Customer mobility and congestion in supermarkets

Fabian Ying,[1] Alisdair O. G. Wallis,[2] Mariano Beguerisse-Díaz,[1] Mason A. Porter,[3] and Sam D. Howison[1]
[1]*Mathematical Institute, University of Oxford, Oxford OX2 6GG, United Kingdom*
[2]*Tesco PLC, Tesco House, Shire Park, Kestrel Way, Welwyn Garden City, AL7 1GA, United Kingdom*
[3]*Department of Mathematics, University of California, Los Angeles, California 90095, USA*

(Received 30 May 2019; revised manuscript received 26 September 2019; published 9 December 2019)

The analysis and characterization of human mobility using population-level mobility models is important for numerous applications, ranging from the estimation of commuter flows in cities to modeling trade flows between countries. However, almost all of these applications have focused on large spatial scales, which typically range between intracity scales and intercountry scales. In this paper, we investigate population-level human mobility models on a much smaller spatial scale by using them to estimate customer mobility flow between supermarket zones. We use anonymized, ordered customer-basket data to infer empirical mobility flow in supermarkets, and we apply variants of the gravity and intervening-opportunities models to fit this mobility flow and estimate the flow on unseen data. We find that a doubly-constrained gravity model and an extended radiation model (which is a type of intervening-opportunities model) can successfully estimate 65%–70% of the flow inside supermarkets. Using a gravity model as a case study, we then investigate how to reduce congestion in supermarkets using mobility models. We model each supermarket zone as a queue, and we use a gravity model to identify store layouts with low congestion, which we measure either by the maximum number of visits to a zone or by the total mean queue size. We then use a simulated-annealing algorithm to find store layouts with lower congestion than a supermarket's original layout. In these optimized store layouts, we find that popular zones are often in the perimeter of a store. Our research gives insight both into how customers move in supermarkets and into how retailers can arrange stores to reduce congestion. It also provides a case study of human mobility on small spatial scales.

DOI: 10.1103/PhysRevE.100.062304

I. INTRODUCTION

Understanding human mobility is important for city planners, policy makers, transportation researchers, and many others. Motivated by practical applications and the desire to explore fundamental phenomena in human sciences, many researchers have developed and analyzed population-level models, such as gravity [1] and intervening-opportunities (IO) models [2,3], to study human mobility [4].

Population-level mobility models characterize the flow of people or other entities between locations using local attributes, such as their populations or the distance between the locations. They have been used for many applications, including modeling commuter flow between locations [5], trade flow between countries [6], and traffic flow inside a city [7]. These applications are diverse, but they are all on large spatial scales, ranging from an intercountry scale of thousands of kilometers (e.g., estimating trade flow [6,8]) to city and regional scales of tens of kilometers (e.g., estimating commuting patterns [9]). For even smaller spatial scales [10] (e.g., building level), the prevalent approach in most studies is to use pedestrian models, such as mobility models for individuals (e.g., random walks [11]) or models of crowd dynamics [12].

We consider the problem of modeling mobility flow between zones inside supermarkets and investigate how the flow changes when we rearrange store layouts. We therefore examine aggregate flow, which (despite the small spatial scales of these systems) makes population-level mobility models more suitable than random walks [13] or crowd-dynamic models. The small, building-level spatial scale may affect the fundamental features of mobility dynamics (and therefore the performance of the models) in important ways. For example, it has been reported that some models (such as radiation models) perform worse on small spatial scales than on large ones [9]. A possible reason for this observation is that the spatial "force" is much smaller on small scales than on large ones due to the smaller cost of making a trip in the former situation, so other nonspatial "forces" that are not captured by these models may instead be the primary factors that underlie the flow. In a supermarket, for example, the distance between two zones (the spatial "force") may be less relevant than the number of their complementary items (a nonspatial "force") for the flow between the two zones. Furthermore, these models are inherently memoryless, as they describe mobility flow from an origin location to a destination location using local attributes of these locations, without considering the location from which (or how) a person who leaves the origin location entered it in the first place. When one models humans walking inside a building (e.g., in a supermarket or a museum), the direction from which a person comes likely influences where that person goes next, so there is memory in the system. For example, Farley and Ring [14] observed that customers in supermarkets tend to move from the entrance unidirectionally along the outer perimeter after entering a store.

In the present paper, we conduct a detailed case study of mobility models in an investigation of congestion in

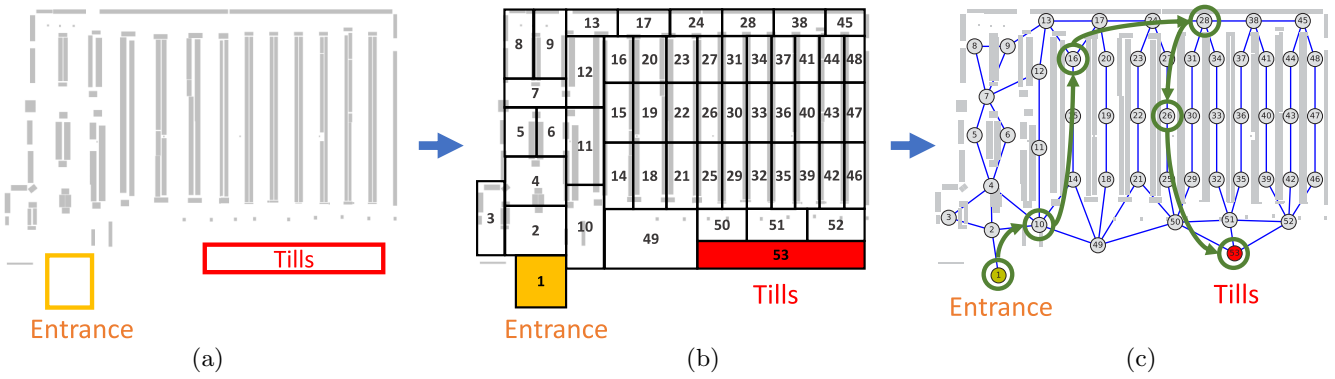

FIG. 1. We divide (a) a store into (b) zones and represent it as (c) a network. We depict the shelves in gray. We encode the ordered shopping basket, which we highlight in green in panel (c), of a customer who purchases an item (e.g., bread) in zone 10, another item (e.g., milk) in zone 16, a third item (e.g., butter) in zone 28, and a fourth item (e.g., pasta) in zone 26 as a shopping journey (1, 10, 16, 28, 26, 53) and then divide the journey into five origin–destination (OD) trips: (1, 10), (10, 16), (16, 28), (28, 26), and (26, 53). Each green arrow (and accompanying arc) in panel (c) represents an OD trip.

supermarkets, a practical problem that is influenced by the layout of a store. Reducing congestion is important not only for improving the shopping experience of customers, but also for reducing the fulfillment time and cost of online orders. (In many supermarkets, staff members go around a store alongside customers and pick up items that were ordered online.) Congestion may delay such orders and thereby incur additional costs to a business and inconvenience customers in a store. In our study, we integrate mobility models with a congestion model—such that each supermarket zone is a queue and we make the simplifying assumption that customers traverse shortest paths between purchases—to estimate congestion in supermarkets. We then use a simple optimization algorithm to find store layouts with low congestion.

Our article has three main contributions. First, we show that several different mobility models can successfully estimate the majority of observed trips in supermarket customer-flow data, demonstrating that these models can work on small (specifically, building-level) spatial scales. Second, we show how to combine these models with a congestion model based on queuing networks to estimate congestion in customer flow. Third, we demonstrate how to optimize store layouts to reduce congestion.

Our article proceeds as follows. In Sec. II, we describe our mathematical setup. In Sec. III, we describe the data set from which we infer the origin–destination (OD) trips in 17 supermarkets. In Sec. IV, we describe the mobility models and goodness-of-fit measures that we use in our investigation. We also describe how we estimate the parameters of our models. In Sec. V, we present our results when applying these models to supermarket store data, using both (in-sample) fitting and (out-of-sample) estimation of customer flows. In Sec. VI, we describe an application of a human mobility model to estimate customer congestion and determine store layouts that reduce it. Specifically, we discuss our congestion model, our optimization method, and the results of the optimization. We conclude and discuss future research directions in Sec. VII, and we give some additional details about our work in appendices.

II. MATHEMATICAL SETUP

In this section, we set up our approach for analyzing mobility flow in supermarkets. We discuss how we discretize space in a supermarket, how we model shopping journeys, and how we characterize flow between zones of a supermarket. We will discuss our data in Sec. III and mobility models in Sec. IV.

