A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco

A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco

Andres Sevtsuk, Rounaq Basu, Xiaojiang Li, Raul Kalvo

Abstract

Big data from smartphone applications are enabling travel behavior studies at an unprecedented scale. In this paper, we examine pedestrian route choice preferences in San Francisco, California using a large, anonymized dataset of walking trajectories collected from an activity-based smartphone application. We study the impact of various street attributes known to affect pedestrian route choice from prior literature. Unlike most studies, where data has been constrained to a particular destination type (e.g. walking to transit stations) or limited in volume, a large number of actual trajectories presented here include a wide diversity of destinations and geographies, allowing us to describing typical pedestrians' preferences in San Francisco as a whole. Other innovations presented in the paper include using a novel technique for generating alternative paths for route choice estimation and gathering previously hard-to-get route attribute information by computationally processing a large set of Google Street View images. We also demonstrate how the estimated coefficients can be operationalized for policy and planning to describe pedestrian accessibility to BART stations in San Francisco using 'perceived distance' as opposed to traversed distance.

Cite this article:

Sevtsuk, A., Basu, R., Li, X., and Kalvo, R. (2021). A big data approach to understanding pedestrian route choice preferences - Evidence from San Francisco. *Travel Behavior and Society.* Online First (June 15, 2021). https://doi.org/10.1016/j.tbs.2021.05.010

Keywords

Pedestrian route choice; Travel behavior; GPS trajectories; Google Street View; Pedestrian accessibility; Path size logit.

1. Introduction

Municipal governments around the world are increasingly promoting walkability and transit-oriented development, which have been shown to be associated with a range of health, economic, social and environmental benefits (Leyden 2003; Speck 2011; Luberoff 2019). From a global competitive perspective, virtually all cities ranked among the most livable in popular indices produced by the Economist[1], Mercer[2] or Monocle[3] offer high-quality public transit systems and boast of high levels of pedestrian activity on their streets. Understanding pedestrian behavior has thus gained renewed interest among urban planners and transportation researchers in recent times. Among numerous areas of research, pedestrian route choice—quantitative analysis of whether and how specific route characteristics affect pedestrians' route preferences—is particularly relevant, since the choice of a particular path over multiple competing alternatives can be seen as direct evidence of how the built environment affects travel behavior, thereby offering clues as to how design, policy, and planning can be leveraged to achieve more sustainable and healthy urban mobility outcomes.

In the past, researchers have studied pedestrian route choice using both stated preference and revealed preference surveys. Stated preference surveys offer hypothetical alternatives, which are constructed using efficient and representative design methods to cover the parameter combination space, and ask respondents to either choose among or rank different routes (Senevirante & Morral 1985; Agrawal et al. 2008; Erath et al. 2015). On the other hand, revealed preference surveys observe behavior directly, by either following pedestrians (Hill 1984; Kim

[1] https://www.eiu.com/topic/liveability
[2] https://mobilityexchange.mercer.com/Insights/quality-of-living-rankings
[3] https://monocle.com/magazine/issues/95/top-25-liveable-cities/

2015), observing their trajectories from a distance (Garbrecht 1971; Whyte 1980), asking people to retrace the routes they walked from memory (Muraleetharan & Hagiwara 2007). In recent years, there has also been an increased use of tracking technologies to record the geographic coordinates of walking trajectories passively, using GPS data (Broach and Dill 2015; Vanky 2017). While the labor-intensive nature of in-person surveys have limited such approaches to relatively small sample sizes, passive tracking technology is facilitating data collection with larger sample sizes and wider geographic coverage (Shoval and Isaacson 2006). This can potentially enable researchers to model pedestrian route choice preferences at the scale of an entire city with relatively low investment (of both cost and effort) in data collection.

In this paper, we analyze pedestrian route choice preferences in San Francisco, California using a large, anonymized dataset of global positioning service (GPS) walking trajectories. Using data collected from an activity-based smartphone application, we analyze how pedestrian route choice is affected by a variety of route characteristics suggested in prior literature. Some of the key innovations presented in the paper include (a) gathering and mapping unique route attribute information by computationally processing a large set of Google Street View images; (b) using a novel technique for generating alternative paths for route choice estimation; and (c) estimating how specific route characteristics affect pedestrian route choice using a large number of actual trajectories (i.e., revealed preference) that do not focus on a particular destination type (e.g. walking to transit), but include a wide diversity of destinations and geographies, thereby describing typical preferences of pedestrians in San Francisco as a whole. We also show how the findings can be operationalized for policy and planning by describing pedestrian accessibility to different destinations using 'perceived distance' as opposed to traversed distance.

2. Literature Review

Several studies have demonstrated that pedestrians do not necessarily choose the shortest route, when alternative route options are available. Although distance and travel time are indeed critical factors determining route choice, their relative importance depends on a variety of other route characteristics. Pedestrians are often willing to deviate to safer, more comfortable, or more interesting routes, so long as the detours compared to shortest paths remain within a reasonable range.

The set of route characteristics that can affect pedestrian path choice can be quite diverse. Guo and Loo (2013) found that the presence of shopfronts, open space, and wider sidewalks significantly increased the likelihood of path choice in New York City and Hong Kong. The importance of local context and culture was highlighted through their finding of the estimated coefficients being different for the same street attributes between the two cities. For instance, pedestrians in Hong Kong showed significantly stronger dislike for longer routes, while New Yorkers showed a stronger preference for retail frontages along routes. Muraleetharan and Hagiwara (2013) analyzed route choice in Sapporo, Japan and assessed how the level of service (LOS)—a metric that characterizes a bundle of different sidewalk qualities, including sidewalk width, presence of obstructions, the density of other pedestrians, and the presence and density of bicyclists on a sidewalk—affected pedestrian path choice. They found that that people were consistently willing to choose longer paths when the overall LOS was higher. At the same time, the relative impact of LOS was notably higher on short routes compared to long routes, since detours along shorter routes produce a lower overall penalty on travel time or distance. Erath et al. (2015) studied pedestrian path choice in Singapore using stated preference surveys and found that pedestrians were most attracted to routes that offered greenery, retail frontages, high-quality pavements, and physical shelter from rain and sun. Contrastingly, loudness, obstructions, road crossings, pedestrian tunnels and bridges were negatively associated with route choice. Another study conducted in Singapore by Olszewski and Wibowo (2005) investigated pedestrian path choice to transit stations. The authors found that besides walking distance, route choice to transit stations was significantly affected by the number of road crossings, traffic conflicts, and the number of ascending steps, typically brought about by elevated road crossings that are common in Singapore. The authors used the concept of "equivalent walking distance" to describe how specific route attributes changed the users' perception of route length. Crossing a single road, for instance, was found to be equivalent to extending the walk by 55m, while the presence of an elevated road crossing (with 32 ascending steps) was perceived as equivalent to extending the walk by 90 meters. Drawing from this study, we too use the concept of "equivalent walking distance" to interpret our findings in subsequent sections. Similar to Erath et al. (2015), we also demonstrate how perceived walking distance can be operationalized for policy and planning to measure subjective pedestrian accessibility.

Ewing and Handy (2009) conducted a survey of an expert panel—composed of both academics and urban design practitioners—asking participants to rate the importance of various urban design qualities related to walkability. Even though they did not estimate path choice directly or probe the specific relevance of urban design factors to actual route choice, the study revealed a rather comprehensive set of qualities that experts in the field tend to associate with walkability. The findings were categorized into five groups—imageability, enclosure, human scale, transparency, and complexity—wherein each group contained a number of indicators. The human scale factors, for instance, captured the dimensions of sightlines, the number of street furniture elements, proportion of first floors with windows, building heights, and the presence of planter boxes on streets. The study demonstrated that the list of urban design qualities hypothesized to affect walkability is potentially large and nuanced.

