DOI: 10.1111/tgis.12624

RESEARCH ARTICLE

Transactions in GIS WILEY

Calibrating the dynamic Huff model for business analysis using location big data

Yunlei Liang[1] | Song Gao[1] | Yuxin Cai[1] | Natasha Zhang Foutz[2] | Lei Wu[3]

[1]GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

[2]School of of Commerce, University of Virginia, Charlottesville, VA, USA

[3]Data Science, WeWork, Palo Alto, CA, USA

Correspondence
Song Gao, GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI, USA.
Email: song.gao@wisc.edu

Abstract

The Huff model has been widely used in location-based business analysis to delineate a trade area containing a store's potential customers. Calibrating the Huff model and its extensions requires empirical location visit data. Many studies rely on labor-intensive surveys. With the increasing availability of mobile devices, users in location-based platforms share rich multimedia information about their locations at a fine spatio-temporal resolution, which offers opportunities for business intelligence. In this research, we present a time-aware dynamic Huff model (T-Huff) for location-based market share analysis and calibrate this model using large-scale store visit patterns based on mobile phone location data across the 10 most populated US cities. By comparing the hourly visit patterns of two types of stores, we demonstrate that the calibrated T-Huff model is more accurate than the original Huff model in predicting the market share of different types of business (e.g., supermarkets versus department stores) over time. We also identify the regional variability where people in large metropolitan areas with a well-developed transit system show less sensitivity to long-distance visits. In addition, several socioeconomic and demographic factors (e.g., median household income) that potentially affect people's visit decisions are examined and summarized.

1 | INTRODUCTION

"Location, location, and location!" Location information is a key component in business intelligence and implementation of crucial, revenue-generating marketing strategies, such as location-based advertisement and services (Fan, Lau, & Zhao, 2015; Gao & Mai, 2017; Huang, Gartner, Krisp, Raubal, & Van de Weghe, 2018; Negash & Gray, 2008). With the increasing use of social media, smart devices, and mobile apps, users share rich multimedia information about their locations and associated activities, such as working, shopping, or dining, in a granular spatio-temporal resolution with unprecedented breadth, depth, and scale. Such location-based profiles provide invaluable sources of information for various business analytics and recommendation systems.

While the original Huff model (Huff, 1964) and its subsequent extensions have been used to understand a brand's trade area, they are largely static. The availability of granular spatio-temporal mobility data has permitted the examination of the dynamics of customer mobility patterns at the individual level. For instance, several studies have examined the effects of sampling locations on calibrating the original Huff model to delineate trade areas using mobile phone data (Lu, Shaw, Fang, Zhang, & Yin, 2017) and social media data (Wang, Jiang, Liu, Ye, & Wang, 2016). In addition, at the aggregated level, there are shifts in the dynamic of stores' trade areas, driven by various potential factors such as seasonality, marketing strategies, geo-socioeconomic changes surrounding the stores, and individuals' dynamic behaviors. Predicting where or which type of location an individual would visit is also about when the individual is regarding the temporal dynamics of human mobility patterns (Gao, 2015; McKenzie, Janowicz, Gao, Yang, & Hu, 2015; Tu et al., 2017; Yang et al., 2016; Ye, Janowicz, Mülligann, & Lee, 2011; Yuan & Raubal, 2012), social relations (Liu, Yin, Lu, & Mou, 2020; McKenzie, Janowicz, Gao, & Gong, 2015), and semantic configuration and regional variability for temporal signatures of points of interest (POIs) (Shi, Chi, Liu, & Liu, 2015; Xu, Belyi, Bojic, & Ratti, 2017). Customers may exhibit different temporal visit preferences to different types of stores, resulting in dynamically shifting trade areas for these stores over different time periods. For example, grocery stores usually have more daily visits on weekends than on weekdays. The traditional Huff model can only provide one static estimate for each store, which ignores the potential temporal information. However, the temporal dynamics of POI visits in cities can be more accurately captured by using large-scale mobile phone location tracking data, facilitating the calibration of a "dynamic" Huff model to better represent dynamic trade areas at a more granular temporal scale.

In this research, therefore, we present a time-aware dynamic Huff model (T-Huff) for business location analysis by augmenting the original Huff Model with a dynamic element to capture the time-varying probability of store visitation at the individual customer level. At the aggregated level, the resulting dynamic market share model is calibrated by large-scale store visits based on mobile phone location tracking data. We aim to answer the following two research questions:

1. How accurate is the dynamic Huff model in predicting the market share of different types of business (e.g., supermarkets and department stores) over time?
2. How do spatial and socioeconomic factors determine the customer choice of particular store visits? Is there any regional variability for store visits in different cities?

The contribution of this research is threefold. First, we propose a dynamic Huff model to estimate hourly store visits from a particular neighborhood over time. Second, by using large-scale individual-level POI visit data across the 10 most populated US cities, we calibrate the T-Huff model parameters using the technique of particle swarm optimization (PSO) and find that the T-Huff model outperforms the static Huff model when estimating store temporal visits, although regional variability persists. Third, we demonstrate that various factors, such as distance, neighborhood total population, and socioeconomic variables (e.g., median household income, race, and ethnic diversity), entail distinct influence on store visits across categories and brands.

The remainder of this article is organized as follows. We review the relevant literature in Section 2 before introducing the formulations of the original and dynamic Huff models in Section 3. Then we present the data and study area under analysis in Section 4 and report the key empirical findings of the proposed model for three top chain-store brands across 10 US cities and discuss the broader implications in Section 5. Finally, we draw conclusions and share our vision for future work in Section 6.

2 | LITERATURE REVIEW

There is a rich tradition in the marketing literature of studying store traffic and its driving factors. For example, Hutchinson (1940) used surveys to measure the amount of traffic passing by a Morgantown, WV shoe store and identified 13 factors which could impact sales, including seasonal variations, weather, general business conditions, purchasing power, special location factors, price levels, and competition. Bennett (1944) studied the out-of-town buying habits in a Maryland town located between Washington, DC and Baltimore, and found that in many categories purchases were made out-of-town in Baltimore because the survey respondents preferred the proximity of the stores in Baltimore as compared with their town in terms of shopping convenience. To understand how opening a branch store will impact the parent store's performance, Blankertz (1951) conducted a study revealing that branch and parent stores did not attract separate customer groups; rather, both drew trade from substantially the same group. Nearby customers in the "buffer" area between branch and parent traveled most frequently inward to the downtown shopping center despite the greater travel and time involved.

Huff (1964) defined a trade area as "a geographically delineated region, containing potential customers for whom there exists a probability greater than zero of their purchasing a given class of products or services offered for sale by a particular firm or by a particular agglomeration of firms." Stanley and Sewall (1976) further suggested a series of modifications to the Huff model to evaluate the potential of prospective retail store locations.

This literature further evolved into more sophisticated location analysis, for instance, to advise store site selections. Rosenbloom (1976) reported on the formation and application of a retail strategy matrix that incorporated three relevant factors: a store's geography, consumer demand, and the area's heterogeneity for identifying and selecting new trade areas for retail stores. He also suggested methods that can be used to adjust the merchandise of existing retail outlets to their trade locations. Ghosh and Craig (1983) presented a procedure to help retailers formulate a strategic location plan in a dynamic environment, which involved a model for assessing site desirability, a criterion for selecting among alternative sites, and a heuristic to facilitate the computational procedure. More broadly, Grether (1983) called for more regional-spatial analysis in marketing research.

The development in this area has also propelled methodological innovation. For example, Fotheringham (1988) proposed a competing destinations model to study hierarchical spatial choices of stores and showed its superior performance as compared to other choice models, such as the nested logit model. Donthu and Rust (1989) used kernel density estimation to estimate the spatial distribution of customers in a market and showed how a density-based product positioning methodology may be applied to site selection for a new or relocated store or distribution center. Rust and Donthu (1995) accounted for geographically localized misspecification errors in store choice models with omitted variables that can be correlated with geographic location. They showed that spatial non-stationarity of the model parameters may also be expressed as an instance of omitted variables and therefore be addressed using their method.