In our investigation, we employ mobility models that require us to discretize space (i.e., a supermarket), which we divide into a discrete number of disjoint locations, with an associated measure of distance between distinct locations. To do this, we manually divide each store into rectangular zones of approximately equal size. (See Appendix A for details.) We then represent a store as a network $\mathcal{G}$ with n nodes (representing the zones) and m edges, which connect neighboring zones (see Fig. 1). We distinguish an entrance zone (labeled 1) and a till zone (labeled n). A store network $\mathcal{G}$, which is embedded in space, is undirected. Although there are distances between supermarket zones, the network $\mathcal{G}$ itself is unweighted. For the location of each node, we use the centroid of its corresponding zone. For each edge (i, j), we assign an edge length l_{ij}, which we take to be the Euclidean distance between its two incident nodes i and j. (The edge length approximates the walking distance between two nodes [15].) We define an $n \times n$ distance matrix $\mathbf{\Lambda}$ that is associated with $\mathcal{G}$. The entry d_{ij} of $\mathbf{\Lambda}$ is equal to the *shortest-path distance* between i and j; this distance is the minimum length of a path between i and j. We define the *length* of each zone as the length of the longer side of the rectangle that encloses the zone.

One customer's *shopping journey* is a sequence of $K + 2$ zones $(s_0, \ldots, s_{K+1})$, where K is the number of items that the customer buys, $s_0 = 1$ (entrance), $s_{K+1} = n$ (tills), and $s_1, \ldots, s_K$ are the zones at which the customer picks up items, which we order by their pick-up times. A customer can purchase multiple items in the same zone, so $s_0, \ldots, s_{K+1}$ may not be distinct. Each consecutive pair (s_k, s_{k+1}) of distinct zones (so $s_k \neq s_{k+1}$) for $0 \leqslant k \leqslant K$ constitutes an *origin–destination (OD) trip* (or simply a *trip*). That is, a trip is

TABLE I. Key quantities and their descriptions.

Symbol	Description
$\mathcal{G}$	Store network
$\mathbf{\Lambda}$	Distance matrix (with entries d_{ij}) associated with the graph $\mathcal{G}$
l_{ij}	Length of edge (i, j) in the graph $\mathcal{G}$
d_{ij}	Shortest-path length between zones i and j in the graph $\mathcal{G}$
l	Mean zone length (excluding entrance and tills)
n	Number of nodes in the graph $\mathcal{G}$
m	Number of edges in the graph $\mathcal{G}$
τ	Time period
ρ	Fraction of baskets in the data set (i.e., the number of baskets in the data set in a time period divided by the total number of baskets during the same time period)
T_{ij}	Number of origin–destination (OD) trips from zone i to zone j
$\boldsymbol{T}$	OD matrix with entries T_{ij}
O_k	Number of OD trips that start at zone k (it equals the sum of the entries in the kth row of $\boldsymbol{T}$)
D_k	Number of OD trips that terminate at k (it equals the sum of the entries in the kth column of $\boldsymbol{T}$)
f_{ij}	Attraction factor of zone j to a customer in zone i
S_{ij}	Number of intervening opportunities of the OD pair (i, j) (it equals $\sum_{\{k:d_{ik}<d_{ij}\}} D_k$)
N	Number of trips (it equals $\sum_{i,j} T_{ij}$)
C	Number of shopping journeys
μ_k, μ	Service rate of zone k (we use μ when it is independent of k)
v_k	Estimated number of visits to zone k
λ_k	Rate of customer arrival at zone k
Q	Total mean queue size
$\lambda_{\max}$	Maximum arrival rate

a segment of a customer's shopping journey that is either between consecutive purchases in different zones, from the entrance to the first purchase, or from the last purchase to the tills. We are interested in the number T_{ij} of OD trips in a store from each origin zone i to each destination zone j (over some duration τ). We do not consider flow within a zone and thus set $T_{kk} = 0$ for $k = 1, \ldots, n$. The $n \times n$ matrix $\boldsymbol{T}$, with entries T_{ij}, is called an *origin–destination (OD) matrix* [4]; its off-diagonal entries record mobility flow between zones. We denote an empirical OD matrix by $\boldsymbol{T}^{\text{data}}$ and an OD matrix from a model by $\boldsymbol{T}^{\text{model}}$. Throughout our paper, we denote an origin node of a trip by i and a destination node of a trip by j; we index other nodes using the symbol k. We summarize and describe our key quantities in Table I.

III. DATA

We use anonymized, ordered customer-basket data from 17 large stores of a major United Kingdom supermarket chain (Tesco) over a common three-month period (91 days). The data consist of a fraction $\rho \approx 0.07$ of all customer baskets in these stores. We summarize the properties of the data in Table II.

For each store, we infer the number T_{ij}^{data} of OD trips from zone i to zone j over the $\tau = 91$ days from the data as follows. Each ordered customer basket is a list of item purchases, which we order by pick-up time. We use item-location data to map each ordered list of purchases to their associated zones in a supermarket. For example, we map a list of purchases (e.g., bread, milk, butter, and pasta) to its corresponding shopping journey (1, 10, 16, 28, 26, 53), where the entrance is in zone 1, bread is in zone 10, milk is in zone 16, butter is in zone 28, pasta is in zone 26, and the tills are in zone 53 [see Fig. 1(c)]. In this example, each item has a unique location, so we can recover the corresponding shopping journey in a straightforward way. However, about 10% of the purchased items have unknown locations and about 8% of the purchased items have multiple locations. We refer to the latter items as *multilocated items*, and we remove items with unknown locations from customer baskets. For each basket with one or more multilocated items, we consider all combinations of possible purchase locations for those items. (For example, there are 2^r combinations for a basket with r multilocated items with two locations each.) For each combination, we calculate the sum of the shortest-path distances between the locations of consecutive purchases in the basket. We then choose a combination of the purchase locations that minimizes this sum. We do not possess data to validate that customers tend to buy multilocated items at locations that minimize the sum of the shortest-path distances between consecutive purchases, but this assumption likely has only a small impact on the results of our analysis (see Sec. V for these results), as only about 8% of the items are multilocated.

TABLE II. Summary of our data set, which comes from 17 Tesco stores. For each quantity, we give the minimum, mean, and maximum values across the 17 stores.

	Min	Mean	Max
Number of zones (n)	61	123	197
Number of edges (m)	128	236	401
Number of baskets	2479	13 672	29 201
Mean zone length, excluding entrance and tills (l)	6.42 m	6.91 m	7.65 m

We decompose each customer shopping journey into its sequence of OD trips and estimate the total number T_{ij}^{data} of trips from zone i to zone j by counting all OD trips (i, j) from the data. [For example, the previous example shopping journey (1, 10, 16, 28, 26, 53) consists of 5 trips: (1, 10), (10, 16), (16, 28), (28, 26), and (26, 53).] Assuming that the observed mobility patterns in our data set are representative of the mobility of all customers, we rescale T_{ij}^{data} by multiplying it by $1/\rho$ (where ρ is the number of baskets in the data set in a time period divided by the total number of baskets during the same time period) to estimate the mobility flow of all customers who visit a store.

IV. MOBILITY MODELS

We examine several mobility models, which we use to estimate $\boldsymbol{T}^{\text{data}}$. Let $O_k^{\text{data}} = \sum_j T_{kj}^{\text{data}}$ and $D_k^{\text{data}} = \sum_i T_{ik}^{\text{data}}$, respectively, be the empirical numbers of trips that depart from and arrive at each node k. (Note that O_k^{data} and D_k^{data} are the row and column sums, respectively, of $\boldsymbol{T}^{\text{data}}$.) We consider a class of models that yield an $n \times n$ OD matrix $\boldsymbol{T}^{\text{model}}$ from O_k^{data}, D_k^{data}, the store network, and either one or zero fitting parameters. The goal of the models is for $\boldsymbol{T}^{\text{model}}$ to be "close" to an empirical OD matrix $\boldsymbol{T}^{\text{data}}$. (We discuss diagnostics for comparing $\boldsymbol{T}^{\text{model}}$ and $\boldsymbol{T}^{\text{data}}$ in Sec. IV E.) The models use $2n$ pieces of information of $\boldsymbol{T}^{\text{data}}$ to estimate the $(n-1) \times (n-1)$ off-diagonal entries of $\boldsymbol{T}^{\text{data}}$.

In our problem, $O_k^{\text{data}} = D_k^{\text{data}}$ for all nodes k, except for $k = 1$ (the entrance) and $k = n$ (the tills), as every customer who finishes a trip in zone k (except for $k = 1$ and $k = n$) continues their journey with a trip that starts from k. Note that $O_1^{\text{data}} \neq D_1^{\text{data}}$, because each shopping journey starts at node 1, so the first trip (which starts at 1) of each shopping journey does not have a preceding trip that ends at 1. Similarly, $O_n^{\text{data}} \neq D_n^{\text{data}}$, because each shopping journey ends at node n, so the last trip (which ends at n) does not have a subsequent trip that starts at n. The quantity O_k^{data} also gives the number of *shopping visits* at k (except for $k = 1$ and $k = n$). It is thus equal to the total number of times that customers visit k to purchase one or more items. The number C of journeys in the data satisfies both $C = O_1^{\text{data}} - D_1^{\text{data}}$ and $C = D_n^{\text{data}} - O_n^{\text{data}}$, as every journey starts at node 1 and ends at node n. Note that D_1^{data} and O_n^{data} need not be equal to 0, as the entrance and till nodes can contain items. Therefore, one is able to determine $\{O_k^{\text{data}}\}_{k=1}^n$ and $\{D_k^{\text{data}}\}_{k=1}^n$ from $\{O_k^{\text{data}}\}_{k=1}^n$ and D_1^{data}. In practice, we estimate these values from sales data (see Appendix D). Therefore, we assume that we know $\{O_k^{\text{data}}\}_{k=1}^n$ and $\{D_k^{\text{data}}\}_{k=1}^n$.

