Title
Improving the LandScan USA Non-Obligate Population Estimate (NOPE)

Authors
Brelsford, Christa
Moehl, Jessica
Weber, Eric
et al.

Publication Date
2022-09-09

DOI
10.25436/E2X307

Peer reviewed

Improving the LandScan USA Non-Obligate Population Estimate (NOPE)

Christa Brelsford[1][0000−0002−3490−8020], Jessica Moehl[1][0000−0001−9579−2562], Eric Weber[1][0000−0002−0098−3874], Kevin Sparks[1][0000−0002−0340−8090], and Amy Rose[1][0000−0003−1597−0301]

Geospatial Science and Human Security Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37830 USA
brelsfordcm@ornl.gov

Abstract. Where do people go when they have nowhere to be? Non-obligate activities are a significant part of our social and cultural lives, but there are no existing large scale data which characterize spatial variability in population allocation for these activities. As large scale population estimates have ever-finer resolutions, gaps in our ability to estimate this population segment have an increasingly large impact on high resolution population estimates. In this paper, we demonstrate an improved method for estimating the spatial allocation of the non-obligate population - people who are not at work, school, or in another residential institution. This method builds upon on anonymized and aggregate data on visits to public places, allocating the non-obligate population proportionally to worker population while accounting for the estimated ratio of visitors to workers in public places.

Keywords: Leisure Activities · Population Allocation · LandScan

DOI: https://doi.org/10.25436/E2X307

1 Introduction

A broad range of analyses and actions including emergency response and hazard mitigation rely on a clear picture of where people are and when they are there [4,6,7]. However, estimating population with high spatial and temporal fidelity is a substantial challenge, even in data-rich environments. Censuses, which are the traditional means for population enumeration, are limited in that they account for people only at the aggregate spatial location of their primary residence. Employment data focusing on workplace locations are becoming more available and there are myriad institutional data that report spatially-explicit information including school enrollments, prison populations, and hotel occupancy. Unfortunately, there are still critical gaps in our understanding of how human behavior in the context of social, cultural, and economic activity translates to physical location over time. One such gap is our lack of understanding of the patterns of what can be described as "non-obligate" public behavior. In other words, how can we account for the segment of the population that engages in activities throughout the course of the day that are not associated with primary activity-based locations such as home, work, and school? A fundamental challenge is the absence of published statistics regarding the number and locations of people when they are engaged in what are treated as secondary activities - leisure, socialization, civic and cultural engagement, and shopping. This is further confounded by the lack of information on the spatial and temporal variability of such activities.

We improve upon the LandScan USA method for allocating the non-obligate population by using newly available information reporting visits to Points of Interest (POI); in which the instantaneous ratio of 'visitors' to 'workers' at POIs is estimated based on visit duration [2]. We find that the preferred model increases the Non-Obligate Population Estimate (NOPE) in urban centers and retail or recreation areas, and reduces the estimated population in outlying counties and industrial areas. This method is used in the LandScan USA 2021 release.

2 Data

We use 2019 population estimates from the American Community Survey data for block group residential population [8]. We use 2019 Longitudinal employer-household dynamics Origin-Destination Employment Statistics (LODES) population data for the block group level worker population [9]. We use [2] for the long-to-short visit ratio, γ_i.

We use information from the American Time Use Survey published in response to the COVID-19 pandemic to estimate the share of population which is likely to be in public engaged in a choice based activity, as opposed to at work or school [3]. These data show that 11% of waking time is spent in a public place that is not work or school, and we use this as our baseline estimate of the total non-obligate population at any given time.

We explore two potential catchment areas for non-obligate travel. The first is simply all 3142 United States counties or county equivalents. The second catchment area is based on Combined, Micro- and Metropolitan Statistical Areas (MSAs). These catchments are a combination of counties and equivalents, and micro- and metropolitan areas as defined by the United States census.For places which are not in an MSA, we define the catchment area for NOPEs to be the county. For places which are in an MSA, the catchment area is the MSA. This results in 2228 MSA based catchment areas, of which 926 are MSAs, and 1302 are non-metro counties or equivalents.

For model validation only, we use data on anonymized and aggregated foot traffic to public places curated by SafeGraph, a geospatial data provider [5]. These data present statistics on visits to public places and the SafeGraph-inferred set of home block groups for weekly visitors to public places based on mobile device records. Brelsford et al. [2] describe the SafeGraph data and it's application for model validation in more detail.

