

Spatial Microsimulation and Activity Allocation in Python: An Update on the Likeness Toolkit

Joseph V. Tuccillo^{‡*}, James D. Gaboardi[‡]



Abstract—Understanding human security and social equity issues within human systems requires large-scale models of population dynamics that simulate high-fidelity representations of individuals and access to essential activities (work/school, social, errands, health). Likeness is a Python toolkit that provides these capabilities for Oak Ridge National Laboratory’s (ORNL) UrbanPop spatial microsimulation project. In step with the initial development phase for Likeness (2021 - 2022), we built out several foundational examples of work/school and health service access. In this paper, we describe expansion and scaling of Likeness capabilities to metropolitan areas in the United States. We then provide an integrated demonstration of our methods based on a case study of Leon County, FL and perform validation exercises on 1) neighborhood demographic composition and 2) visits by demographic cohorts (gender/age) obtained from point of interest (POI) footfall data for essential services (grocery stores). Taking into account lessons learned from our case study, we scope improvements to our model as well as provide a roadmap of the anticipated Likeness development cycle into 2023 - 2024.

Index Terms—activity space, synthetic population, microsimulation, population dynamics

Introduction

Agent-based models (ABMs) of population dynamics are essential for understanding human security and social equity issues within human systems [1], [2], [3]. Such models often rely upon *synthetic populations* – virtual representations of individuals plausibly residing within an area – to assess how individual behaviors and interactions contribute to complex system-level behavior. Applications of ABMs for population dynamics range from urban planning (access to essential services like food and healthcare), public health (facility occupancy, social contact networks), and disaster preparedness (social vulnerability to environmental hazards, evacuation and critical infrastructure planning).

A current challenge for research in population dynamics is to more directly represent population heterogeneity [4]. Individuals exhibit a variety of patterns of life dictating their behavior and interactions [5], which are in turn influenced by social, demographic, and economic characteristics. Incorporating these factors into the design of synthetic populations contributes to more holistic models of how human systems operate.

* Corresponding author: tuccillojv@ornl.gov

‡ Oak Ridge National Laboratory, Geospatial Sciences and Human Security

Copyright © 2023 Oak Ridge National Laboratory This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

To support enhanced ABMs of population dynamics, the UrbanPop spatial microsimulation framework developed by Oak Ridge National Laboratory (ORNL) generates high-fidelity synthetic populations on hundreds of attributes from the American Community Survey (ACS) and its Public-Use Microdata Sample (PUMS) and combines them with nighttime/daytime behavioral routines [3]. Central to UrbanPop’s capabilities is Likeness, a Python toolkit that combines population synthesis, spatial network modeling, and activity allocation [6], [7]. Foundational examples produced for Likeness have explored travel routing from home locations to anchor activities (e.g., work, school), POI occupancy characteristics [6], [7], as well as routing incidental travel from anchor to non-anchor activities (e.g., social, errands, health) [8]. This paper builds from these foundational examples to discuss scaling our methods to larger areas of interest. Moreover, it examines the performance of our activity allocation model relative to real-world POI demographics from digital trace (anonymized mobile device) data on POI visitation.

Expansion and Scaling of Likeness Capabilities

Synchronous with creating foundational examples for Likeness, we have developed an integrated workflow that scales the ecosystem (Figure 1) to support microsimulation for any metropolitan statistical area (MSA) in the United States.

Figure 2 outlines the procedure for agent generation at the MSA level. As discussed in [6], the `livelike.acs.puma` class is the core object for residential population synthesis. It stores all census microdata and geographic (e.g., block group, tract) model constraints to support spatial allocation for a single Public-Use Microdata Area (PUMA) in the United States. Residential synthetic populations for MSAs are generated as collections of PUMAs, which are parsed automatically by combining U.S. Census metropolitan/micropolitan delineation files¹ with the Census 2010 PUMA-to-tract relationship file². PUMAs parsed for the target MSA are converted into `livelike.acs.puma` objects in bulk via `livelike.multi.make_pumas()`. This function accepts the PUMA Federal Information Processing Standards (FIPS) codes (unique IDs) in list form, along with the target ACS year. This information is then used to gather the relevant ACS Summary File (SF) constraints for spatial allocation of PUMS responses to small census areas of roughly 8,000 people

1. <https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html>

2. <https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2010-tract-to-puma-record-layout.html>

