A Multinomial Logit Framework for Secondary Location Choice with Attractiveness and Joint Mode Selection

Andrew Ding[1] and Milos Balać*[2]

[1]MSc Student, Department of Computer Science, ETH Zürich, Switzerland
[2]Senior Researcher, Center for Sustainable Future Mobility, ETH Zürich, Switzerland

Short summary

Agent-based transport simulations typically rely on a synthetic population and their travel demand over an average day. One important part of the travel demand generation is the assignment of secondary activity locations within trip chains. This work proposes an adapted multinomial logit framework for secondary location choice that jointly selects mode and location through importance sampling. The model incorporates location attractiveness from employment counts and distance decay parameters estimated from a national travel survey. Candidate locations are filtered by purpose and are sampled within spatially adaptive buffers, with a mixed local/global sampling scheme to preserve long-distance trip representation. The mode share is calibrated through an iterative ratio adjustment technique. Evaluated against the travel survey distance distributions, the proposed model generally reproduces similar distance distributions up to the 95th percentile of distances.
Keywords: Activity-based modeling, discrete choice, location choice, multinomial logit, transport simulation, eqasim

1 Introduction

Synthetic travel demand is a cornerstone of any agent-based transport simulation. It is generated in a sequence of steps based on various models that assign seuqence, type, location, and duration of all activities for each person. While workplace and education locations are typically obtained from commuting matrices, secondary location choice usually requires a different approach due to a limited available data. Within the activity-based transport modeling literature, there have been various methods to model the choice of secondary activity locations. The relaxation-discretization algorithm, introduced by Hörl & Axhausen (2023), and used in the synthetic demand generation pipeline, eqasim Hörl & Balac (2021), assigns secondary locations by iteratively sampling from fitted distance distributions and discretizing to the nearest candidate location. The CARLA algorithm, introduced by Petre et al. (2025), builds on the relaxation-discretization method by implementing a divide and conquer strategy, recursively splitting the chain into smaller subsegments and utilizing a weighted combination of location attractiveness and distance deviation to select locations. While CARLA introduces location attractiveness into the assignment process, both methods rely on observed distances as direct inputs.

In addition, several studies have utilized attractiveness-weighted distance decay approaches to assign secondary locations. Adiga et al. (2015) used capacity and travel times to select shopping locations between anchor locations in generating a synthetic population of the United States. Barrett et al. (2009) adopted a similar approach, calibrating the distance decay parameter from observed median trip lengths. Tozluoglu et al. (2022) extended this by estimating distance decay parameters specific to trip purpose and transport mode, and employing spatial filtering to constrain the candidate set. Within a Multi-Agent Transport Simulation, MATSim (Horni et al., 2016), Horni et al. (2009) took a different approach, embedding location choice within an iterative coevolutionary simulation, using time-geographic constraints and capacity restraints to define feasible choice sets.

The multinomial logit (MNL) model (Luce, 1959; McFadden, 1973), further developed for travel demand modeling by Ben-Akiva & Lerman (1985), frames location selection as a probabilistic decision in which selection probabilities are derived from the relative utilities of available alternatives, allowing for the incorporation of both distance decay and location attractiveness within a single utility specification. A practical challenge in applying the MNL to location choice is the sheer size of the candidate set, which may contain thousands of locations. McFadden (1977) proved that by

adding a correction term to the utility function, consistent estimates can be obtained by instead selecting from a weighted sample of candidates (Guevara & Ben-Akiva, 2013).
This paper proposes a methodology for secondary location choice built around a multinomial logit (MNL) framework that jointly selects mode and location. The model combines mode- and purpose-specific distance decay parameters with location attractiveness in a single utility function. Candidate locations are filtered by General Classification of Economic Activities (NOGA) codes to match the trip purpose type (e.g. shopping) and sub-type (e.g. groceries), and are sampled using a spatially adaptive scheme that controls the ratio of local to global sampling in order to include long-distance trips. Mode share is achieved through calibrated sampling rates per mode, conditioned on household vehicle and bicycle ownership. For multi-stop chains, an iterative approach decomposes the chain into subproblems, predicting locations in order of decreasing activity duration with each assigned location serving as an anchor for subsequent predictions. The methodology is tested using the 2015 Swiss nation travel survey called Microcensus and business census data (STATENT) containing location, type, and number of employees for each business, and evaluated against observed distance distributions and the existing implementation in eqasim.

2 Methodology

Datasets

Two key datasets were used to implement this method, the 2015 Microcensus (*Verkehrsverhalten der Bevölkerung*, 2017) and 2015 STATENT (*Swiss structural business statistics (STATENT) 2015-2021*, 2023) data. Datasets from 2015 were used, as it is the latest available Microcensus dataset that was not affected by the COVID-19 pandemic. STATENT was used primarily as a proxy for location attractiveness as it was assumed that businesses with more employees would signify a more visited location. The dataset comprises 676,068 establishments and includes information on location, business type (purpose), and number of employees.
The Microcensus dataset provides detailed trip information (such as distance, locations, purposes, modes, etc.), which helps to deduce existing trends within secondary location selection. The Microcensus data was preprocessed by aggregating and generalizing the modes and purposes of trips to specific categories. Household-level attributes including car ownership and bicycle ownership were merged to each trip. Persons with any trip containing an unknown mode or purpose were removed entirely to ensure complete trip chains. Additionally, only trips which have full chains that begin and end at the home location were retained. Finally, weekend trips and loop trips with zero crowfly distance were removed. After preprocessing, a total of 138,740 trip segments from 38,657 individuals remain in the dataset.

Activity chains

An activity chain is a sequence of trips that start at a primary location, visits one or more secondary activity locations, and ends at the same or a different primary location. The primary and secondary activity types are defined in the table below.
For inputs to the model, these activity chains were taken from the Microcensus with the secondary locations removed, leaving only the primary location as known locations.

Table 1: Primary and secondary activity types

Primary	**Secondary**
Home	Shop
Work	Leisure
Education	Other

Candidate locations

STATENT was used to provide locations for businesses such as restaurants, shops, leisure locations, etc. For attractiveness, the total number of employees were used as a proxy variable. STATENT locations were aggregated by coordinates and NOGA 2008 codes, so that a location with many sim-

ilar businesses (i.e., a shopping center with many stores) would be considered as a single candidate, with total employees summed to represent the greater attraction to these locations.
OpenStreetMap (OpenStreetMap contributors, 2017) data was used to provide park locations and population census data, STATPOP (*Statistik der Bevölkerung und Haushalte (STATPOP): Variablenliste*, 2024) data was used for the location of private residences (for visits). For parks and private residences, the area of the park in square meters and the number of residents were used as measures of attractiveness. To prevent large outliers from dominating, these values were logged and rounded to a whole integer.
Transit accessibility was also used as an attractiveness measure for trips which utilize public transit. For each candidate location, a 400 metre buffer was intersected with all bus stops, while an 800 metre buffer was intersected with all rapid transit stops (i.e. rail, tram, metro, and ferry), which are distances commonly used in the literature to represent reasonable walking access to public transit (El-Geneidy et al., 2014). The average hourly departures of all stops falling within these buffers were summed to produce a single transit accessibility score for each location. Candidate locations were assigned a purpose based on either the NOGA code prefix or the location type (e.g., park or residence), and further filtered by more specific categories, such as leisure or shopping. After aggregation and NOGA-based filtering, in total 1,688,054 secondary locations were included.

