Robust population synthesis: a framework for reliability.

Federico Bigi[1], Taha Hossein Rashidi[2], Giovanni Tataranno[3], and Francesco Viti[4]

[1]PhD Student, Faculty of Science, Technology and Medicine (FSTM), University of Luxembourg, Esch-Sur-Alzette, Luxembourg
[2]Associate Professor, University of New South Wales (UNSW Sydney), Sydney, Australia
[3]PhD Student, University of Luxembourg, Luxembourg
[4]Associate Professor, University of Luxembourg, Luxembourg

Short summary

This paper proposes a novel framework for evaluating the reliability of synthetic population generation in agent-based transportation simulations (ABM-TOS). Our methodology categorizes key parameters into distinct categories, including sociodemographic attributes, household features, and travel patterns, to test them against the real data for a comprehensive validation of the generated population. The framework's effectiveness is demonstrated by generating synthetic populations of varying sizes, representing 1% to 30% of Luxembourg's 2021 population. The study showcases the importance of accurate population synthesis in ABM-TOS and highlights potential areas for improvement in future research.
Keywords: agent-based modeling, MATSim, mobility transportation simulation, population synthesis, population generation.

1 Introduction

Agent-based models (ABM) applied in the transportation domain rely strongly on the correct generation of the travel demand, which is represented by a synthetic population of households and individuals characterized by sociodemographic attributes and daily activity patterns that at best reflect the real population in an area of study. The reliability of the synthetic population generation (SPG) process is critical in ABM simulations, which have been extensively used in different research domains beyond transportation (Hörl & Balac (2021); Anderson & van Der Merwe (2021); Coelho et al. (2021); ODonoghue et al. (2013)).
Focusing on ABM for transportation-oriented simulations (ABM-TOS), the population synthesis operation generates the input demand for the simulation environments, a fundamental step for any research investigation. This process usually involves the generation of individual-level data that accurately reflects the demographic and socio-economic characteristics of a given population, as well as information regarding their travel mobility patterns. These synthetic populations are then used to simulate various aspects of urban mobility, including travel behavior and transportation infrastructure use. These processes can generally be divided into two stages: *fitting* and *generation* (Müller & Axhausen (2010); Farooq et al. (2013)). While the fitting stage deals with the adjustment of an initial distribution to match known marginal distributions, the generation stage involves creating synthetic individuals from these distributions (Tanton (2013)).
Usually two types of data sources can be gathered for generating a synthesized population: publicly available data and travel surveys (Farooq et al. (2013)). These data are usually associated with spatial zoning and can be found in the form of individual agent samples and cross-classification tables.
One of the primary obstacles is the unavailability of complete data for the population that one wants to achieve, which can lead to inaccurate models and simulate spatially-dependent phenomena. To overcome this problem, multiple techniques have been applied to this topic, which are explained further in Section 2. The goal of these methods is to extract the underlying picture from the available data and generate a population that shares as much as possible the same characteristics as the real one. Nonetheless, even if the distribution matches the real data, there is no guarantee that the population reflects reality in its entirety. This paper presents a new framework for population synthesis in agent-based traffic model simulations (ABM-TOS), addressing literature gaps. It includes:

- A novel general framework for evaluating key distributions in ABM-TOS;
- A tool for assessing synthetic population quality and 'debugging' eventual inaccuracies;
- The introduction of MOBIUS, a new synthesizer demonstrated in the Luxembourg Scenario.

