MIT Open Access Articles

Predicting pedestrian flow along city streets: a comparison of route choice estimation approaches in downtown San Francisco

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Sevtsuk, Andres and Raul Kalvo. "Predicting pedestrian flow along city streets: a comparison of route choice estimation approaches in downtown San Francisco." Forthcoming in: International Journal of Sustainable Transportation.

As Published: 10.1080/15568318.2020.1858377

Persistent URL: <https://hdl.handle.net/1721.1/128967>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Predicting pedestrian flow along city streets: a comparison of route choice estimation approaches in downtown San Francisco.

By **Andres Sevtsuk¹**, **Raul Kalvo²**

1 Massachusetts Institute of Technology, Department of Urban Studies and Planning.

2 Tampere University.

Abstract

Street attributes are thought to play an important role in influencing pedestrian route choices. Faced with alternatives, pedestrians have been observed to choose faster, safer, more comfortable, more interesting or more beautiful routes. Literature on pedestrian route choice has provided methods for assessing the likelihood of such options using discrete choice models. However, route choice estimation, which is data intensive and computationally challenging, remains infrequently deployed in planning mobility analysis practice. Even when coefficients from previous studies are available, operationalizing them in foot-traffic predictions has been rare due to uncertainty involved in the transferability of behavioral effects from one context to another, as well as computational challenges of predicting route choice with custom attributes. This paper explores a simpler method of route choice prediction, implemented in the Urban Network Analysis toolbox, which assigns probabilities to available route options based on distance alone. We compare the accuracy of distance-weighted approaches to the more detailed utility-weighted approach using a large dataset of observed GPS pedestrian traces that include numerous trips between same intersections pairs in downtown San Francisco as a benchmark. Even though a utility-weighted model matches observed pedestrian flows most accurately, a distance-weighted model is only marginally inferior, on average. However, shortest-distance and highest-utility route predictions are both significantly inferior to the utility-weighted and distance-weighted sample-enumeration methods. Our findings suggest that simplified assumptions can be used to predict pedestrian flow in practice with existing software, opening pedestrian flow predictions to a wider range of planning and transportation applications.

Introduction

Urban planning and transportation studies have long found that street attributes play an important role in influencing pedestrian route choice (Pushkarev and Zupan 1975). Faced with alternative options, pedestrians have been found to prefer routes of higher utility—faster, safer, more comfortable, more interesting or more beautiful routes (Guo and Loo 2013, Ewing et al. 2016; Lue and Miller 2019). A robust literature on pedestrian route choice has provided methods for assessing the likelihood of different route options using discrete choice models (Broach and Dill 2015; Erath et al. 2015; Guo and Loo 2013). But rarely have estimated pedestrian route choice coefficients been used proactively to predict pedestrian flows on unbiased data. Even when coefficients from previous studies suggest how route attributes could affect route choice, operationalizing them in pedestrian flow prediction has been challenged by the uncertainty of using behavioral coefficients from a different study area as well as difficulties involved in assembling data or implementing model results in predictive routing applications with custom attributes. For predictive pedestrian models to gain broader popularity among planning and transport practitioners, simpler approaches are needed.

Distance-based approaches for modeling pedestrian behavior are easier to specify and have recently been implemented in network analysis software tools, such as the Urban Network Analysis toolbox, sDNA, Depthmap, Urbano and MCA (Turner 2001; Porta et al. 2005; Sevtsuk 2011, 2018, 2020b; Dogan et al. 2018; Cooper and Chiaradia 2020). Such tools can greatly simplify the use and broaden the application of pedestrian flow modeling. There is, however, a lack of empirical work to evaluate their accuracy with respect to observed pedestrian flows, or to compare their predictive accuracy with the more detailed and “standard setting” utility-based path choice predictions.

In this paper we operationalize pedestrian path choice coefficients from a prior study carried out in San Francisco by Sevtsuk et al. (2020a) to predict pedestrian flows. Using data from anonymous GPS traces collected from a smart phone app, we examine pedestrian flows between node pairs in the street network that are traversed by a relatively large number of pedestrians. Diverging from the more common revealed route choice preference surveys that record routes for spatially heterogeneous origin-destination pairs, the data we use enables us to focus on

individual users walking between the exact same origin-destination nodes in the street network, collectively producing a route-choice pattern within a fixed environment. This data allows us to compare how accurately different routing predictions match the observed flow patterns—how do the observed flows compare with different prediction approaches? In particular, we explore how the accuracy of utility-weighted pedestrian flow predictions compares against significantly simpler distance-weighted pedestrian flow predictions, which network analysis software can produce without detailed data on route attributes (Sevtsuk 2018). Based on prior literature on pedestrian path choice, we expect a utility-based model to provide a more accurate match. But how much additional accuracy is gained in specifying utility-based predictions, compared to significantly simpler distance-weighted predictions?

Distance-weighted models have been empirically compared to observed pedestrian counts on a district’s street network with rather promising outcomes (Sevtsuk 2018a; Coper et al. 2019, Sevtsuk 2020b, c). But when predicted pedestrian flows are compared to pedestrian counts on particular street segments, it is implicitly assumed that the observed counts capture modeled movement between surrounding origin-destination pairs. This may or may not be a realistic, since trip trajectories themselves remain unobserved in data that counts pedestrians at point locations. Counters, whether human or automated, register all passing foot-traffic, not only flows between node-pairs that are captured in a predictive model.

Pedestrian flow predictions have not, to the best of our knowledge, been compared against pedestrian trajectories that flow from known origins to known destinations, which would allow the modeled trajectories to be compared against observed trajectories. We also have not found any precedents that compare either shortest- or k-shortest-route pedestrian flow predictions (Sevtsuk 2018; Coper et al. 2019, Sevtsuk 2020 b,c) against utility-weighted path choice models that set the theoretical standard for pedestrian path choice estimation. By addressing these gaps, this paper compares observed pedestrian trajectories to multiple prediction approaches including shortest-distance, highest-utility, distance-weighted and utility-weighted pedestrian flow predictions.

The paper is organized as follows. We first discuss pedestrian flow prediction literature and describe commonly used approaches as well as missing gaps. We then introduce the data and our

experimental setup for comparing pedestrian flow models in San Francisco, CA. The results section presents our findings and the discussion comments on implications for both urban planning and transportation practice and points to future research.

Literature

Pedestrian path choice analysis has emerged as an active niche in the broader context of traffic path choice analysis (Lee & Moudon 2006; Ewing & Handy 2009; Guo & Loo 2013; Broach & Dill 2015; Erath et al. 2015; Ewing et al. 2016; Shatu et al. 2019). Pedestrian path choice models assume that when faced with alternative path options, people choose their walking routes based on path attributes—their perceived qualities that can be captured as part of a utility function. Path choice models identify people’s behavioral reactions to route attributes with coefficients, which can explain the magnitude and statistical significance of such attributes on observed path selection behavior. The most fundamental utility attribute is typically path length or travel time, which pedestrians generally tend to minimize. But studies have shown that people are often willing to deviate to longer routes if such detours provide access to preferred path attributes—in particular safer, more comfortable, easier to navigate, and more interesting routes.

