Developing a Park Activity Location Choice Model from Passive Origin-Destination Data Tables

Gregory Macfarlane[a,*], Teresa Tapia, Carole Turley-Voulgaris[b]

[a] *Civil and Environmental Engineering Department, 430 Engineering Building, Provo, Utah 84602*
[b] *Some Other Place*

Abstract

Parks provide important benefits to those who live near them, in the form of improved property values, health outcomes, etc.; nevertheless, measuring and understanding who lives near a park is an open research question. In particular, it is not well understood which park individuals will choose to use when given a choice among a set of nearby parks of varying sizes and at varying distances from their home. In this paper we present a park activity location choice model estimated from a passive origin-destination dataset — supplied by StreetLight Data, Inc. — representing trips to parks and green spaces in Alameda County, California. The estimated model parameters reveal heterogeneous preferences for park size and willingness-to-travel across block-group level socioeconomic segmentation: Specifically, high-income block groups appear more positively attracted to larger parks, and block groups with a high proportion of ethnic minority individuals are more likely to select nearby parks. The findings have importance for understanding recreational access among different populations, and the methodology more generally supplies a potential template for using passive data products within travel modeling.

Keywords: Accessibility Passive Data Location Choice

1. Introduction

Parks and other green spaces generate immense value for the public who are able to access them. The City Parks Alliance [6] categorizes the observed benefits of urban parks as encouraging active lifestyles [1], contributing to local economies, aiding in stormwater management and flood mitigation, improving local air quality, increasing community engagement [14], and enhancing public equity.

*Corresponding Author

Email addresses: `gregmacfarlane@byu.edu` (Gregory Macfarlane), `teresa.tapia@streetlightdata.com` (Teresa Tapia), `cat@example.com` (Carole Turley-Voulgaris)

Nevertheless, understanding and quantifying these benefits depends in many cases on identifying who lives near the parks and is therefore able to access them. Many previous studies [e.g., 22] rely on comparison of total greenspace across metropolitan areas; this methodology may not adequately control for city-level fixed effects and it may ignore the potentially inequitable distribution of park space within a region. Studies focusing on access within metropolitan areas typically assume that people living within a certain distance or travel time threshold have access to a park, or examine the quantity of park space within one's own arbitrarily defined "neighborhood" [18, Stark2014]. But these methods do not account for the fact that some people will travel to other parks to perform recreational activities. A more holistic measure that continuously measures access across multiple preference dimensions is desirable.

An appealing solution would be to examine and model the activity location choices of park users. Such a model would help researchers understand how individuals of different backgrounds and preferences value different park amenities. Further, the logsums of a location choice model provide a continuous measure of accessibility that explicitly accounts for such variation [7]. Unfortunately, park choice models of this form are rare in the literature. Travel demand models built for infrastructure forecasting are a common way to generate such accessibility logsums, but these models group many different kinds of social and recreational trips together [21]. Further, the attraction term for such trip purposes is commonly a function of the retail or service employment or the number of households at the destination; a typical park or green space has neither employees nor residents. Finally, many regional household travel surveys are oriented towards an average weekday travel pattern, and many park trips occur irregularly or on weekends.

In this paper we present a park destination choice model where individuals living in Alameda County, California choose among parks in the same county. The individuals are constructed from passive data that was derived from mobile devices and processed using algorithms developed by StreetLight Data, Inc. The origin location points are inferred residence block groups for unique devices and the destination points are geofenced polygons representing green and open spaces. The individuals' choice of park location is conditioned on the distance from the block group to the parks in the choice set as well as the size of each park; market segmentation allows for heterogeneous responses between ethnic groups and income strata.

The paper proceeds in the following manner: A discussion of prior attempts to study park choice and employ passive origin-destination data in the literature is given directly. The Methodology section presents the data gathering and cleaning efforts as well as the econometric location choice model. The Results section presents the estimated model coefficients and a discussion of the findings, as well as a model validation exercise. After presenting limitations and associated avenues for future research, a final Conclusions section outlines the contributions of this study for recreational trip modeling and location choice modeling more generally.

2. Literature

Understanding who has access to parks is a long-standing question across multiple scientific disciplines. Researchers specializing in recreation management, public health, urban planning, ecology, and civil engineering have all played a role in shaping our collective understanding of park design, access, and use. A complete review of all of these fields is not warranted for the scope of this paper, but some recent findings are worth discussion.