The more recent literature in this domain has focused on location-based competition among stores or chains. A positive association between the number of larger stores and the number and size of smaller stores is reported, implying a mutually beneficial relationship among different types of retailers rather than an overwhelming competitive advantage for larger stores (Miller, Reardon, & McCorkle, 1999). Vitorino (2012) used a strategic model of entry to study the store configurations of all US regional shopping centers and to quantify the magnitude of inter-store spillovers. The author showed that, consistent with the agglomeration and clustering theories, firms may have incentives to co-locate despite potential business stealing effects; and that the firms' negative and positive strategic effects help predict both how many firms can operate profitably in a given market and the firm-type configurations. In the context of retail outlet locations in the fast food industry, both McDonald's and Burger King were shown to be better off avoiding close location competition if the market area is large enough; but in small market areas, McDonald's would prefer to be located together with Burger King; in contrast, Burger King's profits always increased with greater differentiation (Thomadsen, 2007). Regarding customers' location awareness, Jiang et al. (2019) calibrated the Huff model with social media data and found that customers far from the existing retail agglomerations might be more sensitive to the distance.

Furthermore, studying price competition among (gasoline) retailers conditional on geographic locations, Chan, Padmanabhan, and Seetharaman (2007) found that consumers were willing to travel up to a mile for a saving of $0.03 per liter. Talukdar (2008) found price differentials between wealthy and poor neighborhoods to be 10–15% for everyday items. Even after controlling for the store size and competition, prices were found to be 2–5% higher in poor areas, which was explained by access to cars that acted as a key determinant of consumers' price search patterns.

In sum, the original Huff model and its subsequent extensions have been widely used to model a brand or a store's trade area and to predict customer visit probability, but they are largely static. Recent research by McKenzie and Adams (2017) demonstrated that thematic regions can be represented dynamically using place-type specific temporal patterns. Customers have different temporal visit preferences for different types of stores. A dynamic model is thus required to better capture the spatio-temporal characteristics of customers' store visit behaviors.

3 | METHODS

3.1 | The original Huff model

The Huff model was introduced in order to provide a probabilistic analysis of shopping center trade area, that is, the region containing potential customers (Huff, 1963, 1964). The identification of a trade area for a store is crucial as the business owner can estimate how many potential customers will visit the store within the region, and therefore can predict the market sales of the store among competing businesses.

The Huff model, which is essentially a gravity-based spatial interaction model, proposes that there are two major factors affecting the number of potential customers of a store. The first is the merchandise offerings, namely, the ability of the store to fulfill the customers' needs (Huff, 1963). This is also called the attractiveness of the store. If a store has a large number of items, it is able to attract more customers even from distant regions. The second factor is the travel time or travel distance to the store. As the cost of travel to a store increases, the willingness to visit the store could be significantly reduced (Huff, 1963).

Based on those two factors, the probability of one customer traveling to a given store can be expressed as

$$P_{ij} = \frac{\frac{S_j^{\alpha}}{D_{ij}^{\beta}}}{\sum_{j=1}^{n}\left(\frac{S_j^{\alpha}}{D_{ij}^{\beta}}\right)} \tag{1}$$

where P_{ij} is the probability of customer i visiting store j; S_j is the attractiveness of store j; D_{ij} is the physical distance between customer i and store j; and n indicates there are n stores that customer i can visit. The parameters α and β are used to reflect the effects of attractiveness and distance on the model.

3.2 | A time-aware dynamic Huff model

Given that people visit different places of interest at different times (McKenzie, Janowicz, Gao, & Gong, 2015; McKenzie, Janowicz, Gao, Yang, & Hu, 2015), we propose the following time-aware dynamic Huff model:

$$P_{ijt} = \frac{\frac{S_j^{\alpha}}{D_{ij}^{\beta}}}{\sum_{j=1}^{n} \left(\frac{S_j^{\alpha}}{D_{ij}^{\beta}} \right)} P_{jt} \tag{2}$$

$$P_{jt} = \frac{V_{jt}}{\sum_{t=1}^{m} V_{jt}} \tag{3}$$

where P_{ijt} is the probability of customer i visiting store j within a temporal window t (e.g., a hour or a day of week); S_j is the attractiveness of store j; D_{ij} is the physical distance between customer i and store j; P_{jt} is the temporal visit probability for one store j within a temporal window t. V_{jt} is the total visit count for one store j within a particular hour t (in this research), and we sum the counts over one week as $\sum_{t=1}^{m} V_{jt}$ (i.e., m = 168 hr). As shown in Figure 1, even for the same chain-store brand (e.g., Whole Foods), the five branch stores in Los Angeles have distinct temporal visit patterns. The parameters α and β are used to reflect the effects of attractiveness and distance on the model.

We also construct another advanced time-aware dynamic Huff (A-Huff) model which estimates the customer visiting probability at each time-stamp by comparing all possible visits the customer may make at the same time-stamp, which considers the business competition from integrated spatial and temporal aspects. The A-Huff model shares the same parameters with the T-Huff model in Equation (2) but with a different formulation as follows:

$$P_{ijt} = \frac{\left(\frac{S_j^{\alpha}}{D_{ij}^{\beta}} \right) P_{jt}}{\sum_{j=1}^{n} \left(\frac{S_j^{\alpha}}{D_{ij}^{\beta}} \right) P_{jt}} \tag{4}$$

$$P_{jt} = \frac{V_{jt}}{\sum_{t=1}^{m} V_{jt}} \tag{5}$$

In addition to the predicted visiting probability P_{ijt} using the A-Huff model, the actual visiting probability P'_{ijt} for this model is calculated using the formula:

$$P'_{ijt} = \frac{V_{ij} P_{jt}}{\sum_{j=1}^{n} V_{ij} P_{jt}} \tag{6}$$

where V_{ij} is the observed pairwise visits from customer i in a specific neighborhood to store j.

3.3 | Parameter calibration using PSO

Before we use the original Huff and the time-aware dynamic models (T-Huff and A-Huff) to make market share predictions, we need to calibrate the models by adjusting their parameters to make sure that the results approximate or reflect reality. Previously the two parameters (α and β) were often decided arbitrarily, which can lead to inaccurate or even erroneous results (Huff, 2003). A few methods have been used to optimize α and β. Many researchers have used ordinary least squares (OLS) method to estimate the parameters by transforming the Huff model into a logarithm-centering format and estimating the parameters using linear regression (Huff & McCallum, 2008; Nakanishi & Cooper, 1974). Geographically weighted regression was also used to calibrate the Huff model, estimating the parameters for every point within the study area (Súarez-Vega, Gutíerrez-Acuña, & Rodríguez-Díaz, 2015). Recent research has applied optimization algorithms such as the PSO technique to find an optimal or