A number of authors have found perceptions of route safety to affect route choice. Street features related to safety may consist of traffic volumes and traffic crossings along the route (Broach and Dill 2015), proximity to adjacent traffic flow, presence of sidewalks (Muraleethan and Hagiwara 2007), perceived crime rates, or level changes (Olzsewski and Wibowo 2005; Erath et al. 2015). Path choice also depends on comfort-level of alternative options. Comfort variables may include sidewalk width or paver stone quality (Erath et al. 2015; Sevtsuk et al. 2020 a), level-of-crowding indicators (Muraleethan and Hagiwara 2007), shading, wind and other climate related comfort variables (Bosselman et al 1995; Zacharias, J., Stathopoulos 2001; Kim and MacDonald 2015; Erath et al 2015). Third, a number of studies have related pedestrian path choice to navigational attributes (i.e. landmarks) and the geometric characteristics of street networks. Navigational attributes found to affect pedestrian route choice most commonly include the number of turns encountered on a route (Golledge 1995; Conroy-Dalton 2001; Hillier and Iida 2005; Montello 2006; Lue and Miller 2019; Shatu 2019), but may also include the presence of

landmarks or the geometric properties of street networks along the route (Lynch 1960; Golledge 1995; Conroy-Dalton 2001; Montel and Sas 2006; Mohsenin and Sevtsuk 2013; Sharmin & Kamruzzaman 2016; Coutrot et al. 2018, 2020). Fourth, pedestrian path choice can also be affected by the interestingness as well as aesthetic attractiveness of a route. Among others, Guo (2009), Guo and Loo (2013) and Broach and Dill (2015) have empirically found that pedestrians prefer routes that pass by more ground floor amenities and shopfronts. An overview of how different factors have been found to significantly affect pedestrian path choice in previous studies are provided by Ewing et al (2006), Mehta (2008); Ewing & Handy (2009).

However useful utility-based choice models can be in identifying factors that explain path choices on observed trajectories, there has not been much work on validating how well, once calibrated, they can predict pedestrian flows *ex ante* on out of sample data that has not been used for coefficient estimation. One of the earliest examples of examining how observed pedestrian flows correspond to predicted flows was carried out by Garbrecht (1971), who hypothesized that pedestrians assign either equal probabilities to all equidistant routes within regular street grids, or make equally probable turning decisions within the sample-space of equidistant route segments (Garbrecht 1970, 71). Though he did not focus on other utility attributes beyond distance, and only examined pedestrian path choice within a limited context of a gridiron parking lot, Garbrecht's work has inspired our methodology by examining pedestrian route choices between fixed origin and destination locations and treating routes choices probabilistically, rather than using a single lowest-distance or highest-utility path.

Borges and Timmermans (1986) developed a discrete choice model for pedestrian mobility in shopping areas and calibrated the model on 426 respondents, whose actual routes were surveyed in Maastricht, Holland. The model, which used a distance-weighted probability-distribution similar to what we examine below, showed an impressive fit to both pedestrian counts and shop visits, but the authors did not use the fitted model to assess the predictive accuracy on out-of-sample, unbiased data. A number of other pedestrian path choice studies have fitted models to observed route choices, but have also not examined the predictive accuracy of such models at other locations (Hill 1982; Olzsewski and Wibowo 2005; Guo and Loo 2013; Erath et al. 2015; Broach and Dill 2015).

As one exception, Broach and Dill (2016) used a number of path utility variables from a previously calibrated model to assign predicted utilities to both pedestrian and bicycle route choice models in Portland, OR and operationalized the resulting behavioral coefficients to estimate mode choice for pedestrians and bicyclists. Though the authors did find a relationship between available path utility and non-motorized mode choice, the focus here too was not on the accuracy of actual path choice nor on the scrutiny of different path choice models.

A few studies have correlated empirically observed pedestrian counts with predicted flows between assumed origin and destination locations (Hillier and Iida 2005; Sevtsuk 2018; Cooper et al. 2019). Yet these studies too have relied on underlying assumptions about path choice, using shortest paths, k-shortest paths or other route assignment conventions that are assumed rather than estimated as part of the analysis. In none of the studies we examined was the focus on predicting how multiple travelers move between the fixed origin-destination pairs, or on comparing the accuracy of different pedestrian path prediction approaches over street networks with respect to observed data.

A sizable body of pedestrian routing literature has also emerged around agent-based simulation models (Batty 2001; Kerridge et al 2001; Liu et al 2014). These types of models usually focus on visualizing crowd dynamics interactively using graphic agents in a simulation environment. A key benefit of agent-based models is that they allow simulated agents to react to each other, capturing the effects of queuing, overcrowding and other forms of person-to-person interactions that cannot be captured in network-based models. Yet, agent-based pedestrian models have focused less on capturing route utilities that are central to transportation models, and instead typically used the visual environment around agents to guide navigation (Puusepp et al 2017). Agent-based navigation models have more recently also been combined with machine-learning approaches to improve their predictive capacity (Liebig et al 2012; Tribby et al 2012).

An assessment of predictive accuracy of different path choice model thus appears needed. It seems particularly important to examine how utility-based path choice models compare to simpler alternatives. Such a comparison is developed in the following, using observed pedestrian routes in San Francisco, CA as evidence.

Data and methodology

Assessing the accuracy of alternative approaches to pedestrian flow prediction requires a rather specific experimental setup. In order to estimate the accuracy of different path choice assumptions, model results cannot be only correlated with observed pedestrian counts at discrete locations—it is important to examine entire trajectories between given origin and destination locations. Furthermore, multiple uncoordinated individual's trajectories are needed between identical origin-destination pairs for statistical analysis of alternative routing approaches.

Such data has been challenging to come by or collect. Both stated preference and revealed preference path choice surveys can observe individual walking trajectories, but the spatial distribution of trajectories tends to vary, making it difficult to observe trajectories of numerous pedestrians between the same origin-destination pairs. Studies that have documented walking routes around a fixed location (e.g. transit station) have managed to fix one of the two locations, but trips still tend to start or end at different places, making it challenging to observe numerous independent journeys between fixed origin-destination pairs (Olzsewski and Wibowo 2005; Kim 2015).

In this study we obtained data from an activity-based smart phone app in San Francisco, which features hundreds of thousands of unique pedestrian trips, recorded in 2014. The high volume of the data makes it possible to identify unique origin-destination pairs in the city that have been traversed by a relatively large number of individual users. These origin-destination pairs are not necessarily trip starting and ending points, but rather pairs of nodes in the street network that happened to be traversed by a large number of different pedestrians.