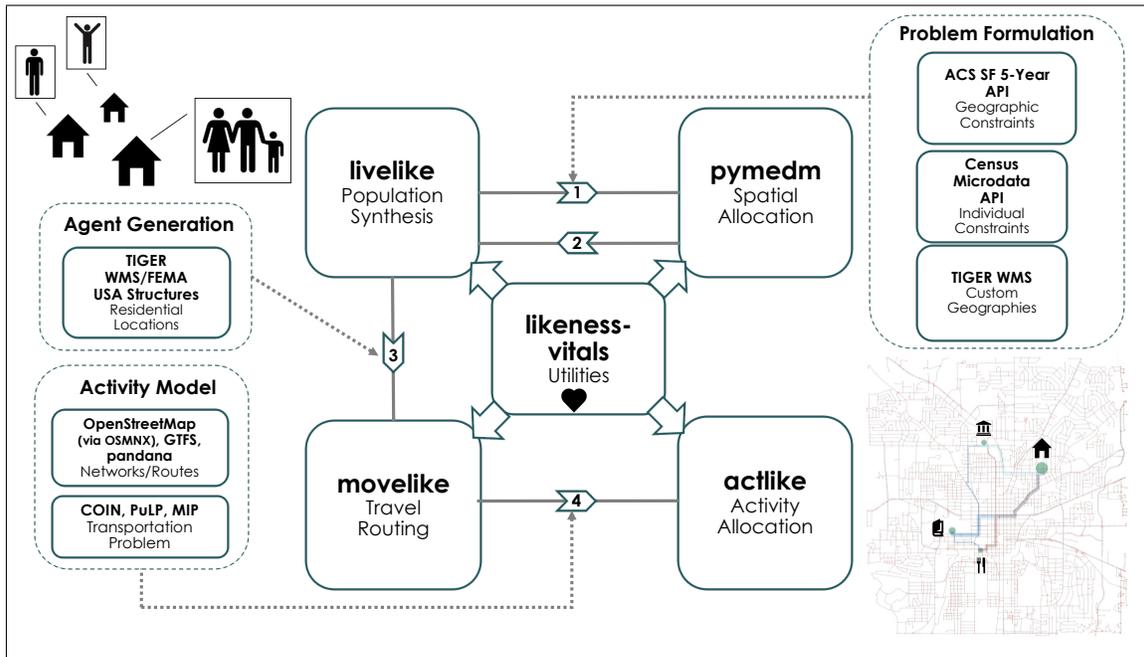


Fig. 1: Likeness ecosystem overview.

or less (block groups, tracts). Spatial allocation is subsequently handled with parallel processing, which is supported by both packages Likeness offers for Penalized Maximum-Entropy Dasymeric Modeling (P-MEDM) [9]: `pymedm` (bleeding-edge, Python-native version, based on `jaxopt` [10]) and `pmedm_legacy` (stable bridge to original R/C++ routine³ via `rpy2`). As demonstrated in [6] and [7], the population synthesis routine also collects diagnostics on how effectively each P-MEDM solution reproduces variable estimates from the ACS SF relative to reported Margins of Error (MOEs). These utilities are available in both `pymedm` and `pmedm_legacy`.

Likeness generates agents for microsimulation in a way that provides realistic home (origin) locations from which to allocate essential activities on transportation networks. Our initial approach, based on census block-level housing density [6], [7], is now implemented in `livelike` as a housing universe generation procedure. Additionally, we are actively developing a method (demonstrated in [Integrated Demonstration: Leon County, Florida](#)) that enhances this capability by matching synthesized households to residential locations from building footprint data. These matches are performed by conflating attributes of synthetic households (e.g., dwelling type, income) with building footprint attributes (e.g., floor area, number of units). These utilities are designed to be agnostic to the building footprint provider and can even support custom building features.

Agent residential locations provided by `livelike` act as origin points for simulating travel to essential activities. The next stage in the Likeness workflow employs network analysis to model the cost of travel to these activities [11], [12], [13] and allocate agents to POIs accordingly based on mathematical programming routines [14], [15], [16], [17]⁴. In our first iteration of Likeness,

both these tasks were accomplished within `actlike`. However, we concluded that the network modeling piece was specialized enough to be split from the `actlike` package, which led to the creation of `movelike`. With a push for varied modes of network traversal, three new modes of travel can now be modeled: walking, biking, and public transit. However, modeling travel behavior via public transportation is less straightforward than for driving, biking, and walking networks due to stricter network topology, including factors like connectivity and directionality of routes. We have made our foray into modeling more realistic public transit behavior within `movelike` through the incorporation of the General Transit Feed Specification (GTFS)⁵. GTFS is a data specification that stipulates the required files, along with their structure and format⁶, for publishing, ingesting, and utilizing public transit datasets. The GTFS datasets can be obtained via services such as *TransitFeeds*⁷ and *The Mobility Database Catalogs*⁸. In our current iteration we utilize GTFS data feeds to implement a pseudo-transit network space by which agents can engage in limited traversal. This is accomplished through a mask of *OpenStreetMap*⁹ (OSM) street segments known to be associated with bus routes. The OSM network is masked by passing a (multi)polygon feature of buffered and unioned bus routes within the study area into `osmnx` [12]. This method demonstrates progression in representing public transit but certainly has room for improvement, which will be discussed in [Development Roadmap: 2023 - 2024](#).

Finally – and at the heart of it all – expansion and scaling of the Likeness ecosystem led to the development of a new package for common utilities, `likeness-vitals`. The `likeness-vitals` package provides support for monitoring

3. <https://bitbucket.org/nnnagle/pmedmrcpp>

4. The PuLP and Python-MIP open-source optimization Python packages are cited here, along with COIN-OR (a consortium that supports various open-source Operations Research projects) and the COIN-OR Branch-and-Cut solver.

5. <https://gtfs.org/>

6. <https://gtfs.org/schedule/reference/#dataset-files>

7. <https://transitfeeds.com/>

8. <https://github.com/MobilityData/mobility-database-catalogs>

9. <https://www.openstreetmap.org/>



Fig. 2: Likeness agent generation procedure for Metropolitan Statistical Areas (MSAs) in the United States.

and timing processes, data manipulation, shared spatial functionality, and Census API access.