We use *doubly-constrained* (also called *production–attraction-constrained*) versions [16] of mobility models. In these models, the mobility flow $\boldsymbol{T}^{\text{model}}$ satisfies

$$O_k^{\text{data}} = O_k^{\text{model}}\,, \tag{1}$$

$$D_k^{\text{data}} = D_k^{\text{model}}\,, \tag{2}$$

where $O_k^{\text{model}} = \sum_j T_{kj}^{\text{model}}$ and $D_k^{\text{model}} = \sum_i T_{ik}^{\text{model}}$. That is, $\boldsymbol{T}^{\text{model}}$ has the same row sums and column sums as $\boldsymbol{T}^{\text{data}}$. This, in turn, implies for each node k that both the number of trips that arrive at k and the number of trips that depart from k are equal to their empirical values. Therefore, for notational simplicity, we drop the superscripts on O_k and D_k for the remainder of our paper. Because $O_k = D_k$ for $k = 2, \ldots, n-1$, the number of people at each node (except for the entrance and till nodes) is also conserved in the models.

For each origin node i, there is a vector of "attraction values" f_{ij} for each possible destination j. We calculate these values from a model-specific function f_{model} that takes O_i, $\{D_k\}_{k=1}^n$, and information (such as the distance between two nodes) from a store network as inputs. The function f_{model} is the same for each origin node i. (Allowing this function to be heterogeneous for different nodes would allow us to incorporate different types of supermarket zones into our models.) The mobility flow in a doubly-constrained model is

$$T_{ij}^{\text{model}} = (A_i O_i) \times (B_j D_j) \times f_{ij}\,, \tag{3}$$

where $A_i, B_j \geqslant 0$ are "balancing factors" to ensure that Eqs. (1) and (2) are satisfied. Given O_k, D_k, and f_{ij}, we determine A_i and B_j using an iterative proportional-fitting procedure [17]. See Appendix B for more details.

We can interpret T_{ij}^{model} as the mean aggregate flow that arises from a continuous-time random-walk model at stationarity. In our models, customers arrive at node 1 (i.e., the entrance) at a rate of $\lambda = C/\tau$. In contrast to a standard random walk on a network [18], customers do not choose a random neighbor at each step. Instead, each customer at node i chooses a random destination j, which need not be adjacent to i, with probability $P_{ij} = T_{ij}/\sum_k T_{ik} = T_{ij}/O_i = A_i B_j D_j f_{ij}$ when $O_i \neq 0$ (and $P_{ij} = 0$ when $O_i = 0$), and it then takes a trip from i to the chosen destination j. That is, the customer traverses some path from i to j. For simplicity, we assume that customers take a shortest path from node i to node j, where we choose this path uniformly at random from all shortest paths between these two nodes. (Other routing models are possible; one possibility is a standard random walk that starts at node i and reaches an absorbing state at node j.) We remove customers who finish a trip at node n at rate λ; this ensures that the mean number of customers in the system is constant. The quantity T_{ij}^{model} thus gives the mean number of trips from zone i to zone j over a time period τ. Our model assumes that there is no memory in customer mobility; the next destination of each customer depends only on its present location.

We present each model in a scale-invariant form, such that the parameters are dimensionless and the attraction values f_{ij} are invariant under the scalings $O_k \mapsto aO_k$ and $D_k \mapsto aD_k$ for $a > 0$ and for all k. Because O_k and D_k scale with the number C of journeys in the data set (and therefore with τ), scale invariance ensures that the model parameters and the transition probabilities P_{ij} are independent of C and τ [19].

In Table III, we summarize the choices of f_{ij} and the number of parameters for each of the four models that we employ. We discuss these models in the following subsections.

A. Gravity models

Gravity models of mobility [1,20–22], which are named after Newton's law of gravity, have been used to model a variety of systems, including human migration [22–24], cargo-ship movement [25], intercity telecommunication [26], spatial accessibility of health services [27], and trade flow [6,8].

TABLE III. Summary of the four mobility models that we employ. The OD matrix of each model is given by Eq. (3), with different functional forms for the attraction values f_{ij}.

Model	f_{ij}	Parameter	Parameter range	References
Gravity	$D_j(d_{ij}/l)^{-\gamma}$	γ	$[0,\infty)$	[1,20,21]
Intervening opportunities (IOs)	$\exp(-\frac{L}{N}S_{ij}) - \exp(-\frac{L}{N}(S_{ij}+D_j))$	L	$[0,N]$	[3,30]
Radiation	$\frac{O_iD_j}{(O_i+S_{ij})(O_i+D_j+S_{ij})}$	–	–	[5,9,36]
Extended radiation	$\frac{[(O_i+S_{ij}+D_j)^\alpha-(O_i+S_{ij})^\alpha][(O_i)^\alpha+N^\alpha)}{((O_i+S_{ij})^\alpha+N^\alpha)[(O_i+S_{ij}+D_j)^\alpha+N^\alpha]}$	α	$[0,\infty)$	[38]

In a gravity model, the mobility flow between two locations depends only on the distance between the locations and on the "mass" (i.e., "population") of the two locations. In our models, each node k has two types of populations: there is an origin population O_k (the number of trips that depart from zone k) and a destination population D_k (the number of trips that arrive at zone k). We use the origin population when we calculate the mobility flow from k (i.e., the outflow from k) and the destination population when we calculate the mobility flow to k (i.e., the inflow to k). Because $O_k = D_k$ for $k \in \{2,\dots,n-1\}$ in our problem, the two population values are the same, except for the entrance ($k=1$) and till ($k=n$) nodes.

We use a doubly-constrained gravity model, with

$$f_{ij} = f_g(O_i, D_j, d_{ij}) = O_iD_j(d_{ij}/l)^{-\gamma}, \tag{4}$$

where $l > 0$ is a spatial normalization factor (which we choose to be the mean zone length, excluding the entrance and till zones) and $\gamma \geqslant 0$ is a dimensionless fitting parameter. We exclude the entrance and till zones in calculating mean zone length, because typically they are much larger than the other zones in a store. See Appendix A for more details.

The expression $(d_{ij}/l)^{-\gamma}$ is an example of a "deterrence function," for which an exponential function is also a common choice [4]. In contrast to other studies on small spatial scales [5,28,29], we find that a power-law deterrence function gives a (slightly) better fit to our data than an exponential deterrence function (see Appendix C).

B. Intervening-opportunities models

In intervening-opportunities (IO) models, which were first proposed by Stouffer in 1940 [2], each location (i.e., node) has opportunities, which (depending on their number and/or quality) give some amount of "popularity" to that location. The key concept in IO models is the notion of *intervening opportunities*. The intervening opportunities S_{ij} of an OD pair (i,j) consist of all opportunities in nodes k that satisfy $d_{ik} < d_{ij}$. (Note that $S_{ij} \neq S_{ji}$ in general.) In IO models, the mobility flow between two locations (i.e., zones in our application) depends on the number of intervening opportunities (rather than on the distance) between the two locations and on the "populations" (which play the same role as in gravity models) of the two locations. A larger number of intervening opportunities of an OD pair (i,j) entails a smaller mobility flow from zone i to zone j, because customers are more likely to find what they are looking for (or to be diverted) before they reach j. Intervening-opportunities models and their variants have been used in many applications, including models of intracity mobility [30], interstate migration [31–33], rioting behavior [34], and the creation of social ties in a city [35].

We measure the number of opportunities at each node k by D_k, the number of trips that arrive at k. The number S_{ij} of intervening opportunities of an OD pair (i,j) is then

$$S_{ij} = \sum_{\substack{k \neq i \\ d_{ik} < d_{ij}}} D_k \,. \tag{5}$$

Therefore, the "opportunities" in our problem are opportunities for customers to stop and purchase something. Note that $S_{ij} \mapsto aS_{ij}$ when we scale $D_k \mapsto aD_k$ for all k, so the number of opportunities scales linearly with C.

In Stouffer's original IO formulation (StIO), the number of people who move a given distance is proportional to the number of opportunities at that distance and is inversely proportional to the number of intervening opportunities. The attraction values f_{ij} are

$$f_{ij} = f_{\mathrm{StIO}}(D_j, S_{ij}) = \frac{D_j}{S_{ij} + cN}\,, \tag{6}$$

for some $c > 0$ (to avoid dividing by 0), where $N = \sum_{i,j} T_{ij}$ is the total number of trips. In Eq. (6), we use cN instead of c to ensure scale invariance. Additionally, f_{ij} is only approximately inversely proportional to S_{ij} (because of the cN term).