Table 2: Candidate location filtering by trip purpose and sub-type.

Purpose	**Sub-type**	**NOGA Prefixes / Data Source**
Shop	Groceries	4711, 472, 4781
	Other retail	451, 4719, 474, 475, 476, 477, 478
Leisure	Restaurant	56
	Active sports	931, 8551
	Cultural	90, 91, 932, 8552
	Religion	9491
	Parks	OpenStreetMap
	Visits	STATPOP
Other	Shopping	47, 451
	Arts, Entertainment and Recreation	9
	Education	85
	Parks	OpenStreetMap
	Restaurants	56

Distance decay estimation

Distance decay parameters were estimated for each mode, purpose pair, and round-trip combination which follows the approach introduced in Tozluoglu et al. (2022). A purpose pair is defined by the purpose of one leg of a trip such as: work → shop. Purpose pairs with fewer than 250 observations were collapsed into a generic "other_pair" category for each mode. The roundtrip indicator denotes if the trip returns back to the first location (see Figure 1). For each of these (mode, purpose pair, roundtrip) combinations, an exponential decay function $f(d) = |\beta| \cdot e^{\beta d}$ was fitted via nonlinear least squares regression to the normalized histogram of observed crowfly distances.



Figure 1: Example activity chains illustrating roundtrip (left) and non-roundtrip (right) classifications. Primary activities (orange) serve as anchors, while secondary activities (blue) are the predicted locations.

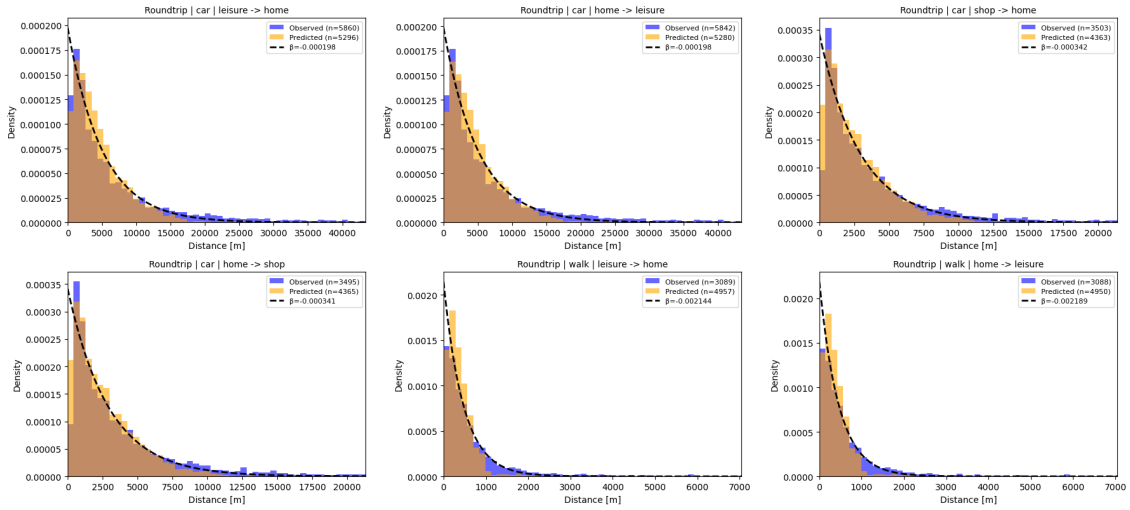


Figure 2: Observed and predicted distance densities with fitted decay parameters (β) for the highest sample size (mode, purpose pair, roundtrip) combos.

Mode calibration

Mode is jointly sampled with location by sampling K_m secondary locations for each mode. K_m is calibrated such that chosen modes match the true mode share in the microcensus, conditioned on the individual's car and bike ownership (see Figure 3). To calibrate these weights an iterative ratio adjustment technique is used. From initial mode weights, the full model is run and the predicted mode shares are determined. Then, based on the predicted shares, the weights are adjusted using the ratio rule: $w_{\text{new}} = w_{\text{old}} \times \left(\frac{\text{target}}{\text{predicted}}\right)^{\alpha}$ where α is the learning rate. This process is repeated for 10 iterations, or if all mode percentages change by only less than 0.25 percentage points.

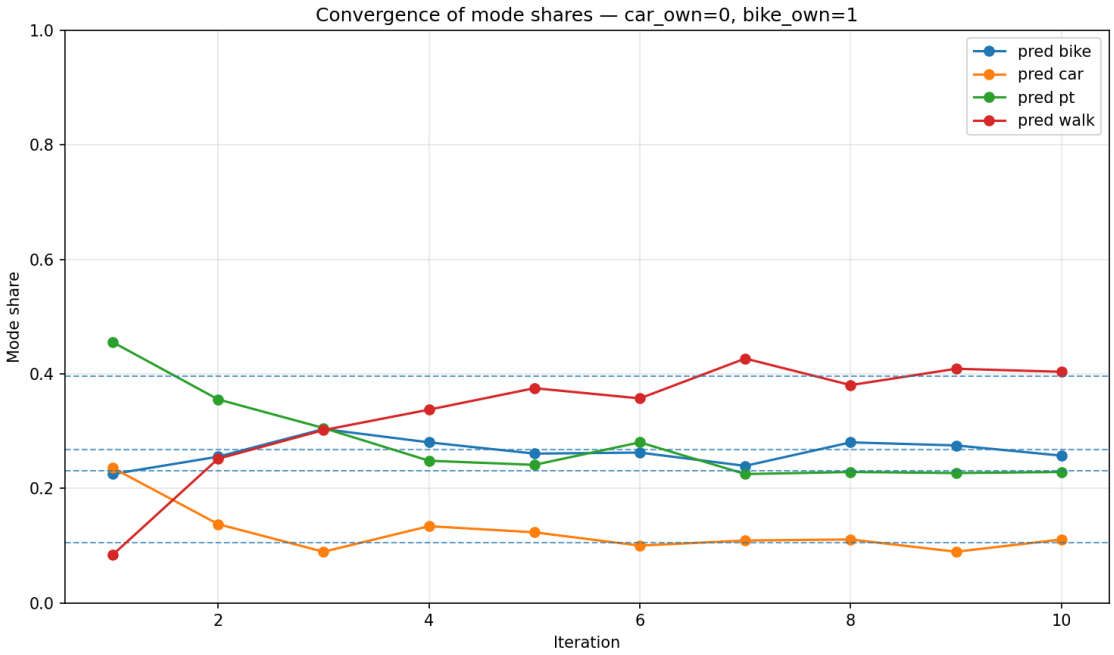


Figure 3: Convergence of Mode Calibration Weights by Iteration, for individuals without car ownership but with access to a bicycle.