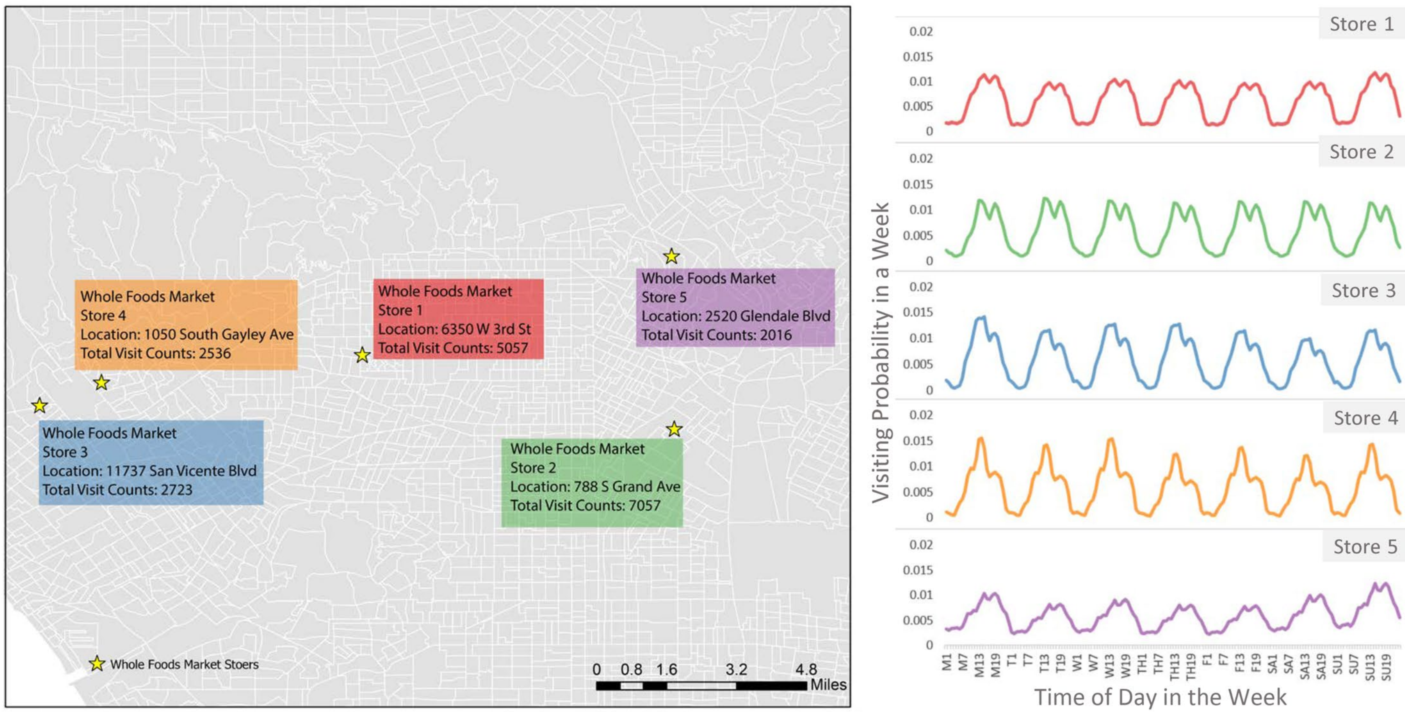


FIGURE 1 Whole Foods Markets in Los Angeles with their temporal visit probability plots

near-optimal solution of the parameters that fit the observation data more accurately (Suhara, Bahrami, Bozkaya, & Pentland, 2019).

In this research, to calibrate both the Huff and the T-Huff model parameters we used the PSO technique, which was introduced by Eberhart and Kennedy (1995), inspired by the foraging behavior of flocks of birds. As a widely used optimization method, PSO makes few or no assumptions (e.g., linearity) about the problem being optimized, so it is appropriate for our problem. Also, we are able to design the objective function based on different needs. In our case, we selected the correlation between the predicted store visit probability and the actual visit probability as the objective function. Compared with the traditional OLS approach, the PSO technique allows more freedom at the optimization design stage and is efficient at finding solutions from a very large candidate solution space, which means that we can try a large number of α and β values and observe the trend of convergence through the optimization process.

To initialize the optimization, a few particles are generated, and each particle represents a potential (α,β) pair. The particles will change their positions (the values of α and β) based on their previous best location and the global best position (Kennedy, 2010; Xiao, Wang, Liu, & Wang, 2013). The particles should then gradually cluster in the area of the optimal solution and return an optimized result. Here the performance of every particle is determined by a pre-defined objective function. The goal of the optimization process is to find the best combination of the parameters that maximize the objective function. The objective function in this study is the Pearson correlation between the estimated probability and the actual probability of pairwise visits from a particular neighborhood to a store. We calibrate the parameters for each specific brand of stores using large-scale anonymous mobile phone location tracking data (in the following section) in order to find the models that can best reflect the particular store visit patterns.

4 | DATA AND STUDY AREA

We collected over 3.6 million POIs with visit patterns in the U.S. from the SafeGraph business venue database (https://www.safegraph.com). The POIs are first classified based on the North American Industry Classification System (NAICS) six-digit sector codes. Among them, we selected two categories of interest: (1) supermarkets and grocery stores (445110); and (2) department stores (452210). There are over 20,000 POIs in total for the two selected categories in the 10 most populated US cities (New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, San Antonio, San Diego, Dallas, and San Jose). In addition to the spatial distribution of the POIs, we also retrieved the fine-resolution visit patterns of all those POIs from the aforementioned SafeGraph database which covers dynamic human mobility patterns of millions of anonymous smartphone users. The SafeGraph data sampling correlated highly with the US Census populations (https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset). These mobile location data consist of "pings" identifying the coordinates of a smartphone at a moment in time. To enhance privacy, SafeGraph excludes census block group (CBG) information if fewer than five devices visited a place in a month from a given CBG. For each POI, the records of aggregated visitor patterns illustrate the number of unique visitors and the number of total visits to each venue during the specified time window (October to December 2018 in our data set), which could reflect the attractiveness of each venue. For example, Figure 2 shows the spatial distributions of CBGs that have visit flows to the five Whole Foods Markets and the 14 Ross Stores in Los Angeles. Furthermore, we computed the average hourly visit probability for each POI over 168 hr (24 hr × 7 days of a week) to show the dynamic visit patterns. For future studies, the hourly visit frequency can also be estimated from other resources, such as the shopper's loyalty card data or the popular times collected by Google Maps or Yelp for business locations. The corresponding demographic and socioeconomic attribute data of all CBGs were collected from the American Community Survey.

FIGURE 2 Spatial distributions of CBGs that have visit flows to: (a) five Whole Foods Markets; and (b) 14 Ross Stores in Los Angeles (note: the number of stores for each brand only reflects the data we have; the geo-visualization is created using the kepler.gl tool)

5 | RESULTS

5.1 | Visit distance distributions

We first analyzed the distribution of the median distance that the visitors traveled from homes to all the stores given a specific NAICS category. The probability density distributions of visit distances across cities showed a variety of heavy-tailed distributions. The mean of the median distance (great circle distance) from the visitors' homes to supermarkets and grocery stores (NAICS: 445110) across these cities is about 7.8 km. However, the median distance distribution does vary over different cities (as shown in Figure 3). Most people in Philadelphia, San Jose, Chicago, and Los Angeles traveled relatively shorter distances (with median 3.8, 4.5, 4.6, and 4.7 km, respectively) than people in other big metropolitan areas in the U.S. such as Dallas and New York (with the largest medians 8.4 and 7.8 km, respectively). As expected, the mean of the median distance from the visitors' homes to department stores (NAICS: 452210) across these cities is about 10.3 km and larger than that to supermarkets and grocery stores.

In addition, the distance decay phenomenon exists in the visit median distance density distribution across all cities (as shown in the log–log plots in Figure 3). The visit probability decreased significantly after about 10 km, which offers insights into location business decision-making. And different cities have varying decay exponents β (Gao, Liu, Wang, & Ma, 2013), which may link to their urban morphology (e.g., size and shape) (Kang, Ma, Tong, & Liu, 2012). The distance decay slopes for supermarkets and grocery stores are steeper than that of department stores in all cities, which demonstrates that there is much less long-distance travel for supermarkets and grocery store visits than for department stores.

5.2 | Huff model calibration for top brands

5.2.1 | Parameter calibration and comparison

Given the variability of store visits for chain-store brands and local brands in our exploratory analysis, we did not calibrate the models for all brands in each POI category. Instead, we only designed comparative experiments for top three chain-store brands with the most stores across the 10 most populated cities in our data set. Take the Whole Foods Markets in Los Angeles as an example: the attractiveness of each Whole Foods store is estimated using the total visit count over the 3 months in the Safegraph data set. Figure 2a shows the flow map from each CBG to the five Whole Foods Markets in the Los Angeles area. It is clear that people in each CBG have a particular store visit preference, and the store visited is usually within a certain spatial proximity to that CBG. The Whole Foods Markets are chain stores that usually have similar product layouts and sizes. Therefore, the major factor affecting the visits of customers is usually the distance from the customer to the store. There are also some other factors. For example, for the two Whole Foods Markets on the left in Figure 2a, we can see a clear delineation of visiting CBGs to the two stores separated by the highway. Even though these two Whole Foods Markets are located closely to each other, they have very distinct visitors due to the infrastructure barrier in that area. Other demographic and socioeconomic factors influencing store visits will be further discussed in Section 5.3.

The model parameter calibration is conducted for each brand of stores in order to find the best (α, β) pair reflecting the effects of attractiveness and distance on the particular brand using observed store visit data. A set of values for α and β is first determined in order to identify a smaller data range for optimization. The results of the correlation for the selected α and β for the original Huff model are shown in Table 1. In general, the model produces very good results, with all Pearson's correlation coefficients >0.6. A higher correlation is obtained with α between 0 and 1 and β between 0 to 2 approximately. Therefore, the bounds for α and β in the PSO optimization are set to between 0 and 2. The optimization is repeated 10 times with 10 particles and is implemented using the Pyswarms open-source library in Python. The highest correlation obtained from the optimization is 0.864 when α= 0.717 and β= 0.805. The α and β values are then fed into the Huff model to estimate the store visit probability.