Integrated Demonstration: Leon County, Florida

Following the workflow described in Section [Expansion and Scaling of Likeness Capabilities](#), we demonstrate the current capabilities of Likeness and validate our activity allocation routine for Leon County, Florida. Leon County, whose primary city is Tallahassee, features a population of just under 300,000 residents, a compact urban footprint, and a diverse array of transportation modes (driving, transit, bike, walking). Our mobility validation exercise is based on grocery store visits from simulated home locations. Grocery stores provide a useful test case, acting both as catchments for the general population as well as points of access to vital services including food and healthcare. We obtained grocery store visits from Foursquare’s Research Visits feed, which provides POI visitation data attributed by demographic cohort (gender by age) for a variety of facility types^{10,11}.

We first simulated a single synthetic population for the Tallahassee Core-Based Statistical Area (CBSA). Spatial allocation (P-MEDM) was constrained on variables across subjects including sampling universe totals (i.e., population, housing units, households), descriptive factors including demographics, socioeconomic status, housing, mobility, and worker and student characteristics. For the remainder of the analysis, we focused on Leon County alone, removing large outlying areas of the MSA (PUMA 1206300, “Apalachee Region (Outside Leon County)”). Agents¹² were generated with the Federal Emergency Management Agency’s (FEMA) open USA Structures database [18], [19]. Our allocation procedure leveraged `livelike` utilities to match synthetic households to single-family residential, multi-family residential, mobile homes, and group quarters housing types.

We used employment and travel mode characteristics to assign agents to transportation networks used to access grocery POIs from home locations. Agents labeled as ‘employed’ possess an associated flag that identifies reported commute mode that can take the following values: ‘car_truck_van’, ‘bicycle’, ‘walked’, ‘wfh’, ‘public_transportation’, ‘other’, and ‘motorcycle’. Because detailed travel modes are unavailable for agents that are not employed (e.g., retired, active military), we rely on private (i.e., household-level) vehicle ownership instead.

Travel modes assigned to each agent were conflated with four transportation network types: ‘walked’, ‘bicycle’, ‘public_transportation’, and ‘drive’. We supported this process by developing a decision tree, visualized in Figure 3, through which we assume:

- Employed agents will use the transportation network that best matches their commute mode to access grocery POIs.

10. <https://location.foursquare.com/places/docs/how-does-places-work>

11. <https://location.foursquare.com/visits/docs/research-feed-schema>

12. Agents less than 20 years old were not included in the Foursquare POI data, thus they are excluded from our analysis.

- Agents that are not employed will use a privately-owned vehicle, and thus the ‘drive’ network, to access grocery POIs when available.
- Agents that are not employed and lack a privately-owned vehicle will use public transportation if they are located in a block group that is served by Tallahassee’s bus network (StarMetro) and opt to walk otherwise.

As demonstrated in Table 1, the ‘drive’ network supports the overwhelming majority of travel to grocery POIs in Leon County, followed by walking, public transportation, and bicycle access.

TABLE 1: Householder agents >20yo per assigned travel mode

Mode Assignment	Agent Count
'walked'	5,325
'bicycle'	1,052
'public_transportation'	3,830
'drive'	101,681
	111,888

Figure 4 shows that Leon County’s agent population is distributed unevenly relative to assigned travel modes. Because Leon County’s infrastructure primarily supports travel by car, agents who drive are distributed closest to the area’s general population density. The spatial distribution of agents who travel by walking also tends to follow Leon County’s settlement patterns, though in more limited numbers than for those who drive. Agents using public transport, meanwhile, are largely present in and near the center of the county, roughly occupying denser urban areas where StarMetro service is available. Bicyclists are distributed similarly to bus takers, but with several individual clusters associated with smaller outlying towns and settlements.

Grocery store POIs with medium to high visit confidence (at least 30 device visits per month, $n = 53$) were obtained from Foursquare for Leon County in January 2023. Destination capacities were estimated based on visit counts weighted by representativeness of the demographic cohort within the state’s 2010 Census population¹³. Destination capacities were estimated by the daily average (mean) for each POI during the collection month (01/2023).

After travel modes were assigned to agents, four network cost matrices were calculated from origin (residential location) to destination (grocery store) POIs in `movelike`. Agents were then allocated to a single probable destination POI based on least cost network travel paths with the `actlike.ActivityAllocation` routine, which solves a modified Transportation Problem¹⁴ [20], [21], [22], [23], where destination POI capacities are scaled [24] by the proportion of assigned travel mode for each scenario. All