In addition to conventional data collection methods, there has been a notable increase in using GPS data for pedestrian route choice studies in recent years (Wolf et al. 2001; Shoval and Isaacson 2006; Cho et al. 2011; Broach and Dill 2015; Vanky 2017; Lue 2019). This has enabled researchers to collect data with larger sample sizes and expand the geographic extent of study areas. Broach and Dill (2015), for instance, studied pedestrian route choice in Portland, Oregon using GPS devices carried by 283 adult participants. They found that route choice was negatively and significantly related to distance, number of turns, elevation gain, sub-standard sidewalk quality, traffic volumes, unsignalized arterial crossings, as well as unmarked collector road crossings. On the other hand, path choice was positively related to the presence of shopfronts along the route (Broach and Fill 2015). Vanky (2017) and Malleson et al. (2018) studied pedestrians in Boston using GPS traces from a smartphone application, outlining broad characteristics of walking activity, but did not estimate a choice model to statistically relate route choice behavior with its determinant factors. GPS data has also been used to study the walking habits of particular groups of pedestrians, such as tourists (Asuakura & Iryo 2007), men versus women (Lue 2019), children and adolescents (Duncan & Mummery 2007; Wiehe 2008), and retail patrons (Moiseeva and Timmermans 2010). The data presented in this paper also uses GPS traces from pedestrians, but with a sample size that is an order of magnitude larger than previous studies.

3. Data

Our analysis is based on a large dataset of pedestrian GPS traces in San Francisco captured continuously from May 2014 to May 2015 using a popular activity-oriented smartphone application, thereby covering a full calendar year. Despite the advantage of providing an unprecedentedly large sample, the anonymous nature of the data does have limitations. The anonymized data contain no identifiable user attributes or information about trip origin type, destination type, or trip purpose. Therefore, we are unable to infer which journeys are first- or last-mile legs of multi-modal trips, or which journeys involve the same individual(s)' traces at different times.

Since raw GPS traces are usually not aligned to streets due to noise in GPS signal quality created by street canyons, building obstructions, or trees, the actual geolocated dot-data from GPS were map-matched to street centerlines, using 2014 TIGER/Line (Topologically Integrated Geographic Encoding and Referencing) geometries from the US Census.[4] Several methods have been proposed for map-matching GPS traces to street geometries, varying in computational complexity (Raymond et al 2012; Wei et al 2013; Chao et al 2020). We applied the widely-used Hidden Markov Map-Matching (HMM) algorithm by Newson & Krumm (2009), which has been successfully deployed on pedestrian GPS data in the past (see e.g. Malleson et al., 2018). The HMM algorithm finds the most likely connected route for each GPS trace by measuring the proximity of each dot along a GPS trace to surrounding street segments and assigning transition probabilities between street segments based on the connectivity of the street network. This process also helped further anonymize the sample, since the side of street a person walked on

[4] https://catalog.data.gov/dataset/tiger-line-shapefile-2014-county-san-francisco-county-ca-all-roads-county-based-shapefile

and the exact trip origin and destination are generalized to include an entire street segment for the start and end of each trip, rather than a precise address point.

Each observation in the data includes a spatial GPS trace, timestamps showing when the trip started and ended, and the movement mode, which is either walk or run (we use only walk trips in this study). Out of a significantly larger raw dataset of about 280,000 traces, 14,760 individual traces were included for estimating a discrete route choice model (Figure 1). First, we eliminated all identical trajectories from the raw data, keeping only unique routes. Since the data were fully anonymized, this helped eliminate the possible bias that many of the duplicate trajectories could be routinely walked by the same individual(s) over the year, without making a conscious route choice[5]. Even with this filter, trips by the same individual(s) over time are likely to persist in the data. Unfortunately, the data does not allow us to include a panel effect correction, as we are unable to detect repeated observations for the same individual(s). Furthermore, keeping only unique trajectories in the data also means that all routes are given equal importance in our analysis, ignoring the fact that a few specific routes may be far more popular than others. However, such spatial heterogeneity in popularity is likely to be correlated with trip purpose (e.g. certain routes may enjoy heavy use by tourists). Since we do not have information on trip purpose, using only unique trajectories is possibly the best tradeoff to obtain ‘general’ route choice preferences of pedestrians in San Francisco.

[5] However, this process could unfortunately also eliminate genuinely unique observations that were chosen by different individuals.

Second, we excluded routes that only featured one or no turns[6], since initial data analysis revealed that many of these included routes that only switched from one side of a divided avenue to another or included simple straight-line journeys. We also excluded routes with more than seven turns, which exploratory data analysis indicated as circuitous trajectories that did not take any obvious path from an origin to a destination. Third, we excluded routes that were shorter than 200 meters, since these typically limited path choice, as well as routes longer than 1,000 meters due to the possibility that longer paths could erroneously indicate movement on other modes, such as bus or streetcar. Fourth, we eliminated routes that deviated more than 50% longer than the shortest paths. The latter were deemed more likely to include additional stops along the way or motivations other than walking directly between a given origin and destination. Last, we included only trajectories that were located in parts of San Francisco that had GIS data with complete street attributes. This contains the vast majority of the city, but excludes Treasure Island and Alcatraz Island, which technically fall under the city's jurisdiction. Within these constraints, the sample still includes a large variety of pedestrian trajectories in all parts of the city (N = 14,760 traces). It is worth noting that routes can also include round trips, where the origin and destination are near each other, so long as the walk produced a continuous GPS trace with no breaks.

[6] A turn is determined as at least a 45-degree change in direction along TIGER street centerlines. We experimented with different turn angles including 20deg, 30deg but found the 45deg angle to yield most reliable results based on visual checking. Turns were counted using a GIS script.