In our investigation, we use Schneider's reformulation of the IO model [3], because it is more popular than Stouffer's IO formulation [4] and is underpinned by a mechanistic model. In this reformulation, the attraction values f_{ij} are

$$f_{ij} = f_{\mathrm{IO}}(D_j, S_{ij}) = e^{-\frac{L}{N}S_{ij}} - e^{-\frac{L}{N}(S_{ij}+D_j)} > 0\,, \tag{7}$$

where $L \in [0,N]$ is a dimensionless fitting parameter. The quantity f_{ij} equals the number of customers at node i who take a trip to node j divided by the number of customers who leave i under the following mechanistic model. Each customer at i considers opportunities in nondecreasing order of distance from i, and they accept each opportunity with probability L/N. They take a trip to the node j that has the first opportunity that they accept. One can show that the number of customers who take a trip from i to j divided by the total number who leave i is equal to the right-hand side of Eq. (7) (see Appendix E). However, because we use a doubly-constrained version of the IO model, f_{ij} gives only the attraction value of zone j to a customer in zone i. In a doubly-constrained model, the quantity $T_{ij}^{\mathrm{model}}/O_i = A_iB_jD_jf_{ij}$ equals the actual number of customers who make a trip from i to j divided by the total number of trips that originate at i.

C. Radiation model

The original radiation model [5], which was proposed as an alternative to gravity models, is a parameter-free variant of the IO model with attraction values

$$f_{ij} = f_{\text{rad}}(O_i, D_j, S_{ij}) = \frac{O_i D_j}{(O_i + S_{ij})(O_i + D_j + S_{ij})} . \quad (8)$$

The radiation model and its variants have been used for studies of commuter flows [5,36], human migration [24], mobile-phone calls [5], and other applications. An advantage of this version of the radiation model is that it has no parameters. However, it does not appear to do a good job of capturing human mobility on small spatial scales [9,29,37].

D. Extended radiation model

Yang *et al.* [38] proposed an extension of the radiation model that includes an exponent $\alpha \geqslant 0$. In this model, the attraction values f_{ij} are

$$f_{ij} = f_{\text{ext}}(O_i, D_j, S_{ij}) = \frac{[(O_i + S_{ij} + D_j)^\alpha - (O_i + S_{ij})^\alpha][(O_i)^\alpha + N^\alpha)}{((O_i + S_{ij})^\alpha + N^\alpha)[(O_i + S_{ij} + D_j)^\alpha + N^\alpha]} . \quad (9)$$

Yang *et al.* claimed that this extended radiation model fits empirical OD matrices better than the original radiation model for intracity commuting flow and observed that their calibrated values of α decreased as they considered systems with smaller spatial scales. When $\alpha = 1$, one recovers a variation of the original radiation model from Eq. (8) (specifically, with each occurrence of O_i replaced by $O_i + N$).

E. Goodness-of-fit measures

1. Common part of commuters (CPC)

The *common part of commuters* (CPC) score is the proportion of trips that the OD matrices $\boldsymbol{T}^{\text{data}}$ and $\boldsymbol{T}^{\text{model}}$ have in common:

$$\text{CPC}(\boldsymbol{T}^{\text{data}}, \boldsymbol{T}^{\text{model}}) = \frac{\sum_i \sum_j \min\left\{T_{ij}^{\text{data}}, T_{ij}^{\text{model}}\right\}}{\sum_i \sum_j \frac{1}{2}\left(T_{ij}^{\text{data}} + T_{ij}^{\text{model}}\right)} . \quad (10)$$

It was introduced in Refs. [37,39] and has been used in studies of human mobility [28,38,40]. The CPC score is based on the Sørensen index [41]; it varies from 0 (when there is no agreement between the model and data) to 1 (when $\boldsymbol{T}^{\text{data}}$ and $\boldsymbol{T}^{\text{model}}$ are identical). Because our models are doubly-constrained, $\sum_{i,j} T_{ij}^{\text{data}} = \sum_{i,j} T_{ij}^{\text{model}}$. Therefore, we interpret the CPC score as the fraction of customers whose trip is assigned correctly by a model.

There are various other goodness-of-fit measures, such as normalized root-mean-squared error, information gain, common part of edges, cosine distance, and Pearson product-moment correlation. However, in past studies, these measures often gave similar results as CPC when comparing the performance of mobility models [28,40], so we primarily use CPC, which has an intuitive interpretation in our modeling context.

2. Error in estimated number of zone visits

In addition to CPC, we also consider an application-specific goodness-of-fit measure NRMSE_v, which measures the normalized root-mean-square error (NRMSE) in the number of visits to each node. When we examine congestion in supermarkets in Sec. VI, we use measures of congestion that depend on the number of visits to each node, so a mobility model should have low values of NRMSE_v for it to be viable for our application to congestion. Given an OD matrix $\boldsymbol{T}$, we estimate the number of visits by assuming that each customer, for an OD trip (i, j), takes a shortest path, which we choose uniformly at random among all shortest paths from i to j. Each customer who takes a trip from i to j visits each node along the chosen shortest path. The estimated number v_k of visits to each node k is the weighted sum of the number T_{ij} of trips with OD pairs (i, j) for all i and j, where the weight ω_{ikj} is the fraction of shortest paths from i to j that traverse k. (We use the convention that the starting and terminal nodes, i and j, are traversed as part of a shortest path from i to j.) That is,

$$v_k = v_k(\boldsymbol{T}) = \sum_{i,j} \omega_{ikj} T_{ij} . \quad (11)$$

The number v_k of visits is closely related to geodesic node betweenness centrality [42], which we recover when $T_{ij} = 1$ for all (i, j). We can compute all v_k values in $O(nm)$ time using a straightforward adaptation of a fast algorithm for computing geodesic betweenness centrality [43].

To measure the model error in the estimated number of visits, we calculate the NRMSE in $v_k(\boldsymbol{T})$ with the formula

$$\text{NRMSE}_v = \left\{ \frac{\sum_{k=1}^{n} [v_k(\boldsymbol{T}^{\text{data}}) - v_k(\boldsymbol{T}^{\text{model}})]^2}{n v_{\max}(\boldsymbol{T}^{\text{data}})^2} \right\}^{\frac{1}{2}} , \quad (12)$$

where $v_{\max}(\boldsymbol{T}) = \max_k [v_k(\boldsymbol{T})]$ is the number of visits to the most-visited node.

Our shortest-path assumption is a modeling choice and seems likely to be somewhat unrealistic for describing the precise trajectories between purchases. For example, in one study [44], it was reported that customer trajectories between purchases are, on average, about four times as long as a shortest path. However, to the best of our knowledge, there does not exist a model that accurately describes how customers move between purchases. We use the shortest-path assumption as a null model, and we note that it is possible to replace this assumption with a more intricate model that estimates the number of visits to each node. Additionally, we use the shortest-path assumption only to estimate the number of visits to each node; we do not use this assumption when we estimate OD matrices.

F. Parameter calibration

Following the approach in Ref. [37], we calibrate the model parameters γ, L, and α of the gravity, IO, and extended radiation models (respectively) for each data set by maximizing the CPC score. We call a parameter value "optimal" when it maximizes the CPC score for a given model and data set.

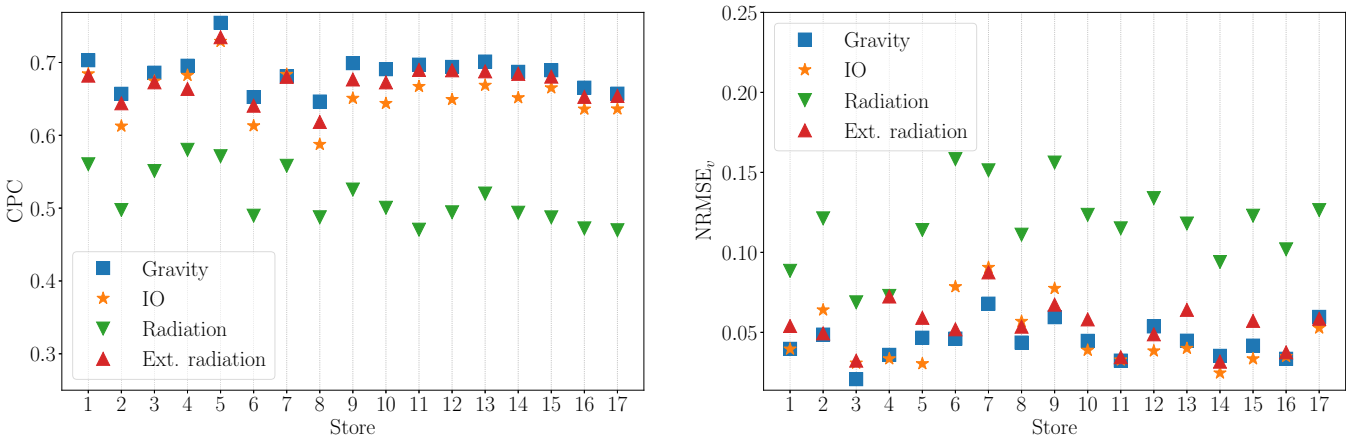

FIG. 2. (a) CPC scores and (b) NRMSE$_v$ values when fitting the gravity, intervening-opportunities, radiation, and extended radiation models to mobility-flow data from 17 supermarkets.

