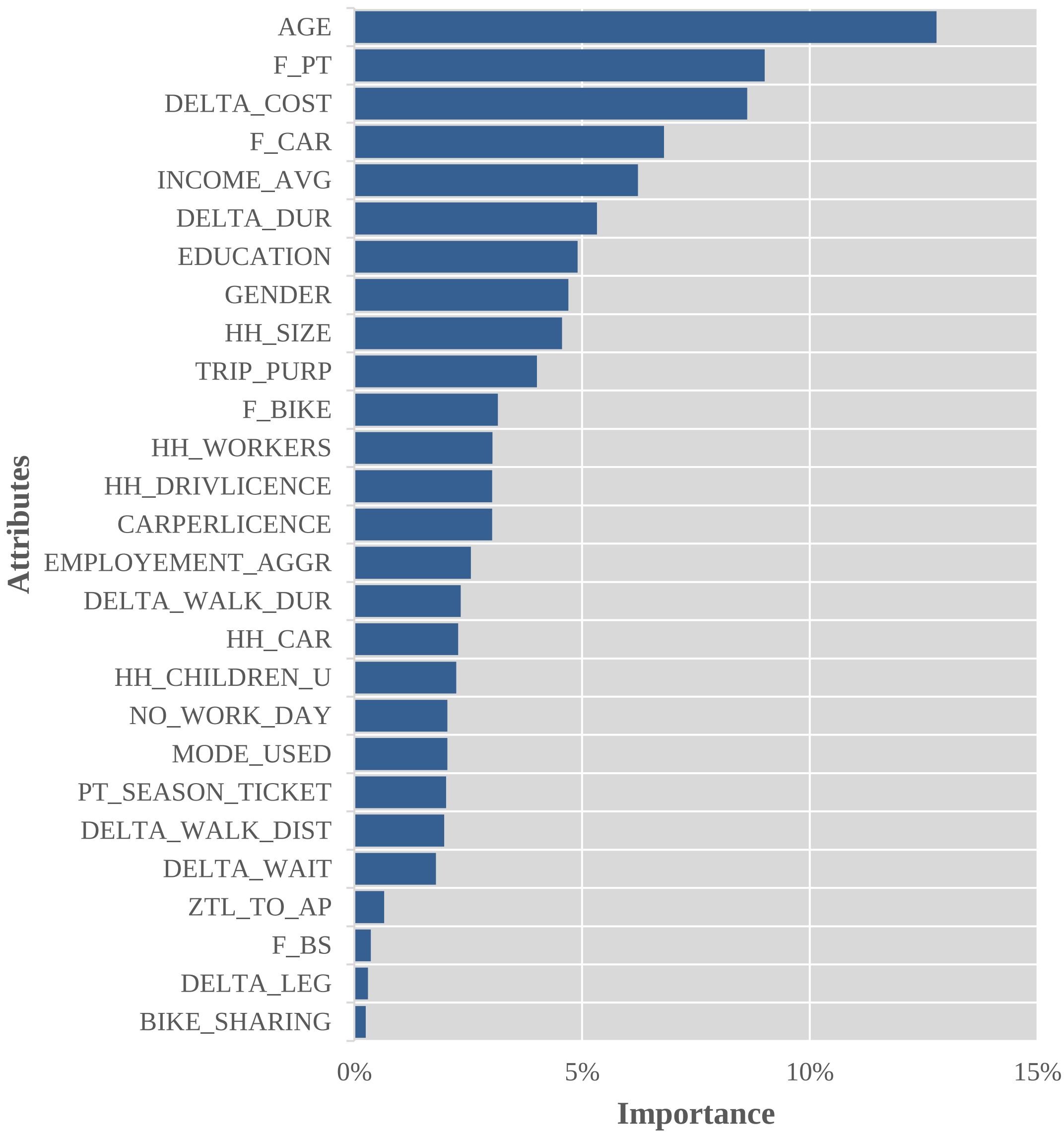

Figure 1. Relative variable importance of Random Forest model

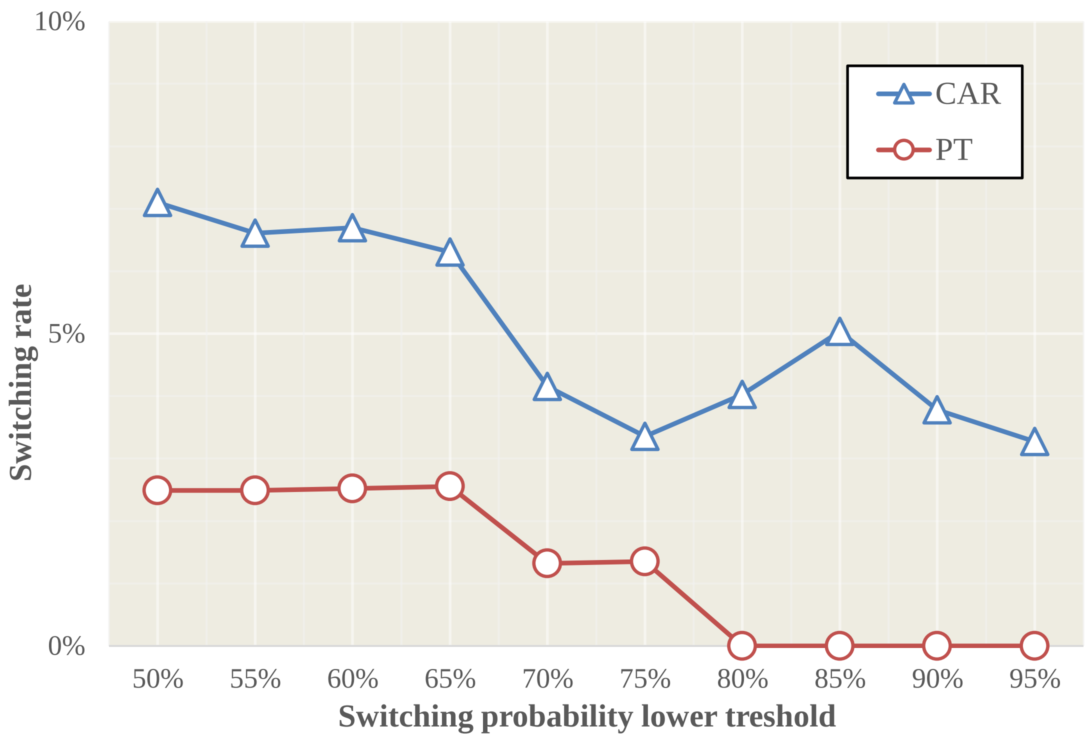

