A Simulation Approach for Modeling High-Resolution Daytime Commuter Travel Flows and Distributions of Worker Subpopulations

April Morton*[1], Jesse Piburn[1], Nicholas Nagle[2], H M Abdul Aziz[1], Samantha Duchscherer[1] and Robert N. Stewart[1]

[1] Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831
[2] University of Tennessee, Knoxville, Department of Geography, 1000 Phillip Fulmer Way, Knoxville, TN 37916
*Email: mortonam@ornl.gov

Abstract

High-resolution daytime travel flows and population distribution data are critical for addressing important issues ranging from business planning to emergency preparedness and response. Though dasymetric modeling approaches and Census datasets have been developed to capture daytime populations and travel flows, they often lack the spatial resolution, demographic detail and uncertainty measures needed by decision-makers. In this paper, we propose a high-resolution simulation approach for estimating and quantifying the uncertainty of worker subpopulation distributions and their commuter travel flows in the United States (U.S.). An early case study and novel visualization illustrate the model's potential utility and impact.

Keywords: Simulation, Population Dynamics, Commuter Travel Flows, Daytime Population Distributions

1. Introduction

In recent years, a stronger requirement for understanding daytime populations, the distribution of populations at times other than when they are expected to be at their residences at night, has emerged from both the business and emergency preparedness and response communities (Bhaduri 2008). Understanding the daytime population distribution not only provides a competitive economic advantage to businesses concerned with targeting specific consumers, but to the emergency preparedness and response community concerned with assessing the at-risk population from threats of technological disasters, natural disasters and terrorist attacks (Bhaduri 2008).

Though the temporal dynamics that affect daytime population distributions are well understood (Quinn 1950), few attempts have been made to incorporate such temporal daytime variability in a high-resolution population distribution model. The LandScan Global model (Dobson et. al. 2000), the earliest example of a population distribution model which considers diurnal change, reports "ambient" or average population counts over a 24-hour period. Since then, Los Alamos National Laboratory has developed a method for estimating urban daytime and nighttime populations at a 250 m resolution (McPherson & Brown 2004) and Oak Ridge National Laboratory has developed LandScan USA (Bhaduri et. al. 2007) to model nighttime, and daytime population counts at a 3 arc second resolution. Furthermore, in 1990, 2000, and again in 2006, the American Association of State Highway and Transportation Officials (AASHTO) partnered with the Census and all states in the United States (U.S.) to develop special census products (AASHTO 2010) for better understanding where people live and work, their journey to work commuting patterns and the modes they use for getting to there.

Despite the impact of such models and datasets, the lack of detailed population demographics within daytime distributions of populations is still troublesome, particularly when understanding who is affected by impactful events is crucial to our ability to respond effectively. Furthermore, many of these existing models lack the spatial resolution and uncertainty measures needed to make meaningful decisions. To move toward more informative and spatially detailed datasets, we propose a simulation approach for estimating and quantifying the uncertainty of worker subpopulation distributions and their commuter travel flows in the U.S. A case study and novel visualization illustrate the model's potential utility and impact. In addition, opportunities for improvement and further development are discussed.

2. Methodology

To produce high-resolution daytime commuter travel flows and distributions, we propose a model that simulates likely place of work Census block groups for sets of workers with known residential locations and workplace industries. The simulation method randomly draws probable work locations from multinomial distributions whose event probabilities describe the likelihoods that randomly chosen workers from given industries, living in known block groups and tracts, work in specific block groups. To infer these event probabilities, we first obtain the journey to work tract-to-tract mean travel flows by industry reported in table B302102 of the Census Transportation Planning Products (CTPP) (AASHTO 2010). We next infer the number of employees by industry in each block group from the 2010 Pitney Bowes Business Points dataset (Pitney Bowes 2010). This is done by converting each business point's North American Industry Classification System (NAICS) code to its corresponding CTPP industry code, and then summing over the number of employees working for each business in every industry and block group.

Given these input datasets, we then calculate, for each tract r, industry o, and block group b contained in tract c, the probability that a randomly selected resident living in tract r and working in industry o works in block group b through the equation

$$p_{rbo} = \frac{f_{rco}}{\sum_{j=1}^{d} f_{rjo}} \frac{e_{bo}}{\sum_{i=1}^{h} e_{io}} \tag{1}$$

where f_{rco} represents the number of workers who travel from tract r to tract c to work in industry o, e_{bo} represents the number of employees who work in block group b in industry o, h represents the number of block groups in tract c and d represents the number of unique non-zero tract-to-tract flows for a resident living in tract r and working in industry o. If there are no flows from a simulated resident's tract for his or her industry, we determine the nearest residential tract with available probabilities for that particular industry and use these probabilities and block groups to simulate the resident's likely place of work block group. For the sake of preserving the uncertainty, we simulate several likely places of work for each resident.

