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

Gregory S. Macfarlane
Civil and Environmental Engineering Department
Brigham Young University, Provo, Utah 84602
Email: gregmacfarlane@byu.edu

Teresa Tapia
Customer Success Manager
StreetLight Data, Inc.
San Francisco, CA 94117
Email: teresa.tapia@streetlightdata.com

Word Count: 5762 words + 4 table(s) $\times$ 250 = 6762 words

Submission Date: August 2, 2019

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: park access, destination choice, passive data

INTRODUCTION

Parks and other green spaces generate immense value for the public who are able to access them. The City Parks Alliance (2019) categorizes the observed benefits of urban parks as encouraging active lifestyles (Bancroft et al., 2015), contributing to local economies, aiding in stormwater management and flood mitigation, improving local air quality, increasing community engagement (Madzia et al., 2018), and enhancing public equity.

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., Richardson et al., 2012) 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. Metropolitan-level efforts 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" (Mitchell and Popham, 2008; Stark et al., 2014). 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 (de Jong et al., 2007). 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 (National Academies of Sciences Engineering and Medicine, 2012). 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.

LITERATURE REVIEW

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 (2019). 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 (Rigolon et al., 2018). 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 (Ussery et al., 2016), 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. (2010) 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 (Kaczynski et al., 2016), 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 (Institute of Transportation Engineers, 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 (between 2 and 11) and may not represent large-sample behavior (Millard-Ball, 2015). 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 (Barnard and Brindle, 1987): 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. (2006) 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, (*23*) 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.

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 (Naboulsi et al., 2016). 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. Bolla and Davoli, 2000; Calabrese et al., 2011), probe GPS data (Huang and Levinson, 2015), and more recently Location Based Services (LBS) (Roll, 2019; Komanduri et al., 2017). 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. A common application for such passive origin-destination matrices is in constructing or validating the trip distribution components of regional travel demand models (Huntsinger and Donnelly, 2014) or in constructing non-behavioral elements of such models; for example, Huntsinger and Ward (2015) demonstrate the use of a passive origin-destination matrix in an external trip model.

Passive origin-destination matrices are beginning to inform trip distribution model development more directly as well. Kressner (2017) proposes one methodology, where passive origin-destination matrices serve as a probabilistic sampling frame for a simulated trip destination choice. Bernardin et al. (2018) 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 (2018) 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 both 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 other contexts, including park choice.

METHODOLOGY

We constructed a dataset on which to estimate park trip destination choices for a sample of observed trips in Alameda County, California. Alameda County is one of the seven counties that constitutes the San Francisco Bay Area metropolitan region in California. Alameda is the seventh most populous county in California with a population of 1.5 million (U.S. Census Bureau, 2019), and has 14 incorporated cities and several unincorporated communities. It is an economically and ethnically diverse county and hence it was an attractive area to use for this study. The racial makeup of Alameda County was (49.7%) White, (11.2%) African American, (1.0%) Native American, (38.7%) Asian, (1.0%) Pacific Islander, and (22.4% Hispanic or Latino (of any race). Alameda County has a diverse set of parks, ranging from local small community parks, urban and transit-accessible parks like the Lake Merritt Recreational area, accessible coastal access, and suburban recreational areas like Lake Chabot.

Data

We constructed an analysis dataset from a publicly-available parks polygons layer, a commercially acquired passive device origin-destination table representing trips between the parks and home block groups, and American Community Survey data for the home block groups.

We obtained a polygons shapefile layer representing open spaces in Alameda County,

TABLE 1 Descriptive Statistics of Study Block Groups

Park Type	N	Median Acres (IQR)	Total Visits
Local Park	441	3.89 (1.31, 9.81)	639,463
Local Recreation Area	57	9.27 (3.58, 53.79)	156,720
State Recreation Area	2	421.05 (325.00, 517.09)	161

California from the California Protected Areas Database (GreenInfo Network, 2019). This dataset was selected because it included multiple different types of open space including local and state parks, traditional green spaces as well as wildlife refuges and other facilities that can be used for recreation. We removed facilities that did not allow open access to the public (such as the Oakland Zoo) and facilities whose boundaries conflated with freeway right-of-way – this prevents trips through the park from being conflated with park trips in the passive origin-destination data. Table 1 shows some descriptive characteristics of the parks data.