V. RESULTS

A. Fit to data

We test the four models in Sec. IV on each of the 17 stores. In our computations, the gravity model consistently achieves the best CPC score across the stores, with a mean of about 0.686 [see Fig. 2(a) and Table IV]. This value is comparable with reported CPC scores in previous studies on mobility systems with larger spatial scales [28,38,40]. The performance of the extended radiation model, with a mean CPC score of 0.672, is almost as successful on average. The IO model consistently yields lower CPC scores than the gravity and extended radiation models. The gravity model also has the best (i.e., lowest) mean value of NRMSE$_v$ across the 17 stores (see Table IV), closely followed by the IO model and then the extended radiation model. In terms of NRMSE$_v$, the relative performance of these three models is store-dependent [see Fig. 2(b)]. In some stores, the gravity model has the lowest value of NRMSE$_v$; in other stores, either the IO model or the extended radiation model achieves the lowest value. For each store, the radiation model performs the worst among the four models, yielding both the lowest CPC scores and the highest values of NRMSE$_v$ across all 17 stores. The poor performance of the radiation model is consistent with other studies of mobility systems on small spatial scales [9,29,37].

TABLE IV. Mean CPC scores and NRMSE$_v$ values from fitting the gravity, IO, radiation, and extended radiation models to mobility-flow data from 17 supermarkets. We list the models in decreasing order of their mean CPC score. We highlight the best value in each column in bold.

Model	Mean CPC	Mean NRMSE$_v$
Gravity	**0.686**	**0.045**
Ext. radiation	0.672	0.054
IO	0.655	0.047
Radiation	0.513	0.116

To further investigate the performance of the models, we examine the results of a single store (which we call "Store A") in more detail. Specifically, we examine the estimated OD matrix $\boldsymbol{T}^{\text{model}}$, the estimated number $v_k(\boldsymbol{T}^{\text{model}})$ of visits, and the distance distribution of the OD trips for each of the mobility models. The results for this store are qualitatively similar to the results for the other stores.

In Fig. 3, we compare the empirical number T_{ij}^{data} of trips with the estimated number T_{ij}^{model} of trips from each mobility model for each OD pair (i, j) of nodes in Store A. To evaluate the quality of T_{ij}^{model}, we create logarithmic bins from 1 to $\max_{i,j}[T_{ij}^{\text{data}}]$. For each bin, we consider all OD pairs whose empirical number of trips lies within the bin. We calculate the mean, median, and interquartile range for the estimated number T_{ij}^{model} of trips for these OD pairs (i, j). See the black box plots in Fig. 3.

On average, the estimated numbers of trips from all four models are close to their empirical numbers, except for OD pairs with a large number of trips. For these OD pairs, the gravity, IO, and the extended radiation models underestimate the number of trips. We tested whether this bias results from the presence of many very short trips between neighboring shelves on different sides of a zone boundary. However, this does not appear to be the case. We also examined a more general type of gravity model to explore whether it can improve the fit and/or remove this bias, but it did not. The underestimation by the models of the number of trips for OD pairs with many trips is an interesting and unexplained feature of the empirical data. The radiation model is effective at estimating the mean number of trips for most of the bins, but its overall performance is poor because of the large variance in its estimates for each bin [see Fig. 3(c)].

In Fig. 4, we compare the estimated number $v_k(\boldsymbol{T}^{\text{model}})$ of visits that we compute from the OD matrix $\boldsymbol{T}^{\text{model}}$ of the models with the number $v_k(\boldsymbol{T}^{\text{data}})$ of visits that we estimate using the empirical OD matrix $\boldsymbol{T}^{\text{data}}$ for Store A. We find that the gravity, IO, and extended radiation models are effective at estimating the number of visits for most nodes, except for some of the ones with a large number of visits. For these

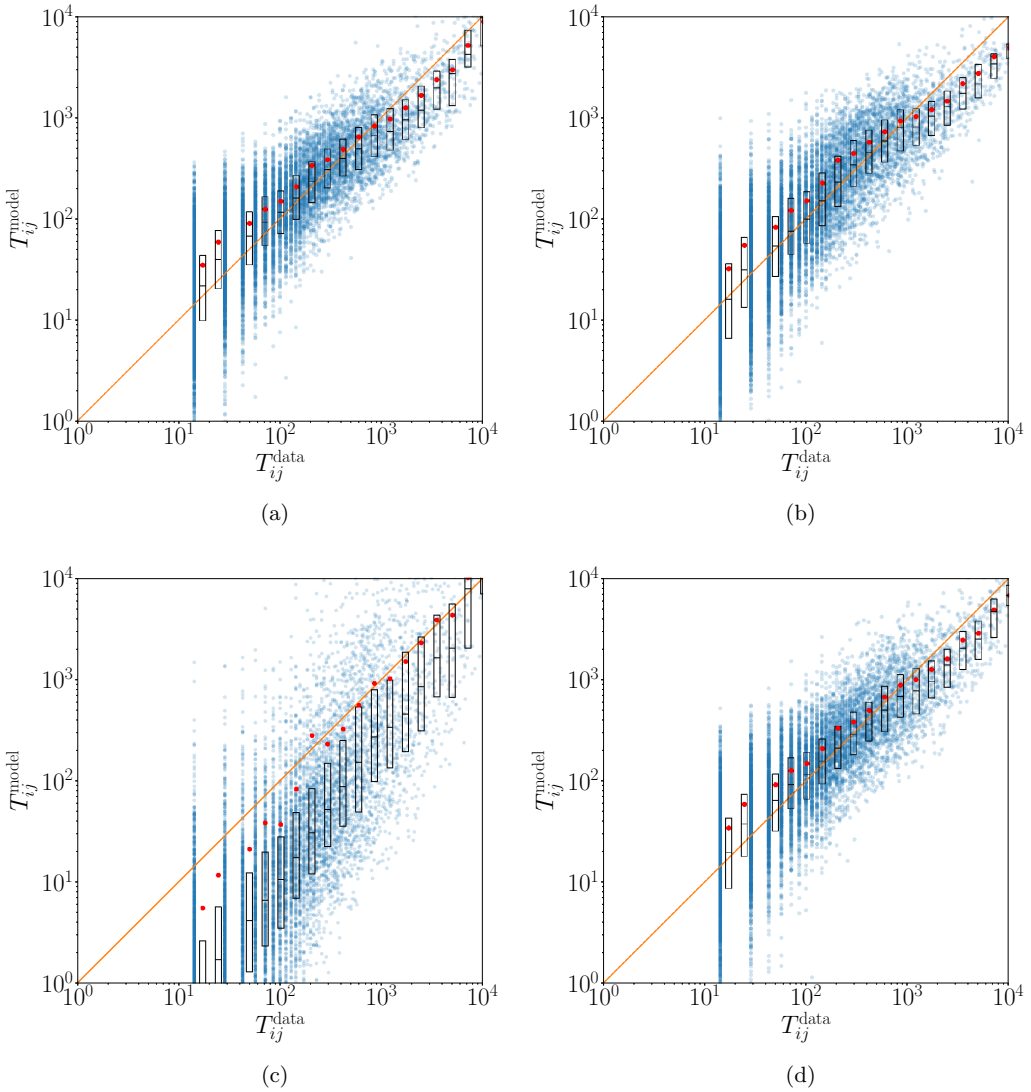

FIG. 3. Comparison (blue dots) between the number of trips in the data (T_{ij}^{data}) and the model estimates (T_{ij}^{model}) for the (a) gravity model, (b) IO model, (c) radiation model, and (d) extended radiation model. We also plot the mean number of trips that are estimated by the model (red dots) for each logarithmic bin of the data. The orange line is the identity line. Each box (in black) extends from the lower to the upper quartile values of the binned model estimate; we draw a horizontal line at the median.

nodes, the three models overestimate the number of visits. The radiation model underestimates the number of visits for most nodes [see Figure 4(c)].

In Fig. 5, we compare the distributions of trip distances in our models with the empirical distribution. The gravity, IO, and extended radiation models have trip-distance distributions that qualitatively resemble the empirical distribution. Among the four models, the trip-distance distribution from the gravity model is closest to the empirical distribution. The IO model underestimates the number of long-distance trips (specifically, those above 60 m), and the extended radiation model overestimates the number of these long-distance trips. The trip-distance distribution of the radiation model is qualitatively different from the empirical distribution (see Fig. 5).

In summary, among the models that we examine, the gravity model best fits the empirical mobility-flow data. On average, it successfully explains about 69% of the OD trips in the data sets. It also is effective at estimating the number of visits to each node, with NRMSE values of about 0.045 (see Table IV). The extended radiation and IO models are close behind; on average, they successfully explain the data of about 65%–67% of the OD trips. For the most part, these three models also yield trip-distance distributions that look similar to the empirical distribution. The radiation model does not fit the data well, so we exclude it from further consideration.