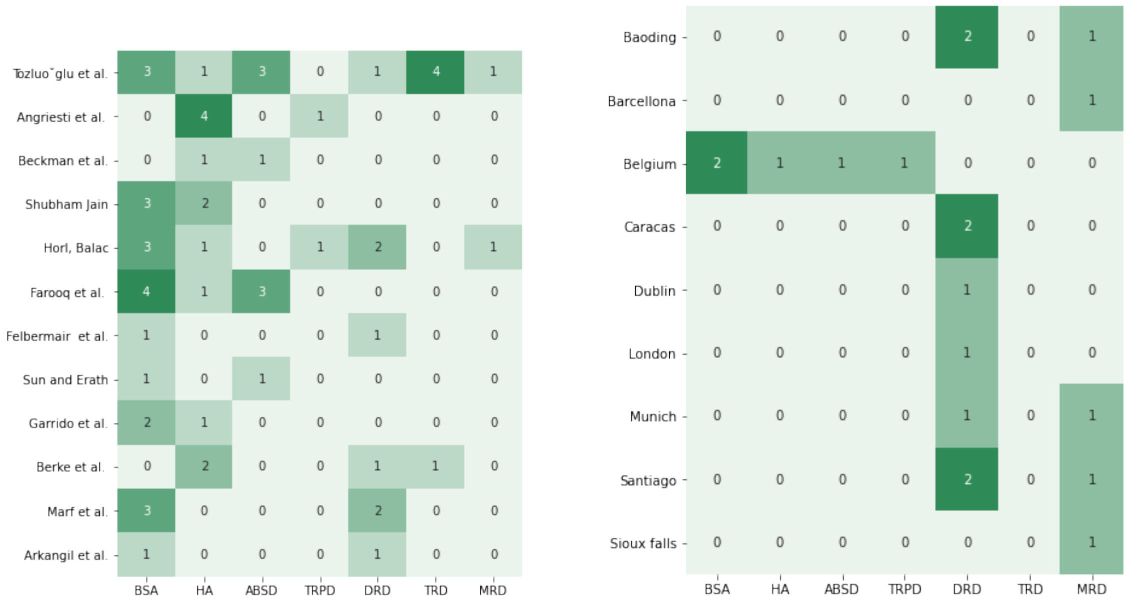
The remainder of this paper is organized as follows. In Section 2, we present a small literature review, analysing different population synthesizer, underlining the absence of a no generalized comprehensive framework for analyzing the synthetic population. Section 3, we delve into each category of framework. In Section 4 we present the case study for MOBIUS, a novel population synthesizer, and the results of the application of the framework to the 3 synthesized populations, together with the detailed categorization choice. Finally, Section 5 concludes the paper and presents the next step of the research.

2 Literature review

In this section, we present the literature regarding different methodologies and techniques widely used in population synthesis for different *ABM* simulations, namely Iterative Proportional Fitting (*IPF*) approach, Markov Chain Monte Carlo (*MCMC*) methods, and the use of Machine Learning (*ML*) models, together with some of the different population synthesizers that use one of the above-mentioned techniques. The Iterative Proportional Fitting (IPF) has been first described by Deming (et al. Deming & Stephan (1940)) and is also known as matrix ranking, RAS method, or matrix scaling. The strength of IPF (Arentze et al. (2007); Durán-Heras et al. (2018)) lies in its ability to preserve the relationships between different attributes in the data while ensuring that the synthetic population aligns with known marginal distributions. An example of the use of IPF methodology can be found in the work of Horl et al. (Hörl & Balac (2021)), which presented a methodology for generating synthetic travel demand using IPF algorithm and open and publicly available data for Paris. Another example comes from Tozluoglu et al. (Tozluoğlu et al. (2022)), which presents the documentation for their population synthesizer, Synthetic Sweden Mobility (SySMo). Another popular approach is through Markov Chain Monte Carlo (MCMC) methods. These are a class of algorithms used in computational statistics for sampling from a probability distribution (Geyer (2011)). MCMC methods work by constructing a Markov chain that has the desired distribution as its equilibrium to the reference distribution. MCMC methods have seen an increase in popularity due to their robustness in handling complex probabilistic models, given their ability to capture the relationship between multi-dimensional data spaces, and correctly reproducing the underlying distributions of the provided data. Farooq et al. (Farooq et al. (2013)) propose a Markov Chain Monte Carlo simulation-based approach for synthesizing populations for use in urban systems evolution microsimulations. Finally, Machine Learning (ML) models are starting to see an increase in population synthesis processes, proposing interesting methodologies given their capability of reproducing complicated distributions underlying the input data. Berke et al. (Berke et al. (2022)) present a framework for generating synthetic mobility data using a deep recurrent neural network (RNN) trained on real location data. The issue with the population synthesizers provided in the literature is that they do not all offer a structured analysis for assessing the quality of the generated population for use in ABM-TOS. To systematically analyze these existing methods, we propose a framework that classifies different types of distributions that are important for any ABM-TOS. This framework includes the following categories, which are further explained in Section 3:

- Basic Sociodemographic Attributes (*BSA*)
- Household Attributes (*HA*)
- Advanced Sociodemographic Distributions (*ASBD*)
- Tripchain Related Distributions (*TRPD*)
- Distance Related Distributions (*DRD*)
- Time-Related Distributions (*TRD*)
- Mode Related Distributions (*MRD*)

In applying our framework to a broad range of existing studies (Tozluoğlu et al. (2022); Agriesti et al. (2021); Beckman et al. (1996); Jain et al. (2015); Hörl & Balac (2021); Farooq et al. (2013); Felbermair et al. (2020); Sun & Erath (2015); Garrido et al. (2020); Berke et al. (2022); Badu-Marfo et al. (2020); Arkangil et al. (2022) and the majority of MATSim scenarios presented in Horni et al. (2016)), we aim to highlight the gaps and inconsistencies in current methods. The analysis is presented in Figure 1.



(a) (b)

Figure 1: (a) Literature validation results classified; (b) MATSim scenarios validation results classified.

Figures 1a and 1b reveal shortcomings in validation analyses for agent-based models (ABM). Most literature, as shown in Figure 1a, focuses on BSA and HA metrics, often neglecting TRD and MRD due to varied objectives of population synthesizers, data exclusion, and post-simulation calculations. In Figure 1b's analysis of ABM-TOS scenarios, with the exclusion of the Belgium scenario, most fail to address key mobility-related distributions (DRD, TRD, MRD), with all the papers omitting analysis regarding the TRD and TRPD distributions. This gap in comprehensive evaluation can lead to inaccuracies in representing population characteristics or behaviors in ABM simulations. To address this, we propose a framework designed to analyze a large set of different distributions that are crucial for mobility simulations.

3 Methodology

What emerges from the literature is the lack of a structured and comprehensive methodology for assessing the performance of population synthesis. This gap motivates our proposal for a new framework designed to analyze all the different distributions that are crucial for mobility simulations.
For our methodological framework, we decided to group the different parameters usually analyzed into different categories, together with new distributions extracted from the literature, that proved to have an impact on the *ABM* and *ABM-TOS*. This is presented in Table 1
For further explanation of the categories (please refer to Bigi et al. (2024)). To quantify the reliability of each parameter, we propose using metrics capable of assessing the similarity of different variable distributions. For our specific analysis, the Hellinger distance, the NRMSE, and the JS divergence were chosen to compare the generated distributions with the reference ones.
The Hellinger distance (Kitsos & Toulias (2017)) is commonly used in statistics and information theory to measure the similarity between two probability distributions and is computed as follows:

$$H(P,Q) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2} \tag{1}$$

where P and Q are the two probability distributions being compared, n is the number of values

Category	Description	Key References	Examples
Basic Sociodemographic Attributes (BSA)	Fundamental sociodemographic parameters.	Hanson & Hanson (1981)	Gender distribution, age.
Household Attributes (HA)	Household interactions and characteristics.	Bradley & Vovsha (2005), Hu et al. (2023)	Household location, car ownership allocation.
Advanced Sociodemographic Distributions (ASBD)	Bivariate distributions combining BSA and HA.	-	Age-household size, age-gender, gender-household size
Trip-chain Related Distributions (TRPD)	Critical for modeling sequences of activities and travel events.	Scheffer et al. (2021)	Activity type, Activity location
Distance-Related Distributions (DRD)	Distance-based KPIs for the population.	-	Total distance traveled, home-primary activity distance
Time-Related Distributions (TRD)	Time-based KPIs for the activity of the population.	-	Duration for primary and secondary activities, activity start times
Mode-Related Distributions (MRD)	KPIs for transportation options and usage patterns.	-	Total distance traveled per mode, modal split