We provided the park boundaries layer to a commercial firm, StreetLight Data Inc., which develops and resells origin-destination matrices derived from passive device location data. The provider employs a proprietary data processing engine (called Route Science) to algorithmically transform observed device location data points (the provider uses in-vehicle GPS units and mobile device LBS) over time into contextualized, normalized, and aggregated travel patterns. From these travel patterns, the Route Science processing algorithms infer likely home Census block group locations for composite groups of people and converts raw location data points into trip origin and destination points (Pan et al., 2006; Friedrich et al., 2010).

For each park polygon, the firm returned a population-weighted estimate of how many devices were observed by home location block group over several months in the period between May 2018 and October 2018. We transformed this table such that it represented the weighted unique devices traveling between block groups and parks. We discarded home location block groups outside of Alameda County; the imputed home locations can be far away from the study area for a small amount of trips and are unlikely to represent common or repeated park activities.

In order to understand the demographic makeup of the home block groups and potentially the characteristics of the people who make each trip, we obtained 2013-2017 five-year data aggregations from the American Community Survey (U.S. Census Bureau, 2009) using the `tidycensus` (Walker, 2019) interface to the Census API for several key demographic and built environment variables: the share of individuals by ethnic group, the share of households by income level, household median income, and the housing unit density. An important attribute of the choice model is the distance from the home block group to the park boundary. Because we have no information on where in the block group a home is actually located, we use the population-weighted block group centroid published by the Census Bureau as the location for all homes in the block group. We then measured the Euclidean distance in miles between the block group and the boundary of each park in the polygons layer.

Model

In random utility choice theory, if an individual living in block group n wishes to make a park trip, the probability that the individual will choose park i from the set of all parks J can be described as a ratio of the park's measurable utility V_{ni} to the sum of the utilities for all parks in the set. In

TABLE 2 Descriptive Statistics for Residence Block Groups

	Median (IQR)
Density: Households per square kilometer	1,352.9 (880.8, 2,187.0)
Income: Median tract income	85,673.0 (58,478.0, 119,375.0)
Low Income: Share of households making less than \$35k	15.1 (7.8, 26.1)
High Income: Share of households making more than \$125k	30.4 (15.5, 46.5)
Black: Share of population who is black	6.8 (1.8, 18.4)
Asian: Share of population who is Asian	21.1 (10.4, 37.8)
Other: Share of population who belong to other minority groups	0.5 (0.0, 1.9)

the common destination choice framework we apply a multinomial logit model (McFadden, 1974; Recker and Kostyniuk, 1978),

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j \in J} \exp(V_{nj})} \tag{1}$$

where the measurable utility V_{ni} is a linear-in-parameters function of the destination attributes. If the park choice can be defined solely by its size and the impedance in reaching the park from block group n, then the measurable utility would be

$$V_{ni} = \beta_1 * size_i + \beta_2 * impedance_{ni} \tag{2}$$

where β_1, β_2 are estimable coefficients giving the relative utility (or disutility) of that attribute to the choice maker, all else equal. It is possible to add additional amenities of the park or the journey to the utility equation. However, as the number of alternatives is large, it is impractical to consider alternative-specific constants or coefficients and therefore not possible to include attributes of the home block group or traveler n directly. We can, however, segment the data and estimate different distance and size parameters for different segments to observe heterogeneity in the utility parameters between different socioeconomic groups.