3. Case Study

To illustrate the utility of the proposed model, we estimate the 2010 block group to block group daytime travel flows and block group daytime distributions of all female workers 16 years and over in the professional, scientific, management, and administrative and waste management services traveling within the Knoxville Metropolitan Statistical Area. We estimate the 2010 residential block group locations of all residents in this subpopulation using the mean counts per block group reported in line 17 of summary table C24030 within the 2008-2012 ACS summary tables dataset (U.S. Census 2012). To capture the uncertainty, we simulate 50 possible commuter travel flows for each worker.

3.1 Results and Discussion

To summarize the model results, we present both a visualization of the mean and standard deviation travel flows on a road network and a map displaying the mean and standard deviation of worker counts in their daytime work block groups.

To visualize the flows, we use the Open Source Routing Machine (Luxen & Vetter 2004) to determine the optimal travel path of each simulated resident on a road network. Since the routing machine requires origin and destination points, rather than polygons, we use the centroid of each worker's origin block group as his or her origin point and randomly choose 30 destination business points within each of his or her simulated destination block groups. Workers are only assigned to business points corresponding to their work industries, and the probabilities that they are assigned to specific business points are based on each business' relative number of employees.

Fig. 1 shows the mean and standard deviation count of workers in the subpopulation who travel to work over the road network during the average weekday. As expected, workers in the subpopulation are much more likely to pass over the main interstates and popular side roads than the less popular back roads. Furthermore, we see that the main interstate 40 and highway 162, both major roads for the commute to Oak Ridge National Laboratory (ORNL), have much higher volumes of traffic then the other road networks. This makes sense, as ORNL is the largest company in the Knoxville MSA that falls in the professional, scientific, management, and administrative and waste management services category. Standard deviations indicate a similar pattern, with higher routes typically having higher standard deviation values.

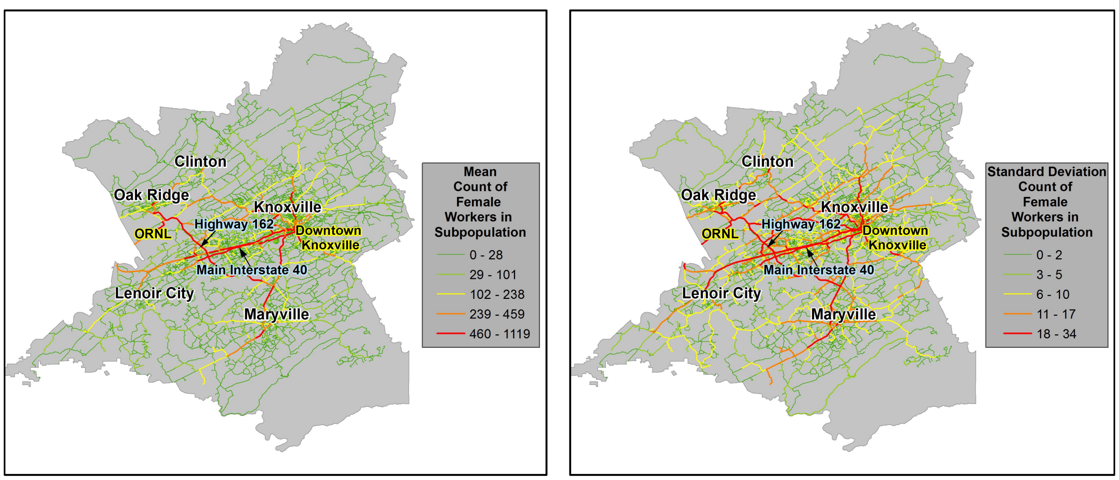

Figure 1: Mean and standard deviation count of female workers 16 years and over in the professional, scientific, management, and administrative and waste management services who travel to work over the road network during the average weekday.

Fig. 2 shows the mean and standard deviation daytime counts of female workers 16 years and over in the professional, scientific, management, and administrative and waste management services. As expected, the same block group containing ORNL has the most workers in our subpopulation. Furthermore, we see that, again, block groups with larger counts generally have larger standard deviations.