Table 1: Table describing the different categories to analyze

in the distributions, and p_i and q_i are the values at the ith index of P and Q, respectively. This metric was chosen mostly because of its sensitivity to variations in the shape of the distributions. The Normalized Root Mean Square Error (NRMSE) was chosen as a second metric to evaluate the difference in magnitude of the selected distributions. As for the Hellinger distance, it is a measure used for assessing the similarity between two probability distributions, and is defined as follows:

$$NRMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{y_{max} - y_{min}}\right)^2} \tag{2}$$

where y_i and $\hat{y}_i$ represent the observed and estimated values, respectively, y_{max} and y_{min} are the maximum and minimum observed values, and N is the total number of observations.
To finalize our analysis with a metric to evaluate the goodness of fit of each distribution, we chose to add to our evaluation methodology the Jensen-Shannon (JS) divergence (Hoyos-Osorio & Sanchez Giraldo (2023)). The JS divergence is defined as the mean of the Kullback-Leibler (KL) divergence (Joyce (2011)) of P from the average distribution M, and the KL divergence of Q from M, where P and Q are the two probability distributions in comparison.

$$JSD(P,Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M) \tag{3}$$

One of the reasons behind choosing these three statistical metrics over others lies in their collective ability to analyze the distributions from different angles. By pairing them, we can assess the error in the magnitude of the generated distributions through NRMSE, the shape via the Hellinger distance, and the overall fit with the JS divergence. Table 2 serves as an example decision matrix, summarizing the implications of various combinations of these three metrics and suggesting corresponding actions to adjust the statistical analysis of distributions.

NRMSE	**Hellinger**	**JS Divergence**	**Interpretation**	**Suggested Action**
High	High	High	Significant divergence	Re-evaluate model assumptions and data sources.
High	High	Low	Shape of distributions is off	Consider transformations or alternative models.
High	Low	High	Magnitude discrepancies	Check for errors in data or scale mismatches.
High	Low	Low	Magnitude is overestimated	Adjust model to reduce estimation errors.
Low	High	High	Good magnitude but poor fit	Refine model to better capture distribution shapes.
Low	High	Low	Minor shape discrepancies	May require slight model adjustments.
Low	Low	High	Good fit with misalignment	Fine-tune model parameters.
Low	Low	Low	Well-aligned distributions	Validate with additional data if possible.

Table 2: Decision Matrix for Statistical Analysis of Distributions

Moreover, having these three metrics bounded between 0 and 1 means that they can be somewhat comparable. When visualized within a heatmap, as seen in Section 54, Figure 2, they can provide a fast and intuitive way to understand the state of the overall population. Furthermore, it helps identify critical areas, connecting apparent unrelated problems, through a global view of the generated data.

4 Case Study and Results

In this section, the process of generating the synthetic population is discussed using data from the Luxmobil Survey 2017 in Luxembourg. Conducted by the Ministry of Mobility and Public Works, the survey included 40,000 residents and 45,000 cross-border workers, with a response rate of 30%, ensuring representative and unbiased results. Approximately 35,000 valid responses were obtained and, after cleaning, approximately 22,000 responses were retained, representing 2.8% of daily travellers. The survey provided a zoning structure at the administrative unit level with 150 different zones. Respondents reported daily trips and modes of transport in travel diaries, along with socio-demographic information. From the travel survey, we extracted the following data for this analysis:

- spatial distribution with zonal information, with related facilities location file;
- zonal household and population distribution categorized by the considered attributes;
- zonal trip chain distribution;
- OD matrices per hour, which we computed from the Luxmobil travel survey and validated through PTV Visum;
- zonal distribution of typical activity time duration and departure time for the first activity;

Algorithm 1 MOBIUS Assignment Loop

for each agent a_i to generate **do**
    Initialize T_i, tripchain of a_i, and $time_{curr}$, current time
    for each $act \in T_i$, with act as single activity in the Tripchain **do**
        if act is the first activity **then**
            Assign T_{dep}, departure time from home, set it as $time_{curr}$
            Estimate Mode of transport and travel time to get tt_{mode}
            Update $time_{curr} += tt_{mode}$
        else
            Assign location for act through the OD matrix sampling, let the location be z_{dest}
            Sample t_{act}, activity duration from distribution of duration of activity in z_{dest}
            Estimate Mode of transport and travel time to get tt_{mode}
            Update $time_{curr} += tt_{mode} + t_{act}$
        end if
    end for
end for
Return a_i with $act \in T_i$ assigned with times and locations.