The logarithm of the sum in the denominator of Equation 1 (called the logsum) provides a measure of the consumer surplus of the choice set (Williams, 1977),

$$CS_n = \ln \sum_{j \in J} \exp(V_{nj}) + C \tag{3}$$

where C is a constant indicating an unknown absolute value. But comparing the relative logsum values across choice makers, $CS_n - CS_{n-1}$ gives an indication of which choice maker has a more valuable choice set. Or, in this case of a park destination choice model, which choice maker has better access to parks.

In the most typical cases, researchers estimate the utility coefficients for destination choice models from household travel surveys. As we have no knowledge of an appropriate survey on park access, we need to synthesize a suitable estimation data set. We do this by sampling `n_obs` random discrete device origin-destination pairs from the commercial passive data matrix, weighted by the volume of the flows. This corresponds to a 4.3% sample of all the observed device origin-destination pairs.

The sampled origin-destination pair gives the home location as well as the "chosen" alternative for a synthetic person. In principle the individual's choice set contains all the parks in our dataset; in practice it can be difficult to estimate choice models with so many alternatives ($|J| = 500$). For this reason we randomly sample 10 additional parks to serve as the non-chosen alternatives for our synthetic choice maker. Such random sampling of alternatives reduces the efficiency of the estimated coefficients but the coefficients remain unbiased (Train, 2009). As the model has no alternative-specific constants, the standard likelihood comparison statistic against the market shares model ρ^2 is not computable. We instead use the likelihood comparison against the equal shares model ρ_0^2.

The resulting analysis dataset therefore contains 2×10^4 choice makers that select between 11 parks including the park they were observed to choose; the measured distance between the choice maker's block group and all parks in the choice set; and the acreage of each park in the choice set. We hold out a random sample of approximately 20% of choice makers for validation purposes. We use the `mlogit` package for R (Croissant, 2019; R Core Team, 2019) to estimate the multinomial logit models.

RESULTS

We estimated multinomial logit park destination choice models including coefficients for the distance between the park and the home block group and the acreage of the park. We applied a Yeo-Johnson transformation (Yeo and Johnson, 2000) to both distance and acreage; the Yeo-Johnson transformation replicates the constant marginal elasticity of a logarithmic transformation while avoiding undefined values ($YJ(0) = 0$). For efficiency and clarity, we call this transformation `log()` in the model results tables. Using a constant marginal elasticity is better reflective of how people perceive distances and sizes; a one-mile increase to a trip distance is more impactful to a one-mile trip than a ten-mile trip.

The results of several estimated models are given in Table 3. The first model — labeled "All" — uses no segmentation and includes the entire estimation sample. The coefficients have rational directionality: the negative coefficient on distance implies that individuals choose parks closer to their home locations, and the positive coefficient on acres implies that individuals choose larger parks, all else equal. The ratio of the distance to size coefficients is 0.21, implying that individuals in this model are willing to travel 4.74 times further to access a park twice as large, all else equal.

The next two models compare models estimated on segments of the estimation sample, where one segment includes block groups where the share of Black or African American and "Other" ethnic group (Native American and Pacific Islander, mostly) individuals exceeds 30% of the population. The "Non-Minority" segment includes the balance of block groups. The difference in the estimated coefficients is striking. For block groups with a high share of minority individuals, the distance coefficient is substantially larger and the acreage coefficient is substantially smaller than for the complementary segment. This suggests that minority individuals may be less likely to travel further, or be more willing to choose a smaller park, than other individuals.

The final set of three models segments the data based on the share of housheolds at different income levels in the estimation dataset. The "High Income" segment includes block groups where more than 50% of households report an income greater than \$125,000 per year; the "Low Income" segment includes block groups where more than 30% of households report an income less than \$35,000 per year, and the "Other Income" segment includes the remainder. In this case, each segment has a signficant and substantive difference in their perception of park acreage, with the high-income segment most interested in seeking out large parks. Curiously, the low-income segment also has a higher sensitivity than the remainder segment to larger parks. The distance coefficient shows a somewhat similar pattern, with the low- and high- income groups more similar to each other (and in this case statistically indistinguishable) than to the remainder group, though in this