The actual geolocated dot data from GPS traces was map-matched to street centerlines, using 2014 TIGER/Line (Topologically Integrated Geographic Encoding and Referencing) geometries from the US Census.¹ We used the Hidden Markov Map-Matching (HMM) algorithm to map-match raw GPS traces to street centerlines (Newson & Krumm, 2009). Though a more detailed sidewalk and crosswalk network geometry was also considered, GPS data in dense urban environments is not accurate enough to distinguish the side of a street that a pedestrian is on with

¹ <https://catalog.data.gov/dataset/tiger-line-shapefile-2014-county-san-francisco-county-ca-all-roads-county-based-shapefile>

certainty. Using a more detailed and denser network of sidewalk segments also makes it more challenging to relate street-based attributes (e.g. tree-cover, highway proximity, turns) to specific segments over others.² We therefore preferred to use street centrelines, which avoid these ambiguities, while maintaining a reasonably accurate distinction of streets chosen according to GPS traces.

We implemented a Python script to identify how many separate trajectories passed through each pair of intersections in the city's street network. In order to ensure that multiple route options were realistically available between each location pair, we eliminated all node-pairs that lay on straight line from each other and only selected node-pairs that were located diagonally across several city blocks so as to offer several plausible route options. We also constrained the shortest path distance between the node-pairs to no less than 200m so as to maximize available route options.³ Within these constraints, we identified ten intersection pairs, shown in Figure 1, which have 50 or more unique pedestrian trajectories. Perhaps not surprisingly, all of the node pairs that met the frequency and choice criteria were located in downtown, where most foot-traffic occurs. The distance between the chosen origin and destination nodes range from 389m (nodes C-C') to 533m (nodes I-I'). Black line weights in Figure 1 indicate the number of observed pedestrians on each street segment between the origin-destination pairs from our GPS data and red numbers next to diagonals indicate their corresponding total numbers for the given node pair. Node pair A-A', for instance, had 94 observed trajectories crossing both nodes, while node pair E-E' had 202 trajectories.

² Sidewalk networks are often represented by multiple lines (i.e. wide sidewalks have two parallel lines with perpendicular connectors, plazas and parks have radiating networks).

³ The filtering brought the initial raw route count of 120,045 unique trajectories in the city down to 26,344 trajectories.

Figure 1. Ten locations of node-pairs where repeating pedestrian flows were analyzed in downtown San Francisco, CA. Numbers next to diagonals indicate numbers of pedestrians observed between the given O-D pair.

Pedestrian flow prediction

Pedestrian flows were predicted for the same ten node pairs using five different prediction approaches, which we refer to as “shortest path”, “equal probability”, “distance-weighted probability”, “utility-weighted probability”, and “highest utility” (Figure 2).

As the name suggests, the “shortest path” approach finds a single, absolute shortest route between each node-pair (Figure 2b). The “equal probability” approach finds all routes within a given “detour ratio” above the shortest path, assigning each of them an equal probability (Figure 2c). The detour ratio describes the ratio between allowable path lengths and the shortest path length for a given location pair. If the “equal probability” approach identifies five satisfactory routes from an origin to a destination, then each of the five routes obtain the same probability of

0.2. Satisfactory routes may not repeat nodes—they thus exclude loops and backtracking along the way. Where segments overlap on alternative routes, their probabilities are summed. If the origin and destination are located in the middle of a segment, then the first and last segment of the journey are inevitably shared by all routes and therefore obtain a probability of one.

Figure 2. a. Actual flows between nodes F-F' and alternative prediction methods: b. Shortest path. c. equal probability for all routes that are up to 20% longer than the shortest route. d. distance weighted probability. e. utility weighted probability. f. highest utility route.

We use a detour ratio of “1.2” in our predictions, which identifies all routes that are up 20% longer than the shortest path. An analysis of all 14,760 unique trajectories used in Sevtsuk et al. (2020 a) suggested that the average chosen route was 9% longer than the shortest available route

between the same origin and destination (detour ratio= 1.09). Our choice of 1.2 reflects the 80th percentile detours observed in the data. We also examined the pedestrian trajectories between the ten origin-destination pairs used in the present analysis (Figure 1) and found that rarely did people choose routes that exceeded 20% detours over the shortest available route.

The “distance-weighted” probability approach generates the same exact set of path options as the “equal probability” approach, but assigns proportionately higher probabilities to shorter route options than longer route options (Figure 2d). This is a sample enumeration technique that treats distance as a disutility factor, where U_{in} represents the inverse of distance (1/route length) for route i (McFadden 1980). The probability $\Pr(i|C_n)$ that a particular route i is chosen from a set C of n alternatives is expressed as:

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

Equation 1

We assume the scale parameter of the Gumbel distribution (μ) to be unity⁴.

The “utility-weighted” approach too uses the same set of +20% detour path options as the “equal probability” and “distance-weighted” approaches above, but assigns probabilities to individual route options based on their quality attributes as well as empirically calibrated behavioral coefficients that describe how pedestrians are likely to react to such quality attributes. To calculate route utilities, here we simply reuse the estimated coefficients for nine different route attributes based on the final utility model presented by Sevtsuk et al. (2020a) in San Francisco, which worked with the same dataset of anonymous GPS traces in the city (Table 1). The Sevtsuk et al. (2020a) model had a relatively high goodness-of-fit outcome for a choice model (adjusted rho-squared of 0.665), suggesting a 66.5% improvement in choice predictions using the estimated model over an initial benchmark of random choices

⁴ Discrete choice models commonly assume that the random component of utility is IID and takes a Gumbel distribution. The Gumbel distribution is a skewed distribution with two parameters, a location μ and scale σ . If we fix $\sigma=1$, this helpful property allows us to formulate the probability of each choice with only the values for each μ_j , a multivariate generalization of the inverse logit function. See Train (2001, Chapter 3) for the derivation.

Variable name	Definition	Model	
		Beta	Robust t
Amenities	Tens of amenities along route, capped at 50	0.201 *** (9.94)	
Green View Index (GVI)	Weighted avg. % GVI along route (using link lengths)	1.020 * (1.92)	
Length	100s of meters of route length	-0.955 *** (-41.35)	
Sidewalk width	Weighted avg. of sidewalk width in 10s of ft (using link lengths)	0.791 *** (8.32)	
Sky View Factor (SVF)	Weighted average % SVF along route (using link lengths)	-0.471 ~ (-1.62)	
Traffic speed	Weighted average speed limit in 10s of mph (using link lengths)	-0.434 *** (-3.35)	
Traffic volume	Weighted avg. of 1000s of veh/h (using link lengths)	-0.759 *** (-3.02)	
Turns	10s of turns along route	-6.690 *** (-42.31)	
Elevation gain	100s of meters of vertical uphill gain along route	-3.980 *** (-7.48)	
Adjusted rho squared		0.665	
Log-likelihood Model		-4668.6	
Akaike Information Criterion		9353.1	

Significance level ~p<0.25, *p<0.1, **p<0.05, ***p<0.01
Cell entries are coefficients, robust t-statistics in parentheses.

Table 1. Parameter estimates for multinomial logit model of pedestrian route choice in San Francisco (n=14,760).