The novel population synthesizer, MOBIUS (Mobility Optimization Based on Iterative User Synthesis), was developed by gathering and scaling data to match the target population. Its macrostructure has been inspired by two state-of-the-art synthesizers (presented in Tozluoğlu et al. (2022) and Hörl & Balac (2021)), respectively. It differs from them in its approach to creating households with attributes like zone, car ownership, and composition size, enhancing attribute inheritance, control, and code efficiency. For each household, agents are generated with attributes appropriate to their age group. The activity generation phase assigns trip chains to agents, using the top 25 chains to eliminate outliers. Location assignments are based on an OD probability matrix, considering time and activity type, with home locations determined by a bounding box approach. Activity destinations are assigned precisely through additional sampling, and leg modes are initially estimated by beeline distance, then fine-tuned using MATSim simulation. The complete process is detailed in Algorithm 1, explaining the synthetic population generation loop.

The MOBIUS synthesizer was applied by generating three synthetic populations representing 1% (6,454 individuals), 10% (64,539 individuals), and 30% (193,617 individuals) of Luxembourg's total 2021 population (approximately 645,390), including resident and cross-border data. Each population size underwent MATSim simulations until equilibrium, determined by modal split alignment over 150 iterations. Further calibration using Cadyts (Chen (2012)) for 300 iterations has been performed, based on 21 traffic counts from October 2021, covering about 30% of total traffic in the dataset. Road capacities were adjusted to align with observed volume-to-capacity ratios. The varying population sizes were chosen to observe metrics convergence, like age distribution errors reducing with sample growth, and to identify areas needing analysis if errors persist.

The results of this analysis, applied to the above-mentioned populations and adopting the three metrics described in the previous section, are summarized in Figure 2, with the chosen metrics for each category presented in Table 3.