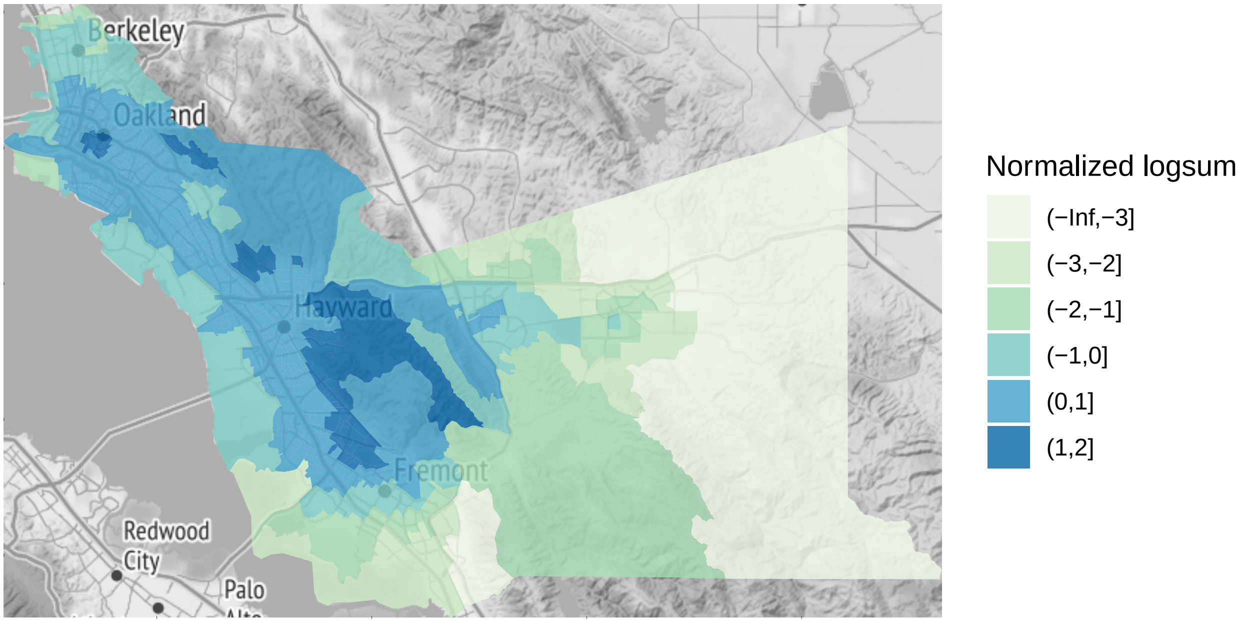

FIGURE 1 Normalized park accessibility logsum values.

case the remainder group is *more* sensitive to distance.

Figure 1 shows the normalized value of the destination choice logsum using the unsegmented model for each block group in Alameda county. According to this measure, the best access is found in the mountains east of Hayward, where there are large mountain parks nearby. The center of Oakland also sees good relative park access, with underdeveloped areas in the south and east of the county having the least access.

Validation

We applied each of the six models (or three models with six total market segments) to the validation holdout sample to examine the predictiveness of the estimated coefficients. The table below shows for each model and segment the total number of "correct" predictions, meaning that the alternative with the highest estimated probability was also the observed choice. The table also shows the mean estimated probability of the observed choice. The results indicate that a model predicting the park choice of residents in high-income neighborhoods is the most likely to result in correct predictions, and residents of low-income neighborhoods are the least predictable. The other model/segments all predict about half of the individuals in their segments correctly.

Figure 2 shows a comparison of the trip length frequency distribution for the observed and predicted choice of the individuals in the holdout sample. As in the numerical analysis of prediction accuracy, the high-income segment shows the best replication of the actual distance traveled; the models for all other segments — perhaps especially low-income and minority neighborhoods — seem to show a tendency to over-predict the utility of nearby parks at the expense of parks in the 3- to 7- mile range.

