Accounting for intra-household joint travel in agent-based transport simulations

Lucas Javaudin*[1,2], Andrea Araldo[1], and Nicolas Coulombel[2]

[1]Télécom SudParis, Institut Polytechnique de Paris, France
[2]LVMT, ENPC, Institut Polytechnique de Paris, Univ Gustave Eiffel, France

Short summary

Intra-household joint home-based tours — trips in which household members depart together, engage in shared activities, and return together — represent a significant share of daily travel, yet are systematically ignored in transport simulations. Conflating joint and solo tours within a single mode choice framework introduces bias in preference parameter estimates. This paper proposes a three-step methodology to integrate joint tours in agent-based transport models: a Random Forest classifier to identify joint tours, a Multinomial Logit model estimating mode choice specific to joint tours, and a Penalized Logistic Regression for driver/passenger assignment. Applied to the Paris region using household travel survey data, the methodology successfully replicates observed joint tour shares and mode distributions in a synthetic population. The proposed framework enables more reliable evaluation of policies whose impacts differ between joint and solo travel, such as HOV lanes or family transit fare discounts.

Keywords: agent-based modelling; family carpooling; intra-household interactions; mode choice; synthetic population.

1 Introduction

In the Paris region, car trips account for 101 million vehicle-kilometres per day, of which 79 million are driven and 22 million are travelled as a passenger (IDF Mobilités-OMNIL-DRIEAT, 2020). A notable share of this — 24 million vehicle-kilometres, comprising 13 million as passenger and 11 million as driver — involves travelling with at least one other household member, accounting for a total of 2.4 million daily trips. This form of intra-household ridesharing, variously referred to as family carpooling, family ridesharing, or fampool, thus constitutes a significant component of mobility in the region.

Among the 11 million fampool kilometres travelled as driver, 7.3 million (67 %) are performed as part of fully-joint home-based tours, defined as sequences of trips that originate and terminate at home (Shiftan, 1998), in which household members depart together, engage in shared activities, and return together (Ho & Mulley, 2015a). For brevity, fully-joint home-based tours are referred to simply as "joint tours" hereafter. Typical examples include family shopping trips, diner trips as a couple, or trips to the park with young children. These tours are governed by fundamentally different behavioural logics than solo travel: walking speeds may be constrained by the presence of young children, and public transit fares are charged per traveller, whereas fuel costs are shared. Nonetheless, this difference is generally ignored in transport models and simulations, where joint and solo tours are handled with the same mode choice framework. This can introduce systematic bias in preference parameter estimates (Ho & Mulley, 2015a,b).

Several studies have examined mode choice in joint travel settings. Dissanayake & Morikawa (2010) propose a combined revealed and stated preference Nested Logit model to jointly estimate vehicle ownership, mode choice, and trip-sharing decisions among couples in Bangkok. DeLoach & Tiemann (2012) apply a Multinomial Logit model to estimate mode shares across solo driver, fampool, formal carpool, and other modes in the United States, with a focus on fuel price sensitivity. Ho & Mulley (2015a) use a Nested Logit model to jointly estimate intra-household travel arrangements and mode choice in Australia. Picard et al. (2018) model the joint mode choice of both spouses in French couples, explicitly accounting for the possibility of travelling together by car.

While these studies provide valuable behavioural insights, they stop short of producing outputs that can be directly integrated into transport simulations. This paper addresses that gap by proposing

a three-step pipeline — joint tour classification, mode choice estimation, and driver/passenger assignment — designed to produce simulation-ready trip assignments from standard survey and synthetic population data. The resulting framework enables reliable evaluation of policies whose impacts differ between joint and solo travel, such as high-occupancy vehicle (HOV) lanes or family public transit fare discounts.

2 Methodology

The proposed methodology to integrate joint home-based tours in a transport simulation is illustrated in Figure 1. It takes two inputs: a *household travel survey*, reporting travel diaries alongside household, person, and trip characteristics; and a *synthetic population*, providing a random reproduction of the studied population with the trips to be simulated. Three statistical models are estimated on the survey data and then applied sequentially to the synthetic population to produce trip assignments ready to be simulated.