13. <https://location.foursquare.com/visits/docs/foursquare-data-normalization>

14. This model is formulated in [6].

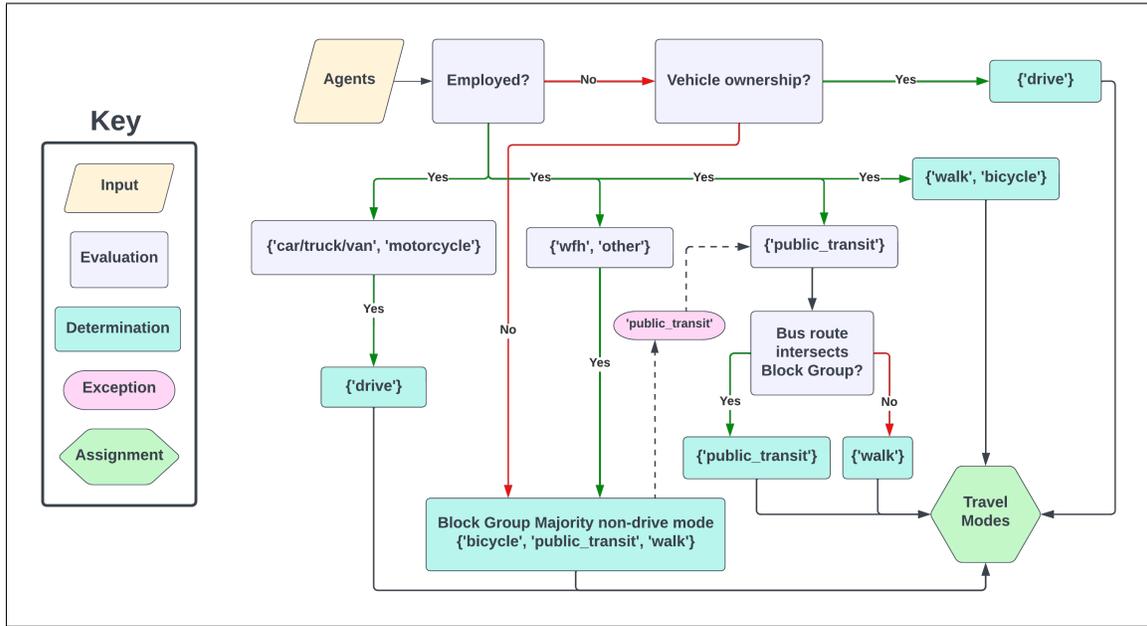


Fig. 3: Agent travel mode assignment decision tree.

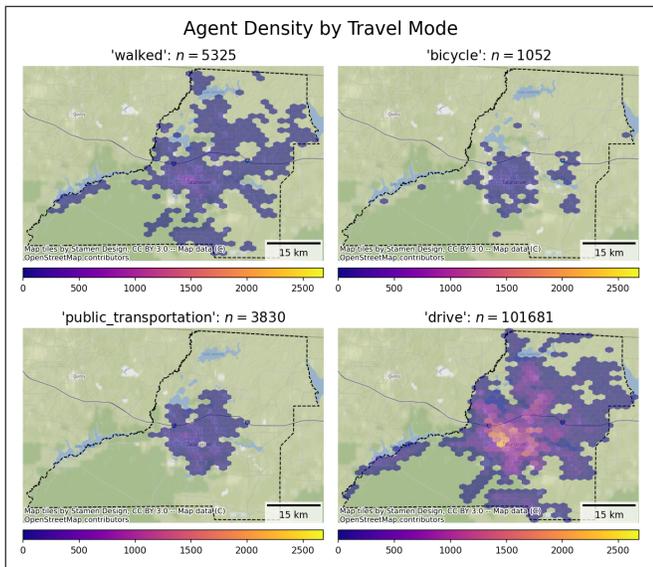


Fig. 4: Synthetic population distribution by travel mode.

models were run consecutively on two machines for benchmarking purposes. These were:

- A personal laptop (macOS) with a 2.3 GHz Quad-core Intel Core i7 processor (32 GB RAM).
- A virtual machine (Ubuntu) with a 2.8 GHz 22-core Intel(R) Xeon(R) processor (86 GB RAM).

The large disparity in problem size seen in Table 1 is even more pronounced in solution runtimes, shown in Table 2. Optimal solutions for non-drive models were found in a maximum time of just over 1 minute on both machines, with the drive model taking more than 17 and 8 hours to solve on the macOS and Ubuntu machines, respectively. Considering the solution time for the drive scenario, there is clearly a need for a more effective solution tech-

nique, which will be further discussed in [Development Roadmap: 2023 - 2024](#).

TABLE 2: Allocation Solution Runtimes (min.)

Mode Assignment	macOS	Ubuntu
	4 cores 32 GB RAM	22 cores 86 GB RAM
'walked'	0.72	1.19
'bicycle'	0.05	0.26
'public_transportation'	0.48	0.69
'drive'	1061.43	483.94

Validation procedures

Our validation procedures were designed to quantify the degree to which Likeness 1) resembles reference population estimates provided by the ACS SF with the synthetic populations (**demographic validation**) and 2) allocates activities matching real-world visitation patterns by demographics segments captured by the Foursquare POI data (**mobility validation**).

To produce our demographic validation, we followed [3] and measured our synthetic populations' degrees of conformity with 90% Margins of Error (MOEs) available from the ACS SF. ACS MOEs provide bounds for the expected ranges of values that our variables of interest could take. Tabulating individual attributes within each block group's synthetic population results in a reconstruction of the ACS SF estimates that can be assessed against the 90% MOEs. Greater conformity with the MOEs ("MOE Fit Rate") indicates a synthetic population that could plausibly resemble that block group's "true" population.