LIMITATIONS AND FUTURE DIRECTIONS

The ideal dataset for estimating individual choices would be a high-quality, large-sample household travel survey of real individuals. Such a survey would give details on whether an observed trip

TABLE 3 Park Destination Choice Estimates

	All	Non-minority N'hood	Minority N'hood	Low Income	Other Income	High Income
log(Distance)	−1.768*	−1.721*	−2.042*	−1.728*	−1.916*	−1.734*
	[−1.796; −1.741]	[−1.751; −1.691]	[−2.121; −1.963]	[−1.763; −1.693]	[−1.982; −1.851]	[−1.795; −1.672]
log(Acres)	0.373*	0.392*	0.263*	0.368*	0.319*	0.448*
	[0.362; 0.383]	[0.381; 0.403]	[0.235; 0.292]	[0.355; 0.381]	[0.295; 0.343]	[0.424; 0.472]
ρ_0^2	0.387	0.391	0.378	0.370	0.364	0.466
AIC	47124.017	39583.700	7364.297	28803.159	9972.092	8259.608
Log Likelihood	−23560.009	−19789.850	−3680.148	−14399.580	−4984.046	−4127.804
Num. obs.	16029	13549	2469	9536	3270	3223

* 0 outside the confidence interval.; 95% confidence interval in brackets

TABLE 4 Model Validation

Model	Segment	%Correct	Mean Prob. of Chosen
All	All	50.77	0.36
Minority	Non-minority N'hood	51.56	0.37
Minority	Minority N'hood	48.53	0.33
Income	Low Income	43.89	0.31
Income	Other Income	50.25	0.36
Income	High Income	57.12	0.44

to a park was actually a recreation trip or rather a different activity entirely. The individual-level demographic data would also be valuable in understanding more clearly the observed heterogeneity in response among different income or ethnic groups. Additionally, the trends and correlations revealed in the presented models may reflect situational inequalities rather than true preferences. For example, the strongly distinct observed parameters on size and distance for minority block groups may indicate that areas with large minority populations tend to have smaller parks that are more geographically distributed relative to other regions of the city. Transit access may also affect park choice and how far people are willing to travel to access a park. Preliminary analysis of our source data indicates a qualitative correlation between good transit access and diverse park use from both a geographic and demographic perspective.

We limited our analysis to home locations and parks in Alameda County, California. It is possible that some Alameda residents visit parks in neighboring counties, just as it is possible that parks in Alameda County attract trips from outside the county borders. This is most likely for block groups and parks on the north and south borders of the county. The lower measured accessibility in the area around Berkeley in the northern part of the county (see Figure 1 is likely affected by the ommission of parks and residents in Contra Costa County.

Using Euclidean distance to represent the distance between the block group centroid and the border of the park leaves something to be desired: Depending on network topography and built environment characteristics, there may be a significant variation in perceived travel times between two parks with similar straight-line distances. That said, a more precise network-based measure might not overcome the inaccuracies resulting from our necessarily measuring distances from the block group centroid. As above, an individual-level survey where the home location is explicitly known would be preferable regardless of the distance method employed.

The activity location data used in this specific analysis treats all days of the week and day periods together; it is likely that weekend park choice is substantially different from weekday choice, as the activities performed may be the same. Also recall that the data consider each device-park pair as a unique trip. Repeated trips to the same park may not be properly considered in the data sample. A more precise time-of-day and day-of-week segmentation is warranted.

We applied a naive random sampling of the alternatives in our model estimation and validation; a more considered approach involving hierarchical destination sampling would yield more efficient estimates and therefore a clearer picture of the role of size and distance on the observed choice. The relatively weak predictive power of such a simple model formulation (size and distance only) indicates that there is potential to examine the role that additional park amenities — ball fields, playgrounds, water features, etc. — play in the relative attractiveness of parks for