A popular measure of park accessibility is the Trust for Public Land's "ParkScore" statistic [25]. ParkScore considers the share of the population that resides within a 10-minute walk of a green space using a sophisticated network routing algorithm, in combination with the total city green space, investment, and amenities weighted against the socioeconomic characteristics of the population outside of the 10-minute walk threshold. The resulting score is a convenient quantitative tool in estimating the relative quality of green space access across cities [23]. It may be less useful at identifying the comparative quality of access within a city, particularly as more than 95% of residents in many large metropolitan areas like San Francisco and New York live within the binary 10-minute walk threshold. The Centers for Disease Control and Prevention (CDC) has developed an "Accessibility to Parks Indicator" along similar lines [26], calculating the share of the population living within a half-mile of a park for each county in the U.S.

There is recognition that park access should in some way be linked with park use. After all, a park that has many visitors must by definition be accessible to those visitors. McCormack et al. [15] provide a comprehensive review of this literature; it is sufficient here to note that most studies find a complicated interplay between park size, maintenance, facilities, and travel distance. Many of these attributes are incorporated into ParkIndex [10], which estimates the resident park use potential within $100m^2$ grid cells, based on a household park use survey in Kansas City.

From a transportation engineering perspective, the park use potential measured by ParkIndex is not dissimilar from a park trip production potential. Along these lines, the question of park use is a destination choice problem, where trip makers consider which park is most attractive to accomplish their recreation activity. The Institute of Transportation Engineers (ITE) Trip Generation Manual [2017] contains trip attraction rates for public parks that use as attraction terms the park acreage, number of picnic tables, employees, and other variables. As with many land uses in Trip Generation, the provided trip generation rates are based on a limited number of observational samples and may not represent large-sample behavior [17]. Moreover, regression-based attraction rates isolated from trip production and travel behavior ignore the geographical and behavioral contexts in which people actually make trips to parks [2]: Though more people may come to larger parks, a park cannot attract more people simply by becoming bigger.

There are limited examples of researchers using a destination choice model to predict recreation attractions. Kinnell et al. [11] apply a choice model to a survey of park visitors in New Jersey, and estimate the relative attractiveness of park attributes including playgrounds, picnic areas, and park acreage weighed against the travel disutility and the relative crime rate at the destination. In a similar study, Meyerhoff et al. [16] model the urban swimming location choice for a surveyed sample. In both studies, the researchers were attempting to ascertain which attributes of a recreation generated the most positive utility, and therefore which attributes should be prioritized for improvement. These studies have not to our knowledge been previously referenced in discussions of park accessibility.

2.1. Passive Location Data

The advent of large-scale mobile networks and the seemingly perpetual association of unique devices with unique users has given researchers a new opportunity to observe the movements and activity location patterns for large subsets of the population [20]. Such passively collected movement data — sometimes referred to as "Big Data" — is passively collected as a by-product of other systems including cellular call data records [e.g., 4, 5], probe GPS data [8], and more recently Location Based Services (LBS) [24, 12]. LBS use a network of mobile applications that obtain the users' physical location. A variety of commercial vendors repackage, clean, and scale these data to population or traffic targets and sell origin-destination matrices to researchers and practitioners at relatively low prices. Monz et al. [19], for example, demonstrate how passive device data can be used to accurately estimate trips to natural recreation areas.

Passive origin-destination matrices are beginning to inform trip distribution model development more directly as well. Kressner [13] proposes one methodology, where passive origin-destination matrices serve as a probabilistic sampling frame for a simulated trip destination choice. Bernardin et al. [3] employ a passive origin-destination matrix as a shadow price reference in an activity-based location choice model, iteratively adjusting the parameters of the choice utilities to minimize the observed error between the matrix and the modeled predictions. A similar method developed by Zhu and Ye [27] uses the passive dataset directly, sampling 10,000 random trips from GPS traces of taxi trips in Shanghai and estimating a destination choice model. Employing the passive data set in this way provides the authors an opportunity to examine the choices of a large sample of a small population (taxi passengers) as well as sufficient data to estimate a "constants-rich" destination choice model with uniquely estimated coefficients for each origin-destination pair. The Zhu and Ye methodology suggests that a similar approach should apply in other contexts, including park choice.

3. Methods

We describe our methods in this chapter.

3.1. Data

4. Applications

Some *significant* applications are demonstrated in this chapter.

4.1. Example one

4.2. Example two

5. Final Words

We have finished a nice book.