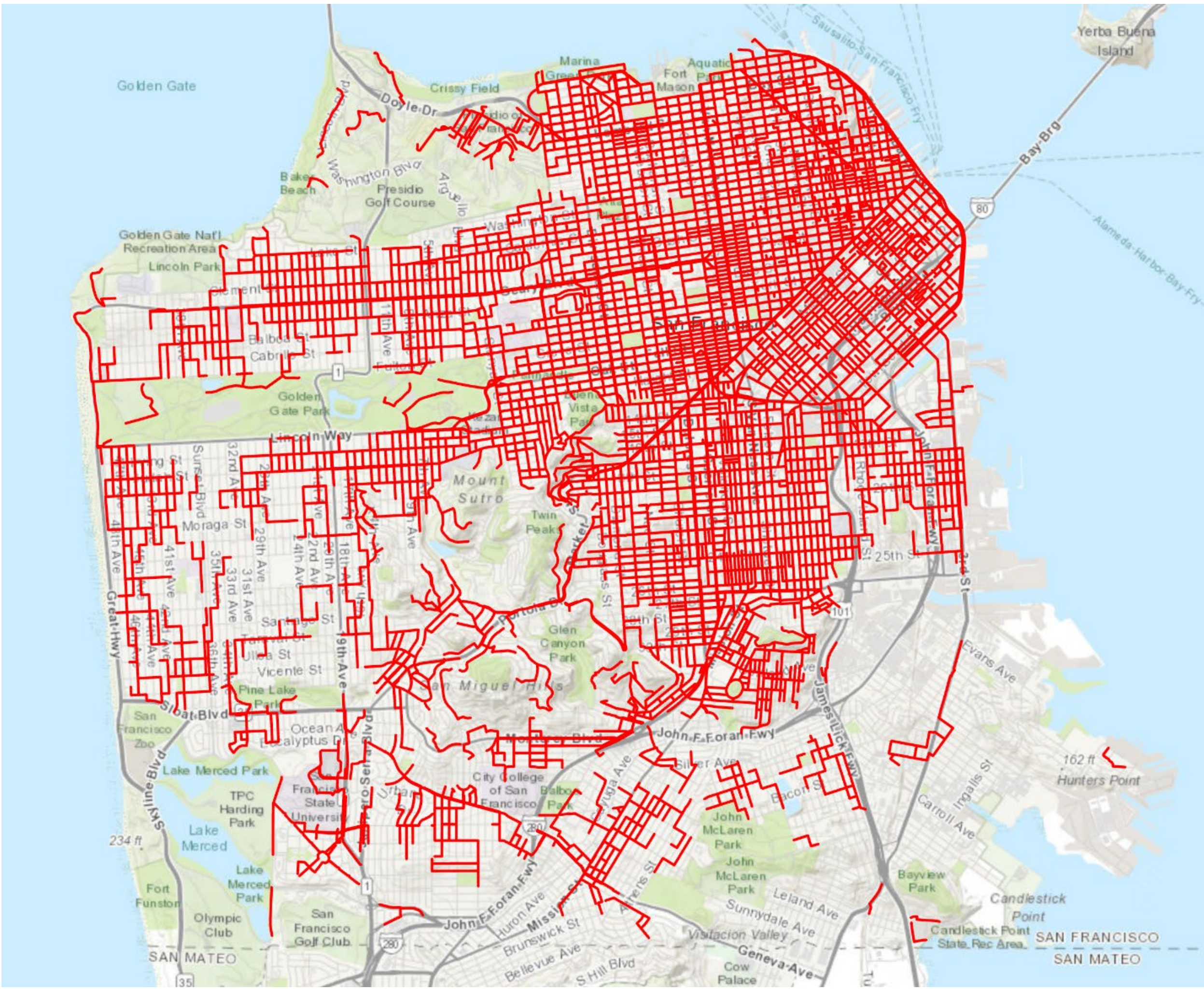

Figure 1. Map-matched GPS traces in San Francisco (N = 14,760).

3.1 Alternative path generation

Path choice estimation requires that in addition to the route actually walked, "plausible" alternative paths be determined as options that were considered but not chosen. The presence of alternatives allows a choice model to evaluate whether and which particular route attributes are systemically related to higher likelihoods of preference. The set of alternatives plays a key role in determining estimated coefficients (Prato 2007; Rieser-Schüsslera et al. 2015). Theoretically, hundreds, or even thousands, of route alternatives are potentially available to decision-makers, but this does not imply that a full sampling approach is necessarily appropriate. Many of the alternatives can overlap with the actual route or with each other, thereby not constituting independent choices, which would violate the independence of irrelevant alternatives (IIA) property in decision theory. With recent increase in computational power, alternative generation has moved from sampling a subset of paths to including virtually all, or close to all, potential route options (Frejinger et al. 2009; Fosgerau et al. 2013; Hassan et al. 2019). However, in practice, a relatively small number of alternatives (e.g. 1-5) is typically used in pedestrian route choice models, whereby alternatives are chosen so that they offer variation in estimated route attributes (Bovy 2009).

A number of different approaches have been developed for alternative route set generation (Prato 2009). The k^{th} shortest path method (Dijkstra, 1959; Gallo and Pallottino, 1988) finds a given number (k) paths that are longer than the shortest path. The link elimination method removes route segments one by one and finds the new shortest path after each removal (Azevedo et al. 1993; Park and Rilett 1997). The branch and bound approach finds all plausible paths up to a certain threshold detour beyond the shortest path (Friedrich et al., 2001) and the constrained enumeration approach finds all routes between origins and destinations that satisfy a set of different constraints—a maximum detour limit, no repetition in segments, directional constraints, overlap constraints, etc. (Prato and Bekhor 2006). We chose to implement the constrained enumeration approach in the following manner. Given the relatively short distance of walking trips in the data (GPS traces were limited to 1km in length), we first found all possible unique routes (forbidding segment repetition) between the same trip origins and destinations that were up to 50% longer than the shortest path—the same maximum detour as in the actual route set.[7] This complete set typically also includes the observed paths, except in cases where the latter contained link repetition (back tracking) or node repetition (looping). Though an exhaustive search can return a very large number of paths between each origin-destination pair (we obtained tens of thousands of alternatives for some O-D pairs), we only drew a random set (k = 6) alternatives from the full set for each pair that satisfied the bounding constraints[8]. The advantage of this approach is that the random draw is performed from a large set of reasonable paths, thereby minimizing the probability of overlap with the actual path as well as other alternatives, while augmenting the diversity of route features in the set.

Once a set of alternative paths were drawn, we checked for overlap with actual routes. This was done by comparing the sequence of nodes in both the original path and each of the alternative paths using the Jaccard Similarity measure[9]. Only alternatives that had less than 25% overlap with the observed paths were included in the final choice set. From those alternatives that remained viable, we assigned up to three alternatives to each actual route. All actual routes obtained at least one alternative path, 55% had two alternative paths, and 27% had three alternative paths. Our final choice set thus contains 14,760 actual routes and 30,949 unchosen, but plausible, alternatives.

[7] Empirical testing with pedestrian route directions revealed that a similar approach has been implemented by Google Maps, where pedestrian route directions are subject to a maximum detour of 50% compared to the shortest path.
[8] Shorter routes can have fewer than six alternatives available due to fundamental geometric constraints in the street network.
[9] Jaccard Similarity is used here to measure similarity in the list of nodes included along the actual path and each of its potential alternatives.

3.2 Street qualities affecting pedestrian path choice

Based on the abovementioned studies (Olszewski & Wibowo 2005; Guo & Loo 2013; Muraleetharan & Hagiwara 2013; Erath et al. 2015; Broach and Dill 2015; Ewing and Handy 2009), we initially identified eleven attributes of streets that we hypothesized to affect pedestrian path choice in San Francisco. Table 1 lists the variables that were included, along with their expected effect on choice probability (denoted through a positive/negative sign), and descriptive statistics. A number of variables were measured as weighted averages along the route, using segment lengths as weights. These include sidewalk width, sky view factor, green view index, speed limit, and traffic volume. If a route included two street segments, for instance, which measured 100m and 200m each, and their sidewalk widths were 6ft and 12ft correspondingly, then the weighted average sidewalk width along the whole walk was measured as (6ft*100m+12ft*200m)/300m= 10ft.

Variable	Description	Exp. Sign	Measurement method	Chosen route (R0)			Alternatives		
				Mean	Min	Max	Mean (R1)	Mean (R2)	Mean (R3)
Length	Route length in meters	-	GIS geometry calculation.	690.2	202.6	1,000	871.2	881.6	899.4
Turns	Number of turns along the route.	-	GIS script, where a turn corresponds to a 45-degree or more change in direction.	3.1	2	7	5.9	6	6.1
Uphill	Uphill ascent along the path in meters.	-	GIS script, which summarizes the vertical gain in meters.	7.8	0	122.8	11.5	11.5	11.7
Highway	% of route within a 75m buffer of a highway.	-	GIS buffer and geometry calculation.	0.5	0	67.3	0.5	0.4	0.5
Amenities	Number of ground floor amenities along the route.	+	GIS count, using NAICS codes for retail (44-45), eating and drinking (722), personal services (811, 812) and entertainment venues (491, 7111, 712, 713, 12131). ESRI business analyst data from 2014.	22.3	0	227	24.2	24	24.9
Public Art	Number of public art pieces along the route.	+	GIS count, using city of SF records.	0.2	0	4	0.3	0.3	0.3
Sidewalk Width	Average width of sidewalks along the route in feet, weighted by segment length.	+	GIS calculation, using city of SF records.	12	0	26.6	11.6	11.5	11.6
Sky View Factor	Average % of sky visibility in Google Street View images along route, weighted by segment length.	-	% of sky view pixels measured by computer vision analysis using Google Street View API.	0.6	0	1	0.6	0.6	0.6
Green View Index	Average % of vegetation in Google Street View images along route, weighted by segment length.	+	% of green pixels measured by computer vision analysis using Google Street View API.	0.1	0	0.6	0.1	0.1	0.1
Speed Limit	Average speed limit along the route in miles per hour, weighted by segment length.	-	GIS calculation, using city of SF records.	25.6	0	45	25.4	25.4	25.4
Traffic Volume	Average traffic volume along the route, in cars per 24 hours, weighted by segment length.	-	GIS calculation, using city of SF records.	3,978	0	54,515	3,964	3,982	3,958
Path Size	Overlap indicator for alternative paths for the same trip (Ben Akiva & Bierlier 1999).	+	Segment-length weighted overlap (Ben-Akiva & Bierlaire 1999).	0.8	0.3	1	0.7	0.7	0.7

Table 1. Descriptive statistics of attributes on observed trajectories (N = 14,760). Bounds (min and max values) and averages are reported for the chosen route, while only average values are reported for the three alternative routes to facilitate comparisons.