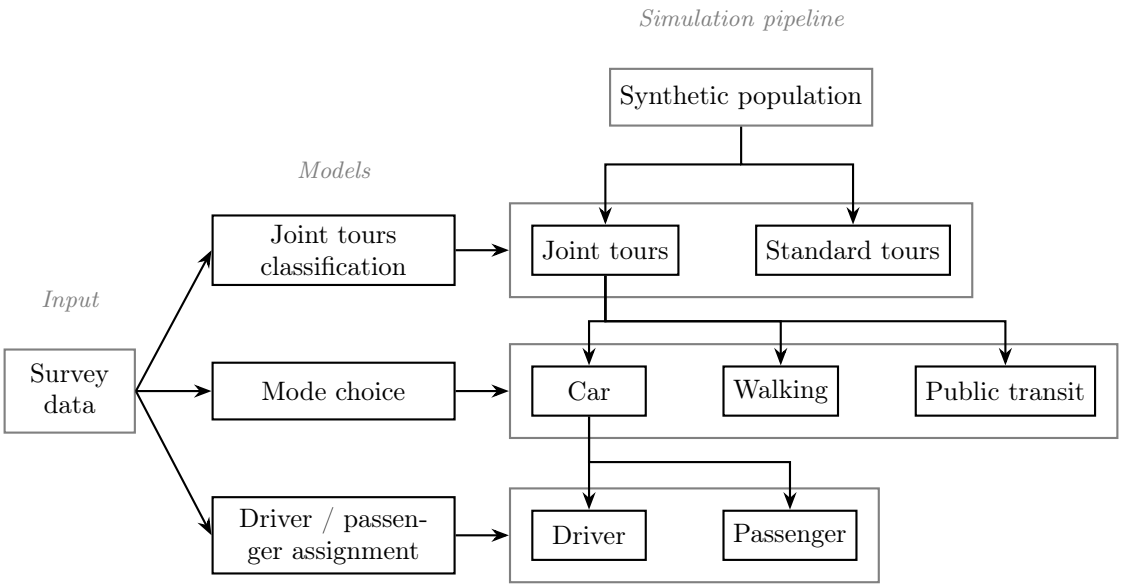


Figure 1: Methodology to integrate joint tours in transport simulations

We assume that household daily activity patterns are fixed, that is, the methodology addresses mode assignment for pre-specified activity chains rather than the generation of those chains. Activity pattern generation is a distinct problem, addressed elsewhere in the literature by, for example, Gliebe & Koppelman (2002), Meister et al. (2005), or Rezvany et al. (2023).

Joint tour classification.

The first model uses a Machine Learning (ML) classifier to determine whether each home-based tour is performed jointly with other household members or alone. Joint tour status is not modelled as the outcome of an individual utility-maximising decision but rather as a latent tour characteristic to be predicted from observable predictors, making a machine learning classifier the more appropriate tool.

When applied to the synthetic population, the classifier produces for each home-based tour a predicted probability of being joint. A Bernoulli draw is then performed for each tour using this predicted probability, determining whether the tour is classified as joint or not.

Mode choice for joint tours.

The second model uses a Multinomial Logit (MNL) model to estimate mode choice among car, walking, and public transit for joint tours. Mode choice is modelled at the tour level, reflecting the empirical observation that travellers rarely mix modes within a tour (see below). The car alternative encompasses both driver and passenger trips, since the mode decision is assumed to be made collectively among tour participants, with the driver/passenger assignment addressed

subsequently by the third model. Crucially, estimating mode choice separately for joint and solo tours avoids the parameter bias introduced by pooled models, which cannot account for the cost-sharing structure of car travel or the per-person fare structure of public transit.

Driver/passenger assignment.

The third model uses an ML classifier to assign each person in car-mode joint tour to either a driver or passenger role. As with joint tour classification, an ML classifier is more appropriate since the goal is to predict the role of each traveller rather than to represent the underlying decision process. In addition, simple natural rules can be used to assign automatically persons to a role, such as assigning as passenger all persons with no driving licence. This step is crucial for transport simulations: only drivers contribute to road congestion, so accurate role assignment is necessary to properly estimate traffic volumes.

Integration into agent-based simulation.

The three models are estimated separately and then applied sequentially to the synthetic population within an agent-based transport simulator. Tours are first partitioned into joint and non-joint groups with an ML classifier. Then, joint and non-joint tours are handled by distinct mode choice models: joint tours use the MNL model presented above, while non-joint tours use the simulator's standard mode choice model. Finally, joint car tours are assigned a driver or passenger role by an ML classifier.

This pipeline enables the simulator to differentiate policy impacts across joint and non-joint travel. For instance, HOV lanes reduce car travel times specifically for joint tours, potentially inducing a modal shift away from public transit for those trips. Conversely, family public transit fare discounts could shift joint tours from car to transit, an effect that cannot be captured in conventional simulations where co-traveller information is absent. Properly modelling these mode shifts in transport simulations is crucial to evaluate urban policies.