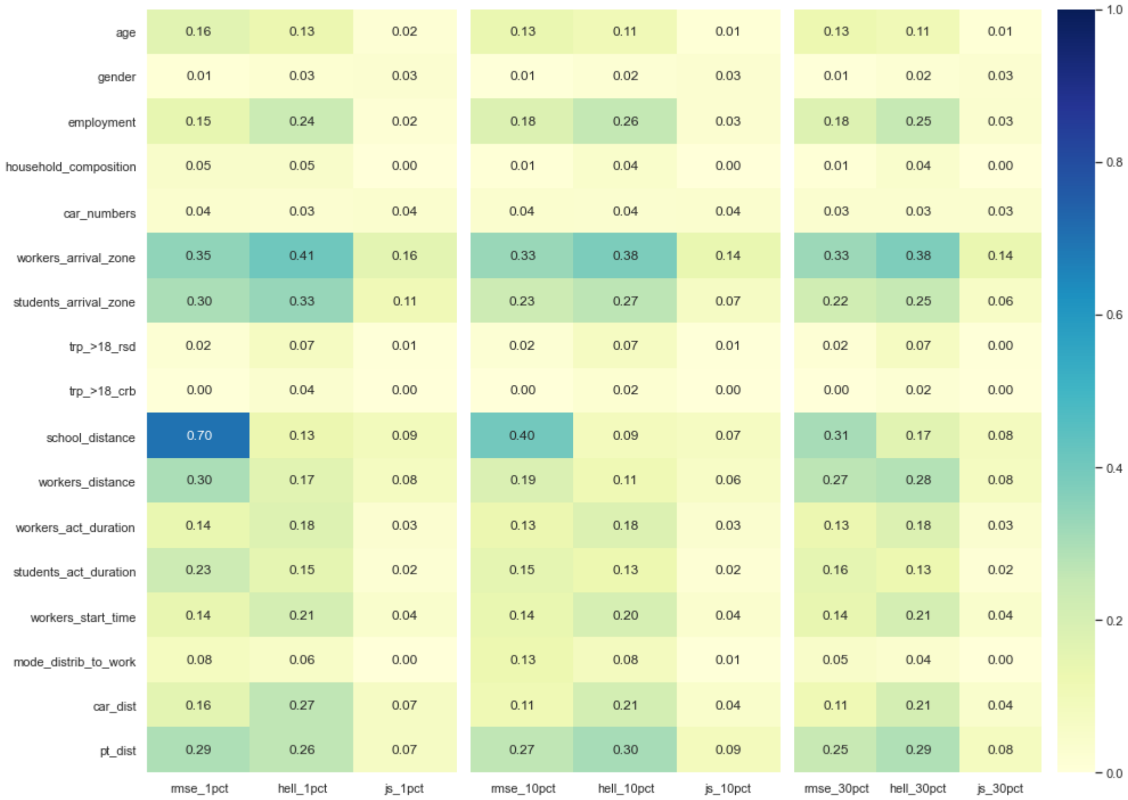


Figure 2: Results for the 1%,10% and 30% Luxembourg generated population.

Category	**Chosen Metrics**
Basic Socio-demographic Attributes (BSA)	Error in age and gender distribution
Household Attributes (HA)	Distribution of household attributes across zones
Advanced Sociodemographic Distribution (ASBD)	OD pair assignment for workers/students (home-work/school) at zonal level
Tripchain Related Distributions (TRPD)	Trip chain assignment for adult cross-border commuters and residents
Distance Related Distributions (DRD)	Beeline school and work distance distribution
Time-Related Distributions (TRD)	Primary activity duration for workers and students
Mode Related Distributions (MRD)	Distribution for primary activity, car and public transport distance distributions

Table 3: Chosen metrics for each category

In the study, most metrics showed trends that were either stable or decreased as the synthesized population size increased. This aligns with expectations and indicates the categories and metric's sensitivity in identifying issues in distribution generation. Specifically, the decreasing NRMSE and JS, along with the stable Hellinger distance in the age category, suggest more alignment between reference and generated distributions as population size grows, despite differences in distribution shapes.

The MOBIUS synthesizer, particularly for Basic Socio-demographic Attributes (BSA) and Household Attributes (HA), performed as expected due to the high number of controlled variables. For

instance, NRMSE for school distance metrics decreased with larger populations, explained by the limited number of schools causing greater discrepancies in smaller samples, which is again expected. On the other hand, worker distance and arrival zone metrics showed stable trends, not significantly improving with larger populations. This could be due to the random sampling in OD segmentation within MOBIUS, where misalignment in departure times can affect agent assignment accuracy.
Moreover, the Mode-Related Distributions (MRD) parameters, particularly mode distribution to work, car distance, and public transport distance, showed significant discrepancies, potentially due to inaccuracies in route estimations and OD-matrix assignment. Calibration with Cadyts in MATSim simulations did not resolve these discrepancies. Furthermore, worker start time and activity duration metrics remained stable across different population sizes, suggesting potential initial estimation inaccuracies and highlighting a limitation of the OD-matrix sampling method.

5 Conclusion

The study highlights the capability of the proposed framework to detect inconsistencies in the distribution of synthetic population parameters. It reveals that for our case study, while the proportion of the generated population affects marginally the accuracy of basic sociodemographic and household attributes, it significantly influences the distributions related to travel distances and times. Even with increased population sizes, notable differences persist in distance by mode and activity duration. This suggests that population synthesizers need improvement, particularly in accurately matching agents to activity locations and optimizing trip-chaining decisions, to more closely replicate observed daily activity-travel patterns.
For agent-based models, generating a high-quality synthetic population necessitates a strategic selection and balancing of control variables. This is vital to ensure that these variables, while generalizing from real data, maintain reliability. This becomes increasingly crucial when these variables are used to derive further distributions, as accurate alignment with real-world data is essential to support model validity and prevent bias. Enhancements in population synthesis, such as those demonstrated by MOBIUS in modal split and travel time, are key to avoiding overfitting and ensuring generality, which allows population expansion and proper analysis. Tools that specialize in complex aspects like routing and traffic conditions are indispensable, as they simplify the modeling process and increase accuracy. Given data limitations, careful consideration is required when excluding or inferring missing data to avoid compounding errors. Adhering to these principles markedly boosts the dependability of the population synthesizer, thereby improving the effectiveness of transportation-focused agent-based simulations.