B. Sensitivity analysis: Parameter dependence of models

We explore how the performance of the gravity, IO, and extended radiation models depends on their respective model parameter values. For each model, let p_{opt} to be the optimal parameter value. We calculate the CPC scores for parameter values between 0 and $10p_{\text{opt}}$ (see Fig. 6). For each of the models, we observe progressively smaller CPC scores for parameter values that are progressively farther away from p_{opt}, so model performance depends on the parameter value. The

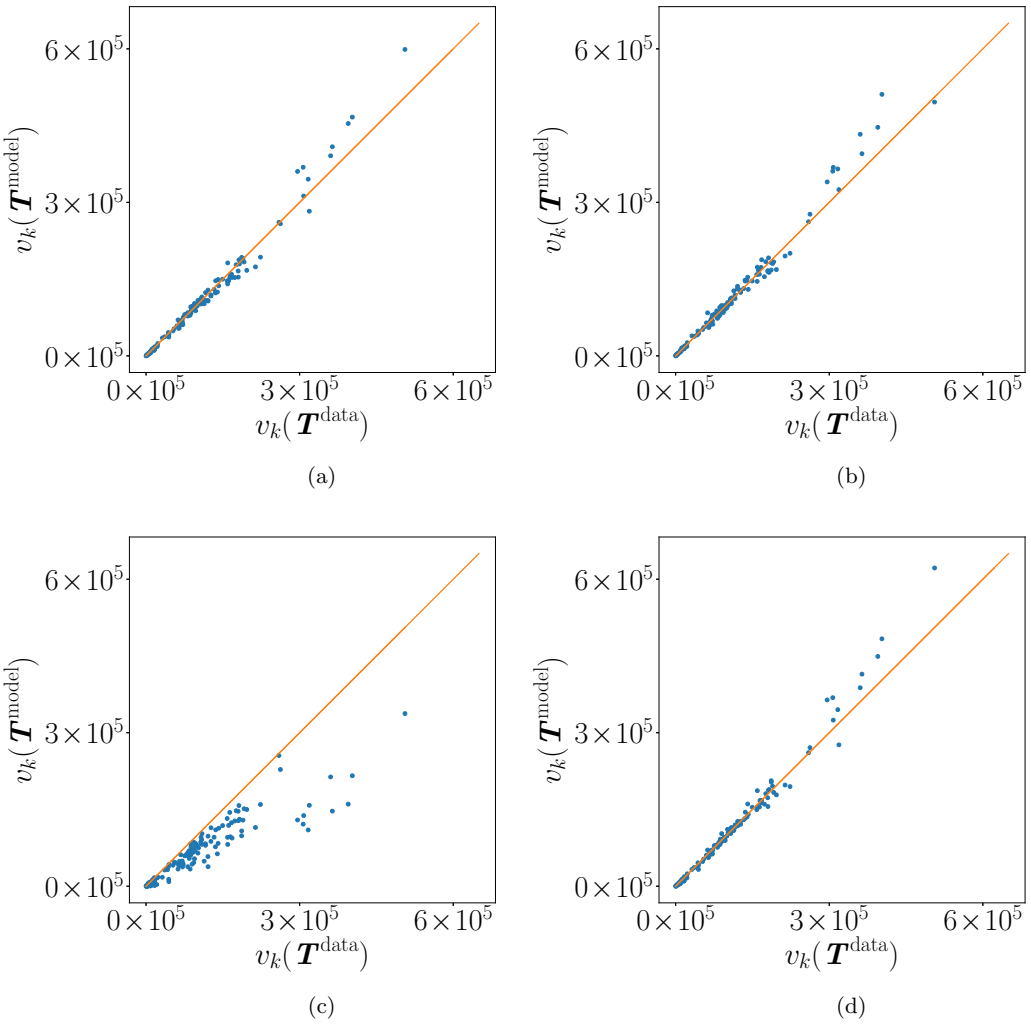

FIG. 4. Comparison of the estimated number v_k of visits for each node k between the data and the mobility models for the (a) gravity model (for which $\text{NRMSE}_v \approx 0.032$), (b) IO model (for which $\text{NRMSE}_v \approx 0.033$), (c) radiation model (for which $\text{NRMSE}_v \approx 0.115$), and (d) extended radiation model (for which $\text{NRMSE}_v \approx 0.034$). The orange line is the identity line. The gravity, IO, and extended radiation models give good fits to the number of visits to each node.

decrease in CPC score with distance from p_{opt} is steepest for the gravity model, second-steepest for the IO model, and shallowest for the extended radiation model. Interestingly, the CPC score for the extended radiation model plateaus as $\alpha \downarrow 0$ at a value close to the maximum CPC score. This suggests that a parameter-free special case of the extended radiation model, which we obtain by setting $\alpha = 0$, may perform well. (We do not explore this special case in this article.)

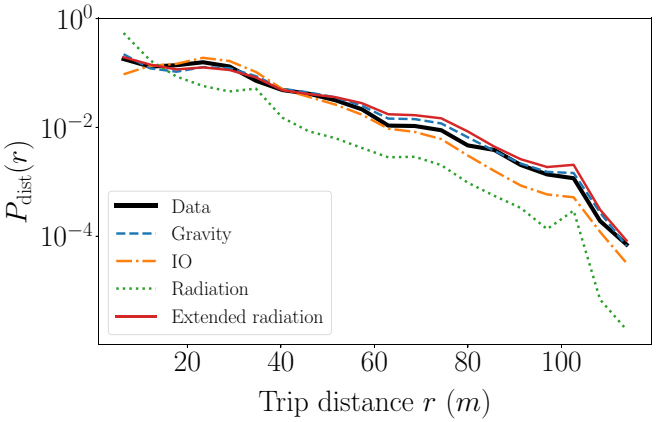

FIG. 5. Probability density function (PDF) of the trip-distance distribution.

C. Evaluation: Model performance on estimating trips in unseen data

We conduct two series of tests to analyze the performance of the gravity, IO, and extended radiation models at estimating mobility flow for a time period for a store for which we do not know the optimal parameter value. As we showed in Sec. V B, the performance of each of these models depends on the value of its associated parameter. In each test, we estimate mobility flow using the optimal parameter value from a different time period of the same store (in our first series of tests) or from a different store for the same time period (in our second series of tests). For each test, we compare the achieved CPC value CPC_{a} with the maximum CPC value CPC_{max} (which we obtain when using the optimal parameter value) of a model.

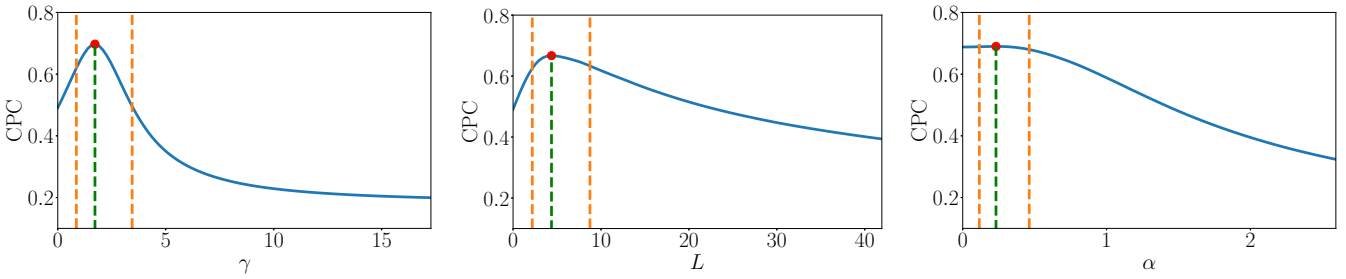

FIG. 6. CPC dependence on the model parameter for the (a) gravity, (b) IO, and (c) extended radiation models. We highlight the maximum CPC score with a red dot. The orange vertical lines are at $0.5p_{\text{opt}}$ and $2p_{\text{opt}}$.

We also compute their ratio

$$R = \frac{\text{CPC}_{\text{a}}}{\text{CPC}_{\text{max}}} \in [0, 1]\,. \tag{13}$$

A value of R that is close to 1 indicates that the estimated mobility flow using a parameter value that is optimal for a different time period or for a different store fits the empirical data just as successfully as using the optimal parameter value. In other words, these tests allow us to investigate whether the optimal parameter values of a model differ significantly between different time periods of the same store or between different stores.

In the first series of tests, we split the 91-day data set of each store into two parts. The first part is the mobility-flow data from the first 60 days; for this subset of the data, we find the optimal parameter value for each model. The second part is the mobility-flow data of the remaining 31 days; we estimate the mobility flow during this period using a mobility model with the empirical values of O_k and D_k (i.e., the number of OD trips that, respectively, start and end at zone k) from this period and the optimal parameter value from the initial 60-day period. We perform one test for each store, so there are 17 tests per model. The mean value of R is about 0.98 for each of the three models (see Table V). Therefore, the model parameter values do not change much across different time periods of the same store. It thus seems that our models are likely not overfitting data, as was also noted recently for these types of mobility models by Hilton *et al.* [45]. This also suggests that to estimate mobility flow of a store during some (sufficiently long) time period, we only need to know the values of O_k and D_k (which we can estimate from purchase data) during that time period and the optimal parameter value for a different (sufficiently long) time period of the same store.

In each test in our second series of tests, we estimate the 91-day mobility flow of one store using a mobility model with the optimal parameter value of another store from the same time period. We perform one test for each ordered pair of distinct stores, so there are $17 \times 16 = 272$ tests in total for each model. The mean value of R is above 0.99 for each of the three models (see Table VI). This suggests that the differences in model parameter values across stores are small and have little effect on the performance of the models. Therefore, we conclude that we can estimate the mobility flow of one store using the optimal parameter value from another store. This also suggests that if we change the layout of a store, the optimal model parameter values should not change appreciably.