Following [3] and [6], we ran our mobility validation using Canonical Correlation Analysis (CCA). CCA, which measures the degree of linear association between two multidimensional datasets [25], is necessary to compare visitation patterns (n grocery store locations by m demographic cohorts). We performed CCA

on both the between-destination (relative prevalence) and within-destination (compositional) characteristics of each POI by demographic group. Both CCA runs were generated from tabulated counts of trips from the observed and synthetic datasets, their key difference being the method of standardization (column-wise for between-destination, row-wise for within-destination). We used the CCA coefficient of determination (R^2) to measure associations between synthetic and observed results. To better understand POI-specific activity allocation performance, we generated an additional local measure of within-destination correspondence. The local within-destination statistic compares the relative sizes of demographic cohorts using Spearman Rank Correlation, a non-parametric measure of the association between the ranks of two variables [26]. We opted for Spearman Correlation due to the relatively small number of demographic cohorts ($n = 11$).

Results

Demographic Validation

TABLE 3: Demographic validation

PUMA	Name	ACS 90% MOE Fit Rate
1206300	Leon County (Central)	0.992
1207300	Leon County (Outer)	0.998
1207301	Apalachee Region (Outside Leon County)	0.994

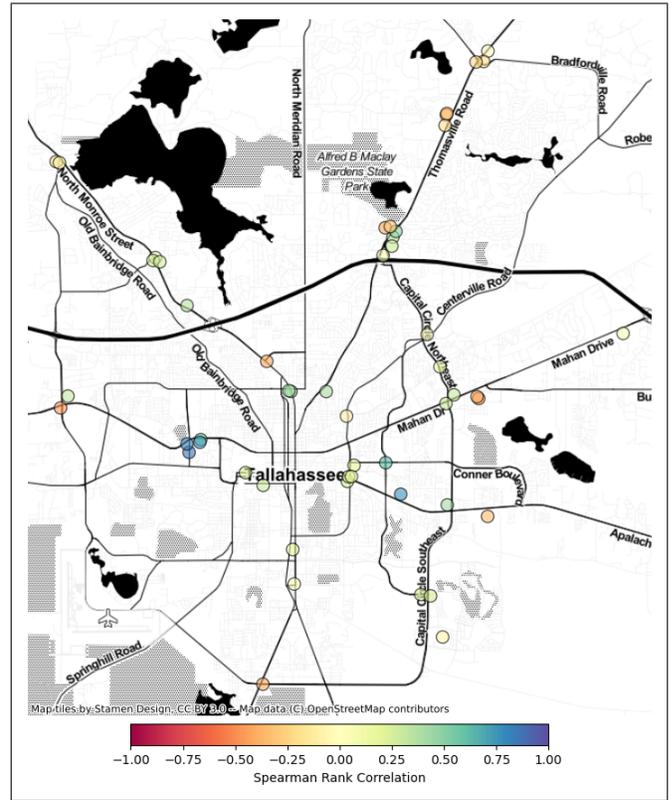


Fig. 6: Local within-POI mobility validation.

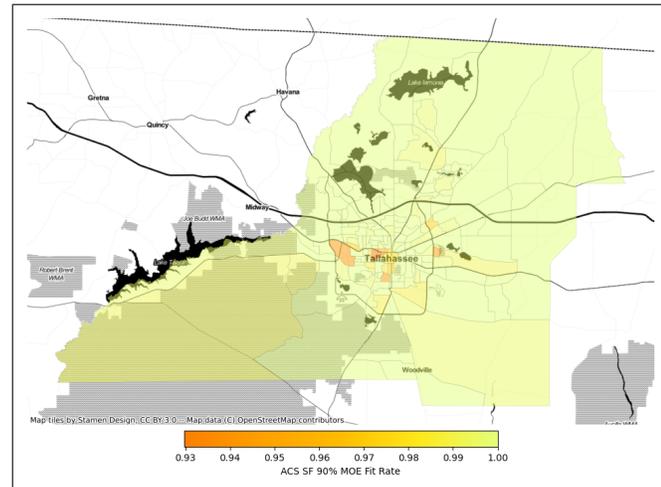


Fig. 5: Local demographic validation for Leon County.

Each of the three P-MEDM runs (one for each PUMA in the Tallahassee CBSA) resulted in a synthetic reconstruction of the ACS 90% MOEs. Overall, these reconstructions matched with at least 99% of the published MOEs from the ACS SF (Table 3). At the more granular level of Leon County block groups (Figure 5), MOE Fit Rates were relatively lower but still in agreement with at least 90% of the ACS SF MOEs in each location. We observed diminished performance in areas with large group quarters populations (e.g., college dormitories, prisons), as well as more sparsely populated rural and peri-urban portions of Leon County.

Mobility Validation

Our mobility simulation more faithfully recreated between-destination demographics ($R^2 = 0.74$) than within-destination

demographics ($R^2 = 0.41$) (Table 3). Inspecting the local within-destination scores, mapped in Figure 6, we observed generally greater correspondence between synthetic and observed POI demographics near Tallahassee’s downtown core, Florida State University, and Florida A&M University, with diminished performance in suburban and outlying areas of Leon County.