References

Agriesti, S., Roncoli, C., & Nahmias-Biran, B.-H. (2021). Assignment of a synthetic population for activity-based modelling employing publicly available data.

Anderson, A., & van Der Merwe, A. (2021, 05). Time-driven activity-based costing related to digital twinning in additive manufacturing. *South African Journal of Industrial Engineering*, *32*, 37 - 43. Retrieved from http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S2224-78902021000100005&nrm=iso

Arentze, T., Timmermans, H., & Hofman, F. (2007). Creating synthetic household populations: Problems and approach. *Transportation Research Record*, *2014*(1), 85-91. Retrieved from https://doi.org/10.3141/2014-11 doi: 10.3141/2014-11

Arkangil, E., Yildirimoglu, M., Kim, J., & Prato, C. (2022). A deep learning framework to generate realistic population and mobility data.

Badu-Marfo, G., Farooq, B., & Paterson, Z. (2020, 04). Composite travel generative adversarial networks for tabular and sequential population synthesis.

Beckman, R. J., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, *30*(6), 415-429. Retrieved from https://www.sciencedirect.com/science/article/pii/0965856496000043 doi: https://doi.org/10.1016/0965-8564(96)00004-3

Berke, A., Doorley, R., Larson, K., & Moro, E. (2022, apr). Generating synthetic mobility data for a realistic population with RNNs to improve utility and privacy. In *Proceedings of the 37th ACM/SIGAPP symposium on applied computing.* ACM. Retrieved from `https://doi.org/10.1145%2F3477314.3507230` doi: 10.1145/3477314.3507230

Bigi, F., Hossein Rashidi, T., & Viti, F. (2024). Synthetic Population: a Reliable Framework for Analysis for Agent-Based Modeling in Mobility. *Transportation Research Record - In publishing.*

Bradley, M., & Vovsha, P. (2005, 09). A model for joint choice of daily activity pattern types of household members. *Transportation*, *32*, 545-571. doi: 10.1007/s11116-005-5761-0

Chen, Y. (2012). *Adding a comprehensive calibration methodology to an agent-based transportation simulation* (Doctoral dissertation). doi: 10.14279/depositonce-3523

Coelho, F., Relvas, S., & Barbosa-Póvoa, A. (2021). Simulation-based decision support tool for in-house logistics: the basis for a digital twin. *Computers Industrial Engineering*, *153*, 107094. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0360835220307646` doi: https://doi.org/10.1016/j.cie.2020.107094

Deming, W. E., & Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, *11*(4), 427 – 444. Retrieved from `https://doi.org/10.1214/aoms/1177731829` doi: 10.1214/aoms/1177731829

Durán-Heras, A., García-Gutiérrez, I., & Castilla-Alcalá, G. (2018). Comparison of iterative proportional fitting and simulated annealing as synthetic population generation techniques: Importance of the rounding method [Article]. *Computers, Environment and Urban Systems*, *68*, 78 – 88. Retrieved from `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85044376181&doi=10.1016%2fj.compenvurbsys.2017.11.001&partnerID=40&md5=39410c7c9def8cbd4fff49761c193e19` (Cited by: 6) doi: 10.1016/j.compenvurbsys.2017.11.001