Most of the variables were operationalized using publicly available GIS data for San Francisco. However, a few variables were quantified using unique and innovative data science approaches, which to our knowledge, have not been used for route choice studies in the past. The Green View Index (GVI) and the Sky View Factor (SVF) capture the extent of visible green space and sky view along the walk (Figure 2). GVI describes the extent of vegetation that is visible to a pedestrian from approximately eye-height on each street (Li et al., 2015a, b). We use SVF as a proxy variable to capture the sense of "enclosure" and shading along a route, which some previous studies have also included or hypothesized (Ewing and Handy 2009; Erath et al. 2015; Li et al., 2018b). Both variables range from 0 to 1, where the maximum value of one indicates totally green or totally enclosed street canyons respectively (Li et al., 2018a). Both metrics were captured by processing a large number of images from Google Street View API on all streets in San Francisco using a deep convolutional algorithm (Li et al. 2015). Google Street View panoramas are captured at 10m intervals on streets that Google's vehicles can publicly access. Each panoramic image contains a continuous "dome" of pictures that cover a 360-degree viewshed horizontally and a 180-degree viewshed vertically. Figure 2 shows the processing of original Street View images to "greenery" and "sky view" categories using a state-of-the-art image segmentation algorithm called PSPNet (Zhao et al. 2017).

Figure 2. The segmentation of street-level images using a deep learning algorithm, demonstrating (a) the raw Google Street View panoramas, (b) the blend of the segmentation results on Google Street View panoramas, and (c) the segmented hemispherical images.

The resulting SVF and GVI indices describe the percent of all pixels in the Street View dome categorized as either "sky" or "green". To produce relevant metrics for the entire route, GVI and SVF were summarized as a weighted average (using segment lengths as weights) along the whole route.

Figure 3 illustrates the mean GVI for street segments across San Francisco. As expected, large parks such as Golden Gate Park, the Presidio and Lake Merced Park have high green view outcomes (GVI > 40%). While streets in typical residential areas and parts of downtown have relatively little vegetation (GVI = 0-5%), some neighborhoods including Noe Valley, the Mission and Castro have notably greener streets (GVI = 5-40%), contributing to their residential appeal.

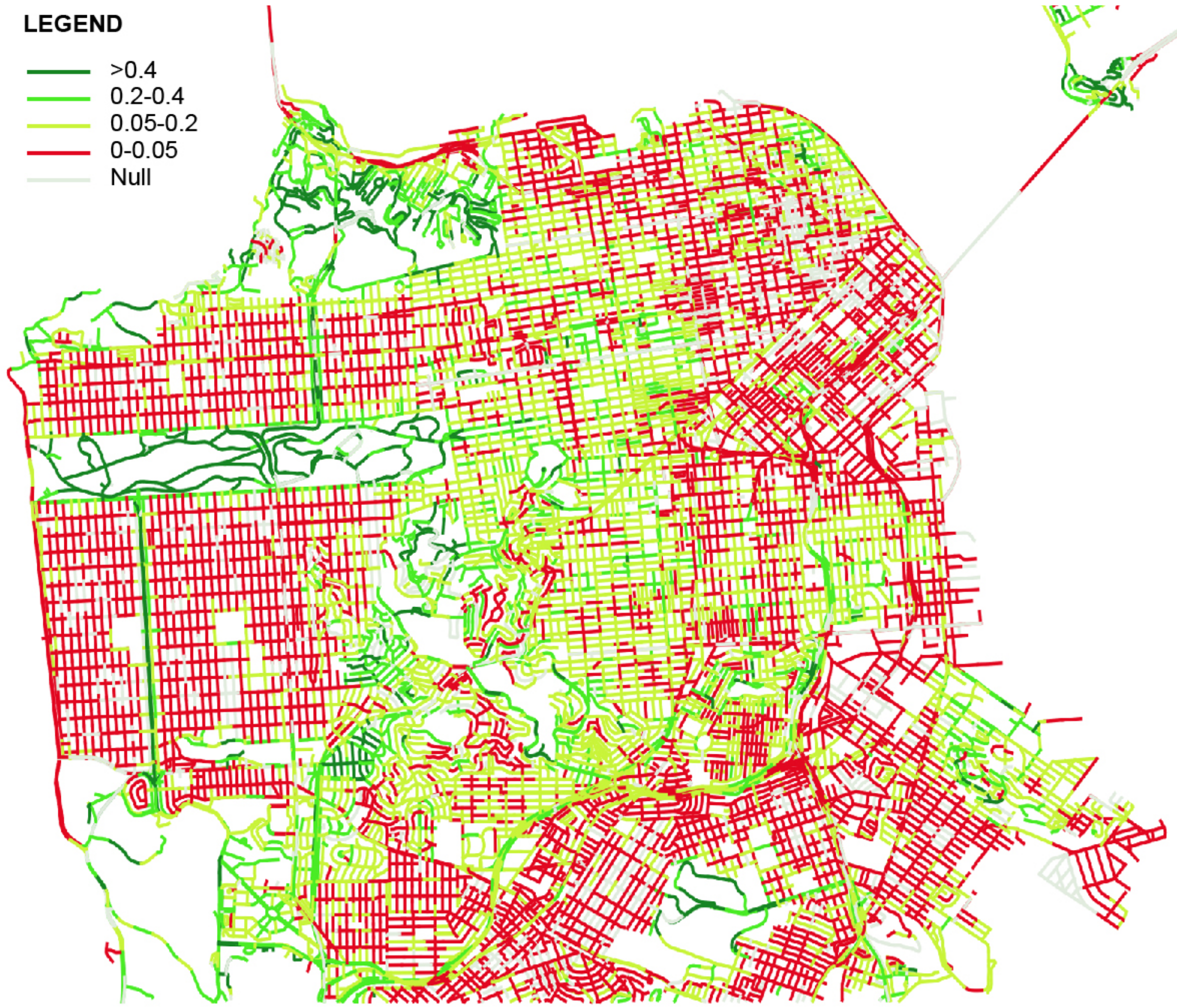

Figure 3. Average Green View Index on San Francisco streets.

4. Results

Route choice parameters were estimated using a path size logit (PSL) model, which presents a multinomial logit (MNL) model that includes a path size correction term to account for the correlation resulting from overlapping alternative routes (Ben-Akiva & Bierlaire 1999):[10]

[10] We used PythonBiogeme for model estimation.

$$U_i = \sum_{k=1}^{11} \beta_{ik} X_{ik} + \beta_{PS} \ln(PS_i) + \varepsilon_i$$

where U_i is the random utility of a route i, β_{ik} represents a vector of preference coefficients for $k = 11$ types of route attributes X_{ik} and PS_i is the path size factor for route i. The error term (ε_i) is assumed to follow a Type-I Generalized Extreme Value (GEV) distribution, also known as a Gumbel distribution. The path size factor is computed as follows:

$$PS_i = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj}}$$

Where Γ_i is the set of links in path i, L_a is the length of link a, L_i is the length of path i, δ_{aj} equals 1 if link a is on path j and 0 otherwise, and $\sum_{j \in C_n} \delta_{aj}$ is the number of paths in choice set C_n that share link a.