The route choice variables included in Sevtsuk (2020a) and reused in our model’s utility function are interpreted as follows.⁵ “Amenities” describes how many retail, food service, personal service and entertainment establishments are found along each route. The green view index (GVI) measures the weighted average percent of greenery seen in Google Street View panoramic images, sampled at 10-meter intervals along the route (Li et al. 2015). The sky view factor (SVF) captures the inverse of the sense of enclosure along the route—it measures the weighted average percent of sky visible in Google Street View panoramic images, sampled at 10-meter intervals along the route. Sidewalk width describes the average width of sidewalks along a route, weighted by segment lengths. Traffic speed and traffic volume describe the weighted average speed limit and the weighted average hourly vehicle counts along the route. Finally, length, turns and elevation gain illustrate route distance, how many unique turns (45 degrees or more) are encountered along the way, and how much uphill elevation is gained on each route. Sevtsuk et al. (2020a) found all variables to be significantly associated with pedestrian path choices in San

⁵ We leave out highway exposure and public art variables from Sevtsuk et al. (2020a) that were found to be insignificant.

Francisco with expected signs (Table 1).⁶ Appendix 1 shows how each of the variables is distributed within the choice sets available between each location pair. While streets in downtown San Francisco are generally highly walkable, the coefficients of variation (CV) illustrate that individual route attributes vary 20% on average across all variable, and over 80% for some attributes at specific location-pairs.

Multiplying the coefficients from Table 1 with the corresponding qualities of each trajectory between our ten origin-destination pairs, and summing across all nine variables produces an estimated utility for each route. Similar to the “distance-weighted” probability above, these utilities can then be used to assign a probability to each route using Equation 1, with the difference that this time the probabilities are determined by nine different route attributes from Table 1, rather than distance alone. This produces a “utility-weighted probability” for each route between an origin-destination pair (Figure 2e).

Finally, the “highest utility” approach only chooses the route with the highest utility score—the route with most positive pedestrian qualities (Figure 2f). This is conceptually similar to the “shortest route”, with the difference that utility maximization is not only performed on distance alone, but on the nine different route attributes simultaneously.

An attractive feature of the first three prediction approaches (Figure 2 b, c, d) is that they rely on relative route length as the sole attribute for probability assignment and can thus be predicted for any origin-destination pair without requiring detailed data on street quality attributes. From a practical perspective, such approaches favor a broader adoption of pedestrian flow prediction in practice—they can be readily applied in existing built environments, where detailed route characteristics have not been measured or where behavioral response coefficients remain unknown. They can also be used in newly planned developments, where detailed sidewalk qualities are yet to be determined. The “utility-weighted” probability and the “highest utility” approaches, on the other hand, are based on empirically calibrated route attributes and are therefore expected to yield more accurate results. A key question we try to address here is how much more accurate? Considerably more work is involved in estimating utility-weighted

⁶ These coefficients do not include variables describing the number of road crossings or signalized delays, which could also impact route choice and would be desirable to add in future work.

coefficients and generating route probabilities based on such coefficients. Does a simpler distance-weighted approach provide a reasonable, second-best alternative to deploy in practice?

To answer these questions, we compared observed pedestrian flows from GPS traces between the ten origin-destination pairs to each of the five prediction approaches across all ten node pairs. This comparative analysis took two forms. First, we compared the results at the street segment level. This is analogous to comparing the predicted flows against pedestrian counts collected on each street segment. Second, we also performed a comparison for whole routes (instead of individual street segments), examining how the predictions matched the entire sequence of segments that pedestrians took between each origin and destination according to their GPS traces. Given different urban forms between the ten origin-destination pairs (Figure 1) the number of unique trajectories available within a 20% detour varied considerably among locations. Between the intersections C-C', for instance, only four unique routes are found, while between intersections I-I', twenty-two unique routes are found using the same detour constraint (Figure 1).

Results

Table 2 presents the results for both segment-level analysis and route-level analysis. It shows the root mean square errors (RMSE) for predicted models compared to observed flows, where smaller RMSE denotes a better fit with actual data. The results are presented for each of the intersection pairs separately, as well as jointly below, where the mean RMSE across all node-pairs is reported. Cells in Table 2 are shaded horizontally across different prediction methods using darker tones to denote a lower RMSE and better fit. A mean pseudo R^2 value, calculated by comparing each model's RMSE to a similar prediction with random values is also presented at the bottom of both tables⁷

The segment-level analysis shows a fairly poor fit for the shortest-path approach (mean RMSE = 0.417, mean pseudo R^2 of 0.023). For nine out of ten node-pairs, pedestrian flow predictions that assumed people only use a single shortest route performed the worst. The only exception was found between intersections J-J', where the shortest path prediction actually matched the actual

⁷ Pseudo $R^2 = 1 - \text{RMSE}(\text{prediction}) / \text{RMSE}(\text{random})$.

route distribution the best. Upon checking our data and juxtaposing the timestamps from GPS walking trajectories with historic Google Street View images, we discovered that there were two large construction projects underway in 2014-15 along the routes between nodes J-J', obstructing pedestrian access on Freemont St, Baele St and Mission St (Figures 1, 3). The Sales Force East tower along with the main Sales Force tower and the Transbay Transit Center were being built on the corner of Freemont St and Mission St. Additionally, "300 Mission St" was being renovated on the corner of Mission St and Baele St. Google Street View images from the time show sidewalk closures around both locations, making walking along Market St and Main St a far superior option. Accordingly, GPS traces show that 127 out of 139 pedestrians took Market St and Main St, avoiding the disruptions along Freemont, Baele and Mission streets. This explains why the "shortest-path" assumption fit the data best here—other route options were temporarily unavailable. At other node pairs and under normal circumstances, pedestrians exhibited greater variability in route choice. We conclude from this that modelling pedestrian flows along shortest paths alone does not reflect observed pedestrian flows accurately.

The equal probability approach, which assumes that people choose indiscriminately between all routes that are up to 20% longer than the shortest route produces a significantly better fit at the segment level (mean RMSE = 0.215, mean pseudo R^2 of 0.496). The fit improves slightly further with a "distance-weighted" approach, where the same set of +20% detour routes are used, but shorter route options are assigned proportionately higher likelihoods than longer route options (mean RMSE = 0.210, mean pseudo R^2 of 0.507).

What about the "utility-weighted" prediction? We find an improvement over the distance-weighted approach, but the improvement is relatively slim—accounting for route attributes reduces the segment-level RMSE by 2.4% on average and improves the pseudo r-squared fit by 2.2% on average, across ten different intersection pairs (mean RMSE = 0.205, mean pseudo R^2 of 0.518). In five out of ten cases the utility-based approach yields the most accurate segment-level flow prediction. But in four out of ten cases, a slightly better fit is found with the distance-weighted model.