It is difficult to pinpoint the inconsistency in recreating POIs’ visitation patterns. In addition to urban density, this is potentially related to diversity of travel modes near the urban core (the increased biking, walking, and public transit use presented in Figure 4), but it requires further investigation. Section [Limitations](#) provides some additional confounding factors that are worthy of exploring relative to these results.

Limitations

We found that overall, Likeness approximates travel to non-anchor (grocery store) POIs modestly well. However, its performance tends to be weaker relative to travel to anchor activities (work/school), demonstrated in [6]. This suggests that approximating destination capacities for activities like grocery store visits provide an added challenge for activity allocation, including when real-world observations from visitation data are used. Additionally, multiple confounding factors related to data inputs and model assumptions may have affected our results:

- **Data-Specific:** Our clearest challenge is temporal mismatch between the synthetic population (2019) and POI visit data (2023). We have yet to increment our ACS year due to issues of non-conforming geography between the 2010 PUMAs and 2020 block groups/tracts. We hope to explore solutions to this problem starting with the

forthcoming 2022 ACS releases, which will adopt 2020 PUMAs across all geographic levels¹⁵. We were also limited to only one month of POI visit data. In future work, we hope to leverage a longer period of record to account for POIs with consistently high rates of visitation.

- **Model-Specific:** Several large assumptions were made that could confound our results, particularly that 1) agents simultaneously travel to grocery stores, 2) agents only select one grocery store to visit, using only one mode of travel, and 3) travel to grocery stores only occurs between home and work. These assumptions can be updated by incorporating information from time-use and travel surveys into Likeness. In this way, we can better reflect the times of day that different demographic cohorts access various activities [27].

We also hope to tighten our assumptions about the feasibility of POI access relative to the various travel modes. For example, in our current assignment process (Figure 3) agents unmatchable to a defined travel mode were considered walkers as a fallback for unavailability of a reachable bus route. Future iterations will refine this decision process by considering a distance threshold for agents to be labeled as either walkers or transit users. For example, we could set a rule that walking agents must be located within a reasonable distance of the closest POI, while agents that use bus service should reside near a bus stop in addition to being close to a bus route. To ensure all agents have at least one feasible POI destination to access, we also plan to incorporate a greater variety of curated locations from ORNL's PlanetSense database [28].

Conclusion and Outlook

This paper presented enhancements and scaling approaches for the Likeness spatial microsimulation toolkit. These include batched population synthesis runs for MSAs in the United States, residential allocation, and large-scale transportation network generation. We demonstrated these new capabilities by developing a mobility validation exercise for Leon County, FL. Our results provided reasonable representation of neighborhood demographics and routing to nearby essential services (grocery stores), with more mixed results related to activity allocation. The activity allocation results, however, do provide new research directions that we plan to explore in our future work. These include temporality of POI travel (travel probabilities relative to demographic cohorts), behavioral factors (willingness to travel given cost and impedances), and the use of multiple travel modes to reach activities.

Given the relative success of our population synthesis procedures in Section [Integrated Demonstration: Leon County, Florida](#), we are interested in also applying Likeness to explore transportation equity in the context of access to essential services like food and healthcare. Such an approach would aim to enhance existing research on transportation and accessibility [29], [30] with a cross-sectional representation of social, demographic, housing, and mobility characteristics. Assuming an agent with some blend of socio-demographic and economic characteristics resides in a particular section of a neighborhood, how many essential services can be readily accessed using their assigned transportation network? Using the Likeness ecosystem, we could develop such a measure for all agents in a synthetic population,

allowing the comparison of accessibility to population metrics like mobility difficulty. These insights could in turn be used to guide urban/regional infrastructure planning, pinpointing areas where drive, transit, or bike/walk services could be improved or expanded.

Development Roadmap: 2023 - 2024

- **Tooling for custom geographic extents.** The MSA-specific workflow demonstrated in this paper is limited in that it does not support custom geographic extents. This prevents analysis, for example, of predominantly rural areas. We are actively developing an approach to create residential synthetic populations for custom areas of interest (AOIs), supported by USA Structures. This functionality will also support the development of synthetic populations with national-scale coverage.
- **Open-sourcing core packages.** Though we have yet to meet our goal of open-sourcing the suite of Likeness packages [6], we are still on track to release the core packages for residential population synthesis, `pymedm` and `livelike`, in 2023. Releases of `movelike` and `actlike` are likely to follow in 2024.
- **Packaging schema.** A further consideration related to open-sourcing is whether we should migrate from a confederated toolkit schema where each module is a semi-independent Python package, as is seen in the modern implementation of PySAL [31], to a single monolithic Python package with submodules. Each schema has benefits, and this decision will require much consideration. With regards to the current confederated schema, the main benefit is modularity and reduced burden for continuous integration testing runtimes. This primary benefit is from a developer standpoint. However, providing a single package to install and use is a clear benefit to the user.
- **Consolidating visualization functionality.** We are in the process of consolidating functionality related to the visualization of input, processing, and results that have been used in an ad-hoc manner in the past. An initial push will be made for the inclusion of “made-to-order” population density hexbin plots and network-space allocation routes.
- **Improving mobility modeling.** Modeling public transit is a key area where we intend to develop increasingly more realistic agent “choices.” As stated previously in [Expansion and Scaling of Likeness Capabilities](#), there is significant potential in exploring further integration of GTFS data for locally-accurate modeling.
- **Optimization bottleneck.** As demonstrated in [Results](#), there is a clear hit in computational performance and runtime when solving `actlike.ActivityAllocation` problems on increasingly larger model instances (e.g., more agents and more POIs). There are two paths to resolving this issue (which may be considered in concert): 1) Reviewing our modified Transportation Problem mixed-integer program (formulated in [6]); and 2) Utilizing a new underlying solver engine, such as HiGHS¹⁶ [32]. In reviewing our formulation we will first investigate the potential for generating fewer constraints in the model. Following this, as stated above, we may consider formulating the model in a new solver. Depending on the outcome