Model

The model follows an iterative multinomial logit (MNL) model estimated through importance sampling. Instead of evaluating all possible locations in the choice set, the model draws a subset of K_m candidates per mode.

In the simple case of a single secondary location (primary → secondary → primary), selection is dependent on if it is defined as a roundtrip or not. First, all locations are filtered by purpose type and sub-type (see Table 2) and are sampled using a buffer distance. Then, to determine the secondary locations, we loop though all modes, sampling K_m secondary locations each, weighted by attractiveness. In the case where the mode is public transport, all locations with zero PT accessibility, are filtered out.

In order to sample locations, a parameter q is used to control the scope of candidate sampling. $q\%$ of candidates are sampled within a distance that covers $q - th$ percentile of visited secondary locations for a given mode and purpose within the Microcensus. For round-trips, the buffer will be a circle around the single anchor location which covers $q\%$ of all locations in the microcensus which have the same mode / purpose combination. For non-roundtrips, the distance between the two anchor activities is multiplied by an empirically calibrated scalar conditioned on mode and purpose pair to produce a buffer radius, similar to the method introduced in Tozluoglu et al. (2022). This scalar is set such that the overlapping region of the two resulting buffers captures at least the q^{th} percentile of observed secondary stops for the given mode and purpose pair. A circle with this radius is then drawn around each anchor, and candidate locations are those falling within the intersection of the two circles. To allow for occasional long trips, we also sample $1 - q\%$ of locations outside the buffer. Since longer trips are typically only to more attractive locations, the out-of-buffer samples are drawn weighted by cubed attractiveness, to emphasize highly attractive locations. For long walking trips (i.e., when anchors are more than 3 km apart), all walk-mode sampling slots (K_{walk}) are reassigned to PT.

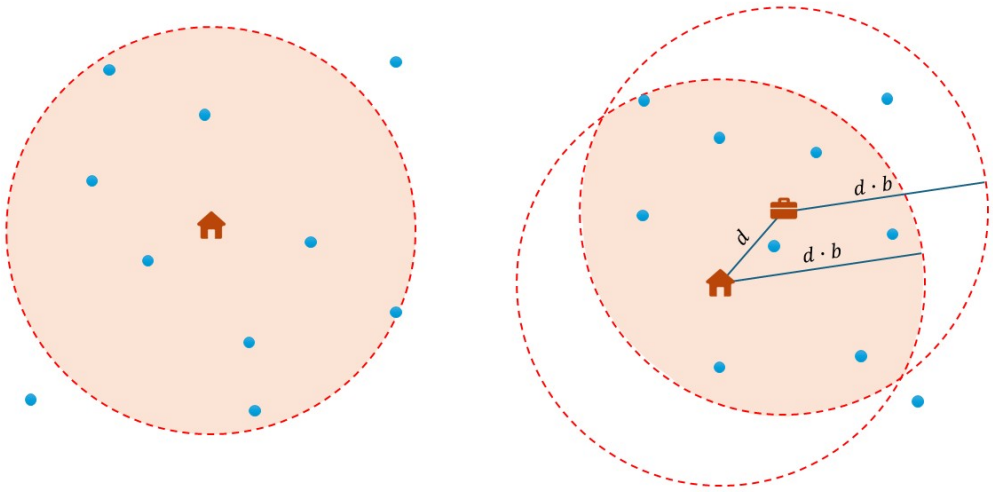


Figure 4: Example buffer distances for roundtrip (left) and non-roundtrip (right), where buffers are constructed to capture the 80th percentile ($q = 0.8$) of observed visited secondary locations within a (mode, purpose pair) combination. d represents the distance between two anchor activities, and b represents the scalar.