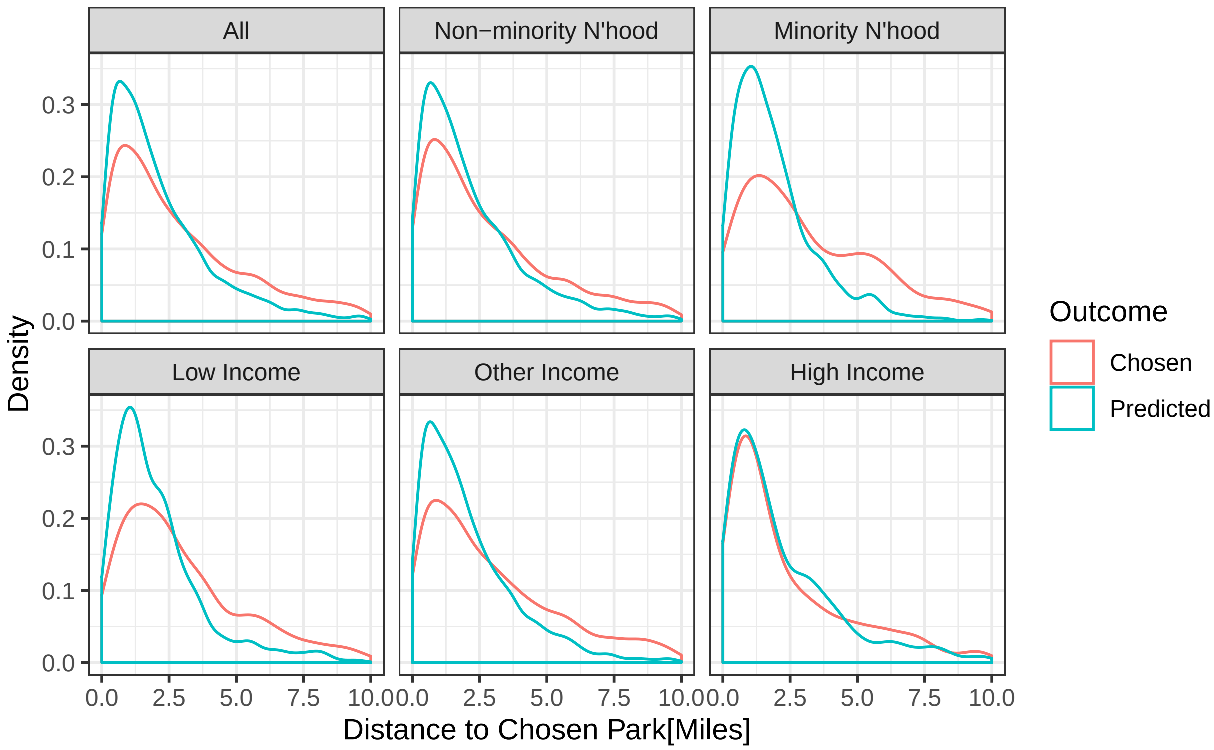

FIGURE 2 Validation trip length frequency distributionns

different market segments. The quality of park maintenance is another important feature identified in the recreation literature (Fletcher and Fletcher, 2003) that is not included here. That said, forecasting the presence or quality of some of these amenities may be infeasible when considering future travel patterns.

CONCLUSIONS

As transportation professionals seek to improve access to parks and better coordinate transportation and land use efforts, it is increasingly important to better understand how, when, and why individuals travel to parks. This intersection between recreation and transportation has received relatively little exploration, partially because travel survey data emphasizes weekday travel and because the role of parks in daily activities can be more complicated than with other land uses. This study contributes to the understanding of recreation access by presenting a method to develop access measures explicitly based on the observed choices of individuals. The resulting access measure is continuously defined and incorporates multiple dimensions of access, including the travel necessary to reach all nearby parks as well as the amenities of each of those parks. Further, the measure we have presented reveals heterogeneous preferences for travel and park size across market segments, a heterogeneity that could perhaps be incorporated into an understanding of accessibility.