VI. REDUCING CONGESTION IN SUPERMARKETS

We now use mobility models to estimate and reduce congestion in supermarkets. Our approach has three components:

(1) a congestion model, based on queuing networks, that estimates congestion from mobility flow $\boldsymbol{T}$;

(2) a mobility model that estimates the change of the flow $\boldsymbol{T}$ with a new store layout; and

(3) an optimization algorithm that finds store layouts with less congestion.

We describe these components in detail in Secs. VI A–VI C.

TABLE V. Mean, median, and minimum values of R (i.e., the ratio of our achieved CPC value to the maximum one) for the 17 stores in our first series of tests. In each test in this series, we use an optimal parameter value from one time period of a store to estimate mobility flow for a different time period of the same store.

Model	Mean R	Median R	Minimum R
Gravity	0.975	0.975	0.959
IO	0.978	0.977	0.963
Ext. radiation	0.975	0.973	0.960

TABLE VI. Mean, median, and minimum values of R (i.e., the ratio of our achieved CPC value to the maximum one) for the 272 tests in our second series of tests. In each test in this series, we use the optimal parameter value from one store to estimate mobility flow for a different store.

Model	Mean R	Median R	Minimum R
Gravity	0.996	0.998	0.976
IO	0.994	0.996	0.926
Ext. radiation	0.999	1.000	0.980

A. Congestion model

In our congestion model, each node acts as a queue. We suppose that the congestion model, which resembles the one in Ref. [46], is in a stationary state. (A key difference is that our model uses continuous time, whereas the model in Ref. [46] uses discrete time.) We take four inputs: (1) an OD matrix $\boldsymbol{T}$ with entries T_{ij}, which we calculate using one of the doubly-constrained mobility models in Sec. IV; (2) a time period τ over which we measure or estimate $\boldsymbol{T}$; (3) a store network $\mathcal{G}$, with its associated distance matrix $\boldsymbol{\Lambda}$; and (4) service rates μ_k for each node k. As we describe later in this section, we can estimate these rates from the mean customer dwell time at k.

In Sec. IV, we interpreted $\boldsymbol{T}$ as the flow from a random-walk model in which new customers arrive at a store's entrance and take trips to random destinations based on a transition matrix $\boldsymbol{P}$ whose entries are $P_{ij} = T_{ij}/\sum_k T_{ik}$ if $\sum_k T_{ik} > 0$ and $P_{ij} = 0$ otherwise. In this section, we instead interpret $\boldsymbol{T}$ as the mean mobility flow during a time period of length τ under the following model. New customers arrive at each node i (not just at the entrance) of a network (i.e., a supermarket) according to a Poisson process with rate $\sum_k T_{ik}/\tau$. Each customer chooses a random destination j with probability P_{ij}. Customers traverse the network by taking a shortest path from zone i to zone j. Customers queue at each node that they visit (for both traversal and shopping) to be served, where each node k is a single-server queue with exponential service rate μ_k [47]. After a customer is served at the destination node j, we remove it from the network. The quantity T_{ij} is then the mean number of customers who take a trip from node i to node j during a time period of length τ.

We can view the model in the formulation of the present section as a "decomposed' variant of the model in Sec. IV. There are n independent random walks, each of which starts at a different node in a network and ends after taking exactly one trip to a random destination, instead of a single random walk that always starts at the entrance node and terminates at the exit node after taking one or more trips. In the new formulation, the mean rate at which customers finish a trip at node k is the same as the mean rate at which new customers start a trip at k. By contrast, in the random-walk perspective of Sec. IV, the exact number of customers who finish at k is equal to the number of customers who start a trip at k during any time period. In other words, customers are "conserved" at each node only in a stochastic sense (i.e., on average during some period of time), rather than in an absolute sense.

We calculate v_k for each node k from $\boldsymbol{T}$ using Eq. (11). We need to separately consider situations with $\mu_k > v_k/\tau$ and $\mu_k < v_k/\tau$.

When $\mu_k > v_k/\tau$ for all k, the quantity v_k is the mean number of customer visits to k during a time period of length τ. The arrival rate λ_k at each node k is then $\lambda_k = v_k/\tau$. We call this situation a *free-flow state*. In this state, the queue size at each node k at stationarity is a geometric random variable with mean $\lambda_k/(\mu_k - \lambda_k)$ and is independent of the queue sizes of the other nodes [47]. The total mean queue size Q in this state is [48]

$$Q = \sum_{k=1}^{n} \frac{\lambda_k}{\mu_k - \lambda_k} \,. \tag{14}$$

When $\mu_k < v_k/\tau$ for some k, node k cannot serve customers sufficiently fast, and the number of customers who wait at the queue keeps increasing. Our system is in a *congested state* and cannot be stationary.

If we have information about the mean customer dwell time w_k at each node k, we can infer the empirical service rate μ_k of each node k using Little's law, which states that the mean queue size is equal to the mean dwell time multiplied by the rate of arrivals:

$$w_k \lambda_k = q_k \,, \tag{15}$$

where

$$q_k = \frac{\lambda_k}{\mu_k - \lambda_k} \tag{16}$$

is the mean queue size at k [48]. Combining Eqs. (15) and (16), we obtain the following formula for the empirical service rate:

$$\mu_k = \frac{1}{w_k} + \lambda_k \,. \tag{17}$$

Because we do not have empirical data for the service rate, we assume for simplicity that the service rates are homogeneous. That is, $\mu_k = \mu > 0$ for all k, so we are in a free-flow state if $\mu > \lambda_{\max}$, where $\lambda_{\max} = \max_k[\lambda_k]$.

We use the maximum arrival rate $\lambda_{\max}$ and the total mean queue size Q as proxies to measure congestion. The measure $\lambda_{\max}$, which does not depend on any parameters, is the minimum service rate that ensures that the system is in a free-flow state. It is also closely related to the traffic capacity ρ_c in the traffic-dynamics model of Arenas *et al.* [49] that has been used to model traffic on transportation and communication networks [46,49–53]. The traffic capacity ρ_c is an important performance measure of the traffic-dynamics model in [49]. It represents the maximum rate at which walkers (which, in our case, represent customers) arrive at a network from outside the system before the system reaches a congested state. In the traffic-dynamics model of Ref. [49], one fixes the service rate μ but varies the rate at which walkers arrive from outside the system (i.e., the *external arrival rate*). By contrast, we fix the external arrival rate and vary μ.

The total mean queue size Q, which measures congestion in a free-flow state, is equal to the total mean number of customers in a store. By Little's law [54], a store layout that minimizes Q also minimizes the mean trip time. Unlike $\lambda_{\max}$, the total mean queue size Q depends on μ, so we perform separate optimizations for different values of μ.

The measures $\lambda_{\max}$ and Q are correlated with each other, as Q is a sum that is dominated by the terms from nodes k with large values of λ_k, so store layouts with smaller values of Q often also have smaller values of $\lambda_{\max}$.

B. Mobility model

We focus on the gravity model to estimate changes in the OD matrix $\boldsymbol{T}^{\text{model}}$ when changing a store's layout, because it provides the best fit to the data, both in terms of the CPC score and in the estimated number v_k of visits (see Table IV). We assume that we can swap the locations of nodes (which