The "highest-utility" route prediction in the final column on the right indicates as poor of a fit as the "shortest path" approach on average (mean RMSE = 0.436, pseudo R^2 of -0.022). The one

exception here is noted at node pairs A-A', where the "highest-utility" flows are second-best to the "utility-weighted" approach. The "highest-utility" path at location A-A' includes an unusually pedestrian friendly environment along Powell St, where no car traffic is allowed and some of the city's most pedestrian-friendly urban design interventions have been implemented. We discuss the trip distribution between intersections A-A' further below.

SEGMENT ANALYSIS

		RMSE	RMSE	RMSE	RMSE	RMSE
Intersection pair	# unique segments	Shortest path	Equal probability	Distance weighted	Utility weighted	Highest utility
A-A'	23	0.577	0.319	0.316	0.126	0.169
B-B'	14	0.593	0.229	0.262	0.061	0.281
C-C'	16	0.544	0.208	0.207	0.109	0.320
D-D'	31	0.290	0.166	0.162	0.160	0.428
E-E'	30	0.343	0.109	0.109	0.117	0.446
F-F'	26	0.380	0.135	0.126	0.086	0.375
G-G'	21	0.438	0.166	0.145	0.163	0.438
H-H'	23	0.302	0.200	0.178	0.183	0.403
I-I'	31	0.433	0.262	0.249	0.289	0.515
J-J'	16	0.250	0.314	0.316	0.489	0.761
Mean RMSE		0.417	0.215	0.210	0.205	0.436
Mean pseudo R ²		0.023	0.496	0.507	0.518	-0.022

n=231 segments

ROUTE ANALYSIS

		RMSE	RMSE	RMSE	RMSE	RMSE
Intersection pair	# unique paths	Shortest path	Equal probability	Distance weighted	Utility weighted	Highest utility
A-A'	11	0.369	0.198	0.197	0.090	0.093
B-B'	5	0.452	0.178	0.189	0.060	0.232
C-C'	4	0.483	0.166	0.165	0.106	0.329
D-D'	12	0.179	0.132	0.130	0.113	0.255
E-E'	8	0.373	0.127	0.127	0.121	0.322
F-F'	15	0.239	0.069	0.065	0.043	0.202
G-G'	9	0.303	0.100	0.093	0.096	0.303
H-H'	10	0.230	0.107	0.099	0.105	0.327
I-I'	22	0.213	0.100	0.097	0.112	0.240
J-J'	8	0.250	0.178	0.175	0.243	0.402
Mean RMSE		0.289	0.131	0.129	0.116	0.268
Mean pseudo R ²		0.461	0.755	0.759	0.784	0.499

n=104 paths

Table 2. Segment-level (top) and route-level (bottom) analysis results.

Figure 3. Google Street View images depicting the construction of Sales Force Tower on Fremont and Mission between 2014 and 2015.

Route-level analysis in the bottom half of Table 2 suggests similar results—the “highest-utility” route alone does not indicate a good fit with observed routes (mean RMSE 0.268, mean pseudo R^2 of 0.499). The “shortest-path” prediction is equally low (mean RMSE 0.289, mean pseudo R^2 of 0.461). The “utility-weighted” probability approach provides the most accurate explanation of routes chosen (mean RMSE 0.116, mean pseudo R^2 of 0.784). Among five out of ten node pairs, the “utility-weighted” probability approach shows the best fit to observed trajectories. However, the “distance-weighted” approach and the “equal-probability” approach are only 3.2% and 3.7% worse in terms of pseudo R^2 values (mean pseudo R^2 of 0.759 and 0.755 respectively). Among four out of ten node-pairs, the distance-weighted approach actually outperforms the utility-weighted approach.

At node-pair J-J', where we found the construction disruptions above, the best route-level fit is achieved not by the shortest path approach, but the distance-weighted approach (Figure 1). Even though the most trafficked individual segments were used by all routes that followed Market St and Main St—outweighing segment-level analysis in favor of the short path approach—some diversity of choice is still seen at the full route level analysis.

Overall, both segment- and route-level analysis show that singular shortest-path and highest-utility predictions performed the poorest, corroborating a key principle of discrete choice theory—different people have different route preferences. The differences between equal-probability, distance-weighted and utility-weighted predictions are almost negligible—the pairwise t-tests showed that errors among the three methods are statistically not significantly different from each other. While “utility-weighted” predictions were most accurate on average, the far simpler distance-weighted and equal probability predictions match the observed flows in the context of downtown San Francisco almost as well.

Theoretically, we should expect distance-based approaches (“equal probability” and “distance-weighted” probability) to perform worse than the “utility-weighted” approach when street attributes in the choice set vary significantly. Since, by definition, the “equal-probability” and “distance-weighted” approaches ignore all street attributes beyond distance, they should under-predict pedestrian flows on streets with more pedestrian-friendly attributes (higher utility) and over-predict pedestrian flows on streets with inferior pedestrian qualities (lower utility).

To check this hypothesis, we produced scatter-plots between actual (y-axis) and predicted (x-axis) values using the distance-weighted probability approach. Figure 4 shows the plots for node-pair A-A', capturing trips between the Market St and Powell St intersection and the Geary St and Mason St intersection (Figure 1). The diagonal trend line in Figure 4 is drawn at a 45 -degree angle according to axis scales, indicating a theoretically perfect match between actual and predicted values. Segments and routes that fall below the trend line are underpredicted, while segments and routes above the trend line are overpredicted. We also divided the segments and routes into three groups: high pedestrian quality segments (75th %-ile utility or above), shown in green, low pedestrian quality segments (25th %-ile utility or below), shown in red, and average quality segments that fall in between, shown without a fill color. Note that the notion of “quality”

here is not determined *a priori*, but rather relative to available route alternatives between the same node pair. Similar scatter-plots for all node pairs are provided in the Appendix 2.

Figure 4. Scatter plots showing actual pedestrian activity (y-axes) and predicted activity (y-axis) between nodes A-A' for both segments (left) and routes (right). Red dots indicate low-utility, green dots high-utility segments and routes.

As expected, the distance-weighted approach generally underpredicts foot-traffic on relatively high-quality streets and overpredicts footfall on low-quality streets. However, both Figure 4 and the related Appendix 2 also suggests a few exceptions, where some high-utility streets too are over-predicted or low-utility streets underpredicted. These likely result from both missing variables in the utility function as well as idiosyncrasies in people's behavior captured in the model's error term.

Figure 5 illustrates the most underpredicted (left) and overpredicted (right) street segments for the same routes between intersections A-A' by the “distance-weighted” prediction model. The most underpredicted segment is on Powell St between Market St and Ellis St— the distance-weighted model predicts fewer users there than actually observed. This is a tree-lined, car-free street with ample sidewalks and a wide selection of amenities along its edges—qualities that contribute to its high predicted utility value (See Figure 5 left and Table 1). The most overpredicted segment by the “distance-weighted” model for this node pair was Cyril Magnin St between Market St and Ellis St (right in Figure 5), which features a wider vehicular road,

relatively narrow sidewalks and minimal amenities along street edges. The monotonous façade of the Hilton hotel takes up the whole length of the block on the left side of the photo.