15. <https://www.census.gov/programs-surveys/acs/news/data-releases/2022/release.html>

16. <https://highs.dev/>

of these experiments we may consider a new underlying optimization problem or implement a heuristic.

Acknowledgements

The authors would like to thank Ty Frazier for his contributions to scoping the incorporation of General Transit Feed Specification (GTFS) data during an earlier phase of this project.

The authors would also like to thank the two volunteer reviewers for their constructive feedback and the SciPy Proceedings editorial team for their ongoing support.

Notice: Research reported in this publication was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1-TR001409, KL2-TR001432 & TL1-TR001431. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-publicaccess-plan>).

REFERENCES

- [1] T. C. Germann, M. Z. Smith, L. R. Dauelsberg, G. Fairchild, T. L. Turton, M. E. Gorris, C. W. Ross, J. P. Ahrens, D. D. Hemphill, C. A. Manore, *et al.*, “Assessing k-12 school reopenings under different covid-19 spread scenarios—united states, school year 2020/21: A retrospective modeling study,” *Epidemics*, vol. 41, p. 100632, 2022. <https://doi.org/10.1016/j.epidem.2022.100632>.
- [2] J. Ozik, J. M. Wozniak, N. Collier, C. M. Macal, and M. Binois, “A population data-driven workflow for covid-19 modeling and learning,” *The International Journal of High Performance Computing Applications*, vol. 35, no. 5, pp. 483–499, 2021.
- [3] J. V. Tuccillo, R. Stewart, A. Rose, N. Trombley, J. Moehl, N. N. Nagle, and B. Bhaduri, “UrbanPop: A spatial microsimulation framework for exploring demographic influences on human dynamics,” *Applied Geography*, vol. 151, p. 102844, 2023. <https://doi.org/10.1016/j.apgeog.2022.102844>.
- [4] United States Department of Energy, “Foundational Science for Biopreparedness and Response: Report from the March 2022 Roundtable.” https://science.osti.gov/-/media/Initiatives/pdf/Biopreparedness_Roundtable_Report_092722.pdf, Mar. 2022. <https://doi.org/10.2172/1868508>.
- [5] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A.-L. Barabási, “Returners and explorers dichotomy in human mobility,” *Nature communications*, vol. 6, no. 1, p. 8166, 2015. <https://doi.org/10.1038/ncomms9166>.
- [6] J. V. Tuccillo and J. D. Gaboardi, “Likeness: a toolkit for connecting the social fabric of place to human dynamics,” in *Proceedings of the 21st Python in Science Conference* (M. Agarwal, C. Calloway, D. Niederhut, and D. Shupe, eds.), pp. 125–135, 2022. <https://doi.org/10.25080/majora-212e5952-014>.
- [7] J. V. Tuccillo and J. D. Gaboardi, “Likeness: a toolkit for connecting the social fabric of place to human dynamics.” <https://doi.org/10.25080/majora-212e5952-02d>, Aug. 2022. <https://doi.org/10.25080/majora-212e5952-02d>.
- [8] J. D. Gaboardi and J. V. Tuccillo, “Spatial Microsimulation and Activity Allocation for Examining COVID-19 Vaccine Access Profiles,” Mar. 2023. <https://doi.org/10.5281/zenodo.7768810>.
- [9] N. N. Nagle, B. P. Buttenfield, S. Leyk, and S. E. Spielman, “Dasymetric modeling and uncertainty,” *Annals of the Association of American Geographers*, vol. 104, no. 1, pp. 80–95, 2014.
- [10] M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert, “Efficient and Modular Implicit Differentiation.” arXiv preprint arXiv:2105.15183, 2021.
- [11] OpenStreetMap contributors, “Planet dump retrieved from <https://planet.osm.org>.” <https://www.openstreetmap.org>, 2023.
- [12] G. Boeing, “OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,” *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, Sept. 2017. <https://doi.org/10.1016/j.compenvurbysys.2017.05.004>.
- [13] F. Foti and P. Waddell, “A Generalized Computational Framework for Accessibility: From the Pedestrian to the Metropolitan Scale,” in *Transportation Research Board Annual Conference*, pp. 1–14, 2012.
- [14] S. Mitchell, M. O’Sullivan, and I. Dunning, “PuLP: A Linear Programming Toolkit for Python.” <https://optimization-online.org/?p=11731>, 2011.
- [15] H. G. Santos and T. A. Toffolo, “Mixed Integer Linear Programming with Python.” <https://www.python-mip.com/>, 2020.