One important point should be noted: tour classification is performed at the individual level, as in Ho & Mulley (2015a), meaning the model does not specify which household members travel together. This is a practical necessity, as activity chains in the synthetic population are generated independently across household members. Nevertheless, the model correctly replicates aggregate joint tour shares and mode distributions, which are the quantities of primary interest for policy evaluation.

3 Results and discussion

Case study.

The methodology is applied to the Paris region, home to approximately 12 million inhabitants. The household travel survey used for model estimation is the *Enquête Globale Transport* (EGT), conducted in 2010 (DRIEA, IDFM, 2010). The EGT records household composition, individual sociodemographic characteristics, and all trips made by respondents on the day prior to the interview.

The data comprises 12 452 fully responding households, representing 25 686 individuals and 97 191 trips. Of the 36 791 home-based tours identified, 4066 (11.1 %) are classified as joint with another household member, defined as two or more persons reporting identical activity timings, locations, and purposes.

Table 1 presents mode shares separately for joint and non-joint home-based tours. Only 3 % of home-based tours involve a mode change across trips within the tour, confirming that tour-level mode choice modelling is appropriate. Mode distributions differ substantially between the two groups: car passenger tours are far more prevalent among joint tours than solo tours (33 % vs. 6 %), while public transit and bicycle tours are considerably less common (9 % vs. 26 % and <1 % vs. 3 %, respectively). Walking shares are similar across both groups. These patterns confirm that joint tours warrant a dedicated mode choice model. We restrict the study of the joint tour mode choice model to car, walking, and public transit, since together they account for 97 % of joint tours.

The predictor variables used for the three models are reported in Table 2. The synthetic population

Table 1: Mode shares for joint and non-joint home-based tours

Mode	Joint tours	Non-joint tours
Walking	32 %	34 %
Car driver	23 %	26 %
Car passenger	33 %	6 %
Public transit	9 %	26 %
Bicycle	<1 %	3 %
Motorcycle	<1 %	1 %
Other	<1 %	<1 %
Mode combinations	3 %	3 %

is generated following the methodology of Hörl & Balac (2021).

Table 2: Predictors used in the three models

Variable	Joint tour classif.	Mode choice	Driv./pass. assignment
Tour level			
Number of activities	✓		
Activity duration	✓		
Trip distance	✓		
Activity start time	✓		
Activity end time	✓		
Leisure activity (0/1)	✓	✓	
Shopping activity (0/1)	✓	✓	
Work activity (0/1)	✓		
Education activity (0/1)	✓		
Other activity (0/1)	✓		
Travel time by mode		✓	
Person level			
Age	✓		✓
Age gap to oldest			✓
Gender	✓		✓
Driving licence (0/1)	✓		✓
Professional occupation (4 groups)	✓		✓
Household level			
Household structure (5 groups)			✓
Nb. persons	✓		
Nb. minors	✓	✓	✓
Nb. adult women			✓
Nb. adult men			✓
Nb. cars	✓	✓	✓
Nb. driving licences	✓		✓

Joint tour classification

The first model aims to classify home-based tours as being joint or solo, by learning from the EGT data. A Random Forest classifier is selected, outperforming seven alternative classifiers in cross-validation, with hyperparameters tuned through the same procedure. The Random Forest

naturally handles the mix of binary, categorical, and continuous features in the dataset. Note that for 7110 tours where the person lives alone, the tour is automatically classified as solo, and only the remaining 29 681 tours are passed to the classifier.

Table 3: Performance metrics for the Random Forest model, compared to a stratified random baseline (mean ± std over 5 folds).

Metric	**Random Forest**	**Baseline**
Brier score (smaller is better)	0.071 ± 0.001	0.234 ± 0.004
Average Precision (larger is better)	0.679 ± 0.011	0.138 ± 0.002
F1 score (larger is better)	0.547 ± 0.019	0.133 ± 0.017

After training the classifier on the survey, we apply it to the tours of the synthetic population. Figure 2 shows that the distribution of joint tours in the synthetic population replicates well the one observed in the survey across four key variables: age, household size, first activity start time, and activity duration. The share of total travelled distance attributable to joint tours is also well reproduced: 7.5 % in the survey versus 9.0 % in the synthetic population.