We find a similar outcome at the route level (Figure 4 right). The most under-predicted route (lone green dot far above the trend line) follows Powell St and Geary St, which includes some of the most outstanding pedestrianization interventions in the city—parklets replacing on-street parking spaces, continuous active commercial building-fronts, generous sidewalks, reduced traffic and trees along the way. This route has the highest predicted utility among 11 available route alternatives between intersections A-A’, according to coefficients in Table 1. The absence of any quality information beyond distance, makes its positive pedestrian appeal invisible to a “distance-weighted” prediction and thus produces a clear underprediction.

Figure 5. Left: Powell St is the most underpredicted street segment between intersections A-A’. Right: Cyril Magnin St is the most overpredicted street segment between intersections A-A’.

Discussion

We have compared five different pedestrian flow prediction approaches to observed pedestrian flows based on GPS walking traces in downtown San Francisco. Our findings show that while the “utility-weighted” prediction model slightly outperformed the significantly simpler “distance-weighted” and “equal probability” prediction methods, the differences were minor and

statistically not significantly different. In the case of the “utility-weighted” method, the slightly higher fit with actual data is likely attributable to the inclusion of detailed route attributes and empirically specified behavioral response coefficients. On average, the “distance-weighted” and “equal probability” approaches were only 2% and 4% less accurate than the utility-weighted approach respectively. At the route-level, the utility-weighted approach produced a pseudo R^2 of 0.784, compared to a distance-weighted R^2 of 0.759 and an equal-probability R^2 of 0.755. It is possible that the differences in means could become significant if a much large sample of segments and routes were used, but the accuracy levels would likely still remain close under similar environmental conditions. We thus conclude that both the “distance-weighted” probability assignment and “equal probability” assignment provide reasonable alternatives for pedestrian flow prediction. When investing significant efforts to modeling route choice according to specific route qualities and parameters is not feasible, these two considerably simpler models can be used to predict pedestrian flow.

The “shortest-path” and “highest-utility” path approaches performed significantly worse. A key advantage of all three best performing methods is they all assume that pedestrians may choose among a number of different route options (up to 20% longer than the shortest available path in this case). Rather than favoring a single best route, this produces a probability distribution across multiple routes—a feature that is clearly important among observed GPS traces as well. Even though the distance-based probability approaches ignore utility attributes beyond distance, they too rely on probabilistic sample enumeration, thereby assigning weighted probabilities to a range of different route options, not just the shortest path.

However, using a “distance-weighted” or “equal probability” assignment with reasonable detours for pedestrian flow prediction does not come without hazards and shortcomings. The approach is likely to produce larger inaccuracies in contexts, where available pedestrian routes have more uneven quality characteristics—some routes with longer distance penalties, some with more ground-floor amenities, some with more trees, some with notably higher traffic levels than others and so on. In the context of downtown San Francisco, where our data was modeled, such quality differences were generally not drastic, albeit still notable in our analysis. In other contexts, the perceived quality range among available route alternatives can be much wider and thus produce greater errors for a “distance-weighted” model. Negative pedestrian attributes, such as crossing

heavy traffic arteries, traversing polluted or damaged streets, or passing through streets with high crime rates could lead the equal-probability and distance-weighted methods to greater levels of over-prediction than observed in our study. Positive pedestrian route attributes, such as walking along car-free pedestrian or bike trails, passing through a main street or a scenic boardwalk could conversely lead to higher levels of underprediction than we found. Therefore, in urban environments, where pedestrian qualities vary starkly from one street to another—i.e. a vibrant main street surrounded by quiet suburban streets; a pleasant community path winding through an otherwise highly trafficked industrial district—a utility-weighted model will likely perform significantly better than a distance-weighted model. The difference could be stronger in favor of a utility model if the users are well familiar with the area and know all the choices from experience. Also, the fewer route alternatives there are (i.e. less connected street networks), the more the differences between routes are likely to matter, favoring a utility-weighted distribution.

Yet in a few cases, we also observed a “distance-weighted” approach producing a better fit with actual data than a “utility-weighted” approach. How do we explain that? This is likely a result of both missing variables in our utility function, as well as behavioral idiosyncrasy in pedestrian path choices. First, our model included nine route quality variables, based on a prior study by Sevtsuk et al. (2020). The actual set of route quality attributes that impact the perceived utility of pedestrian routes could potentially be much longer. We did not control, for instance, for additional fixed effects like street crossings, pavement quality, the presence of street furniture, lighting, or signal delays which other studies have found to influence pedestrian route choice (Broach and Dill 2015; Erath et al. 2015). Perhaps more importantly, we also did not control for dynamic route qualities, such as crowdedness, weather conditions, daylight, noise, or the presence of parked vehicles on the street. Each of these missing variables could alter perceived route utilities, partly explaining why our “utility-weighted” approach can at times fail to detect most preferred routes. Rarely can a utility-weighted model include all of these variables. Given that “distance-weighted” and “equal-probability” predictions are oblivious to both the inclusion and omission of such variables, can in this case work in their favor. Nevertheless, it would be interesting to test some of these additional variables (e.g. number of crossings, signal delays) in a path utility function in future research.

While the utility-weighted model was calibrated on a large dataset of actual walking trajectories, its application at a concrete location can mask a significant error margin, depending on how closely the specific site users behave like “average” users in the city, on whom the utility coefficients were calibrated. Such idiosyncrasies in route preferences, can also produce an advantage for a distance-weighted model, which assigns a narrower range of probabilities to all routes within the specified detour ratio. If downtown streets attract a significantly wider range of demographic users than an “average” pedestrian at the city-wide level, we would expect more deviations from utility-weighted predictions. Distance-weighted models are less sensitive to such deviations.

Future work could examine other route-choice prediction methods that are still based on route geometries alone. While we have focused on equal-probability and distance-weighted methods, these could potentially be combined with other geometry attributes, such as number of turns or node crossings, which also do not require detailed site data, to test more diverse, yet scalable routing alternatives.

The findings of this study have implications for both transportation modeling, urban design and planning practice. Simplified, distance-based pedestrian flow predictions can be used to support decision-making and investments into pedestrian spaces in both existing built environments and in newly planned areas. Municipal neighborhood plans and business-district plans can benefit from simplified pedestrian flow models to identify where investment into better sidewalks, public spaces, street-furniture or landscaping could benefit most users⁸. A better understanding of dominant pedestrian flows, and how newly proposed projects can alter these, can also inform zoning for commercial ground floor uses for shops, services, restaurants and amenities. A comparison of estimated pedestrian flows before and after proposed infrastructure investments (e.g. safe crossings, complete street designs) could support cost-benefit analyses prepared by transportation consultants for public funding applications. A comparison of pedestrian flow estimates between alternative development scenarios (e.g. urban design competition entries)

⁸ This can be readily performed with the “Betweenness” tool in the Urban Network Analysis toolbox software plugin for Rhinoceros 3D (Sevtsuk 2018b).

could enable stakeholders to better understand how proposed projects affect, and desirable benefit, walking activity on city streets.