References

[1] Carolyn Bancroft, Spruha Joshi, Andrew Rundle, Malo Hutson, Catherine Chong, Christopher C. Weiss, Jeanine Genkinger, Kathryn Neckerman, and Gina Lovasi. Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review. *Social Science & Medicine*, 138:22–30, aug 2015. ISSN 02779536. doi: 10.1016/j.socscimed.2015.05.034. URL https://linkinghub.elsevier.com/retrieve/pii/S0277953615003160.

[2] P. O. Barnard and R. E. Brindle. A review and critique of current methods used to predict traffic generation with some accompanying suggestions on alternative approaches. *Transportation Planning and Technology*, 11(4): 273–288, jun 1987. ISSN 0308-1060. doi: 10.1080/03081068708717349. URL http://www.tandfonline.com/doi/abs/10.1080/03081068708717349.

[3] Vincent L Bernardin, John L Bowman, Mark Bradley, Jason Chen, Nazneen Ferdous, and Yuen Lee. Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study. In *TRB Annual Meeting*. Transportation Research Board, 2018. URL https://trid.trb.org/view/1495930.

[4] R. Bolla and F. Davoli. Road traffic estimation from location tracking data in the mobile cellular network. In *2000 IEEE Wireless Communications and Networking Conference. Conference Record (Cat. No.00TH8540)*, pages 1107–1112. IEEE, 2000. ISBN 0-7803-6596-8. doi: 10.1109/WCNC.2000.904783. URL http://ieeexplore.ieee.org/document/904783/.

[5] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, 10(4):36–44, 2011. doi: 10.1109/mprv.2011.41. URL http://dx.doi.org/10.1109/mprv.2011.41http://hdl.handle.net/1721.1/101623http://creativecommons.org/licenses/by-nc-sa/4.0/.

[6] City Parks Alliance. Why City Parks Matter, 2019. URL https://cityparksalliance.org/about-us/why-city-parks-matter/.

[7] Gerard de Jong, Andrew Daly, Marits Pieters, and Toon van der Hoorn. The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41(9):874–889, nov 2007. ISSN 0965-8564. doi: 10.1016/J.TRA.2006.10.002. URL https://www.sciencedirect.com/science/article/pii/S0965856407000316.

[8] Arthur Huang and David Levinson. Axis of travel: Modeling non-work destination choice with GPS data. *Transportation Research Part C: Emerging Technologies*, 58:208–223, sep 2015. ISSN 0968-090X. doi: 10.1016/J.TRC.2015.03.022. URL https://www.sciencedirect.com/science/article/pii/S0968090X15001072.

[9] Institute of Transportation Engineers. *Trip Generation Manual*. Institute of Transportation Engineers, Washington, D.C., 10th editi edition, 2017. ISBN 9781933452647.

[10] Andrew T. Kaczynski, Jasper Schipperijn, J. Aaron Hipp, Gina M. Besenyi, Sonja A. Wilhelm Stanis, S. Morgan Hughey, and Sara Wilcox. ParkIndex: Development of a standardized metric of park access for research and planning. *Preventive Medicine*, 87:110–114, jun 2016. ISSN 0091-7435. doi: 10.1016/J.YPMED.2016.02.012. URL https://www.sciencedirect.com/science/article/pii/S0091743516000517.

[11] J. C. Kinnell, M. F. Bingham, A. F. Mohamed, W. H. Desvousges, T. B. Kiler, E. K. Hastings, and K. T. Kuhns. Estimating Site Choice Decisions for Urban Recreators. *Land Economics*, 82(2):257–272, may 2006. ISSN 0023-7639. doi: 10.3368/le.82.2.257. URL http://le.uwpress.org/cgi/doi/10.3368/le.82.2.257.

[12] Anurag Komanduri, Laura Schewel, Dan Beagan, and Dale Wong. Using Big Data to Develop Comparative Commercial Vehicle Metrics for Port Traffic at Major Ports in the U.S. In *TRB Annual Meeting*. Transportation Research Board, 2017. URL https://trid.trb.org/View/1438797.

[13] Josephine D Kressner. Synthetic Household Travel Data Using Consumer and Mobile Phone Data. Technical report, Transportation Research Board, 2017.

[14] Juliana Madzia, Patrick Ryan, Kimberly Yolton, Zana Percy, Nick Newman, Grace LeMasters, and Cole Brokamp. Residential Greenspace Association with Childhood Behavioral Outcomes. *The Journal of Pediatrics*, dec 2018. ISSN 0022-3476. doi: 10.1016/J.JPEDS.2018.10.061. URL https://www.sciencedirect.com/science/article/pii/S0022347618315683.