With the growing availability of passive transportation data, there is a correspondingly increased opportunity to explore such data to develop a better understanding of travel patterns in more careful detail than is possible with household travel surveys. Capturing a sufficiently large survey to study trip patterns to a single park is an enormous undertaking, and doing such an exercise for an entire park system is prohibitively expensive and time-consuming. Passive data sets therefore enable analyses that would be unlikely or impossible by other means. Challenges to the representativeness and comprehensiveness of passive data products are in many cases fair, but this should not preclude their use in cases where traditional techniques are not practicable.

REFERENCES

Bancroft, C., Joshi, S., Rundle, A., Hutson, M., Chong, C., Weiss, C. C., Genkinger, J., Neckerman, K. and Lovasi, G. (2015), '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.

Barnard, P. O. and Brindle, R. E. (1987), '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.

Bernardin, V. L., Bowman, J. L., Bradley, M., Chen, J., Ferdous, N. and Lee, Y. (2018), Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study, *in* 'TRB Annual Meeting', Transportation Research Board.
URL: *https://trid.trb.org/view/1495930*

Bolla, R. and Davoli, F. (2000), 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)', IEEE, pp. 1107–1112.
URL: *http://ieeexplore.ieee.org/document/904783/*

Calabrese, F., Di Lorenzo, G., Liu, L. and Ratti, C. (2011), 'Estimating Origin-Destination Flows Using Mobile Phone Location Data', *IEEE Pervasive Computing* **10**(4), 36–44.

City Parks Alliance (2019), 'Why City Parks Matter'.
URL: *https://cityparksalliance.org/about-us/why-city-parks-matter/*

Croissant, Y. (2019), *mlogit: Multinomial Logit Models*.
URL: *https://cran.r-project.org/package=mlogit*

de Jong, G., Daly, A., Pieters, M. and van der Hoorn, T. (2007), 'The logsum as an evaluation measure: Review of the literature and new results', *Transportation Research Part A: Policy and Practice* **41**(9), 874–889.

Fletcher, D. and Fletcher, H. (2003), 'Manageable Predictors of Park Visitor Satisfaction: Maintenance and Personnel', *Journal of Park and Recreation Administration* **21**(1), 21–37.

Friedrich, M., Immisch, K., Jehlicka, P., Otterstätter, T. and Schlaich, J. (2010), 'Generating Origin-Destination Matrices from Mobile Phone Trajectories', *Transportation Research Record: Journal of the Transportation Research Board* **2196**, 93–101.

GreenInfo Network (2019), 'California Protected Areas Database'.
URL: *https://www.calands.org/cpad/*

Huang, A. and Levinson, D. (2015), 'Axis of travel: Modeling non-work destination choice with GPS data', *Transportation Research Part C: Emerging Technologies* **58**, 208–223.

Huntsinger, L. F. and Donnelly, R. (2014), Reconciliation of Regional Travel Model and Passive Device Tracking Data, *in* 'TRB Annual Meeting', Transportation Research Board.
URL: *https://trid.trb.org/view/1287620*

Huntsinger, L. F. and Ward, K. (2015), 'Using Mobile Phone Location Data to Develop External Trip Models', *Transportation Research Record: Journal of the Transportation Research Board* **2499**(1), 25–32.

Institute of Transportation Engineers (2017), *Trip Generation Manual*, 10th editi edn, Institute of Transportation Engineers, Washington, D.C.

Kaczynski, A. T., Schipperijn, J., Hipp, J. A., Besenyi, G. M., Wilhelm Stanis, S. A., Hughey, S. M. and Wilcox, S. (2016), 'ParkIndex: Development of a standardized metric of park access for research and planning', *Preventive Medicine* **87**, 110–114.

Kinnell, J. C., Bingham, M. F., Mohamed, A. F., Desvousges, W. H., Kiler, T. B., Hastings, E. K. and Kuhns, K. T. (2006), 'Estimating Site Choice Decisions for Urban Recreators', *Land Economics* **82**(2), 257–272.