More generally, access to software-assisted and relatively easy-to-use pedestrian flow models can also enable municipal authorities to implement pedestrian impact assessments, similar to traffic impact assessments that are required from developers in cities around the world. This will require combining pedestrian flow predictions with activity scheduling, destination choice and mode choice models. A simpler approach to modelling pedestrian route choice in the broader context of activity modeling can be a step towards making pedestrian impact assessments more feasible and more commonly used in planning and transportation practice.

Appendix (1)

Descriptive statistics of route attributes for each of the ten choice-sets among locations A-A' to J-J'.

Location pair	Length	Turns	ZGain	Sdwlk W	GVI	SVF	Trfc Speed	Trfc Vol
A-A' (n=87)								
Min	438.7	2.0	13.1	13.2	0.003	142.76	21.6	3237.8
Max	509.5	7.0	17.4	15.7	0.033	0.43	24.2	7117.4
Mean	458.2	4.7	14.9	14.6	0.013	0.38	23.1	5270.8
Std. Dev.	20.2	1.6	2.0	0.9	0.008	0.04	0.8	1018.7
CV (Std Dev / Mean)	4%	34%	13%	6%	60%	11%	3%	19%
B-B' (n=27)								
Min	391.9	3.0	7.3	11.7	0.028	183.57	20.5	2828.6
Max	488.3	5.0	9.7	15.6	0.045	0.48	23.2	7262.5
Mean	451.4	4.3	8.0	13.5	0.037	0.46	21.9	5343.0
Std. Dev.	34.1	1.0	0.9	1.3	0.005	0.01	1.0	1660.6
CV (Std Dev / Mean)	8%	23%	11%	9%	15%	3%	5%	31%
C-C' (n=22)								
Min	389.7	2.0	1.6	7.6	0.002	128.46	22.6	3198.4
Max	392.5	4.0	5.4	9.9	0.041	0.39	25.0	5261.1
Mean	391.0	2.7	3.6	8.8	0.011	0.37	23.5	3806.1
Std. Dev.	1.2	1.0	1.4	0.7	0.014	0.02	0.9	790.0
CV (Std Dev / Mean)	0%	35%	39%	8%	125%	6%	4%	21%
D-D' (n=102)								
Min	440.9	2.0	1.6	7.2	0.000	105.00	19.6	4619.5
Max	505.2	4.0	3.6	15.3	0.025	0.40	25.0	9748.6
Mean	457.5	3.0	2.8	10.5	0.020	0.33	23.1	7055.8
Std. Dev.	17.9	1.0	0.7	2.6	0.006	0.05	1.9	1765.8
CV (Std Dev / Mean)	4%	34%	26%	25%	31%	14%	8%	25%
E-E' (n=76)								
Min	469.7	2.0	1.1	6.0	0.004	115.04	24.1	4462.6
Max	487.7	2.0	1.2	14.0	0.036	0.39	25.0	6544.5
Mean	475.0	2.0	1.1	9.8	0.016	0.34	24.5	5361.2
Std. Dev.	6.4	0.0	0.0	2.6	0.009	0.05	0.4	756.9
CV (Std Dev / Mean)	1%	0%	3%	26%	58%	14%	2%	14%
F-F' (n=112)								
Min	463.0	2.0	5.0	10.0	0.012	104.20	22.7	3273.4
Max	529.9	6.0	9.4	11.4	0.053	0.32	25.0	8841.8
Mean	500.9	4.6	7.4	10.6	0.031	0.28	24.0	6094.6
Std. Dev.	29.6	1.6	1.3	0.5	0.012	0.04	0.7	1536.2
CV (Std Dev / Mean)	6%	34%	17%	4%	39%	14%	3%	25%
G-G' (n=64)								
Min	462.9	2.0	0.0	9.8	0.000	94.14	23.0	1892.1

Max	525.4	8.0	0.6	11.4	0.027	0.24	25.0	8428.9
Mean	497.8	4.4	0.3	10.8	0.015	0.21	24.2	5520.0
Std. Dev.	30.4	2.1	0.3	0.5	0.012	0.02	0.7	1466.7
CV (Std Dev / Mean)	6%	48%	88%	5%	81%	10%	3%	27%
H-H' (n=86)								
Min	370.5	3.0	3.5	12.0	0.000	90.19	23.7	2545.7
Max	442.5	8.0	4.3	18.7	0.032	0.27	25.0	3742.5
Mean	414.9	4.9	3.9	17.3	0.025	0.25	24.5	2878.8
Std. Dev.	25.6	1.6	0.3	2.0	0.008	0.01	0.5	351.4
CV (Std Dev / Mean)	6%	34%	7%	11%	34%	5%	2%	12%
I-I' (n=196)								
Min	523.4	2.0	0.6	8.4	0.002	116.76	23.0	2761.4
Max	626.4	4.0	1.3	14.0	0.064	0.34	25.0	7005.8
Mean	577.7	3.7	1.0	11.5	0.013	0.26	24.3	5318.5
Std. Dev.	44.1	0.7	0.2	1.8	0.012	0.03	0.7	1067.6
CV (Std Dev / Mean)	8%	19%	19%	16%	88%	12%	3%	20%
J-J' (n=52)								
Min	417.7	1.0	0.1	9.0	0.016	126.81	24.4	1869.8
Max	479.3	7.0	0.2	16.1	0.029	0.35	25.0	3856.3
Mean	434.6	4.4	0.2	14.8	0.023	0.32	24.7	2565.4
Std. Dev.	21.6	1.6	0.0	1.7	0.004	0.02	0.3	805.0
CV (Std Dev / Mean)	5%	37%	18%	12%	19%	5%	1%	31%

Appendix (2)

Scatter plots showing actual pedestrian activity (y-axes) and predicted activity (y-axis) between nodes all node pairs for both segments (left) and routes (right). Red dots indicate low-utility, green dots high-utility segments and routes.