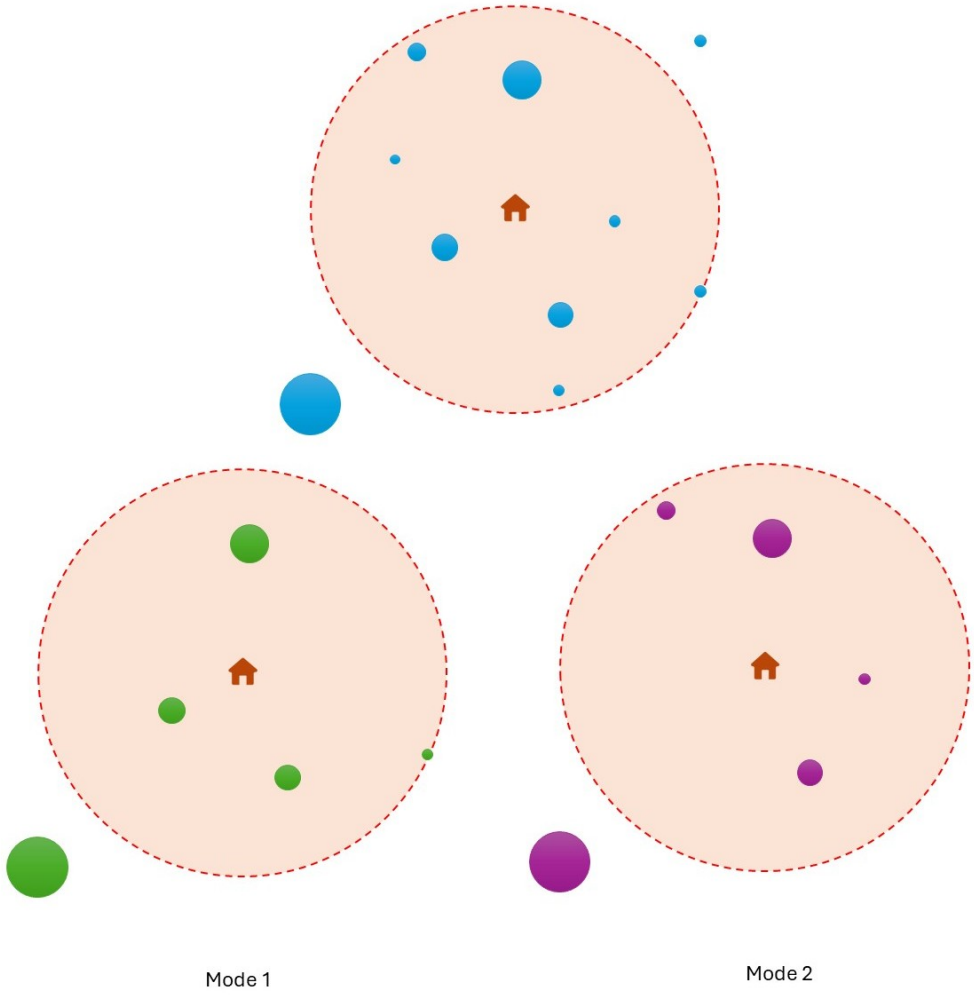


Figure 5: Example roundtrip sampling with $q = 0.8$ and $K_{m_1} = K_{m_2} = 5$. Marker size indicates attractiveness; locations are sampled proportionally to attractiveness, with 80% of samples are drawn within the buffer and 20% from outside.

Once the candidate set of locations are determined, we compute the utilities across all sampled secondary locations across all modes. For each mode, there will be K_m sampled locations, which allow certain modes to be sampled more often than others in order to match the true mode distribution. The utility formula is as follows:

$$V_i = A_i + \beta_1 d_{1,i} + \beta_2 d_{2,i} - \ln(\pi_i) \tag{1}$$

where $A_i = \ln(1+\text{emp}_i)$ is the log-attractiveness for all non pt-modes, else $A_i = \ln(1+\text{emp}_i) * (1+pt_{attr})$ where pt_{attr} is public transport accessibility, d_1 and d_2 are the distances from the proposed location and the two anchors, and β_1 and β_2 are the corresponding distance decay parameters. π_i is the importance sampling probability where $\pi_i = A_i / \sum_j A_j$, which is used to draw candidate i per mode. The $-\ln(\pi_i)$ term corrects for the non-uniform sampling of candidates (McFadden, 1977; Guevara & Ben-Akiva, 2013). From the computed utilities, the final secondary location is selected using the MNL choice probability:

$$P_i = \frac{e^{V_i}}{\sum_j e^{V_j}} \tag{2}$$

where the sum is over all sampled candidates across all modes. A single location is then drawn from this distribution as the selected secondary location.

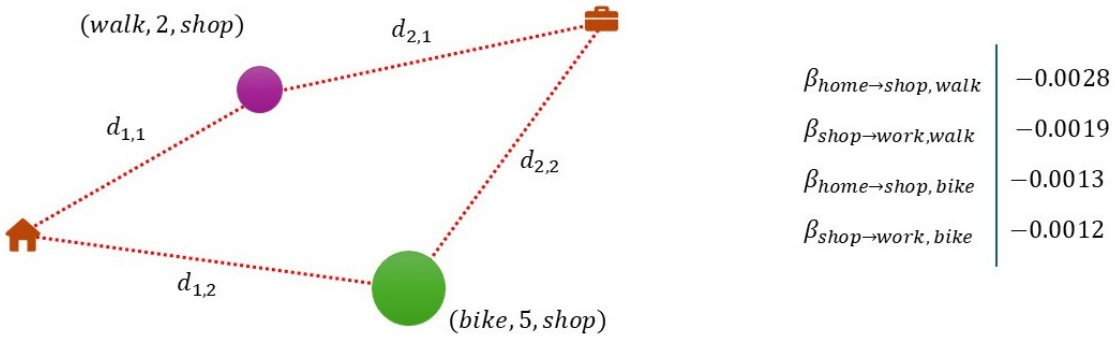


Figure 6: Example utility calculation for a non-roundtrip scenario. Secondary locations are labeled with (mode, attractiveness, purpose). For each sampled location, the appropriate distance decay parameters (β) must be looked up independently for each leg.

In the case with multiple secondary locations within a chain, an iterative approach is utilized. In our approach, we first jointly select the mode and location for the stop in the chain with the longest activity duration. If two or more stops have the same activity duration, then the location that is closer to the middle is selected. If the locations are both in the middle, then the location closer to the start is selected. Once this secondary location and mode is jointly selected, it is then treated as an anchor and the location with the next longest activity duration is selected, using now the two nearest anchors in the chain, similar to the approach described in Petre et al. (2025). In addition, once the mode is selected in the first location, this mode is used as a known parameter for the remainder of the selections. In this case, the remaining locations utilize the initially set parameter for the number of candidate locations.

3 Results and discussion

Mode share

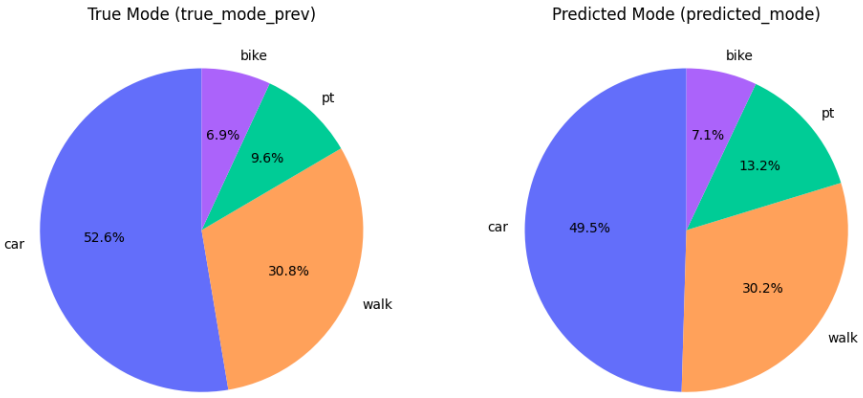
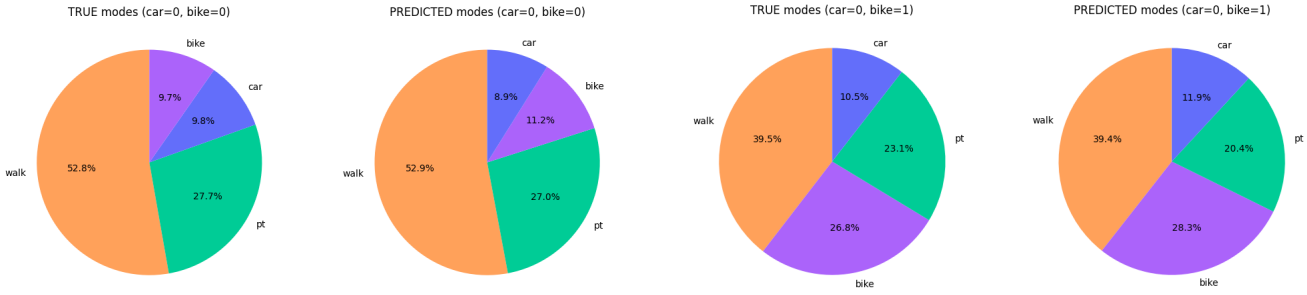