The probability $\Pr(i|C_n)$ that a particular route *i* is chosen from a set *C* of *n* alternatives is expressed as:

$$\Pr(i|C_n) = \frac{e^{\mu(U_{in})}}{\sum_{j \in C_n} e^{\mu(U_{jn})}}$$

Since our data were anonymous (i.e., without any identifiable pedestrian characteristics), our model assumes identical preference coefficients for all pedestrians. Moreover, we cannot include a panel effect correction to account for repeated choices by the same individual, as we do not have the necessary information to infer this within the dataset. We also assume the scale parameter of the Gumbel distribution (μ) to be unity for the model to be identifiable.

Model results are presented in Table 2, where we present five different specifications. All models converged with reliable estimates, which we report in addition to 'robust' t-statistics computed using heteroskedasticity-adjusted standard errors. Likelihood-ratio tests were used to compare various models (with different specifications) against one another in an effort to arrive at a robust yet parsimonious "final" model. First, since distance or travel time are known to be critical variables in path choice, model one (M1) illustrates that distance alone explains 56% of variation in route choice behavior. This confirms that trip distance is indeed a key pedestrian path choice variable in our data in San Francisco. While the first model includes only the effect of distance through an MNL specification, the second model does the same using a Path Size Logit (PSL) specification (Ben-Akiva and Bierlaire, 1999). Even though our choice sets were constructed to not include more than 25% overlap with the observed routes, the PSL model does control for the possible overlap (at most 25%), therefore producing more accurate coefficient estimates. Model two with PSL (M2) shows a slight increase in the adjusted pseudo-ρ^2, which increases from 0.56 to 0.58; the route length coefficient also changes from -1.47 to -1.55, implying a stronger negative effect of route length on route choice probability.

Model three (M3) offers a full PSL specification with all the variables described in Table 1. Most variables in M3 have expected signs, but three remained insignificant. The 'public art' variable, which counts how many pieces of public art a route passes, was found to have a statistically insignificant effect.[11] This is explained by the relatively small number of public art installations on the city's streets—only 1.5% of all street segments feature public art. However, roughly 25% routes in our data did pass by at least one piece of public art, which suggests that the presence of

[11] We also tested public art variable alone and found no significant effect either.

public art installations does not significantly affect an average pedestrian's route choice in San Francisco (rather than pointing to a low-sample occurrence of public art in our dataset).

We also found Highway Exposure—measured as the percentage of a route that falls within a 75-meter buffer of a highway—to be statistically insignificant in M3 (with a robust t-statistic of -1.41). Walking next to highways is a relatively rare condition in San Francisco—only 6% of routes in our data indicate walking next to or under a highway. Though the coefficient has an expected negative sign, which suggests that pedestrians prefer routes that avoid proximity to highways, we dropped the variable in the following model due to statistical insignificance. However, we are unable to conclude that proximity to highways does not affect pedestrian route choice; rather, our data does not have adequate representation of routes with highway exposure to provide conclusive evidence of an effect (or the lack thereof).

Third, GVI also remained insignificant in M3 (robust t-statistic of -1.43). While the presence of green views along the route does have an expected positive sign, this effect cannot be confirmed at the 95% confidence level after controlling for other route attributes.

Multinomial choice estimates for pedestrian path attributes in San Francisco (n=14,760).

	M1			M2			M3			M4			M5		
Variable	Beta		t-stat	Beta		t-stat	Beta		t-stat	Beta		t-stat	Beta		t-stat
Length (100s of meters)	-1.47	***	-79.3	-1.55	***	-78.68	-1.04	***	-42.91	-1.05	***	-42.97	-0.955	***	-41.35
Turns (10s)							-6.55	***	-40.86	-6.54	***	-40.89	-6.69	***	-42.31
Elevation gain (100s of meters)							-4.08	***	-7.66	-4.04	***	-7.56	-3.98	***	-7.48
Higway exposure (% of route within 75m highway)							-1.04	~	-1.41						
Amenities (10s, capped at 50)							0.194	***	9.41	0.193	***	9.39	0.201	***	9.94
Public art (nr of installations along route)							-0.036		-0.86						
Sidewalk width (weighted avg. in 10s of ft)							0.871	***	9.04	0.879	***	9.15	0.791	***	8.32
Sky View Factor SVF (weighted avg. %)							-0.823	**	-2.72	-0.848	***	-2.81	-0.471	~	-1.62
Green View Index GVI (weighted avg. %)							0.787	~	1.43						
Traffic speed (weighted avg. in 10s of mph)							-0.583	***	-4.23	-0.591	***	-4.29	-0.434	***	-3.35
Traffic volume (weighted avg. in 1000s of veh/h)							-0.666	**	-2.62	-0.646	***	-2.51	-0.759	***	-3.02
Path Size				1.95	***	22.01	1.84	***	18.77	1.83	***	18.75			
Adjusted rho squared	0.562			0.582			0.679			0.679			0.665		
Log-likelihood Model	-6,118.7			-5,834.4			-4,471.6			-4,474.4			-4,668.6		
Akaike Information Criterion	12,239.4			11,672.9			8,967.3			8,966.7			9,353.1		

Significance level ~p<0.2, *p<0.1, **p<0.05, ***p<0.01
Cell entries are coefficients, robust t-statistics in parentheses.
All weighted avg. variables calculated using using link lengths as weights.

Table 2. Parameter estimates for different specifications of multinomial and path size logit models of pedestrian route choice in San Francisco (N = 14,760).

Next, we excluded these statistically insignificant effects to construct a more parsimonious model. Model four (M4), whose coefficients are interpreted below, shows a relatively high goodness-of-fit (with an adjusted rho-squared of 0.679), suggesting a 68% improvement in choice predictions using the estimated model over an initial benchmark of random choices. This exceeds typical choice model expectations, where rho-square values between 0.2 and 0.4 are considered to be exceptional (McFadden 1980)—likely due to the use of a much larger dataset than most previous path choice models have been able to use. We also illustrate an additional model (M5) in Table 2, which includes the same set of variables as M4, but using an MNL rather than a PSL specification to show how the exclusion of the path size correction can affect our parameter estimates.

All variables in M4 had expected signs. First, route length had a negative effect on choice, as expected. The number of turns and elevation gain along the route also had negative and highly significant coefficients, indicating that pedestrians prefer routes that require less ascent and are cognitively easier to navigate (i.e., require less turns). The presence of amenities and increased sidewalk width, on the other hand, had a positive effect on route choice. People prefer routes with wider sidewalks and more retail, food and service amenities along the way. The sky view factor (SVF) coefficient was negative, suggesting that people prefer routes that are more enclosed by built edges and trees rather than exposed to the sky. Traffic speed and traffic volume along the route also had negative effects, suggesting that both faster driving speeds and higher vehicle volumes along roads decrease pedestrian route choice probability along the corresponding sidewalks. Finally, the path size factor (overlap correction) also has a positive and significant effect.