Also, Table 2 shows the Pearson correlation results with the same selected α, β values using the T-Huff model. In general, the T-Huff model exhibits higher correlations for all selected α and β than the original Huff model, which reflects that the T-Huff model might provide a more accurate estimation of the dynamic visit probability in most cases. The highest correlation obtained from the optimization procedure is 0.890 with α = 0.787 and β = 0.765.

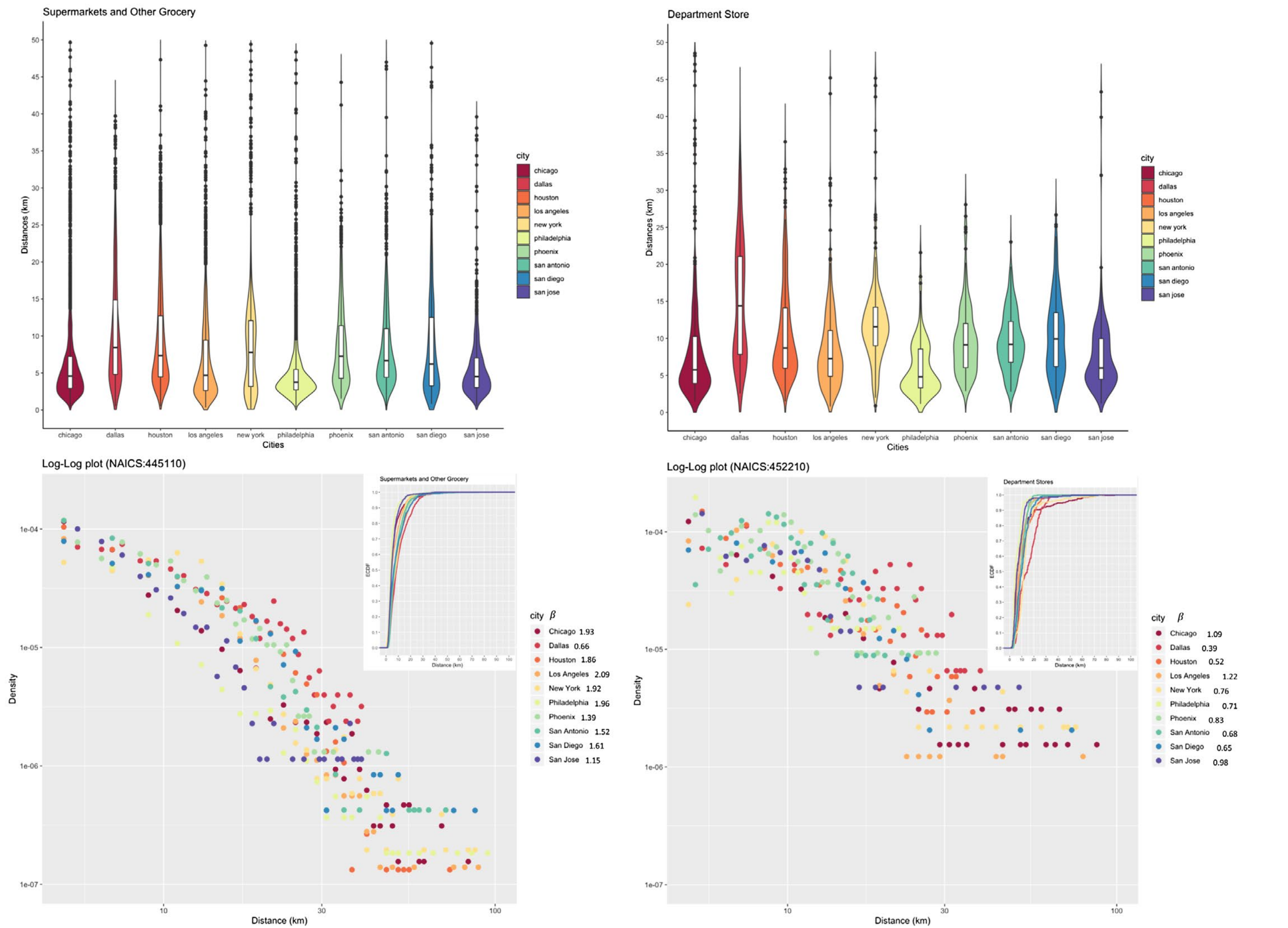


FIGURE 3 Probability density distribution, empirical cumulative distribution, and log–log plots of visitors' distance from home to supermarkets and grocery stores (NAICS: 445110) and to department stores (NAICS: 452210) in the 10 most populated cities in the USA

TABLE 1 Model parameter calibration results with Pearson's correlation for Whole Foods using the original Huff model

	β				
α	**0.1**	**0.5**	**1**	**2**	**5**
0.1	0.807	0.844	0.845	0.817	0.769
0.5	0.808	0.854	0.858	0.825	0.774
1	0.791	0.846	0.862	0.828	0.778
2	0.747	0.797	0.834	0.822	0.776
5	0.683	0.709	0.740	0.773	0.752

TABLE 2 Model parameter calibration results with Pearson's correlation for Whole Foods using the T-Huff model

	β				
α	**0.1**	**0.5**	**1**	**2**	**5**
0.1	0.847	0.874	0.873	0.844	0.791
0.5	0.848	0.882	0.884	0.852	0.796
1	0.835	0.877	0.888	0.855	0.801
2	0.789	0.832	0.861	0.847	0.799
5	0.694	0.716	0.744	0.775	0.761

TABLE 3 Model parameter calibration results with Pearson's correlation for Whole Foods using the M-Huff model

	β				
α	**0.1**	**0.5**	**1**	**2**	**5**
0.1	0.618	0.646	0.647	0.626	0.589
0.5	0.619	0.654	0.657	0.632	0.593
1	0.606	0.648	0.660	0.634	0.596
2	0.572	0.610	0.639	0.629	0.594
5	0.523	0.543	0.566	0.592	0.576

In addition to the original static Huff model and the dynamic Huff models (T-Huff and A-Huff), another model called the M-Huff model is constructed for comparison. The synthetic M-Huff model assumes that the hourly visit probability for one CBG to one store is distributed evenly over the 168 hr in one week (using the mean visit probability) and therefore the model assigns the visit probability equally to each time window (every hour in this study). The correlation is then calculated between this equally distributed visit probability and the actual hourly visit probability from the SafeGraph data set. Table 3 shows the correlation results for selected α, β from the M-Huff model. The highest correlation from the optimization is 0.662 with α = 0.723 and β = 0.806. It is clear that the correlations drop dramatically compared with the results of the original Huff and T-Huff models, which means that the assumed equally distributed hourly visit probability cannot serve as a good representation of the actual dynamic visit patterns. In other words, the store visit patterns do have temporal variation and it is necessary to consider such variation in market-share models.

5.2.2 | Visit spatial pattern comparison

Figure 4 shows two maps of the estimated market share from the original Huff model and the actual market share generated from the SafeGraph POI visit dataset. Here the market share means the probability that people from