Farooq, B., Bierlaire, M., Hurtubia, R., & Flötteröd, G. (2013). Simulation based population synthesis [Article]. *Transportation Research Part B: Methodological*, *58*, 243 – 263. Retrieved from `https://www.scopus.com/inward/record.uri?eid=2-s2.0-84885798000&doi=10.1016%2fj.trb.2013.09.012&partnerID=40&md5=96cd6174f3d8a1865481e53efb034a1d` (Cited by: 101; All Open Access, Green Open Access) doi: 10.1016/j.trb.2013.09.012

Felbermair, S., Lammer, F., Trausinger-Binder, E., & Hebenstreit, C. (2020). Generating synthetic population with activity chains as agent-based model input using statistical raster census data. *Procedia Computer Science*, *170*, 273-280. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1877050920304695` (The 11th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops) doi: https://doi.org/10.1016/j.procs.2020.03.040

Garrido, S., Borysov, S. S., Pereira, F. C., & Rich, J. (2020). Prediction of rare feature combinations in population synthesis: Application of deep generative modelling [Article]. *Transportation Research Part C: Emerging Technologies*, *120*. Retrieved from `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85091974069&doi=10.1016%2fj.trc.2020.102787&partnerID=40&md5=6574bebeac59544154e1b47acdbb92e6` (Cited by: 8; All Open Access, Green Open Access) doi: 10.1016/j.trc.2020.102787

Geyer, C. (2011, May 10). Introduction to markov chain monte carlo. In *Handbook of markov chain monte carlo* (pp. 3–48). CRC Press. doi: 10.1201/b10905-2

Hanson, S., & Hanson, P. (1981). The travel-activity patterns of urban residents: Dimensions and relationships to sociodemographic characteristics. *Economic Geography*, *57*(4), 332-347. Retrieved from `https://www.tandfonline.com/doi/abs/10.2307/144213` doi: 10.2307/144213

Horni, A., Nagel, K., & Axhausen, K. (2016). *The multi-agent transport simulation matsim.* doi: 10.5334/baw

Hoyos-Osorio, J., & Sanchez Giraldo, L. (2023, 05). *The representation jensen-shannon divergence.*

Hu, Y., van Wee, B., & Ettema, D. (2023). Intra-household decisions and the impact of the built environment on activity-travel behavior: A review of the literature. *Journal of Transport Geography*, *106*, 103485. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0966692322002083` doi: https://doi.org/10.1016/j.jtrangeo.2022.103485

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. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X21003016` doi: https://doi.org/10.1016/j.trc.2021.103291

Jain, S., Ronald, N., & Winter, S. (2015, 12). Creating a synthetic population: A comparison of tools.

Joyce, J. (2011, 01). Kullback-leibler divergence. In (p. 720-722). doi: 10.1007/978-3-642-04898-2_327

Kitsos, C., & Toulias, T. (2017, 01). Hellinger distance between generalized normal distributions. *British Journal of Mathematics Computer Science*, *21*, 1-16. doi: 10.9734/BJMCS/2017/32229

Müller, K., & Axhausen, K. (2010, 09). Population synthesis for microsimulation: State of the art.

ODonoghue, C., Morrissey, K., & Lennon, J. (2013, 01). Spatial microsimulation modelling: a review of applications and methodological choices. *International Journal of Microsimulation*, *7*, 26-75. doi: 10.34196/ijm.00093

Scheffer, A., Connors, R., & Viti, F. (2021). Trip chaining impact on within-day mode choice dynamics: Evidences from a multi-day travel survey. *Transportation Research Procedia*, *52*, 684-691. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352146521001289` (23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020, Paphos, Cyprus) doi: https://doi.org/10.1016/j.trpro.2021.01.082

Sun, L., & Erath, A. (2015, 10). A bayesian network approach for population synthesis. *Transportation Research Part C Emerging Technologies*, *61*, 49-62. doi: 10.1016/j.trc.2015.10.010

Tanton, R. (2013, 01). A review of spatial microsimulation methods. *International Journal of Microsimulation*, *7*, 4-25. doi: 10.34196/ijm.00092

Tozluoğlu, , Dhamal, S., Liao, Y., Yeh, S., Sprei, F., Dubhashi, D., . . . Barrett, C. (2022). *Synthetic sweden mobility (sysmo) model documentation.*