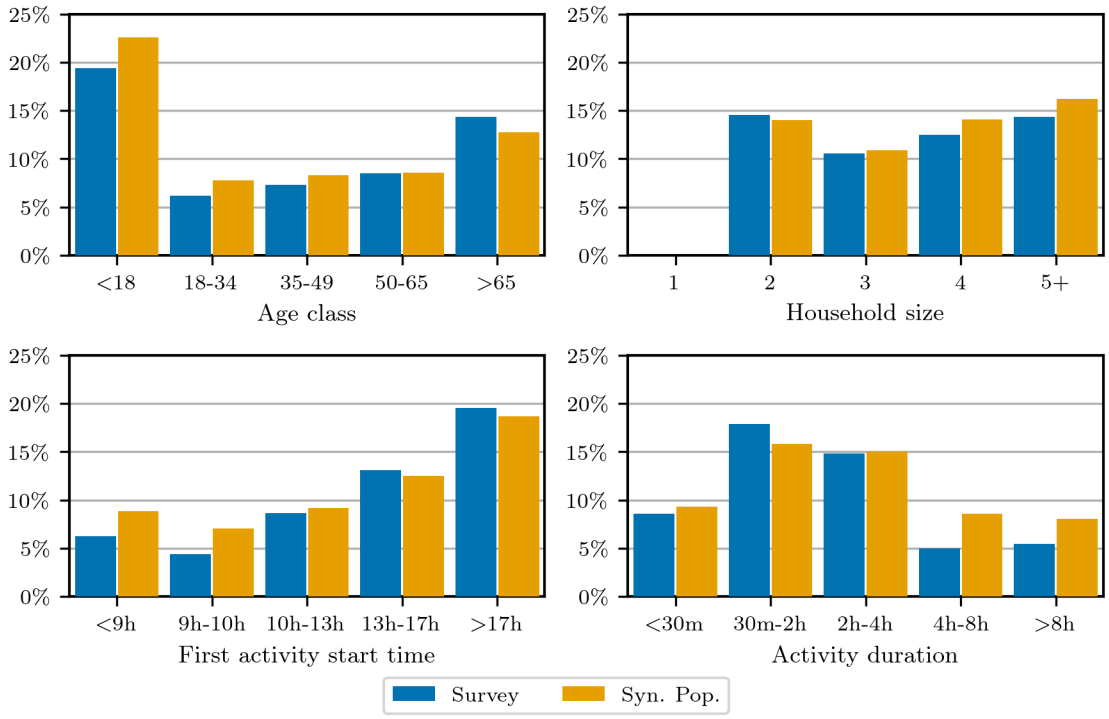


Figure 2: Ratio of joint tours in the travel survey and the synthetic population over four variables

The strongest predictors of joint touring, from permutation importance analysis, are age, total distance between activities, start time of the first activity, total activity duration, and household size. These results are broadly consistent with behavioural expectations: larger households with young children are more likely to take part in joint activities, later trips may reflect leisure purposes that are more typically shared, and activity duration may proxy for the type of activity undertaken.

To assess temporal stability, the model trained on EGT 2010 data was applied to EGT 2020 home-based tours as test set, yielding a still low Brier score of 0.101 suggesting that the determinants of joint touring are relatively stable over time.

Mode choice for joint tours.

For the 3920 joint tours in the EGT data whose mode is walking, car, or public transit, we estimate a Multinomial Logit model of mode choice using the Biogeme Python package (Bierlaire, 2023). Travel time data are derived from three sources: walking travel times are computed from shortest-path distances on the OpenStreetMap network assuming a speed of 3.5 km/h; public transit travel

times are derived from GTFS data; and car travel times are taken from a METROPOLIS2 simulation (Javaudin, 2024), accounting for congestion at the trips' departure times. Walking is used as the reference alternative. The estimation results are presented in Table 4.

Table 4: Multinomial Logit estimation results for mode choice of joint tours

	Alternative		
Variable	Car	Public transit	Walking
Alternative-specific constant	−3.78* (17.3)	−3.62* (14.2)	–
Travel time (hour)	−2.64* (8.09)	−1.52* (6.91)	−4.50* (18.2)
Leisure activity (0/1)	−0.847* (5.89)	−0.737* (3.88)	–
Shopping activity (0/1)	−0.033 (0.229)	−0.607* (3.19)	–
Number of cars	1.27* (15.8)	−0.485* (3.65)	–
Number of minors	−0.077 (1.60)	0.101 (1.66)	–
Model fit			
Number of observations		3920	
Null log-likelihood		−4270.1	
Final log-likelihood		−1772.5	
Adjusted ρ^2		0.762	

Each cell reports the estimated coefficient with the t-statistic in parentheses.
*$p < 0.05$

The alternative-specific constants for both car and public transit are negative relative to walking. For car, this likely reflects fixed costs associated with parking and vehicle access at the destination. For public transit, it may partly capture the fare cost, which is not included as a separate variable in the utility function.

