Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

GEOPOPS: AN OPEN-SOURCE PACKAGE FOR GENERATING GEOGRAPHICALLY REALISTIC SYNTHETIC POPUATATIONS

Alisa Hamilton[1], Sasha Tulchinsky[2], Gary Lin[3], Cliff Kerr[4], Eili Klein[2,5], and Lauren Gardner[1]

[1]Dept. of Civil and Systems Eng., Johns Hopkins University, Baltimore, MD, USA
[2]One Health Trust, Washington, DC, USA
[3]Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA
[4]Institute for Disease Modeling, Seattle, WA, USA
[5]Dept. of Emergency Medicine, Johns Hopkins University, Baltimore, MD, USA

ABSTRACT

Synthetic populations with spatially connected individuals are useful for modeling infectious diseases, particularly when assessing the impact of interventions on geographic and demographic subgroups. However, open-source tools that incorporate geospatial data are limited. We present a method to generate synthetic contact networks for any US region using publicly available data. GeoPops is an open-source package that 1) synthesizes a set of agents approximating US Census distributions, 2) assigns agents to homes, schools, and workplaces, and 3) connects agents within these locations. We compare GeoPops to two other synthesis tools that generate comparable network structures by creating Maryland populations with each package and assessing accuracy against US Census data. We simulate COVID-19 transmission on each population and compare simulations to observed data for the first wave of the pandemic. Our study highlights the utility of spatially connected synthetic populations and builds the capacity of modelers to better inform epidemic decision making.

1 INTRODUCTION

When used with agent-based models (ABMs), spatially explicit synthetic populations allow demographic and geographic subgroup outcome analysis, which can inform more efficient and equitable policy responses (e.g., targeted vaccinations). Several methods exist for generating agents within households that match Census-based demographic distributions (hierarchical sampling (HS), iterative proportional fitting (IPF), and combinatorial optimization (CO)), and published applications show relative accuracy for each. Fewer published methods exist for assigning agents to daily activities based on geographic locations and connecting them within these settings to approximate transmission pathways. Because model outcomes depend heavily on contact structure, understanding how populations and networks are built is essential for interpreting results. In this study, we present the open-source package GeoPops (Hamilton et al. 2025) (formerly GREASYPOP-CO (Tulchinsky et al. 2024)) and compare it to two other synthetic population generation packages.

2 METHODS

GeoPops uses CO to sample households from Public Use Microdata Samples (PUMS) to match marginal demographic targets from the American Community Survey (ACS) at the census block group level. Individuals are then assigned to schools and workplaces based on enrollment data and commute flows. Contact networks are generated using stochastic block models to capture assortative mixing patterns. We compare our method to two other packages that generate similar network structures: Geo-Synthetic-Pop (Jiang et al. 2024) which uses HS, and UrbanPop (Tuccillo et al. 2023) which uses a form of IPF called

Penalized Maximum-Entropy Dasymetric Modeling. All three packages use similar data sources to assign agents to workplaces and schools but different graph algorithms to connect agents within these locations. Geo-Synthetic-Pop uses small-world networks, and UrbanPop uses a custom grouping process. We generate synthetic populations for Maryland using each package and compare population characteristics to US Census benchmarks using the Freeman-Tukey statistic. For each population's home, school, and workplace networks, we conduct a basic network analysis. We then run a COVID-19 ABM built using the open-source software Starsim (Kerr et al. 2024) on each generated population and compare simulations to observed data, stratified by age, race/ethnicity, and county, for the first wave of the pandemic. Finally, we assess the usability and flexibility of each package, discussing documentation clarity, computational efficiency, and adjustability regarding geographic and demographic granularity and network generation.

3 RESULTS

All three packages generated Maryland populations that approximated US Census data for their specified targets, with UrbanPop achieving the best overall fit. GeoPops and UrbanPop incorporated richer demographic detail, including race, income, and workplace category. GeoPops and UrbanPop had more assortative school and workplace networks with higher mean degrees compared to Geo-Synthetic-Pop. In COVID-19 simulations, GeoPops and UrbanPop produced more realistic trends for overall and age- and county-stratified outcomes. GeoPops and UrbanPop were also able to capture outcomes by race/ethnicity similar to observed data. Due to lower network connectivity, Geo-Synthetic-Pop required more initial infections to start and sustain an outbreak and resulted in a smaller and later epidemic peak. Geo-Synthetic-Pop was the easiest to use with prebuilt populations. GeoPops had the clearest documentation and allowed users to generate populations at finer geographic scale but required longer runtime. UrbanPop offered the greatest flexibility for users but required more advanced coding experience.

4 DISCUSSION

Comparing synthetic population packages is inherently challenging, as they are often built for specific purposes. To ensure a meaningful comparison, we focused on packages with applications for COVID-19 modeling that include home, school, and workplace networks. While populations with greater geographic and demographic detail offer advantages—particularly for evaluating interventions by subgroup—more streamlined models may be preferable for rapidly assessing overall outcomes during an emerging crisis. However, it is important to allow users to adjust parameters that greatly impact transmission dynamics (e.g., network degree and assortativity). To support both high-resolution analyses and time-sensitive response modeling, it is essential to improve the accessibility and usability of these tools. This includes providing open-source code, clear documentation, and interfaces that accommodate users with varying levels of technical expertise. Expanding the capacity and reach of synthetic population tools will help ensure that modelers are better equipped to inform public health decisions in both routine and emergency contexts.

REFERENCES

Hamilton, Alisa, Sasha Tulchinsky, Gary Lin, Cliff Kerr, Eili Klein, and Lauren Gardner. 2025. *GeoPops: Geographically Realistic Synthetic Population Generator*. Johns Hopkins University and One Health Trust. https://github.com/ACCIDDA/GeoPops.

Jiang, Na, Fuzhen Yin, Boyu Wang, and Andrew T. Crooks. 2024. "A Large-Scale Geographically Explicit Synthetic Population with Social Networks for the United States." *Scientific Data* 11 (1): 1204. https://doi.org/10.1038/s41597-024-03970-1.

Kerr, Cliff, Robyn Stuart, Romesh Abeysuriya, et al. 2024. "Starsim: A Fast, Flexible Toolbox for Agent-Based Modeling of Health and Disease." https://starsim.org/.

Tuccillo, Joseph, Robert Stewart, Amy Rose, et al. 2023. "UrbanPop: A Spatial Microsimulation Framework for Exploring Demographic Influences on Human Dynamics." *Applied Geography* 151 (February): 102844. https://doi.org/10.1016/j.apgeog.2022.102844.

Tulchinsky, Alexander, Fardad Haghpanah, Alisa Hamilton, Nodar Kipshidze, and Eili Klein. 2024. "Generating Geographically and Economically Realistic Large-Scale Synthetic Contact Networks: A General Method Using Publicly Available Data." *arXiv*, ahead of print, June 20. https://doi.org/10.48550/arXiv.2406.14698.