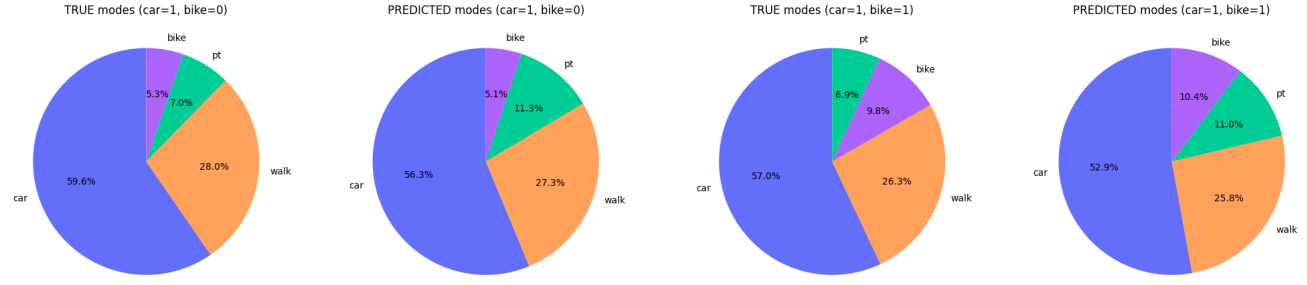
Figure 7: Mode share across all predictions

The calibrated mode shares are shown in Figures 7 and 8. In general, the calibration generally produces consistent mode share distributions compared to the Microcensus data. Some deviations exist when splitting the plots by mobility tool ownership, notably the over prediction of public transit in cases (c) and (d). This may be attributed to the fact that we reallocate walk trips to public transit if the distance between two anchors is greater than 3 km, resulting in the over prediction.



(a) Own neither car/bike

(b) Own bike, no car



(c) Own car, no bike

(d) Own bike and car

Figure 8: Mode share comparisons by ownership

Distance distributions by purpose pair

Figure 9 shows the predicted vs observed distance distributions by purpose pair. The dashed lines represent the mean and the dot-dashed lines represent the median (both values are for the clipped 95th percentile data). Generally, the predicted values closely match the observed values. The main exception is the leisure → home & home → leisure purpose pairs, where the model overpredicts

short distances. As leisure activities encompasses visits, while an individual may visit a friend in another town; due to the high density of nearby residences the model is often drawn toward closer homes, which could explain the underestimation of trip distances for this purpose pair. This effect is shown more clearly in Figure 10.

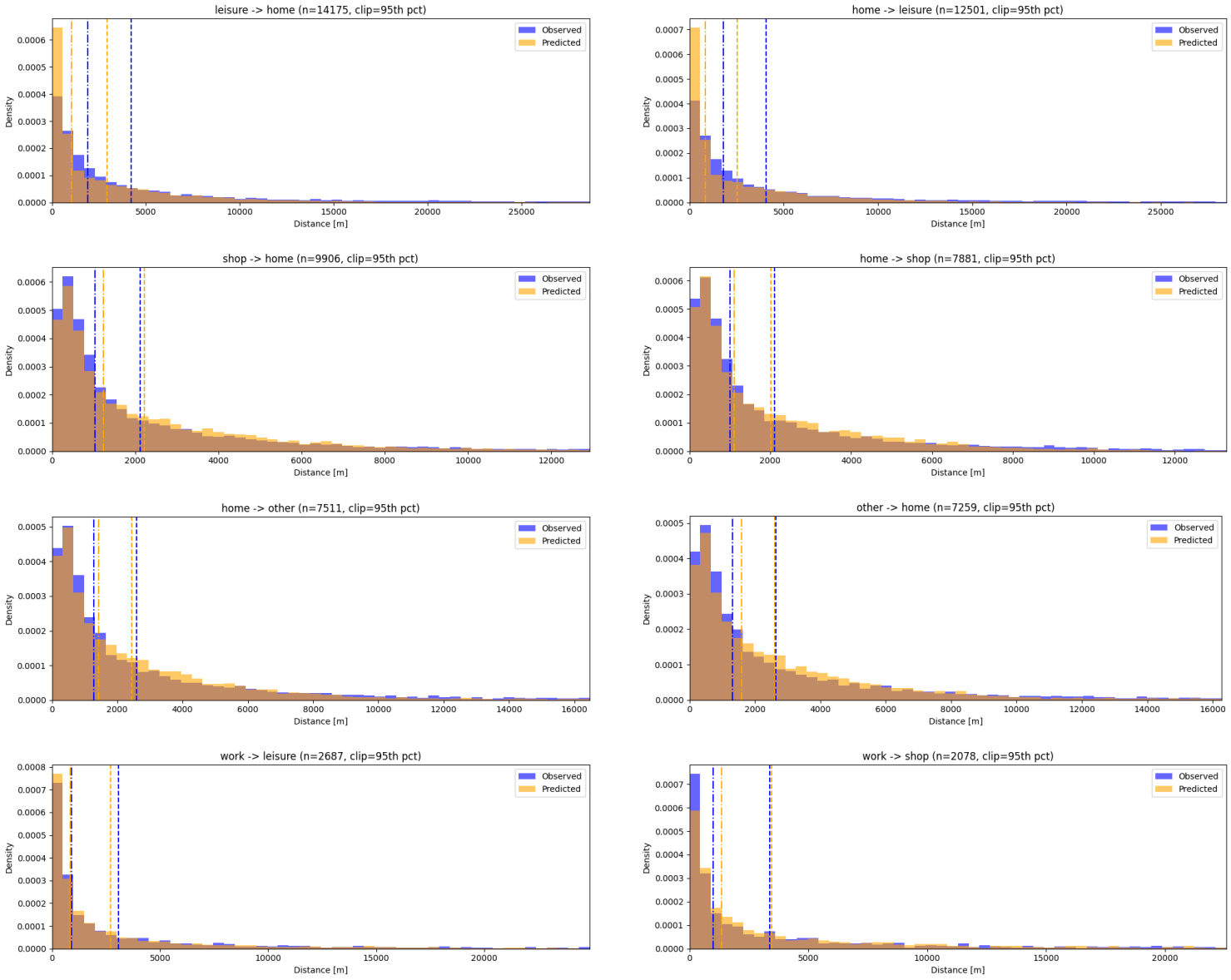


Figure 9: Predicted vs. observed distance distributions by purpose pair. Only purpose pairs with $n > 2000$ are shown.