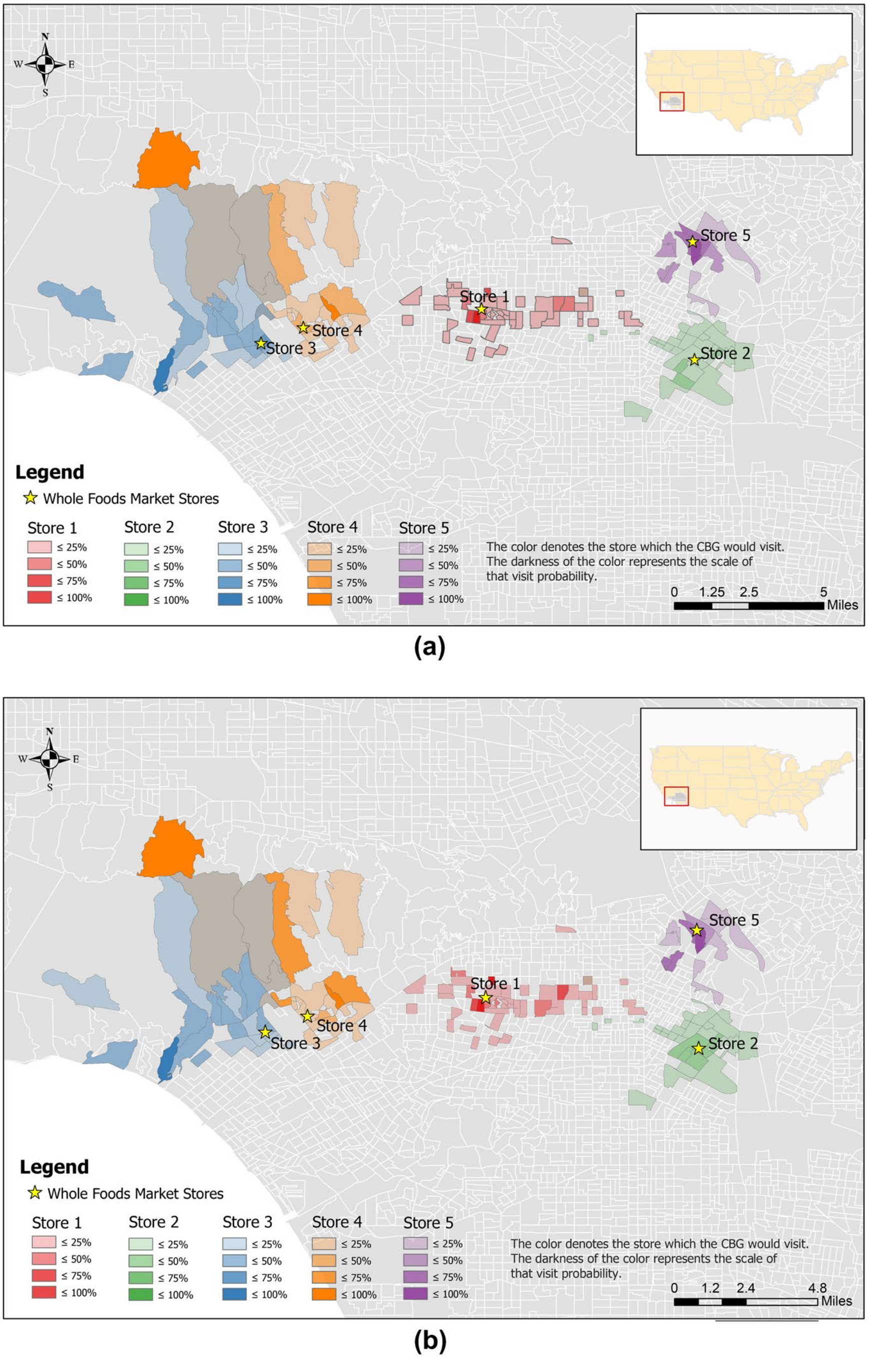


FIGURE 4 (a) Estimated market share of five Whole Foods Market stores in Los Angeles using the original Huff model; and (b) Actual market share derived from the SafeGraph visit database

a CBG will visit that particular store. For every CBG, it has a corresponding visit probability for each store, and the color hue of each CBG represents the store that people from this CBG would visit. The saturation of the color indicates the magnitude of the probability. By comparing the two maps, we find that the spatial distributions of trade areas are very similar (with high correlation of store visit probabilities). This means that the estimated result from the original Huff model can project the total visit probability with high accuracy. The result also supports our earlier statement that the large portion of visitors of each Whole Foods Market are usually within close proximity of that store. People may be reluctant to go to another Whole Foods Market that is far away from them. This is a characteristic of the chain stores: that the location of a store is very important to the performance of that store. As the chain stores may not be very different from each other with regard ti their products, the spatial proximity between the store and the customer becomes a primary factor affecting people's choice.

Figure 5 shows the histograms of hourly visit probability on Sunday 3:00–3:59 p.m. and Monday 11:00–11:59 a.m. Figure 6 maps the difference between the estimated and the actual market share of the Five Whole Foods for two different time windows obtained from the dynamic Huff model. Here we pick two different hours (Sunday 3:00–3:59 p.m. and Monday 11:00–11:59 a.m.) to compare how the POI visit probability may differ at different times of day and on different days of the week (McKenzie, Janowicz, Gao, Yang, et al., 2015). The data classification intervals for the visit probability mapping are determined by geometrical intervals as the probability distributions for all CBGs to all Whole Foods stores in the two hours both follow a right-skewed distribution.

From the T-Huff model, as the visit probability is assigned to a specific hourly window, it has a much smaller range compared with that of the original Huff model. Therefore, the ranges of the probability differences are also smaller, usually between −0.003 and 0.003 from the maps in Figure 6. Also, we can see that most of the prediction errors are between −0.001 and 0.001. The prediction for Monday 11a.m. has a better accuracy than that for Sunday 3 p.m. as we can see that there are fewer dark red or dark green areas on the map for Monday 11a.m. One reason could be that there is larger variability of visits on Sunday 3 p.m.

5.2.3 | Brand comparison and regional variability

The same process of model parameter calibration using PSO for three brands (Whole Foods, Trader Joe's and Ross stores) is conducted for the 10 U.S. cities. Three types of comparisons are examined: first, the performance of the four Huff models; second, how the models perform for the three brands; and third, discovering whether there exists regional variability among the same type of stores across different cities. Table 4 shows the number of stores for the three brands in each city. Table 5 shows the highest correlation coefficients from the PSO for the four Huff models and three brands in the 10 cities. Tables 6 and 7 show the corresponding α and β values for each optimal solution.

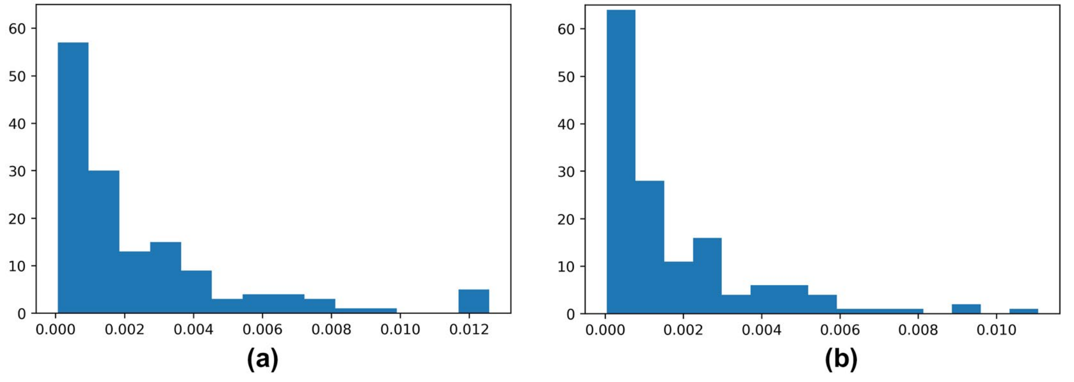
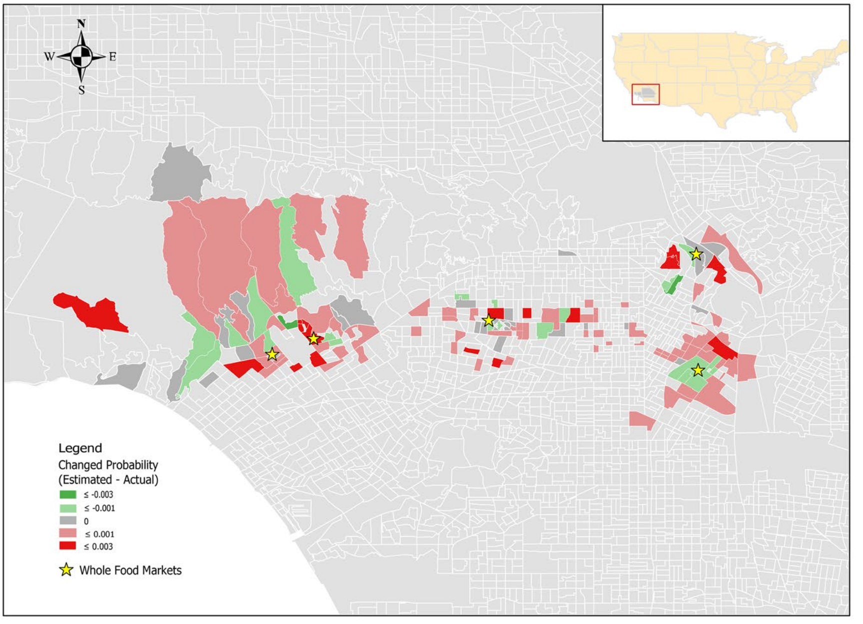
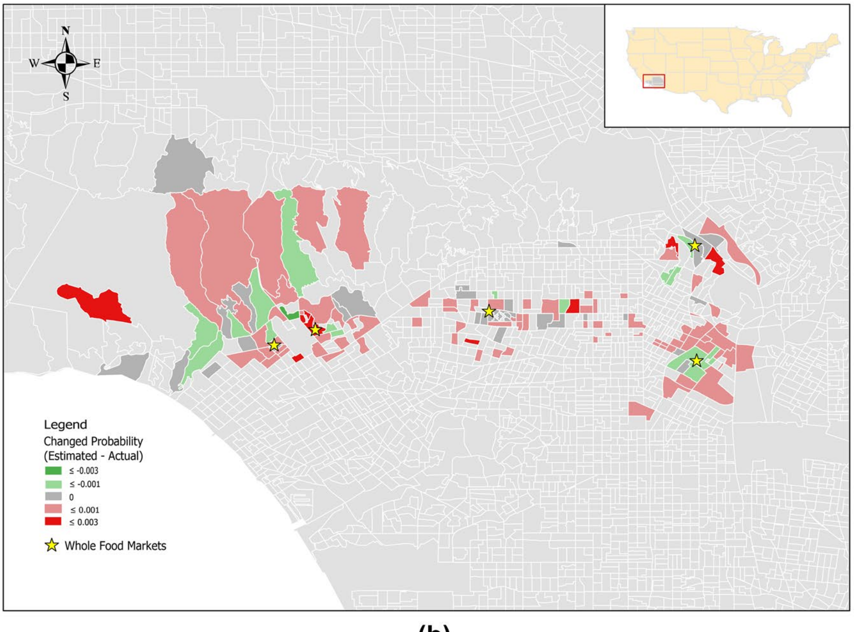