- [16] R. Lougee-Heimer, “The Common Optimization INterface for Operations Research: Promoting open-source software in the operations research community,” *IBM Journal of Research and Development*, vol. 47, no. 1, pp. 57–66, 2003. <https://doi.org/10.1147/rd.471.0057>.
- [17] J. Forrest, T. Ralphs, H. G. Santos, S. Vigerske, J. Forrest, L. Hafer, B. Kristjansson, jpfasano, EdwinStraver, M. Lubin, Jan-Willem, rougee, jponcal1, S. Brito, h-i gassmann, Cristina, M. Saltzman, tostost, B. Pitrus, F. Matsushima, and to st, “coin-or/cbc: Release releases/2.10.10,” Apr. 2023. <https://doi.org/10.5281/zenodo.7843975>.
- [18] H. L. Yang, J. Yuan, D. Lunga, M. Laverdiere, A. Rose, and B. Bhaduri, “Building extraction at scale using convolutional neural network: Mapping of the United States,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2600–2614, 2018. <https://doi.org/10.1109/jstars.2018.2835377>.
- [19] H. L. Yang, M. Tuttle, M. Laverdiere, T. Hauser, B. Swan, E. Schmidt, J. Moehl, A. Reith, J. McKee, and M. Whitehead, “ORNL USA Structures 2022.” https://figshare.com/collections/ORNL_USA_Structures_2022/6139131/2, Sept. 2022. <https://doi.org/10.6084/m9.figshare.c.6139131.v2>.
- [20] F. L. Hitchcock, “The Distribution of a Product from Several Sources to Numerous Localities,” *Journal of Mathematics and Physics*, vol. 20, no. 1-4, pp. 224–230, 1941. <https://doi.org/10.1002/sapm1941201224>.
- [21] T. C. Koopmans, “Optimum Utilization of the Transportation System,” *Econometrica*, vol. 17, pp. 136–146, 1949. <https://doi.org/10.2307/1907301>.
- [22] H. J. Miller and S.-L. Shaw, *Geographic Information Systems for Transportation: Principles and Applications*, ch. 6: Network Flows and Facility Location. New York: Oxford University Press, 2001.
- [23] H. J. Miller and S.-L. Shaw, “Geographic Information Systems for Transportation in the 21st Century,” *Geography Compass*, vol. 9, no. 4, pp. 180–189, 2015. <https://doi.org/10.1111/gec3.12204>.
- [24] R. Lovelace and D. Ballas, “‘Truncate, replicate, sample’: A method for creating integer weights for spatial microsimulation,” *Computers, Environment and Urban Systems*, vol. 41, pp. 1–11, Sept. 2013. <https://doi.org/10.1016/j.compenvurbysys.2013.03.004>.
- [25] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004. <https://doi.org/10.1162/0899766042321814>.
- [26] J. H. Zar, “Significance testing of the spearman rank correlation coefficient,” *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 578–580, 1972. <https://doi.org/10.1080/01621459.1972.10481251>.
- [27] C. M. Macal, N. T. Collier, J. Ozik, E. R. Tataru, and J. T. Murphy, “Chisim: An agent-based simulation model of social interactions in a large urban area,” in *2018 winter simulation conference (WSC)*, pp. 810–820, IEEE, 2018. <https://doi.org/10.1109/wsc.2018.8632409>.
- [28] G. S. Thakur, B. L. Bhaduri, J. O. Piburn, K. M. Sims, R. N. Stewart, and M. L. Urban, “Planetsense: a real-time streaming and spatio-temporal analytics platform for gathering geo-spatial intelligence from open source data,” in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4, 2015.
- [29] M. W. Horner, M. D. Duncan, B. S. Wood, Y. Valdez-Torres, and C. Stansbury, “Do aging populations have differential accessibility to activities? Analyzing the spatial structure of social, professional, and business opportunities,” *Travel Behaviour and Society*, vol. 2, no. 3, pp. 182–191, 2015. <https://doi.org/10.1016/j.tbs.2015.03.002>.
- [30] B. S. Wood and M. W. Horner, “Understanding Accessibility to Snap-Accepting Food Store Locations: Disentangling the Roles of Transportation and Socioeconomic Status,” *Applied Spatial Analysis and Policy*, vol. 9, pp. 309–327, 2016. <https://doi.org/10.1007/s12061-015-9138-2>.

- [31] S. J. Rey, L. Anselin, P. Amaral, D. Arribas-Bel, R. X. Cortes, J. D. Gaboardi, W. Kang, E. Knaap, Z. Li, S. Lumnitz, T. M. Oshan, H. Shao, and L. J. Wolf, “The PySAL Ecosystem: Philosophy and Implementation,” *Geographical Analysis*, vol. 54, no. 3, pp. 467–487, 2022. <https://doi.org/10.1111/gean.12276>.
- [32] Q. Huangfu and J. A. J. Hall, “Parallelizing the dual revised simplex method,” *Mathematical Programming Computation*, vol. 10, no. 1, pp. 119–142, 2018. <https://doi.org/10.1007/s12532-017-0130-5>.