In discrete choice models, a ratio of two coefficients appearing in the same utility function can be used to illustrate trade-offs or a marginal rate of substitution between one variable and another (Olszewski and Wibowo 2005). This is often operationalized through the value-of-travel-time (VOTT) measure, which is commonly found in the marketing literature. Building on this concept, we can, for instance, talk about a trade-off between elevation gain and distance, and illustrate how much additional distance an average pedestrian considers walking to avoid each additional meter of elevation gain. To interpret such trade-offs, model coefficients need to be de-transformed to reflect desired unit increases in corresponding route characteristics. For instance, our de-transformed coefficients in Table 3 suggest that the tradeoff between a one-meter elevation gain, $\beta_{ELEVATION}$, and walking distance in meters, $\beta_{DISTANCE}$, is:

$$\frac{\beta_{ELEVATION}}{\beta_{DISTANCE}} = \frac{-0.0404}{-0.0105} \approx 3.8$$

This ratio indicates that the effort of walking one vertical meter uphill is perceived as equivalent to 3.8 meters of walking on flat ground. Similarly, the trade-off between individual amenities, $\beta_{AMENITIES}$, and walking distance, $\beta_{DISTANCE}$, can be found as:

$$\frac{\beta_{AMENITIES}}{\beta_{DISTANCE}} = \frac{0.0193}{-0.0105} \approx -1.8$$

This suggests that an average pedestrian is willing to extend the walk by 1.8 meters in order to pass by one additional ground floor amenity. A route that passes 30 amenities is perceived as 54 meters shorter than a route with no amenities, keeping all other route attributes constant.

Variable	**Length equivalent interpretation (meters):**	**M4 (with Path Size)**	**M5 (w/o Path Size)**	**Absolute Diff (%)**
Turns	One extra turn along the route is perceived as:	62.3	70.1	12%
Elevation gain	One meter of elevation gain along the route is perceived as:	3.8	4.2	8%
Amenities	Passing one extra amenity is perceived as:	-1.8	-2.1	15%
Sidewalk width	A 10ft increase of avg sidewalk width along the route is perceived as:	-83.7	-82.8	1%
Sky View Factor (SVF)	An average increase in SVF by 10% along the route is perceived as:	8.1	4.9	39%
Traffic speed	A 10mph increase in avg traffic speed along route is perceived as:	56.3	45.4	19%
Traffic volume	A 1000 cars per hour increase in avg traffic volume along the route is perceived as:	61.5	79.5	29%
			Mean	**17.7%**

Table 3. Interpretation of Model 4 and 5 coefficients from Table 2 as "Equivalent walking distance". Although they share the same specification otherwise, M4 includes the path size correction, while M5 does not. The final column in this table (Absolute Diff %) denotes the absolute percentage difference between the equivalent walking distance values calculated using parameter estimates from M4 and M5.

Table 3 extends the concept of "equivalent walking distance" to all seven variables other than distance in our final model M4, providing a consistent way of interpreting model results in comparable distance units. These results (from column M4 with Path Size) are further interpreted in the next paragraph. In order to gauge how sensitive the results are to the inclusion of the path size correction, Table 3 also compares the same transformations to the coefficients obtained from M5, where path size was not included in estimation (column M5 w/o Path Size). We find from the final column in Table 3 that omitting the path size correction can affect the distance-equivalence results by 17.7% on average, even if alternative routes are pre-selected to have no more than 25% overlap with the observed routes.

The number of turns along a route indicate a substantial effect on perceived distance—each additional turn is perceived as equivalent to 62.3 meters of extra walking distance. A negative effect was expected, given that turns can represent both the cognitive complexity of a route as well as the physical inconvenience of crossing more roads (Hill 1982; Sadalla and Montello 1989; Golledge 1995). Having sidewalks that are 10ft wider, on the other hand, reduces the perceived length of a walk by 83.7 meters, on average. As we already discussed above, the presence of amenities also has a positive effect. The effect of street enclosure suggests that the less enclosed, the longer the route feels. A 10% increase in sky view (SVF) along the route is perceived as 8.1 meters of additional walking distance. A route, on which traffic speeds are 10mph higher, on average, is perceived as 56 meters longer. An additional negative effect of traffic is captured in vehicular throughput, where an increase of 1,000 vehicles per hour along the entire route is perceived as equivalent to 61 meters of extra walking distance.

The concept of "equivalent walking distance" is theoretically universal and can be applied to any route. The results presented above are based on a utility function calibrated on actual pedestrians' route choices (i.e., revealed preference data) in San Francisco. They therefore reflect environmental characteristics of a particular study area (city of San Francisco), the preferences of its particular population, as well as the specific sample of data used in our study. However, as a potentially representative and robust sample, the reported coefficients can be expected to represent a typical pedestrian's route choice in San Francisco and therefore also operationalized for route choice prediction.

The concept of "perceived distance" can be used, for instance, to illustrate pedestrian accessibility to amenities (e.g. shops, services, parks) or to public transit stations. Towards this end, original segment lengths of the street network can be transformed according to their physical characteristics and behavioral response coefficients from above to reflect their shortest "perceived length". We have implemented such transformations for the variables used in this study. Figure 4 presents 15-minute walksheds (1km network radius) around Bay Area Rapid Transit (BART) stations in San Francisco using all variables from Model 4 in Table 2. The blue polygon illustrates the objective 15-minute walkshed, where locations on the street network that are up to 1,000 meters away from a station are included. The overlapping red polygons illustrate the "perceived" 15-minute walkshed for the same stations, where distance is transformed according to the estimated coefficients from M4.

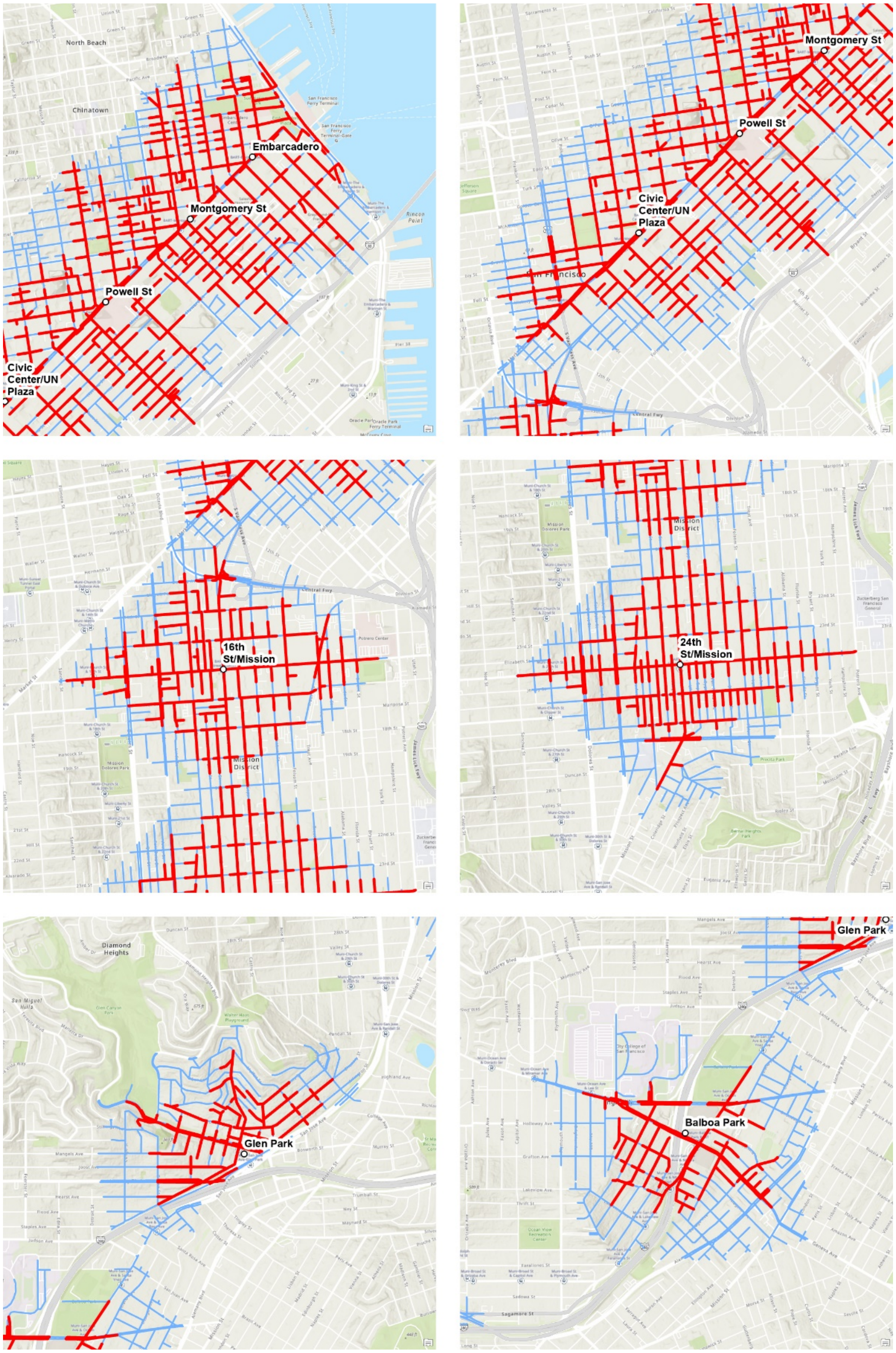

Figure 4. Objective 1,000m distance walksheds (blue) and perceived 1,000m distance walksheds (red) around BART stations in San Francisco.