Travel time has a negative effect on utility for all three modes, as expected. The disutility of travel time is largest for walking (4.50 per hour) and smallest for public transit (1.52 per hour), suggesting that travellers are more tolerant of time spent in a vehicle than time spent on foot, a finding consistent with the broader mode choice literature.

Performing at least one leisure activity during the tour reduces the utility of car and public transit relative to walking. This is consistent with some leisure tours that are pedestrian in nature, such as a walk to a park, where walking is itself part of the activity. Performing at least one shopping activity reduces the utility of public transit relative to walking, plausibly reflecting the inconvenience of carrying purchased goods on board. The number of cars owned by the household has a positive effect on car utility and a negative effect on public transit utility, consistent with the role of vehicle availability as a key determinant of mode choice. The number of minors in the household does not reach statistical significance for either alternative, suggesting that, conditional on the other covariates, household composition does not substantially affect mode choice for joint tours.

Driver/passenger assignment.

Among the 2235 joint car tours in the EGT data, 40 % are performed as driver and 60 % as passenger. For 914 of them, the role can automatically be assigned: persons with no driving licence are assigned as passengers and persons who hold the only driving licence in the households are assigned as drivers.

For the remaining 1321 joint car tours, a ML model is used to classify the persons as either driver or passenger, based on the predictors listed on Table 2. Eight classifiers were evaluated using the same cross-validation protocol as for joint tour classification. The Logistic Regression with L2 penalty achieves the best performance in terms of Brier score and Average Precision (Table 5), and is therefore selected for driver/passenger assignment.

Figure 3 presents the standardised coefficients of the Logistic Regression model. Among the strongest predictors, being a woman substantially increases the probability of being a passenger, consistent with empirical evidence that men are more likely to drive in couple travel. Additionally, belonging to a couple-with-children household increases the probability of being a driver, suggest-

Table 5: Performance metrics for the Logistic Regression classifier, compared to a stratified random baseline (mean ± std over 5 folds).

Metric	Logistic Regression	Baseline
Brier score (smaller is better)	0.145 ± 0.017	0.482 ± 0.047
Average Precision (larger is better)	0.833 ± 0.033	0.553 ± 0.032
F1 score (larger is better)	0.825 ± 0.029	0.559 ± 0.028

ing that in such households one parent is often traveling with the children but without the other spouse, and must therefore take the driver role.

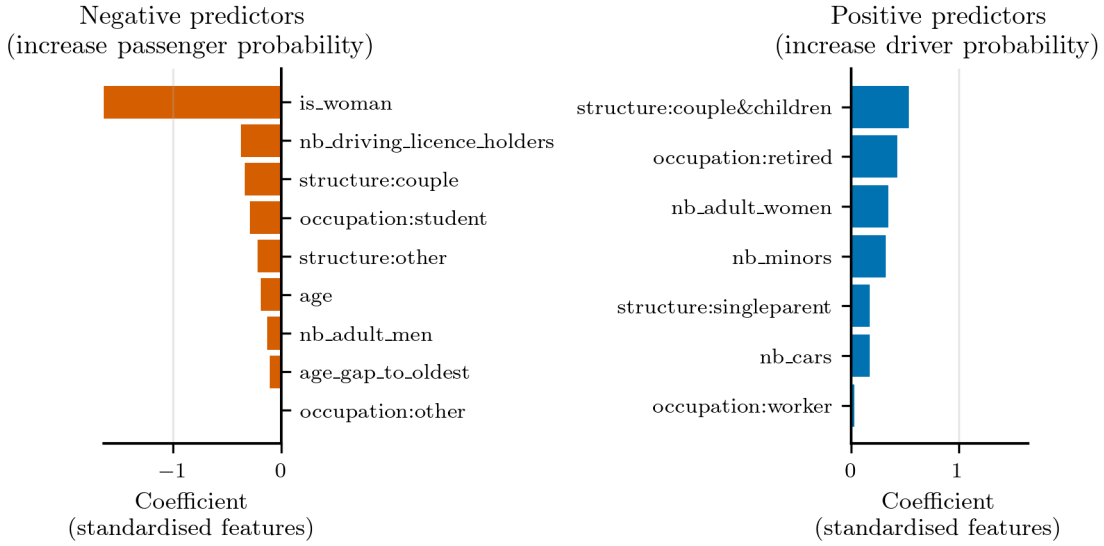


Figure 3: Predictors coefficients of the Logistic Regression

4 Conclusions

This paper proposes a methodology to integrate intra-household fully-joint home-based tours in agent-based transport models. Three statistical models are estimated sequentially on household travel survey data and applied to a synthetic population: a Random Forest classifier to identify joint tours, a Multinomial Logit model for mode choice specific to joint tours, and a Penalized Logistic Regression for driver/passenger assignment.