Another limitation of the current model is that it cannot accurately predict extreme outlier trips, as these locations will be severely penalized due to the large distance. As shown in Figure 11, locations far from the anchors are heavily penalized by the distance component of the utility function, causing predicted distances to fall off more sharply than observed. As a result, the model is able to reasonably predict up to approximately the 95th percentile of the observed distance distribution.

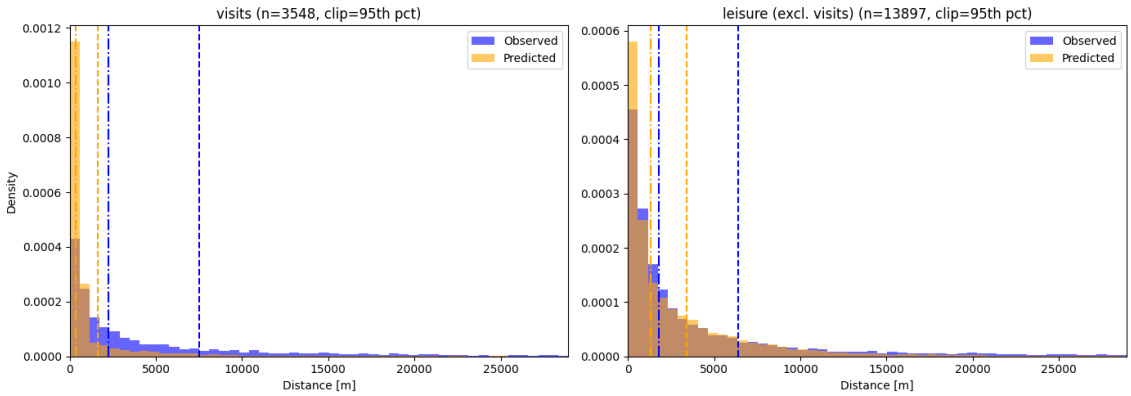


Figure 10: Distance distributions for visits, and leisure activities excluding visits.

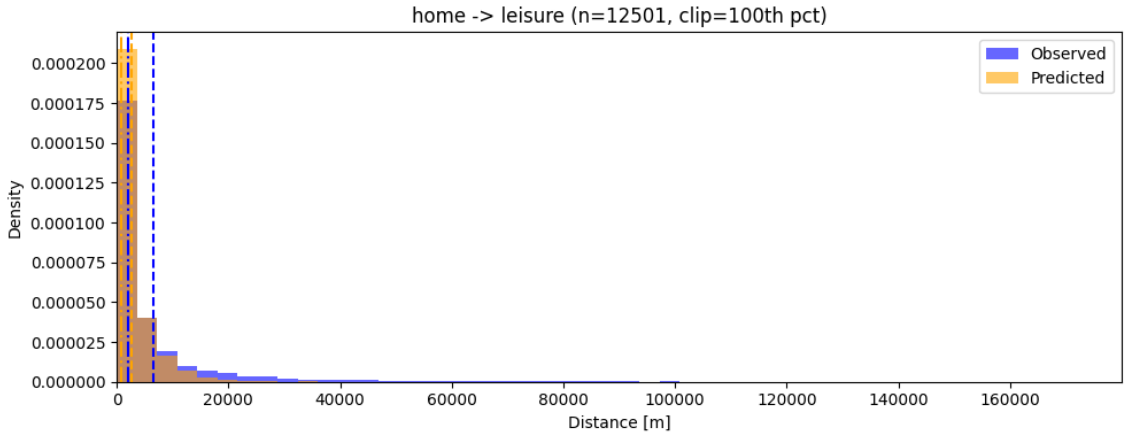


Figure 11: Full predicted vs. observed distance distribution for the home → leisure purpose pair.