The map shows that objective distance can significantly over-estimate the perceived extent of the 15-minute walkshed around all stations. The difference between objective and perceived distance is largest in areas that feature unfavorable walking conditions—topography, circuitous routes (more turns), lack of amenities, narrower sidewalks, higher traffic volumes etc. For instance, the cumulative length of streets captured in the objective 1,000m catchment area around Balboa Station is 34.43km, but less than a third of that (10.34km) according to the perceived extent of the same distance. Perceived catchment areas are more similar to the objective 1,000m ones around downtown stations, where streets have wider sidewalks and more ground floor amenities, but even there they tend to reach only 60-70% of the absolute walkshed. Understanding how route features transform perceived walking distances can help planners estimate more realistic catchment areas for transit stations as well as other types of pedestrian destinations according to expected behavioral responses to streets that link them.

5. Discussion and Conclusion

This paper presents the results of a pedestrian route choice model using an unprecedented large number of GPS traces in San Francisco. Using route choice data that covered all parts of the city allowed us to discuss how street environments affect a typical pedestrian in San Francisco, rather than a specifically recruited sample of users or in a specific geographic neighborhood of the city. Eleven different route characteristics were tested, eight of which showed a significant relationship to pedestrian route choice. The estimated model demonstrated over two-thirds (67.9%) improvement in choice predictions over an initial benchmark of random choices.

In addition to detailed route characteristics measured in GIS, GVI and SVF were operationalized by processing a large number of Google Street View API images with a state-of-the-art feature detection algorithm that classified the extent of greenery and the extent of sky exposure at 10-meter intervals along all streets. We found both variables to reliably capture the anticipated effects on pedestrian routes, though only SVF was found to be significantly related to pedestrian route choice. Google Street View data offer potential capabilities to capture more detailed visual characteristics of streets in future work, but significant computation is required in harvesting such data. Google Street View data thus appears better justified for variables that do not have reasonable alternatives using simpler data sources. For instance, vegetation cover along a route (GVI) can be measured in simpler ways, using GIS data on tree canopies. However, our SVF metric, which captures the sense of street enclosure, does not have obvious alternatives and has been difficult to include in past research (Ewing and Handy 2009).

A particularly challenging but inevitable element of all path choice studies involves the generation of alternative routes. In this study, we explored generating path alternatives using pedestrian directions from Google Maps API, but found these alternatives to be too biased towards more attractive or pleasant trajectories. We ultimately implemented a software solution to source route alternatives using the constrained enumeration approach that returns all possible non-looping route options between trip origins and destinations that can deviate up to a given percent longer than the shortest path. From a potentially large set of routes between each O-D pair, a user can indicate how many random paths to draw for inclusion in the choice set. This approach was implemented as part of the XXX software plugin, which we have made freely available for choice set generation online (Ref XXX, masked to preserve author anonymity during the review process).

The methodology we used does not come without shortcomings. Even after cleaning and filtering, data from GPS trajectories includes biases and unknowns. First, given the anonymous nature of our data, trip purpose remains unknown to us, thus preventing us from exploring systematic differences in pedestrian route choice preferences by trip purpose or type of user. We do not have information on whether the pedestrians were going to work, conducting errands or taking a leisurely stroll. We also could not detect gender, age, race or class differences among travelers. Trip purpose and user type can have a significant influence on route choice, which remains unobserved in our study. This is a documented shortcoming of passive route data collection (Shoval and Isaacsion 2006). Second, GPS tracking requires an open sky environment to detect signals from multiple satellites simultaneously. Signal can be lost in dense built environments where building forms obstruct sky view. This problem may be resolved if auxiliary signal transmitters are distributed in the environments where satellite signals are weak or unavailable (Akura and Irya 2005), but we cannot estimate whether and how much signal obstructions might have affected the raw data we used. Third, using anonymous GPS trajectories from smartphones always comes with the caveat that raw data may not cover complete activity trajectories—users could decide to turn off the app temporarily during their walks. Although certainly plausible, we think that people are unlikely to repeatedly turn their smartphone GPS transceivers on and off during the same walk. If GPS is toggled on or off once during the walk, the recorded trajectory still contains a valid, but partial route. Fourth, and most importantly, using an activity-based app to trace walking routes inevitably comes with a selection bias. Not everyone is equally likely to install or use such an app—our results could therefore be biased towards younger, wealthier, more technologically savvy, and more active users that chose to use the activity-based app that generated the data. Despite these issues, we believe anonymized GPS traces offer ample potential for better understanding spatial mobility choices at significantly larger scales than traditional survey methods allow.

One of the shortcomings of not just this study, but most pedestrian route choice studies to date, has been the relative lack of application of route choice coefficients to predictive purposes. We hope to follow up on this study, where calibrated model results are used to empirically validate both individual pedestrians' route choices as well as aggregate pedestrian flows in key areas of the city. Similar to Guo and Loo (2013), it would be also interesting to repeat a similar project in a different city using the same variables. This would allow the analysis to reveal whether and how pedestrians' response to analogous environmental features may vary in different cities or urban cultures. While the 'equivalent walking distance' construct we presented earlier can help in cross-cultural comparisons, this will have to remain a subject of future work.

References

Agrawal, Asha Weinstein, Marc Schlossberg and Katja Irvin (2008). 'How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference', Journal of Urban Design 13(1): 81–98.

Asakura, Y., & Iryo, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice*, *41*(7), 684–690.

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science.* (pp. 5–33). Springer. https://link.springer.com/chapter/10.1007/978-1-4615-5203-1_2

Bovy, P. H. L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews*, *29*(1), 43–68. https://doi.org/10.1080/01441640802078673

Broach, J., & Dill, J. (2015). Pedestrian Route Choice Model Estimated from Revealed Preference GPS Data. In *Transportation Research Board 94th Annual Meeting*. Washington, D.C. Retrieved from https://trid.trb.org/view/1338221

Calvin P Tribby, Harvey J Miller, Barbara B Brown, Carol M Werner, K. R. S. (2017). Analyzing walking route choice through built environments using random forests and discrete choice techniques. *Environment and Planning B: Urban Analytics and City Science.*, *44*(6), 1145–1167.

Chao, P., Xu, Y., Hua, W., & Zhou, X. (2020). *A Survey on Map-Matching Algorithms BT - Databases Theory and Applications* (R. Borovica-Gajic, J. Qi, & W. Wang (eds.); pp. 121–133). Springer International Publishing.