FIGURE 5 Histograms of visit probability on: (a) Sunday 3:00–3:59 p.m.; and (b) Monday 11:00–11:59 a.m.



(a)



(b)

FIGURE 6 Maps of the visit probability changes between the estimated market share using the T-Huff model and the actual market share derived from the SafeGraph visit database on: (a) Sunday 3:00–3:59 p.m.; and (b) Monday 11:00–11:59 a.m.

By looking at each row, we can compare the performance of four models. The optimal correlations are generally high for the original Huff model, the T-Huff model and the A-Huff model across all stores and cities. But the correlation for the M-Huff model is always much lower than those of the other three models, which indicates that

TABLE 4 The number of stores for the three brands in 10 cities

	Whole Foods	Trader Joe's	Ross stores
Los Angeles	5	11	14
Houston	7	3	24
Chicago	10	5	12
Philadelphia	2	1	8
New York	8	5	0
San Antonio	1	2	15
Dallas	4	4	7
San Diego	1	6	7
San Jose	2	4	6
Phoenix	3	1	15

Note: The number only reflects the data we have.

the temporal variation cannot be ignored or simply considered as equally distributed. The T-Huff and A-Huff models have slightly higher correlations than the traditional Huff model, which may show that the temporal variation is important and can help improve the estimation accuracy. The result for the T-Huff model is the highest among the four models for each brand and city in most cases, which shows that adding the temporal visiting information in this model yields the best performance in our study. By comparing the parameters in Tables 6 and 7, the optimal α and β remain similar for each brand in each city among four models. This indicates that for each particular type of POI in each city, the optimization process is able to find consistent parameters among four models that reflect the impacts of attractiveness and distance specifically for each brand in that city.

We also compare the results row by row to detect any changes over different cities. The parameter changes in Tables 6 and 7 reflect different local patterns. From the table we observe that even for the same brand, the models produce very different parameters across cities, which indicates that people's visit behaviors are affected by regional differences (McKenzie, Janowicz, Gao, et al., 2015), which may link to the size and shape of a city, POI co-location patterns, and urban spatial structure (Gao, Janowicz, & Couclelis, 2017; Kang et al., 2012; Yue et al., 2017).

For example, β is the exponent of distance in the Huff models and reveals the impact of distance decay on visit activities; we can use β to compare different spatial interaction patterns (Liu, Sui, Kang, & Gao, 2014). In general, a larger β means the activities are more affected by the change of distances. Usually, with more spatial interactions in a city, we can expect a smaller β as people are less spatially separated with the support of modern multi-mode transportation (Liu et al., 2014; McKenzie, 2014; Su, Li, Xu, Cai, & Weng, 2017). Comparing the β changes over different cities, it is clear that New York has a very small β for both Whole Foods and Trader Joe's compared with other cities. This indicates the POI visit patterns for people in New York are less influenced by distance. This is reasonable as the well-developed transportation makes people in such a large metropolitan area more connected to each other and long distance will have a less negative impact in terms of preventing people from traveling to other places. We also use the average β for each city to reflect the effects of distance to cities. The top cities with smallest β in Table 6 are New York, San Diego, Philadelphia, and Chicago. Except for San Diego which has a very small β for Whole Foods (there is only one Whole Foods Market in San Diego in our data set), the other three cities are all cities with well-developed public transit systems. The mixed mode of private driving and public transportation may make distance less sensitive for traveling and leads to small β for those cities.

Next, we compare the parameter differences over different types of stores and find some distinct patterns between the supermarkets and grocery stores on the one hand and the department stores on the other hand. Here the supermarkets and grocery stores are represented by two brands, Whole Foods and Trader Joe's, and the department stores are represented by, Ross Stores. Of the nine cities that have Ross Stores, six have smaller average

TABLE 5 Optimized correlation for three brands

	Whole Foods				Trader Joe's				Ross Stores			
	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff
Los Angeles	0.864	0.662	0.890	0.878	0.875	0.588	0.910	0.900	0.854	0.664	0.881	0.863
Houston	0.827	0.567	0.874	0.838	0.682	0.391	0.827	0.776	0.821	0.544	0.864	0.845
Chicago	0.869	0.580	0.904	0.892	0.892	0.578	0.919	00.913	0.933	0.670	0.946	0.940
Philadelphia	0.869	0.612	0.899	0.892	0.956	0.770	0.962	0.959	0.892	0.637	0.917	0.904
New York	0.949	0.602	0.968	0.955	0.847	0.567	0.902	0.863	NA	NA	NA	NA
San Antonio	0.888	0.434	0.931	0.923	0.644	0.406	0.748	0.721	0.935	0.650	0.942	0.942
Dallas	0.901	0.573	0.932	0.908	0.948	0.603	0.963	0.960	0.953	0.610	0.965	0.960
San Diego	0.825	0.614	0.853	0.833	0.919	0.604	0.938	0.929	0.917	0.648	0.928	0.925
San Jose	0.959	0.687	0.964	0.964	0.927	0.563	0.952	0.947	0.903	0.582	0.930	0.924
Phoenix	0.966	0.589	0.979	0.967	0.959	0.571	0.970	0.970	0.900	0.615	0.924	0.910
Average	0.892	0.592	0.919	0.905	0.865	0.564	0.909	0.894	0.901	0.624	0.922	0.913

Abbreviation: NA, no data available.

TABLE 6 Optimized Huff model coefficients β for three brands

	Whole Foods				Trader Joe's				Ross Stores			
	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff
Los Angeles	0.80	0.81	0.76	0.85	0.56	0.60	0.51	0.59	0.93	1.10	1.01	0.93
Houston	0.91	1.01	0.90	0.92	0.74	0.74	0.70	0.97	0.67	0.66	0.63	0.63
Chicago	0.60	0.57	0.66	0.56	0.64	0.58	0.68	0.54	0.73	0.57	0.65	0.68
Philadelphia	0.64	0.54	0.55	0.56	0.59	0.44	0.54	0.57	0.46	0.41	0.68	0.62
New York	0.44	0.44	0.44	0.49	0.06	0.06	0.18	0.13	NA	NA	NA	NA
San Antonio	0.76	0.70	0.77	0.82	0.72	0.76	0.55	0.67	0.78	0.84	0.82	0.84
Dallas	0.93	0.93	0.99	1.01	0.88	0.83	0.72	0.77	0.52	0.55	0.61	0.57
San Diego	0.04	0.03	0.03	0.14	0.93	0.84	0.93	0.94	0.43	0.44	0.49	0.45
San Jose	0.84	0.95	0.84	0.78	0.98	0.85	0.82	0.72	0.78	0.80	0.78	0.75
Phoenix	1.58	1.83	1.63	1.83	1.36	1.23	1.05	0.97	0.70	0.65	0.69	0.65

Abbreviation: NA, no data available.