References

- Adamek, K., Vasan, N., Elshaer, A., English, E., & Bitsuamlak, G. (2017). Pedestrian level wind assessment through city development: A study of the financial district in Toronto. *Sustainable Cities and Society*, 35, 178–190.
- Batty, M. (2001). Editorial: agent based pedestrian modeling. *Environment and Planning B: Planning and Design*, volume 28, pages 321–326.
- Borgers, A., & Timmermans, H. (1986). A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis*, 18(2), 115–128.
- Bosselmann, P., Arens, E., Dunker, K., & Wright, R. (1995). Urban Form and Climate: Case Study, Toronto. *Journal of the American Planning Association*, 61(2), 226–239.
- 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.
- Broach, J., & Dill, J. (2016). Using Predicted Bicyclist and Pedestrian Route Choice to Enhance Mode Choice Models. *Transportation Research Record*, 2564(1), 52–59. <https://doi.org/10.3141/2564-06>
- Conroy-Dalton, R. (2001). *Spatial Navigation in immersive virtual environments: Vol. PhD*. University College London.
- Cooper, C. H. V., & Chiaradia, A. J. F. (2020). sDNA: 3-d spatial network analysis for GIS, CAD, Command Line & Python. *SoftwareX*, 12, 100525. <https://doi.org/https://doi.org/10.1016/j.softx.2020.100525>
- Cooper, C. H. V., Harvey, I., Orford, S., & Chiaradia, A. J. F. (2019). Using multiple hybrid spatial design network analysis to predict longitudinal effect of a major city centre redevelopment on pedestrian flows. *Transportation*.
- Coutrot, A., Manley, E., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2020). Cities have a negative impact on navigation ability: evidence from 38 countries. *BioRxiv*, 2020.01.23.917211. <https://doi.org/10.1101/2020.01.23.917211>
- Coutrot, A., Schmidt, S., Coutrot, L., Pittman, J., Hong, L., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *BioRxiv*, 305433. <https://doi.org/10.1101/305433>
- Dogan, T., Samaranayake, S., & Saraf, N. (2018). Urbano- A New Tool to Promote Mobility-Aware Urban Design, Active Transportation Modeling and Access Analysis for Amenities and Public Transport. *Proceedings of SimAUD 2018*.

- Erath, Alexander L.; van Eggermond, Michael A.B.; Ordóñez Medina, Sergio A.; Axhausen, K. W. (2015). *Modelling for walkability Understanding pedestrians' preferences in Singapore*. Singapore. Retrieved from
- Ewing, R., & Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *Journal of Urban Design*, 14(1), 65–84.
- Ewing, R., Hajrasouliha, A., Neckerman, K. M., Purciel-Hill, M., & Greene, W. (2016). Streetscape Features Related to Pedestrian Activity. *Journal of Planning Education and Research*, Vol. 36((1)), 5–15.
- Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and Measuring Urban Design Qualities Related to Walkability. *Journal of Physical Activity and Health*, 3(s1), S223–S240.
- Garbrecht, D. (1970). Frequency Distribution of Pedestrians in a Rectangular Grid. *Journal of Transport Economics and Policy*, 4(1), 66–88.
- 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. (2009). Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transportation Research Part D*, 14(5), 343–352.
- 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.
<https://doi.org/https://doi.org/10.1016/j.jtrangeo.2012.11.013>
- Hillier, B., & Iida, S. (2005). Network and Psychological Effects in Urban Movement. *International Conference on Spatial Information Theory COSIT 2005.*, 475–490.
- Kerridge, J., Hine, J., & Wigan, M. (2001). Agent-Based Modelling of Pedestrian Movements: The Questions That Need to Be Asked and Answered. *Environment and Planning B: Planning and Design*, 28(3), 327–341. <https://doi.org/10.1068/b2696>
- 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.
- Kim, H., & Macdonald, E. (2015). Wind and the city: An evaluation of San Francisco's planning approach since 1985. *Environment and Planning B: Urban Analytics and City Science*, 44(1), 10–32.

- Lee, C., & Moudon, A. V. (2006). The 3Ds+R: Quantifying land use and urban form correlates of walking. *Transportation Research Part D: Transport and Environment*, 11(3), 204–215.
- Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675–685.
- Liebig, T., Xu, Z., May, M., & Wrobel, S. (2012). Pedestrian Quantity Estimation with Trajectory Patterns. In P. A. Flach, T. De Bie, & N. Cristianini (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 629–643). Springer Berlin Heidelberg.
- Liu, S., Lo, S., Ma, J., & Wang, W. (2014). An Agent-Based Microscopic Pedestrian Flow Simulation Model for Pedestrian Traffic Problems. *IEEE Transactions on Intelligent Transportation Systems*, 15(3), 992–1001.
- Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and Society*, 14, 34–42.
- Lynch, K. (1960). *The image of the city*. Technology Press.
- McFadden, D. (1980). Econometric Models for Probabilistic Choice among Products. *Journal of Business*, 53(3).
- Mohsenin, M., & Sevtsuk, A. (2013). The impact of street properties on cognitive maps. *Journal of Architecture and Urbanism, Volume 37*(Issue 4), 301–309.
- Montello, D. R., & Sas, C. (2006). *Human factors of wayfinding in navigation* (W. Karwowski (ed.); 2nd ed., pp. 2003–2008). Taylor & Francis.
- Mehta, V. (2008). Walkable streets: pedestrian behavior, perceptions and attitudes. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1(3), 217–245.
- 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.

- Porta, S., Crucitti, P., & Latora, V. (2005). The network analysis of urban streets: a primal approach. *Environment and Planning B*, 35(5), 705–725.
- Pushkarev, B., & Zupan, J. (1975). *Urban Space for Pedestrians*. MIT Press.
- Puusepp, R., Lööke, T., Cerrone, D., & Männigo, K. (2017). Simulating Pedestrian Movement. In *Humanizing Digital Reality* (p. pp 547-557).
- Sevtsuk, A., & Mekonnen, M. (2012). Urban Network Analysis Toolbox. *International Journal of Geomatics and Spatial Analysis*, 22(2), 287–305.
- Sevtsuk, A. (2018a). Estimating Pedestrian Flow on Street Networks: Revisiting the Betweenness Metric. Paper presented at the American Association of Geographers 2018 conference in New Orleans, LA.
- Sevtsuk, A. (2018b). *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>
- Sevtsuk, A., Li, X., Basu, R., Kalvo, R. (2020 a). A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. Upcoming.
- Sevtsuk, A. (2020b). Estimating pedestrian flows on street networks: revisiting the betweenness index. Upcoming.
- Sevtsuk, A., Basu, R., & Chancey, B. (2020 c). We shape our buildings, but do they then shape us? A longitudinal analysis of pedestrian flows and development activity in Melbourne. *Transportation Research Board 2021*, 1–20.
- Sharmin, S., & Kamruzzaman, M. (2018). Meta-analysis of the relationships between space syntax measures and pedestrian movement. *Transport Reviews*, 38(4), 524–550.
- Shatu, F., Yigitcanlar, T., & Bunker, J. (2019). Shortest path distance vs. least directional change: Empirical testing of space syntax and geographic theories concerning pedestrian route choice behaviour. *Journal of Transport Geography*, 74, 37–52.
- Train, K. (2002). *Discrete Choice Methods with Simulation*. Cambridge University Press. 378 p.
- Tribby P. C., Miller J. H., Brown B. B., Werner M. C. (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.
- Turner, A. (2001). Depthmap: A Program to Perform Visibility Graph Analysis. *3rd International Symposium on Space Syntax. Georgia Institute of Technology, 7-11 May 2001.*

Zacharias, J., Stathopoulos, T., & Wu, H. (2001). Microclimate and Downtown Open Space Activity. *Environment and Behavior*, 33(2), 296–315.