Distance distribution by mode/roundtrip

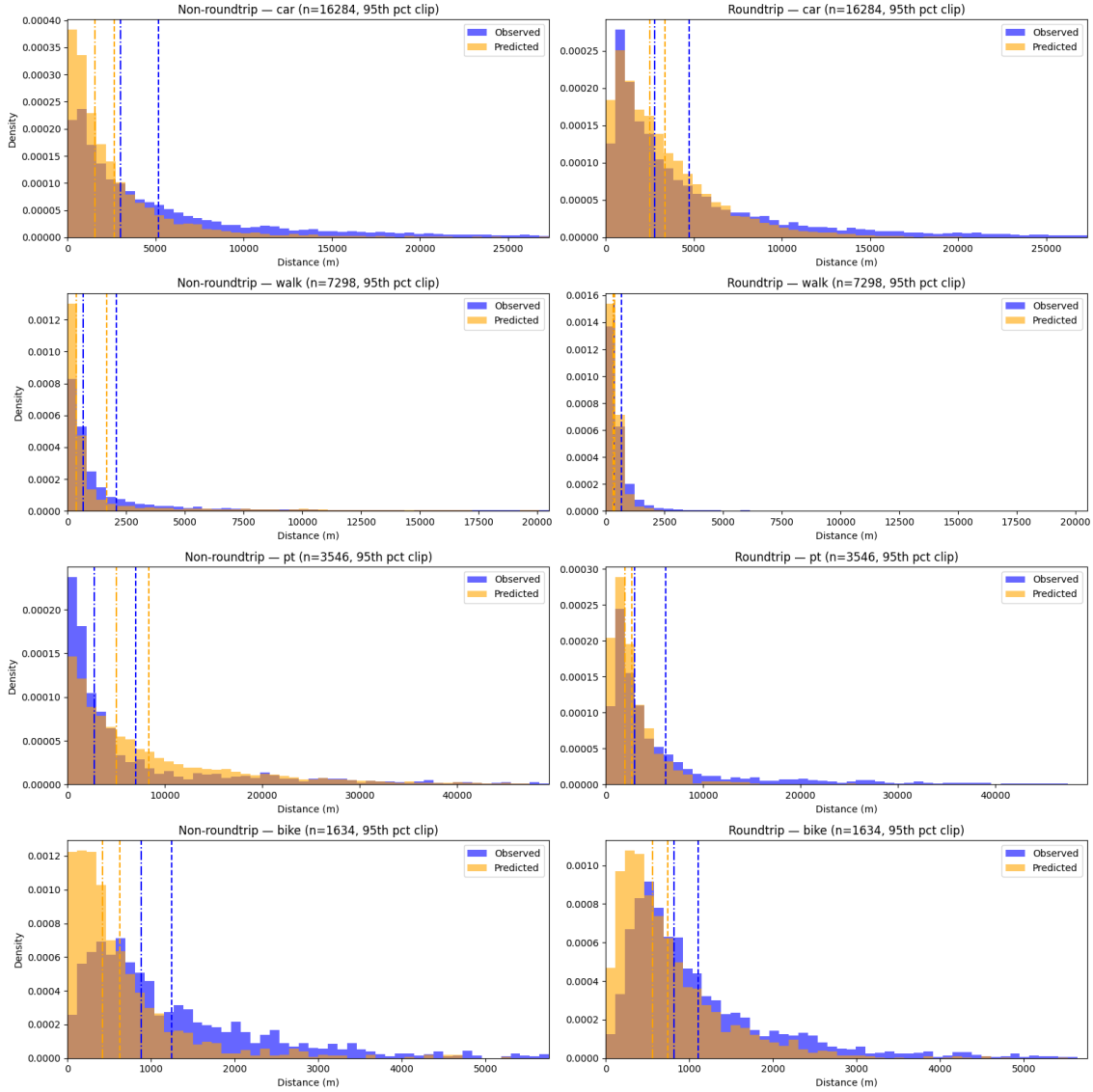


Figure 12: Predicted vs. observed distance distributions by mode / roundtrip combination.

Figure 12 shows the predicted vs observed distance distributions split by mode and round-trip combination. The fit is noticeably weaker compared to the purpose pair distributions, likely due to the greater uncertainty in mode selection. For the car, walk and public transit modes, the predicted results seem to generally follow the observed distribution, albeit overpredicting shorter trips and underpredicting longer trips in the tail of the distribution. Bike trips show the largest discrepancies, where it appears to severely overpredict short trips. This is likely because the model applies the same distance decay function shape to all modes, and therefore cannot capture the fact that below a certain distance, walking is preferred over cycling.

Comparison to the current approach in eqasim

Here we compare our approach to the current secondary location assignment method used in eqasim as developed by Hörl & Axhausen (2023). To do this we generated a 1% sample of the synthetic Swiss population, containing 79,386 activity chains (102,915 secondary stops), stripped the secondary stop locations, and ran the chain-based model on these chains. To compare the two approaches, we spatially joined both sets of predicted coordinates to the secondary locations dataset using a nearest neighbor match (50m threshold), allowing us to retrieve the attractiveness of each selected location.
On average, the predictions made by our proposed methodology have a higher attractiveness score

compared to the original model. Selected locations from our model have an average attractiveness of 1.78 while the current eqasim model has an average attractiveness of 1.21, which is a 47% increase.

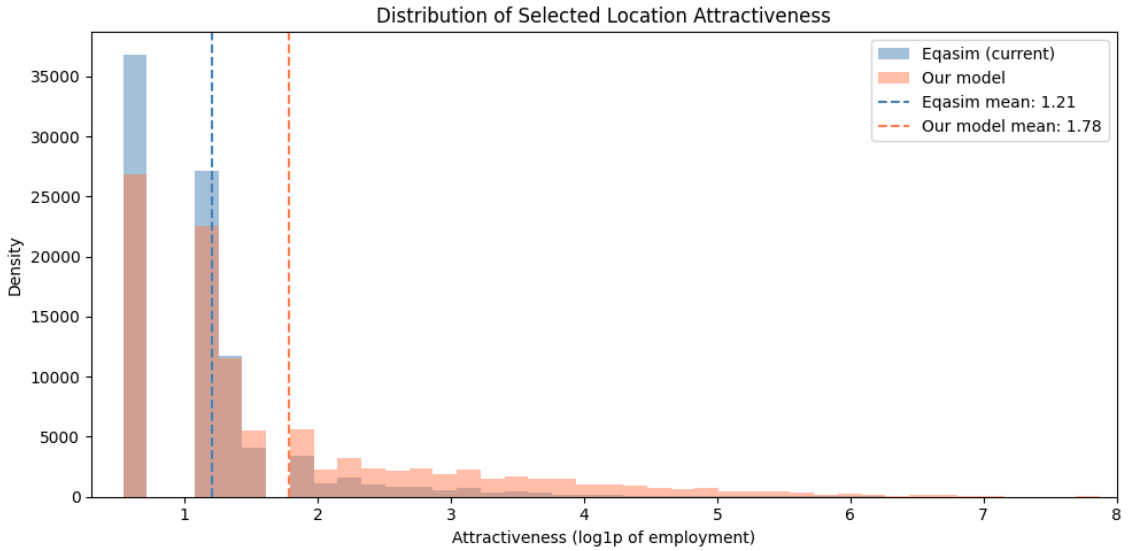


Figure 13: Comparison of Selected Location Attractiveness between the two Models.

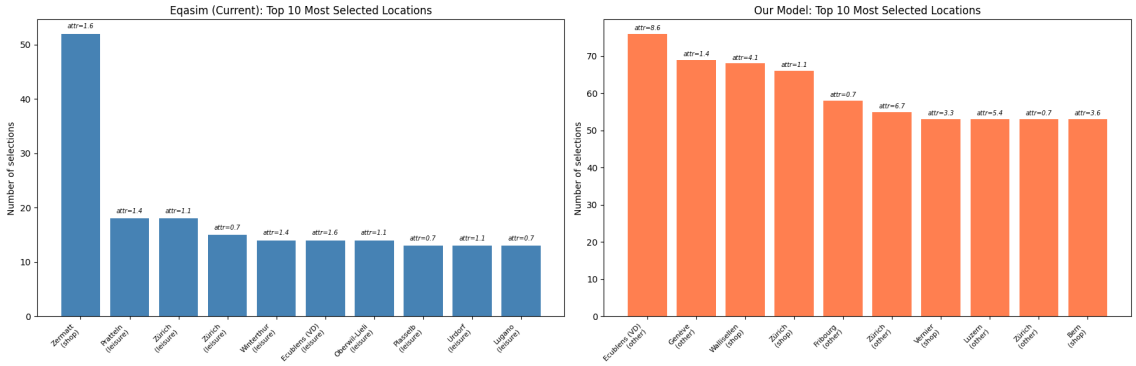


Figure 14: Top Selected Locations and their Attractiveness per Model.