TABLE 7 Optimized Huff model coefficients α for three brands

	Whole Foods				Trader Joe's				Ross Store's			
	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff	Huff	M-Huff	T-Huff	A-Huff
Los Angeles	0.72	0.72	0.79	0.69	0.44	0.44	0.45	0.54	0.59	0.77	0.73	0.62
Houston	0.82	0.85	0.94	0.87	0.98	0.95	0.93	0.99	0.62	0.61	0.52	0.48
Chicago	0.48	0.49	0.55	0.46	0.46	0.44	0.43	0.34	0.39	0.37	0.38	0.40
Philadelphia	0.88	0.98	0.84	0.83	0.36	0.35	0.31	0.37	0.60	0.54	0.80	0.74
New York	0.39	0.30	0.28	0.33	0.26	0.12	0.20	0.17	NA	NA	NA	NA
San Antonio	0.78	0.89	0.72	0.71	0.83	0.90	0.68	0.69	0.58	0.60	0.66	0.67
Dallas	0.33	0.37	0.40	0.43	0.84	0.83	0.73	0.73	0.43	0.47	0.47	0.45
San Diego	0.06	0.16	0.01	0.23	0.99	0.82	0.84	0.90	0.46	0.48	0.46	0.37
San Jose	0.58	0.69	0.70	0.66	0.49	0.48	0.53	0.52	0.72	0.76	0.81	0.66
Phoenix	0.64	0.61	0.64	0.56	0.36	0.39	0.42	0.46	0.79	0.73	0.75	0.68

Abbreviation: NA, no data available.

β for Ross Stores than for Whole Foods and Trader Joe's. As we showed in Section 5.1, the department stores have a smoother distance decay slope than the supermarkets and grocery stores, which means that distance has a greater effect on visits to supermarket and grocery stores. From our result, a majority of the cities showed the same trend that distance plays a more important role when people visit supermarkets and grocery stores. This corresponds to the daily experience as customers tend to go to the closest supermarkets or grocery stores as the goods in those types of stores are generally similar. Therefore, distance becomes the major factor to consider when deciding which store to visit, and this is also validated by our data-driven analytical results.

5.3 | Location business insights

In addition to the store attraction and distance, we conduct a multiple linear regression (MLR) analysis to discover potential factors explaining why people from certain neighborhoods often go to a particular POI with regard to the characteristics of that neighborhood and the POI attraction. Specifically, we take factors from the demographic and socioeconomic aspects into consideration to detect whether people from a certain socioeconomic neighborhood will have common mobility patterns in terms of the places they often visit. The dependent variable is the pairwise visit count from a CBG community to a store, and the independent variables with low-multicollinearity are *store total visit counts*, *distance* between a store and a customer's most frequently visited home (work) CBG, the *total population* of a CBG, the *median age* and the *median household income* of people living in that CBG, and the Shannon *entropy* based on the natural logarithm to measure the racial and ethnic diversity of each CBG community (Shannon, 1948). A higher entropy value means a higher racial and ethnic diversity, while a lower entropy value indicates a larger proportion of a dominant racial or ethnic group in a CBG (Prestby, App, Kang, & Gao, 2020). Table 8 shows the MLR coefficients of those variables estimated using the OLS approach and their statistical significance for explaining the overall variability of the visit probability to three brands' stores (i.e., Whole Foods, Trader Joe's and Ross) across the 10 cities. The experiments demonstrate that store attractiveness measured by the total visit counts and median household income are significant positive factors that drive visits from CBGs to the stores of all three brands. Distance plays a significant negative role for both Whole Foods and Ross Stores but not for Trader Joe's. Race and ethnic diversity (entropy measure) have a significant positive influence for Ross and

TABLE 8 Regression coefficients of influential variables for explaining the total visit variability for the three brands' stores

	Whole Foods		Trader Joe's		Ross Stores	
	Coefficients	Sig.	Coefficients	Sig.	Coefficients	Sig.
Intercept	3.399e+01	0.0016**	1.135e+00	0.9341	1.867e+01	1.73e−06***
Total visit counts	1.323e−03	0.0451*	4.112e−03	0.0007***	3.980e−03	<2e−16***
Distance	−9.218e−01	1.76e−07***	−1.436e−02	0.2943	−5.551e−01	<2e−16***
Total population	8.147e−04	0.0592****	3.739e−03	7.31e−06***	4.132e−03	<2e−16***
Median household income	1.431e−04	7.37e−07***	1.411e−04	0.0001***	4.369e−05	0.0423*
Median age	−2.488e−01	0.1609	−2.579e−01	0.27144	−3.155e−01	0.0014**
Entropy	−6.418e−01	0.8994	1.567e+01	0.0168*	7.337e+00	0.0002***

Significance level: ***p <0 .001; **p <0 .01; *p <0 .05; ****p<0.1.

TABLE 9 R^2 for the regression models of three brands across the most populated US cities

City	Whole Foods	Trader Joe's	Ross stores
Los Angeles	.265	.168	.242
Houston	.096	.119	.101
Chicago	0.131	0.391	0.089
Philadelphia	0.289	0.187	0.101
New York	0.293	0.431	NA
San Antonio	0.381	0.202	0.126
Dallas	0.272	0.070	0.203
San Diego	0.240	0.436	0.224
San Jose	0.165	0.222	0.205
Phoenix	0.222	0.567	0.160
R^2 mean	0.235	0.279	0.161
$R^2 SD$	0.085	0.164	0.059

Abbreviation: NA, no data available.

Trader Joe's store visits. The median age of people in CBGs seems not to play a significant role except for Ross Stores, where all factors are significant.

Furthermore, we investigate whether the customer visit patterns for the three brands and the performance of influential factors are different across these U.S. cities. Table 9 shows the R^2 values for the three brands' store visits in the MLR models. Overall the regression models perform better in supermarkets and grocery stores (the mean R^2 value for Trader Joe's is 0.279 and for Whole Foods is 0.235) than in department stores (the mean R^2 value for Ross Stores is 0.161). However, there exists large regional variability of the MLR model performance in explaining the store visit patterns. The standard deviation of R^2 for Trader Joe's (0.164) is the largest among the three brands. The regression model has a higher goodness of fit for the Trader Joe's stores in Phoenix, San Diego, and New York (all with $R^2 > .4$) but a very low R^2 value in Dallas (0.07). Given the large size and socioeconomic complexity of these highly populated cities, there might exist other indicative features that we need to further investigate in the future.

6 | CONCLUSIONS AND FUTURE WORK

In this research we present a time-aware dynamic Huff model (T-Huff) that incorporates the hourly temporal variability of store visits to delineate the dynamic trade areas for different types of business POI. To calibrate the model parameters, we apply the PSO technique with hourly POI visit probability derived from a large-scale mobile phone location data set across the 10 most populated U.S. cities. To answer the two research questions that we posed at the beginning of this research:

1. The calibrated time-aware dynamic Huff model (T-Huff) is more accurate than the original static Huff model without temporal variation in predicting the market share of different types of business (e.g., supermarkets versus department stores) over time.
2. Spatial proximity, demographic and socioeconomic factors (e.g., median household income) have significant impacts on the customer choice of particular store visits. There exists regional variability for store visit patterns across different cities with varying calibrated Huff model parameters and different goodness-of fit-values in

MLR models. The performance variability of models may link to different spatial socioeconomic structure and transportation infrastructure in those large cities.

In sum, our time-aware dynamic Hull models and analytical workflow using location big data can be applied to other categories of business stores for location-based marketing and dynamic trade area analyses.

One limitation of our current analysis is the lack of street-network distance and centrality measures that may influence the spatial distribution of business stores (Porta et al., 2009). In addition, the travel time and traffic congestion contexts for certain routes in people's minds may also impact their accessibility and decision-making (McKenzie, 2014; Stanley & Sewall, 1976; Su et al., 2017). We will consider street-network measures and traffic information in the modeling framework in future work.

ACKNOWLEDGMENTS

We would like to thank SafeGraph Inc. for providing anonymous location visit data support. Support for this research was provided by the University of Wisconsin—Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. This research was also supported by The We Company (WeWork) through academic collaboration with the Geospatial Data Science (GeoDS) Lab at the University of Wisconsin—Madison.

ORCID

Song Gao http://orcid.org/0000-0003-4359-6302

REFERENCES

Bennett, V. W. (1944). Consumer buying habits in a small town located between two large cities. *Journal of Marketing*, *8*(4), 405–416.

Blankertz, D. F. (1951). Customer patronage of a parent and new branch store. *Journal of Marketing*, *16*(2), 152–157.