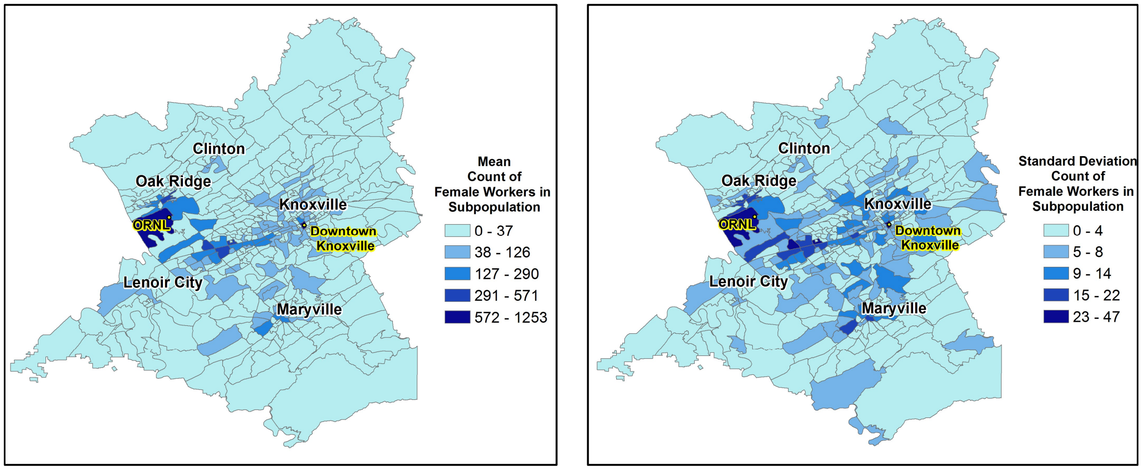

Figure 2: Mean and standard deviation daytime counts, aggregated to Census block groups, of female workers 16 years and over in the professional, scientific, management, and administrative and waste management services

It's important to note that this model is still in its beginning stages and that there are several opportunities for improvement. Perhaps the most obvious one is pursuing a creative way to validate the results, as ground truth data is often unavailable due to privacy issues. Furthermore, since the ACS and CTPP also publish margins of error associated with the mean residential counts and worker flows, it's possible to extend the framework by modeling the residents and event probabilities as random variables themselves. In addition, one could explore dasymetric techniques to apply the framework to more specific subpopulations and further refine the visualization approach for worker travel flows by exploring ways to improve upon the assumption that workers start their travel routes at block group centroids.

4. Conclusion

In this paper, we presented a high-resolution simulation approach for estimating and quantifying the uncertainty of specific worker subpopulation distributions and their commuter travel flows in the United States (U.S.). An early case study and visualizations of the results illustrate the model's potential utility and impact for targeted economic activities and emergency response efforts. Future work includes determining unique ways of validating the model, accounting for other sources of uncertainty, using the model to estimate the distribution of more specific subpopulations, and further developing the visualization approach for commuter travel flows.

5. Copywrite

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-public-access-plan).

6. Acknowledgements

This material is based upon work supported by the U.S. Department of Energy for the development of a high performance, data-driven simulator of the American population for modeling urban dynamics; Oak Ridge National Laboratory's Laboratory Directed Research and Development (LDRD).

7. References

AASHTO. 2010. Census transportation planning products. [Online]. [Accessed 6 Oct. 2016]. Available from: http://data5.ctpp.transportation.org/ctpp

Bhaduri, B., Bright, E., Coleman, P. and Urban, M.L. 2007. LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal, 69(1-2), pp.103-117.

Bhaduri, B. 2008. Population distribution during the day. In: Encyclopedia of GIS (pp. 880-885). Springer US.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C. and Worley, B.A. 2000. LandScan: A global population database for estimating populations at risk. Photogrammetric engineering and remote sensing, 66(7), pp.849-857.

Luxen, D. and Vetter, C. 2011. Real-time routing with OpenStreetMap data. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York, NY, pp. 513–516.

McPherson, T., and Brown, M. 2004. Estimating daytime and nighttime population distributions in U.S. cities for emergency response activities. In: Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone. Seattle, WA.

Pitney Bowes. 2010. U.S. Business Points.

Quinn, J.A. 1950. The daytime population of the Central Business District of Chicago.

US Census Bureau. 2009. A compass for understanding and using American community survey data: What researchers need to know.

U.S. Census Bureau. 2012. 2008 – 2012 American community survey summary tables. [online]. [Accessed 6 Oct. 2016]. Available from: http://factfinder2.census.gov.