3 Methods

We compare three potential models for the the spatial allocation of NOPE in public spaces. Model 1 represents the NOPE allocation model implemented in LandScan USA [1] products between 2007 and 2020. In model 1, we distribute NOPEs within the county proportionally to the block group level LODES worker population. Model 2 includes one methodological change: allowing the non-obligate population to travel within their metropolitan area, rather than merely within their home county. Model 3 represents an additional incremental change: allocating NOPEs proportionally to the product of the measured short-to-long visit ratio (γ_i) and LODES worker population, and also allowing travel within metropolitan based catchment areas.

$$\text{Model 1: } n_i = w_i \frac{N_c}{W_c} \tag{1}$$

$$\text{Model 2: } n_i = w_i \frac{N_m}{W_m} \tag{2}$$

$$\text{Model 3: } n_i = \gamma_i * w_i \frac{N_m}{\sum_{i \in m} (\gamma_i * w_i)} \tag{3}$$

where n is the fine-scale NOPE, w is LODES measured workers, and N or W index total known NOPEs or workers in a larger area. Subscripts $_i$, $_c$, and $_m$ index block groups, counties, and metro based catchment areas, respectively. Finally, γ_i is the 2019 average block group level short-to-long visit ratio presented in [2], where the authors propose that γ_i is an acceptable measure of the typical ratio of workers at a place to non-obligate visitors to a place.

Our core analytical goal is to select the most appropriate model among models 1, 2, and 3. There is no existing large scale measurement of the spatial properties of non-obligate behavior, and so we need to use other validation strategies.

To select between Model 1 and Model 2, we measure the increase in captured shares of visits between using MSAs and counties as a catchment area. In city level analyses we explore the spatial characteristics of within-county vs within-MSA visits. We look for the existence of edge effects across counties within MSAs, in comparison to across MSA boundaries.

To select between Model 3 and Models 1 or 2, we highlight the largest changes in NOPE by block group between Models 3 and 1. In these 'large change' block groups, we test if the categories of POIs in these locations are consistent with correcting the errors we expect to occur if NOPE are allocated only with workers.

4 Results

Figure 1 shows the NOPE population density for Tennessee. As expected, the large spatial scale patterns in NOPE closely mirror residential population density. Both are clustered in cities, towns, and population centers. However, at

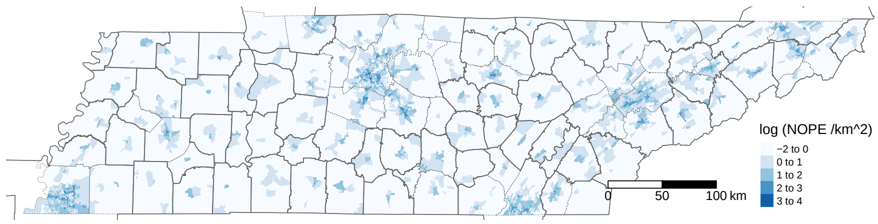

Fig. 1. Log_{10} of NOPE population density (NOPE per square km) based on Model 3.

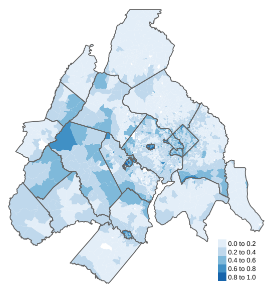

Fig. 2. The share of visits originating in the Washington DC Metro area which leave their origin county but remain in the Metro area. This is the share of total measured visits which are excluded when county-based catchment areas are used.

finer scales NOPE are more concentrated in public parts of cities- along transportation corridors and in shopping, recreation, and entertainment spaces.

What is the appropriate scale to model NOPE travel behavior? Figure 2 shows a block group level map describing the local share of visits with a destination in the same county or metro area based on the travel patterns data. This figure shows the share of visits originating in each block group that cross a county boundary but not the MSA boundary, at the block group level for the Washington DC metropolitan area. In some of the smaller counties, over 60 percent of all identified visits cross a county boundary, but remain in the MSA. This gives a quantitative description of how the changed NOPE catchment area allows actual behavior patterns to be statistically represented in models 2 and 3.

Figure 3a shows the difference in NOPEs when a county (model 1) vs MSA (model 2) based catchment area is used for the spatial allocation of NOPEs in Tennessee. Note that for counties which are not in a multi-county MSA, there is no change in NOPEs. However, for counties which are, there is a meaningful shift in the NOPEs from outlying counties into the central counties of their MSA. This is consistent with our understanding of leisure and recreational behavior: people in outlying counties are more likely to seek activities “in the city” than people in the central parts of cities are to seek out activities in the urban and suburban surrounds.