Chan, T. Y., Padmanabhan, V., & Seetharaman, P. (2007). An econometric model of location and pricing in the gasoline market. *Journal of Marketing Research*, *44*(4), 622–635.

Donthu, N., & Rust, R. T. (1989). Estimating geographic customer densities using kernel density estimation. *Marketing Science*, *8*(2), 191–203.

Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan (pp. 39–43). Piscataway, NJ: IEEE.

Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*, *2*(1), 28–32.

Fotheringham, A. S. (1988). Consumer store choice and choice set definition. *Marketing Science*, *7*(3), 299–310.

Gao, S. (2015). Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spatial Cognition & Computation*, *15*(2), 86–114.

Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, *21*(3), 446–467.

Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, *17*(3), 463–481.

Gao, S., & Mai, G. (2017). Mobile GIS and location-based services. In B. Huang, T. J. Cova, & M.-H. Tsou (Eds.), *Comprehensive geographic information systems* (Vol. *1*, pp. 384–397). Amsterdam, Netherlands: Elsevier.

Ghosh, A., & Craig, C. S. (1983). Formulating retail location strategy in a changing environment. *Journal of Marketing*, *47*(3), 56–68.

Grether, E. (1983). Regional-spatial analysis in marketing. *Journal of Marketing*, *47*(4), 36–43.

Huang, H., Gartner, G., Krisp, J. M., Raubal, M., & Van de Weghe, N. (2018). Location based services: Ongoing evolution and research agenda. *Journal of Location Based Services*, *12*(2), 63–93.

Huff, D., & McCallum, B. M. (2008). *Calibrating the Huff model using ArcGIS business analyst* (White Paper). Redlands, CA: Esri.

Huff, D. L. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, *39*(1), 81–90.

Huff, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing*, *28*(3), 34–38.

Huff, D. L. (2003). Parameter estimation in the Huff model. *ArcUser (Oct.-Dec.)*, *2003*(4), 34–36.

Hutchinson, K. D. (1940). Traffic and trade correlations: A technique in store location. *Journal of Marketing*, *5*(2), 137–142.

Jiang, W., Wang, Y., Dou, M., Liu, S., Shao, S., & Liu, H. (2019). Solving competitive location problems with social media data based on customers' local sensitivities. *ISPRS International Journal of Geo-Information*, *8*(5), 202.

Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics & Its Applications*, *39*(4), 1702–1717.

Kennedy, J. (2010). Particle swarm optimization. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 760–766). Berlin, Germany: Springer.

Liu, K., Yin, L., Lu, F., & Mou, N. (2020). Visualizing and exploring POI configurations of urban regions on POI-type semantic space. *Cities*, *99*, 102610.

Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE*, *9*(1), e86026.

Lu, S., Shaw, S.-L., Fang, Z., Zhang, X., & Yin, L. (2017). Exploring the effects of sampling locations for calibrating the Huff model using mobile phone location data. *Sustainability*, *9*(1), 159.

McKenzie, B. S. (2014). Access to supermarkets among poorer neighborhoods: A comparison of time and distance measures. *Urban Geography*, *35*(1), 133–151.

McKenzie, G., & Adams, B. (2017). Juxtaposing thematic regions derived from spatial and platial user-generated content. In *Proceedings of the 13th International Conference on Spatial Information Theory*, L'Aquila, Italy. Wadern, Germany: Schloss Dagstuhl—Leibniz-Zentrum für Informatik.

McKenzie, G., Janowicz, K., Gao, S., & Gong, L. (2015). How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment & Urban Systems*, *54*, 336–346.

McKenzie, G., Janowicz, K., Gao, S., Yang, J.-A., & Hu, Y. (2015). POI pulse: A multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica*, *50*(2), 71–85.

Miller, C. E., Reardon, J., & McCorkle, D. E. (1999). The effects of competition on retail structure: An examination of intra-type, intertype, and inter-category competition. *Journal of Marketing*, *63*(4), 107–120.

Nakanishi, M., & Cooper, L. G. (1974). Parameter estimation for a multiplicative competitive interaction model-least squares approach. *Journal of Marketing Research*, *11*(3), 303–311.

Negash, S., & Gray, P. (2008). Business intelligence. In F. Burstein & C. W. Holsapple (Eds.), *Handbook on decision support systems* (Vol. *2*, pp. 175–193). Berlin, Germany: Springer.

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., & Scellato, S. (2009). Street centrality and densities of retail and services in Bologna, Italy. *Environment & Planning B*, *36*, 450–465.

Prestby, T., App, J., Kang, Y., & Gao, S. (2020). Understanding neighborhood isolation through spatial interaction network analysis using location big data. *Environment & Planning A*, *52*, 1–5.

Rosenbloom, B. (1976). The trade area mix and retailing mix: A retail strategy matrix. *Journal of Marketing*, *40*(4), 58–66.

Rust, R. T., & Donthu, N. (1995). Capturing geographically localized misspecification error in retail store choice models. *Journal of Marketing Research*, *32*(1), 103–110.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shi, L., Chi, G., Liu, X., & Liu, Y. (2015). Human mobility patterns in different communities: A mobile phone data-based social network approach. *Annals of GIS*, *21*(1), 15–26.

Stanley, T. J., & Sewall, M. A. (1976). Image inputs to a probabilistic model: Predicting retail potential. *Journal of Marketing*, *40*(3), 48–53.

Su, S., Li, Z., Xu, M., Cai, Z., & Weng, M. (2017). A geo-big data approach to intra-urban food deserts: Transit-varying accessibility, social inequalities, and implications for urban planning. *Habitat International*, *64*, 22–40.

Súarez-Vega, R., Gutíerrez-Acuña, J. L., & Rodríguez-Díaz, M. (2015). Locating a supermarket using a locally calibrated Huff model. *International Journal of Geographical Information Science*, *29*(2), 217–233.

Suhara, Y., Bahrami, M., Bozkaya, B., & Pentland, A. (2019). Validating gravity-based market share models using large-scale transactional data. Preprint, arXiv:1902.03488.

Talukdar, D. (2008). Cost of being poor: Retail price and consumer price search differences across inner-city and suburban neighborhoods. *Journal of Consumer Research*, *35*(3), 457–471.

Thomadsen, R. (2007). Product positioning and competition: The role of location in the fast food industry. *Marketing Science*, *26*(6), 792–804.

Tu, W., Cao, J., Yue, Y., Shaw, S.-L., Zhou, M., Wang, Z., & Li, Q. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, *31*(12), 2331–2358.

Vitorino, M. A. (2012). Empirical entry games with complementarities: An application to the shopping center industry. *Journal of Marketing Research*, *49*(2), 175–191.

Wang, Y., Jiang, W., Liu, S., Ye, X., & Wang, T. (2016). Evaluating trade areas using social media data with a calibrated Huff model. *ISPRS International Journal of Geo-Information*, *5*(7), 1–15.

Xiao, Y., Wang, F., Liu, Y., & Wang, J. (2013). Reconstructing gravitational attractions of major cities in China from air passenger flow data, 2001–2008: A particle swarm optimization approach. *Professional Geographer*, *65*(2), 265–282.

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2017). How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS*, *21*(3), 468–487.

Yang, X., Fang, Z., Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., & Lin, Y. (2016. Understanding spatiotemporal patterns of human convergence and divergence using mobile phone location data. *ISPRS International Journal of Geo-Information*, *5*(10), 177–195.

Ye, M., Janowicz, K., Mülligann, C., & Lee, W.-C. (2011). What you are is when you are: The temporal dimension of feature types in location-based social networks. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, IL (pp. 102–111). New York, NY: ACM.

Yuan, Y., & Raubal, M. (2012). Extracting dynamic urban mobility patterns from mobile phone data. In N. Xiao, M. P. Kwan, M. F. Goodchild, & S. Shekhar (Eds.), *Geographic information science: GIScience 2012* (Lecture Notes in Computer Science, Vol. *7478*, pp. 354–367). Berlin, Germany: Springer.

Yue, Y., Zhuang, Y., Yeh, A. G., Xie, J.-Y., Ma, C.-L., & Li, Q.-Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, *31*(4), 658–675.

How to cite this article: Liang Y, Gao S, Cai Y, Foutz NZ, Wu L. Calibrating the dynamic Huff model for business analysis using location big data. *Transactions in GIS*. 2020;24:681–703. https://doi.org/10.1111/tgis.12624