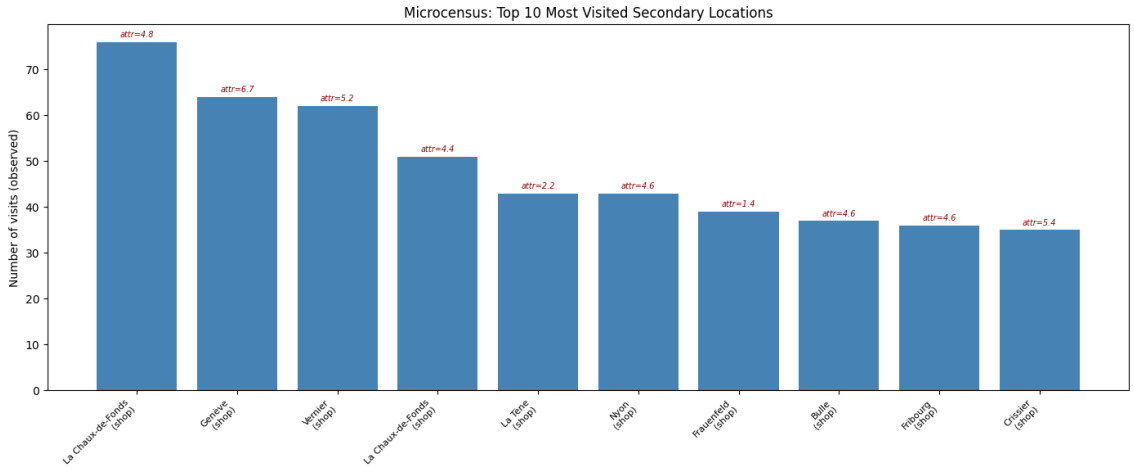


Figure 15: Top Selected Observed Locations and their Attractiveness in the Microcensus.

The top 10 most visited Microcensus destinations have a mean attractiveness of 4.39. The current eqasim model, produces a mean attractiveness of only 1.14 among its top 10 selected destinations,

which is similar from the overall average attractiveness of 1.21 across all selected locations. The negligible difference indicates that the model's most frequently chosen destinations are no more attractive than a randomly selected destination, suggesting that its location choices are independent of attractiveness. Our model achieves a mean attractiveness of 3.56 for the top 10 most visited location, substantially closer to the observed Microcensus average. This further strengthens the idea that by incorporating attractiveness into destination choice, selection patterns better reflect real-world behavior, as people tend to gravitate toward more attractive destinations.

4 Assumptions and limitations

A major assumption made in this methodology is the use of total number of employees as a proxy for attractiveness for STATENT locations, area as a proxy for parks, and number of residents as a proxy for residences. These assumptions do not translate one to one between the different measures of attractiveness, and it does not consider other factors which commuters take into account such as brand, pricing, or current promotions. Future work may include exploring an embedding approach for location attractiveness, which can vary attractiveness of locations based on the socio-economic characteristics of an agent.
In addition, Euclidean distance was used for distance decay, which does not necessarily represent true distances one may have to travel in order to reach a destination. Utilizing skim matrices to estimate travel time between locations (Adiga et al., 2015) may more accurately portray the decision making process of an individual selecting a secondary location.
Finally, the model makes the assumption that the mode is fixed after the first secondary location in a chain is selected. This limits the kinds of trips, especially those trips which combine multiple modes such as public transport and walking within a single trip.

5 Conclusions

This paper presented a new methodology for secondary location choice in activity based travel demand models, built around a multinomial logit framework that jointly selects mode and location using importance sampling and with the incorporation of location attractiveness.
Results show that the model produces distance distributions that closely match observed patterns from the 2015 Microcensus for most purpose pairs, with reasonable accuracy up to approximately the 95th percentile of trip distances. Mode share calibration, while computationally expensive, achieves similar distributions to the training data across vehicle ownership categories. Compared to the existing eqasim approach, the proposed model selects locations with 47% higher average attractiveness, and its most frequently chosen locations correspond to well-known destinations such as shopping centers and major employment hubs.
Future work should explore richer representations of attractiveness such as potentially through learned embeddings that condition on agent socio-economic characteristics. Utilizing travel time instead of Euclidean distance would further enhance the location selection process. Despite these limitations, the proposed approach offers a flexible and extensible framework for secondary location choice that can accommodate new candidate locations with attractiveness measures.

References

Adiga, A., Agashe, A., Arifuzzaman, S., Barrett, C. L., Beckman, R. J., Bisset, K. R., ... others (2015). Generating a synthetic population of the united states. *Network Dynamics and Simulation Science Laboratory, Tech. Rep. NDSSL*, 15–009.

Barrett, C. L., Beckman, R. J., Khan, M., Kumar, V. S. A., Marathe, M. V., Stretz, P. E., ... Lewis, B. (2009). Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 winter simulation conference (wsc)* (p. 1003-1014). doi: 10.1109/WSC .2009.5429425

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, *41*(1), 193–210.

Guevara, C. A., & Ben-Akiva, M. E. (2013). Sampling of alternatives in multivariate extreme value (mev) models. *Transportation Research Part B: Methodological*, *48*, 31–52.

Hörl, S., & Axhausen, K. W. (2023). Relaxation–discretization algorithm for spatially constrained secondary location assignment. *Transportmetrica A: Transport Science*, *19*(2), 1982068.

Hörl, S., & Balac, M. (2021). 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.

Horni, A., Nagel, K., & W Axhausen, K. (2016). *The multi-agent transport simulation matsim*. Ubiquity Press.

Horni, A., Scott, D. M., Balmer, M., & Axhausen, K. W. (2009). Location choice modeling for shopping and leisure activities with matsim: combining microsimulation and time geography. *Transportation Research Record*, *2135*(1), 87–95.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

McFadden, D. (1977). Modelling the choice of residential location.

OpenStreetMap contributors. (2017). *Planet dump retrieved from https://planet.osm.org* . `https://www.openstreetmap.org`.

Petre, F., Bienzeisler, L., & Friedrich, B. (2025). Introducing a novel location-assignment algorithm for activity-based transport models: Carla. *arXiv preprint arXiv:2509.18191*.

Statistik der bevölkerung und haushalte (statpop): Variablenliste (No. 32266057). (2024). Neuchâtel. Retrieved from `https://dam-api.bfs.admin.ch/hub/api/dam/assets/32266057/master`

Swiss structural business statistics (statent) 2015-2021 (No. 32088042). (2023). Neuchâtel: Federal Statistical Office (BFS). Retrieved from `https://dam-api.bfs.admin.ch/hub/api/dam/assets/32088042/master`

Tozluoglu, Ç., Dhamal, S., Liao, Y., Yeh, S., Sprei, F., Dubhashi, D., ... Barrett, C. (2022). Synthetic sweden mobility (sysmo) model documentation. *URL: https://research. chalmers. se/publication/531094/file/531094_ Ful ltext. pdf*.

Verkehrsverhalten der bevölkerung (No. 1840477). (2017). Neuchâtel: Federal Statistical Office (BFS). Retrieved from `https://dam-api.bfs.admin.ch/hub/api/dam/assets/1840477/master`