Figure 2. Switching rates predicted using binomial logit

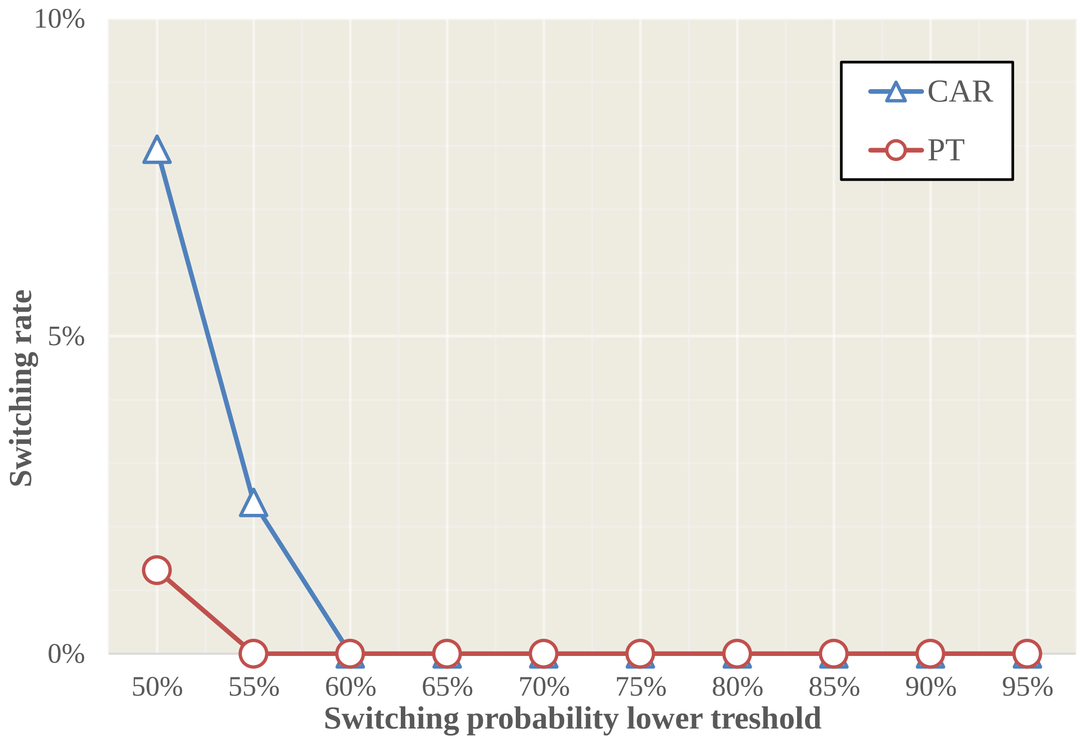

Figure 3. Switching rates predicted using Random Forest