Komanduri, A., Schewel, L., Beagan, D. and Wong, D. (2017), 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.
URL: *https://trid.trb.org/View/1438797*

Kressner, J. D. (2017), Synthetic Household Travel Data Using Consumer and Mobile Phone Data, Technical report, Transportation Research Board.

Madzia, J., Ryan, P., Yolton, K., Percy, Z., Newman, N., LeMasters, G. and Brokamp, C. (2018), 'Residential Greenspace Association with Childhood Behavioral Outcomes', *The Journal of Pediatrics* .

McCormack, G. R., Rock, M., Toohey, A. M. and Hignell, D. (2010), 'Characteristics of urban parks associated with park use and physical activity: A review of qualitative research', *Health & Place* **16**(4), 712–726.

McFadden, D. L. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–142.

Meyerhoff, J., Dehnhardt, A. and Hartje, V. (2010), '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.

Millard-Ball, A. (2015), 'Phantom trips: Overestimating the traffic impacts of new development', *Journal of Transport and Land Use* **8**(1), 31–49.

Mitchell, R. and Popham, F. (2008), 'Effect of exposure to natural environment on health inequalities: an observational population study', *The Lancet* **372**(9650), 1655–1660.

Naboulsi, D., Fiore, M., Ribot, S. and Stanica, R. (2016), 'Large-Scale Mobile Traffic Analysis: A Survey', *IEEE Communications Surveys & Tutorials* **18**(1), 124–161.

National Academies of Sciences Engineering and Medicine (2012), *Travel Demand Forecasting: Parameters and Techniques*, NCHRP 716, National Academies Press, Washington, D.C.

Pan, C., Lu, J., Di, S. and Ran, B. (2006), 'Cellular-Based Data-Extracting Method for Trip Distribution', *Transportation Research Record: Journal of the Transportation Research Board* **1945**(1), 33–39.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.r-project.org/*

Recker, W. W. and Kostyniuk, L. P. (1978), 'Factors influencing destination choice for the urban grocery shopping trip', *Transportation* **7**(1), 19–33.

Richardson, E. A., Mitchell, R., Hartig, T., de Vries, S., Astell-Burt, T. and Frumkin, H. (2012), 'Green cities and health: a question of scale?', *Journal of Epidemiology and Community Health* **66**(2), 160 LP – 165.

Rigolon, A., Browning, M. and Jennings, V. (2018), '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.

Roll, J. (2019), Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon, Technical report, Oregon Department of Transportation.
URL: *https://trid.trb.org/View/1630252*

Stark, J. H., Neckerman, K., Lovasi, G. S., Quinn, J., Weiss, C. C., Bader, M. D., Konty, K., Harris, T. G. and Rundle, A. (2014), 'The impact of neighborhood park access and quality on body mass index among adults in New York City', *Preventive Medicine* **64**, 63–68.

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

Trust for Public Land (2019), '2019 ParkScore® Index | The Trust for Public Land'.
URL: *https://www.tpl.org/parkscore*

U.S. Census Bureau (2009), A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know, Technical Report May, U.S. Goverenment Printing Office, Washington, D.C. :.

U.S. Census Bureau (2019), 'QuickFacts: Alameda County, California'.
URL: *https://www.census.gov/quickfacts/fact/table/alamedacountycalifornia/PST045218*

Ussery, E. N., Yngve, L., Merriam, D., Whitfield, G., Foster, S., Wendel, A. and Boehmer, T. (2016), '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.

Walker, K. (2019), *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*.
URL: *https://cran.r-project.org/package=tidycensus*

Williams, H. C. W. L. (1977), 'On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit', *Environment and Planning A: Economy and Space* **9**(3), 285–344.

Yeo, I.-K. and Johnson, R. A. (2000), 'A New Family of Power Transformations to Improve Normality or Symmetry', *Biometrika* **87**(4), 954–959.

Zhu, J. and Ye, X. (2018), 'Development of destination choice model with pairwise district-level constants using taxi GPS data', *Transportation Research Part C: Emerging Technologies* **93**, 410–424.