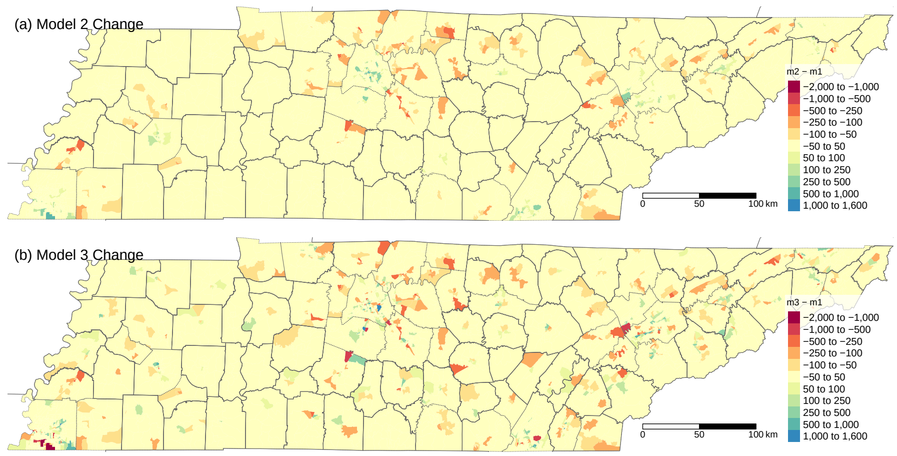

Fig. 3. The difference between Model 2(a) or Model 3(b) and Model 1.

Figure 3b shows the difference between model 3 and model 1 for all of Tennessee, and the state's three largest cities. The spatial shift of NOPEs from outlying counties into central urban counties is visible, as is the within county shift away from industrial areas and towards retail areas. Within individual cities, we can see that changes relative to the baseline model are most significant in blocks with extreme γ_i values, for example shopping malls and industrial work spaces not open to the public.

Table 1 shows TN block groups with the most extreme differences between Model 2 and Model 1 statewide. Most of these block groups are located in or near Tennessee's three largest cities. Block groups where Model 3 predicts more shoppers than Model 1 are dominated by retail spaces - with the exception of a block group in Cleveland, TN, where the 2 POI's are churches. Block groups where Model 3 predicts fewer shoppers than Model 1 are dominated by industrial facilities and office buildings. This is consistent with our expectations - Model 1 does not distinguish between commercial and industrial activities to allocate shoppers across space, while Model 3 does so, albeit imperfectly.

These examples show that the changes from Model 1 to Model 3 are consistent with expectations. Model 3 puts fewer visitors in places we don't expect the non-obligate population to visit regularly, such as manufacturing, distribution, and warehousing, as well as office space. Model 3 puts more visitors in places we do expect the non-obligate population to visit frequently: entertainment, shopping, cultural engagement, and commercial travel. This suggests that Model 3 is a significant improvement over model 1 for the spatial allocation of the non-obligate population.

Table 1. Summary Data from the Tennessee block groups with the greatest positive or negative extremes between Model 3 and Model 1 visitors.

Tract	County	M3 NOPE	M3 - M1	n poi	Dominant POI categories
		Smallest Model 3 - Model 1			
226	Shelby	1917.3	-1487.1	179	Industrial
9801	Shelby	1631.3	-1090.2	26	Memphis Airport
402	Rutherford	509.9	-942.6	115	Factory
42	Shelby	1213.3	-854.2	84	Office & Hotels
9801	Anderson	540.0	-774.3	9	Industrial
102.01	Maury	243.9	-772.0	43	GM Factory
		Largest Model 3 - Model 1			
195	Davidson	2342.0	883.2	279	Entertainment
9801	Davidson	1892.8	1023.3	87	Nashville Airport
104	Bradley	1388.5	1116.8	2	Church
211.13	Shelby	1658.6	1119.2	116	Large Mall
165	Davidson	3671.2	1289.1	12	Vanderbilt
503.07	Williamson	3174.0	1584.4	441	Large Mall

References

1. Bhaduri, B., Bright, E., Coleman, P., Urban, M.L.: LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal **69**(1-2), 103–117 (Sep 2007). https://doi.org/10.1007/s10708-007-9105-9, http://link.springer.com/10.1007/s10708-007-9105-9
2. Brelsford, C., Moehl, J., Weber, E., Sparks, K., Tuccillo, J.V., Rose, A.: Spatial and Temporal Characterization of Activity in Public Space, 2019-2020. Nature Scientific Data (2022). https://doi.org/10.1038/s41597-022-01480-6
3. Bureau of Labor Statistics: May to December 2019 and 2020 Results. News Release USDL-21-1359, US Department of Labor (Jul 2021), https://www.bls.gov/news.release/archives/atus_07222021.pdf
4. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences **111**(45), 15888–15893 (Nov 2014). https://doi.org/10.1073/pnas.1408439111
5. SafeGraph: SafeGraph Data Consortium (2020), https://www.safegraph.com/covid-19-data-consortium
6. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. pp. 851–860 (2010)
7. Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., Shibasaki, R.: Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1231–1239 (2013)
8. U.S. Census Bureau: 2015-2019 American Community Survey 5-Year Estimates. Tech. rep. (2019), https://data.census.gov
9. U.S. Census Bureau: LEHD Origin-Destination Employment Statistics Data (2002-2018) (2021), https://lehd.ces.census.gov/data/#lodes