Properly representing joint tours in transport simulation has direct policy relevance. Policies whose impacts differ systematically between joint and solo travel, such as HOV lanes or family public transit fare discounts, cannot be reliably evaluated in conventional simulations where joint-trip information is absent. The proposed methodology provides the modelling infrastructure needed to assess such policies more accurately.

Several limitations point to directions for future work. First, tour classification is performed at the individual level, meaning the model replicates aggregate joint tour shares without specifying which household members travel together. Extending the methodology to explicitly match household members would enable a richer representation of intra-household coordination. Second, the methodology assumes fixed activity patterns, and integrating joint tour generation with activity scheduling models remains an open challenge.

Acknowledgements

This work has benefited from invaluable advise and feedback from Martin Briand, Pierre Freval, and Tatiana Seregina. Any errors or omissions that may remain are the sole responsibility of the author.

This work has been carried out at the Energy4Climate Interdisciplinary Center (E4C) of IP Paris, which is in part supported by 3rd Programme d'Investissements d'Avenir [ANR-18-EUR-0006-02]. This work was supported by funding from the French National Research Agency (ANR) under the France 2030 program (grant reference: "ANR-24-PEMO-0003").

References

Bierlaire, M. (2023). *A short introduction to Biogeme* (Tech. Rep. No. TRANSP-OR 230620). Lausanne: Transport and Mobility Laboratory, ENAC, EPFL.

DeLoach, S. B., & Tiemann, T. K. (2012, May). Not driving alone? American commuting in the twenty-first century. *Transportation*, *39*(3), 521–537. doi: 10.1007/s11116-011-9374-5

Dissanayake, D., & Morikawa, T. (2010, May). Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested Logit model: Case study in Bangkok Metropolitan Region. *Journal of Transport Geography*, *18*(3), 402–410. doi: 10.1016/j.jtrangeo.2009.07.003

DRIEA, IDFM. (2010). *Enquête globale transport (EGT) - 2010.* ADISP. doi: 10.13144/lil-0883

Gliebe, J. P., & Koppelman, F. S. (2002, February). A model of joint activity participation between household members. *Transportation*, *29*(1), 49–72. doi: 10.1023/A:1012995509777

Ho, C., & Mulley, C. (2015a, February). Intra-household Interactions in tour-based mode choice: The role of social, temporal, spatial and resource constraints. *Transport Policy*, *38*, 52–63. doi: 10.1016/j.tranpol.2014.12.001

Ho, C., & Mulley, C. (2015b, January). Intra-household interactions in transport research: A review. *Transport Reviews*, *35*(1), 33–55. doi: 10.1080/01441647.2014.993745

Hörl, S., & Balac, M. (2021, September). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, *130*, 103291. doi: 10.1016/j.trc.2021.103291

IDF Mobilités-OMNIL-DRIEAT. (2020). *Enquête globale transport (EGT) - 2018-2020.* PROGEDO-ADISP. doi: 10.13144/lil-1581

Javaudin, L. (2024). *Development of a dynamic transport simulator for policy evaluation: Applications to ride-sharing and low emission zone in paris* (Unpublished doctoral dissertation). CY Cergy Paris Université.

Meister, K., Frick, M., & Axhausen, K. W. (2005, September). A GA-based household scheduler. *Transportation*, *32*(5), 473–494. doi: 10.1007/s11116-005-5325-3

Picard, N., Dantan, S., & de Palma, A. (2018, March). Mobility decisions within couples. *Theory and Decision*, *84*(2), 149–180. doi: 10.1007/s11238-017-9648-2

Rezvany, N., Bierlaire, M., & Hillel, T. (2023, December). Simulating intra-household interactions for in- and out-of-home activity scheduling. *Transportation Research Part C: Emerging Technologies*, *157*, 104362. doi: 10.1016/j.trc.2023.104362

Shiftan, Y. (1998, January). Practical Approach to Model Trip Chaining. *Transportation Research Record*, *1645*(1), 17–23. doi: 10.3141/1645-03