Cho, G.-H., Rodriguez, D. A., Evenson, K. R. (2011). Identifying Walking Trips Using GPS Data. *Medicine & Science in Sports & Exercise*, *43*(2). Retrieved from https://journals.lww.com/acsm-msse/Fulltext/2011/02000/Identifying_Walking_Trips_Using_GPS_Data.23.aspx

Dill, J. (2015). Where do people walk? *Active Living Research Conference, San Diego*, 20 p. Retrieved from https://activelivingresearch.org/where-do-people-prefer-walk-pedestrian-route-choice-model-developed-gps-data

Duncan, M. J., & Mummery, W. K. (2007). GIS or GPS? A Comparison of Two Methods For Assessing Route Taken During Active Transport. *American Journal of Preventive Medicine*, *33*(1), 51–53. https://doi.org/https://doi.org/10.1016/j.amepre.2007.02.042

Erath, A., Eggermond, M. van, Ordonez, S., & Axhausen, K. (2015). Modelling for Walkability: understanding pedestrians' preferences in Singapore. *IATBR 2015*. Retrieved from https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/293220/1/v570.pdf

Ewing, R., & Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *Journal of Urban Design*, *14*(1), 65–84.

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70–80. https://doi.org/https://doi.org/10.1016/j.trb.2013.07.012

Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, *43*(10), 984–994. https://doi.org/https://doi.org/10.1016/j.trb.2009.03.001

Garbrecht, D. (1971). Pedestrian paths through a uniform environment. *Town Planning Review*, *41*, 78–84.

Golledge, R. G. (1995). *Path Selection and Route Preference in Human Navigation: A Progress Report* (G. Goos, J. Hartmanis, & J. van Leeuwen (eds.); pp. 207–222). Springer.

Guo, Z., & Loo, B. P. Y. (2013). Pedestrian environment and route choice: evidence from New York City and Hong Kong. *Journal of Transport Geography*, *28*, 124–136.

Hassan, M. N., Rashidi, T. H., & Nassir, N. (2019). Consideration of different travel strategies and choice set sizes in transit path choice modelling. *Transportation*. https://doi.org/10.1007/s11116-019-10075-x

Hill, M. R. (1982). *Spatial Structure and Decision Making Pedestrian Route Selection Through an Urban Environment: Vol. PhD*. University Microfilms International.

Johnson, G. T., & Watson, I. D. (1984). The determination of view-factors in urban canyons. Journal of Climate and Applied Meteorology, 23(2), 329–335.

Kim, H. (2015). Walking distance, route choice, and activities while walking: A record of following pedestrians from transit stations in the San Francisco Bay area. *URBAN DESIGN International*, *20*(2), 144–157. https://doi.org/10.1057/udi.2015.2

Leyden, K. M. (2003). Social capital and the built environment: the importance of walkable neighborhoods. *American Journal of Public Health*, *93*(9), 1546–1551.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015a). Assessing street-level urban greenery using Google Street View and a modified green view index. Urban Forestry & Urban Greening, 14(3), 675-685.

Li, X., Zhang, C., Li, W., Kuzovkina, Y. A., & Weiner, D. (2015b). Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. Urban Forestry & Urban Greening, 14(4), 751-759.

Li, X., Ratti, C., & Seiferling, I. (2018). Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. Landscape and Urban Planning, 169, 81-91.

Li, X., Santi, P., Courtney, T. K., Verma, S. K., & Ratti, C. (2018b). Investigating the association between streetscapes and human walking activities using Google Street View and human trajectory data. Transactions in GIS, 22(4), 1029-1044.

Luberoff, D. (2019). Reimagining and Reconfiguring New York City's Streets. In D. E. Davis & A. Altshuler (Eds.), *Transforming Urban Transportation*. Oxford University Press.

Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and Society*, *14*, 34–42. https://doi.org/https://doi.org/10.1016/j.tbs.2018.09.008

Malleson, N., Vanky, A., Hashemian, B., Santi, P., Verma, S. K., Courtney, T. K., & Ratti, C. (2018). The characteristics of asymmetric pedestrian behavior: A preliminary study using passive smartphone location data. Transactions in GIS, 22(2), 616-634.

McFadden, D. (1980). Econometric Models for Probabilistic Choice among Products. *Journal of Business*, *53*(3).

Moiseeva, A., & Timmermans, H. (2010). Imputing relevant information from multi-day GPS tracers for retail planning and management using data fusion and context-sensitive learning. *Journal of Retailing and Consumer Services*, *17*(3), 189–199.

Muraleetharan, T., & Hagiwara, T. (2007). Overall Level of Service of Urban Walking Environment and Its Influence on Pedestrian Route Choice Behavior: Analysis of Pedestrian Travel in Sapporo, Japan. *Transportation Research Record*, *2002*(1), 7–17.

Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 336–343). Seattle, WA: ACM

Olszewski, P., & Wibowo, S. S. (2005). Using Equivalent Walking Distance to Assess Pedestrian Accessibility to Transit Stations in Singapore. *Transportation Research Record: Journal of the Transportation Research Board.*, *No. 1927*, 38–45.

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations. Journal of Climatology, 1(3), 237–254.

Prato, C. G., & Bekhor, S. (2007). Modeling route choice behavior: How relevant is the composition of choice set? *Paper Presented at the 86th Annual Meeting of the Transportation Research Board, Washing- Ton, D.C., January 2007.*

Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100. https://doi.org/https://doi.org/10.1016/S1755-5345(13)70005-8

Raymond, R., Tetsuro, Morimura, Osagami, T., & Hirosue, N. (2012). Map matching with Hidden Markov Model on sampled road network. Pattern Recognitions (ICPR) 2012 21st International Conference on Pattern Recognition, 1–4. https://www.researchgate.net/publication/261161841_Map_matching_with_Hidden_Markov_Model_on_sampled_road_network

Rieser-Schüsslera, N., Balmerb, M., & Axhausen, K. W. (2013). Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science*, *9*(9), 825–845.

Sadalla, E. K., & Montello, D. R. (1989). Remembering Changes in Direction. *Environment and Behavior*, *Vol. 21*(No. 3), 346–363.

Senevirante, P. N., & Morral, J. F. (1985). Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and Technology*, *Vol. 10*, 147–159.

Sevtsuk, A., & Mekonnen, M. (2012). Urban Network Analysis Toolbox. *International Journal of Geomatics and Spatial Analysis*, *22*(2), 287–305.

Sevtsuk, A. (2018). *Urban Network Analysis. Tools for Modeling Pedestrian and Bicycle Trips in Cities.* Harvard Graduate School of Design. http://cityform.mit.edu/projects/una-rhino-toolbox

Shoval, N., & Isaacson, M. (2006). Application of Tracking Technologies to the Study of Pedestrian Spatial Behavior*. *The Professional Geographer*, *58*(2), 172–183.

Speck, J. (2013). *Walkable City. How downtown can save America one step at a time.* North Point Press. 320p.

Vanky, A. (2017). *To and fro: digital data-driven analyses of pedestrian mobility in urban spaces* (Massachusetts Institute of Technology, PhD dissertation, Department of Urban Studies and Planning). Retrieved from https://dspace.mit.edu/handle/1721.1/111372

Wei, H., Wang, Y., Forman, G., & Zhu, Y. (2013). Map matching: comparison of approaches using sparse and noisy data. *SIGSPATIAL'13: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 444–447. https://doi.org/10.1145/2525314.2525456

Whyte, W. H. (1980). *The Social Life of Small Urban Spaces*. New York City.: Conservation Foundation.

Wiehe, S. E., Carroll, A. E., Liu, G. C., Haberkorn, K. L., Hoch, S. C., Wilson, J. S., & Fortenberry, D. J. (2008). Using GPS-enabled cell phones to track the travel patterns of adolescents. *International Journal of Health Geographics*, *7*(22). Retrieved from https://www.academia.edu/9563811/Using_GPS-enabled_cell_phones_to_track_the_travel_patterns_of_adolescents

Wolf, J., Guensler, R., & Bachman, W. (2001). Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data. *Transportation Research Record*, *1768*(1), 125–134. https://doi.org/10.3141/1768-15

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890).