[15] Gavin R. McCormack, Melanie Rock, Ann M. Toohey, and Danica Hignell. Characteristics of urban parks associated with park use and physical activity: A review of qualitative research. *Health & Place*, 16(4):712–726, jul 2010. ISSN 1353-8292. doi: 10.1016/J.HEALTHPLACE.2010.03.003. URL https://www.sciencedirect.com/science/article/pii/S1353829210000316.

[16] Jürgen Meyerhoff, Alexandra Dehnhardt, and Volkmar Hartje. Take your swimsuit along: the value of improving urban bathing sites in the metropolitan area of Berlin. *Journal of Environmental Planning and Management*, 53 (1):107–124, jan 2010. ISSN 0964-0568. doi: 10.1080/09640560903399863. URL http://www.tandfonline.com/doi/full/10.1080/09640560903399863.

[17] Adam Millard-Ball. Phantom trips: Overestimating the traffic impacts of new development. *Journal of Transport and Land Use*, 8(1):31–49, 2015. doi: 10.5198/jtlu.2015.384. URL https://www.jstor.org/stable/pdf/26202700.pdf?refreqid=excelsior{%}3A6bf53b09919c223966f1dc8468372150.

[18] Richard Mitchell and Frank Popham. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*, 372(9650):1655–1660, nov 2008. ISSN 0140-6736. doi: 10.1016/S0140-6736(08)61689-X.

[19] Christopher Monz, Milan Mitrovich, Ashley D'Antonio, and Abigail Sisneros-Kidd. Using Mobile Device Data to Estimate Visitation in Parks and Protected Areas: An Example from the Nature Reserve of Orange County, California. *Journal of Park and Recreation Administration*, 0 (0), sep 2019. ISSN 21606862. doi: 10.18666/JPRA-2019-9899. URL https://js.sagamorepub.com/jpra/article/view/9899.

[20] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161, 2016. ISSN 1553-877X. doi: 10.1109/COMST.2015.2491361. URL http://ieeexplore.ieee.org/document/7299258/.

[21] National Academies of Sciences Engineering and Medicine. *Travel Demand Forecasting: Parameters and Techniques*. NCHRP 716. National Academies Press, Washington, D.C., 2012. ISBN 978-0-309-28255-0. doi: 10.17226/14665. URL http://www.nap.edu/catalog/14665.

[22] Elizabeth A Richardson, Richard Mitchell, Terry Hartig, Sjerp de Vries, Thomas Astell-Burt, and Howard Frumkin. Green cities and health: a question of scale? *Journal of Epidemiology and Community Health*, 66(2): 160 LP – 165, feb 2012. doi: 10.1136/jech.2011.137240.

[23] Alessandro Rigolon, Matthew Browning, and Viniece Jennings. Inequities in the quality of urban park systems: An environmental justice investigation of cities in the United States. *Landscape and Urban Planning*, 178:156–169, oct 2018. ISSN 0169-2046. doi: 10.1016/J.LANDURBPLAN.2018.05.026. URL https://www.sciencedirect.com/science/article/pii/S0169204618304316.

[24] Josh Roll. Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon. Technical report, Oregon Department of Transportation, 2019. URL https://trid.trb.org/View/1630252.

[25] Trust for Public Land. 2019 ParkScore Index, 2019. URL https://www.tpl.org/parkscore.

[26] Emily Neusel Ussery, Leah Yngve, Dee Merriam, Geoffrey Whitfield, Stephanie Foster, Arthur Wendel, and Tegan Boehmer. The National Environmental Public Health Tracking Network Access to Parks Indicator:

A National County-Level Measure of Park Proximity. *Journal of park and recreation administration*, 34(3):52, 2016. doi: 10.18666/JPRA-2016-V34-I3-7119. URL http://www.ncbi.nlm.nih.gov/pubmed/28868528http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5580831.

[27] Jiayu Zhu and Xin Ye. Development of destination choice model with pairwise district-level constants using taxi GPS data. *Transportation Research Part C: Emerging Technologies*, 93:410–424, aug 2018. ISSN 0968-090X. doi: 10.1016/J.TRC.2018.06.016. URL https://www.sciencedirect.com/science/article/pii/S0968090X18306454?via{%}3Dihub{#}b0085.