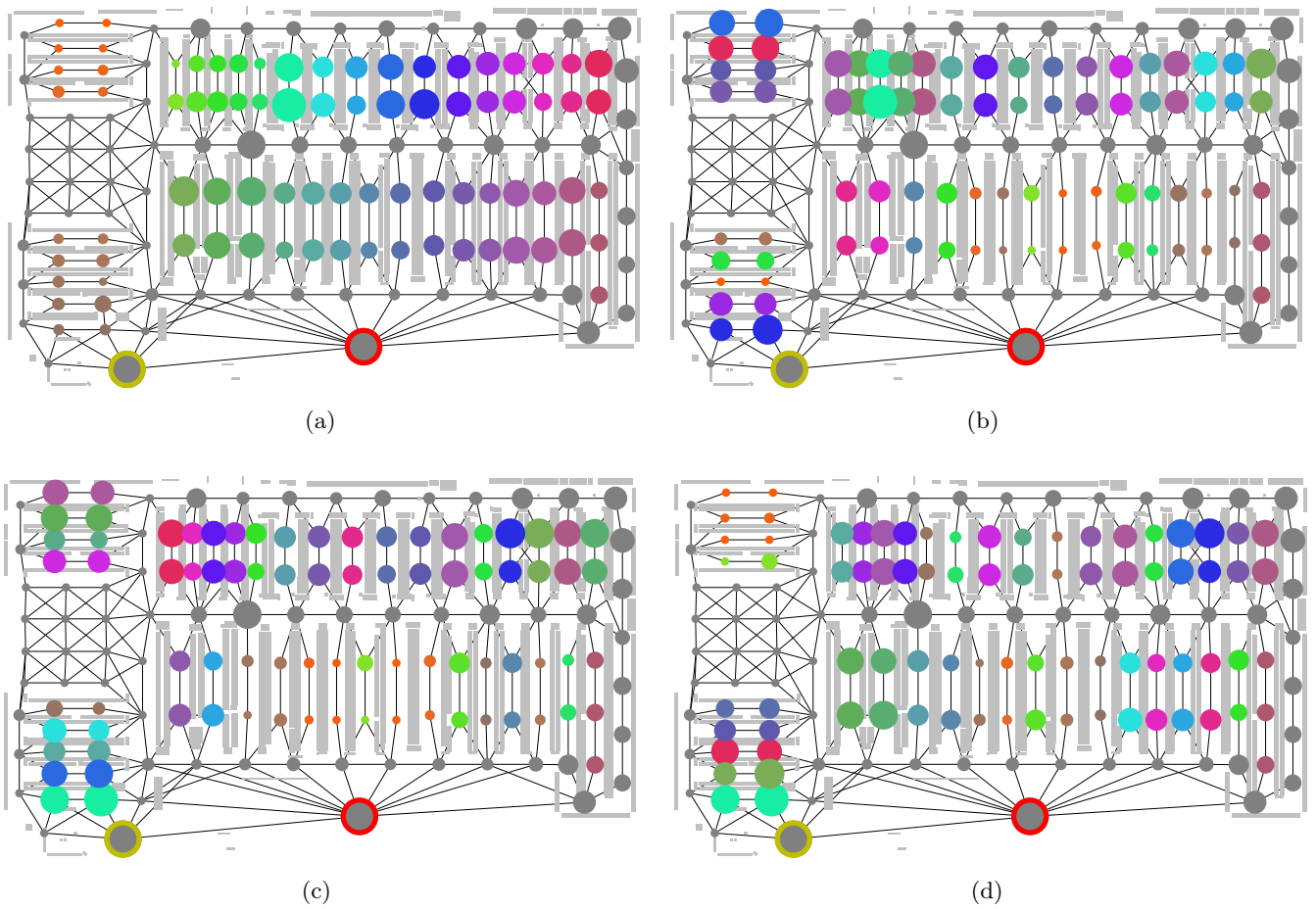

FIG. 7. Location of popular nodes before and after optimization. We show (a) the original store layout, (b) the optimized store layout when minimizing $\lambda_{\max}$, (c) the optimized store layout when minimizing Q with $\mu = 7500$, and (d) the optimized store layout when minimizing Q with $\mu = 15\,000$. Nodes of the same color are in the same aisle, and gray nodes are not part of any aisle. The size of each node k is proportional to $O_k + D_k$ (i.e., to the sum of the numbers of trips that start and end at zone k). We circle the entrance and till nodes in yellow and red, respectively.

corresponds to swapping the contents of their shelves), but that we cannot change the store network topology or edge distances in any other way. To ensure that similar items stay with one another in the same aisle, we add a further constraint (which we call the *aisle constraint*) that we can only swap an aisle (which consists of a set of nodes) with another aisle with the same number of nodes. However, we do allow permuting of nodes within the same aisle. (In Appendix G, we also report our results when we relax the aisle constraint. These results are of similar quality to our more constrained approach in this section.) We do not consider adding or removing nodes or edges, as such changes are often costly. We highlight the nodes of an aisle in Fig. 7(a) by coloring them. Nodes of the same color are in the same aisle, and gray nodes are not part of any aisle.

Crucially, we need a hypothesis for how O_k and D_k (i.e., the numbers of trips to and from a node k) change when we change the location of a node k. We assume that O_k and D_k depend only on the items inside a zone and not on the zone's location. Therefore, when we change the location of a node k, we assume that the node has the same values of O_k and D_k in the new location. In other words, we are assuming that the number of shopping visits at node k (i.e., the number of times that customers visit k to purchase items) is independent of its location. This is a key assumption of our model. For nodes with many essential items, such as bread and milk, this assumption seems justifiable, as customers buy such items regardless of their location in a store. However, we anticipate that this assumption breaks down for nodes in which most items are either less essential or purchased with less (or no) planning.

C. Optimization algorithm

We use a simulated-annealing (SA) algorithm [55] to find permuted layouts of a store with smaller values of one of the two objective functions ($\lambda_{\max}$ or Q). Our SA algorithm swaps two aisles, which we choose uniformly at random from all pairs of aisles with the same number of nodes and centroids that are less than 25 m apart. After swapping the two aisles, we permute the nodes within each aisle, where we choose the permutation uniformly at random from all possible permutations. We list the parameters of the SA algorithm in Appendix F.

TABLE VII. Minimum and mean values of objective functions of the final store layouts from 20 runs of the SA algorithm for optimizing Store A. For each objective function, we show the original value of the objective function, its minimum final value across 20 runs of the optimization algorithm, and its mean final value across the 20 runs. In parentheses, we indicate the amount of improvement as a percentage.

Objective function	Original	Minimum value	Mean value
$\lambda_{\max}$	6575	5009 (−23.8%)	5042 (−23.3%)
Q (with $\mu = 7500$)	38.10	29.10 (−23.6%)	29.28 (−23.1%)
Q (with $\mu = 15\,000$)	12.78	11.86 (−7.2%)	11.93 (−6.7%)

D. Optimization results

We optimize a store's layout (specifically, the layout of Store A) with the SA algorithm for three examples, in which we minimize

(1) $\lambda_{\max}$,
(2) Q with $\mu = 7500$, and
(3) Q with $\mu = 15\,000$.

For $\mu = 7500$, each node serves incoming customers at a rate of 7500 customers per day, which amounts to 12.5 customers per minute in a store that is open for 10 hours. The maximum arrival rate $\lambda_{\max}$ in the original store layout is 6575, so a service rate of $\mu = 7500$ is an example in which the most popular nodes have long mean dwell times. For example, in a store that is open for 10 hours, the most popular node has a mean dwell time of about 38 sec [which we calculate using Eq. (15)]. Our example with $\mu = 15\,000$, a value much larger than the original $\lambda_{\max}$, corresponds to a scenario in which customers typically have short mean dwell times in all nodes. (In this case, the most popular node has a mean dwell time of about 4.2 sec.) We perform each optimization 20 times and report our results in Table VII. For all three examples, the SA algorithm produces store layouts with objective-function values that are significantly smaller than their values in the original store layout. The relative reduction in Q is smaller for $\mu = 15\,000$ than it is for $\mu = 7500$. This is not surprising, because when $\mu = 7500$, a larger fraction of the total mean queue size Q comes from nodes with the highest arrival rates in the original network. For example, the sum of the mean queue sizes of the three nodes with the highest arrival rates contribute 39% of the value of Q for $\mu = 7500$. By contrast, these three nodes contribute only 15% of the value of Q for $\mu = 15\,000$. Therefore, store layouts that lower the arrival rates of the most congested nodes (while the arrival rates of the other nodes do not increase much) tend to have lower values of Q for $\mu = 7500$. For $\mu = 15\,000$, because the most congested nodes contribute less to Q than for they do for $\mu = 7500$, achieving a major relative reduction in Q requires reducing the arrival rates of a larger number of nodes. This is potentially a very difficult task. Therefore, our observation of a lower reduction in Q for $\mu = 15\,000$ than for $\mu = 7500$ is consistent with expectations.

We measure the popularity of node k by $O_k + D_k$ (i.e., by the sum of the numbers of trips that start and end at k). For each node except the entrance and till nodes, $O_k + D_k$ is equal to twice the number of shopping visits to that node. In the networks with the smallest Q with $\mu = 7500$ and $\lambda_{\max}$, our optimization tends to move popular nodes from the center of Store A towards the left and top of the store [see Fig. 7(b)]. By contrast, when $\mu = 15\,000$, many popular nodes remain in the center of Store A [see Figs. 7(c) and 7(d)]. However, our optimization moves some of them to the store's bottom-left area, which previously was not a popular area.

VII. CONCLUSIONS AND DISCUSSION

We employed several population-level mobility models to investigate customer mobility flow between zones in supermarkets, whose spatial scales are much smaller than in previous uses of these models. We estimated origin–destination (OD) matrices, which describe empirical mobility flow, for 17 supermarkets from anonymized and ordered customer-basket data (where a customer's "OD trip" is either a trip between consecutive purchases, a trip from the entrance to the first purchase, or a trip from the last purchase to the tills). We fit the mobility models to empirical distributions of customer OD trips and examined the adjustment of store aisles to reduce congestion in supermarkets.

Among the models that we studied, the gravity model gave the best fit to the empirical mobility flow (it successfully estimated about 69% of the flow on average), and the extended radiation and intervening-opportunities (IO) models were almost as successful. This illustrates that one can successfully use population-level mobility models for applications on spatial scales of tens to hundreds of meters.

In our investigation, we estimated the number v_k of visits to each node k from mobility flow by assuming that each customer traverses a shortest path, and we found that our estimations from the OD matrices from the gravity, IO, and extended radiation models agree well with the total number of visits that we estimated from empirical OD matrices. (We used the shortest-path assumption only to estimate the number of visits; our estimates of mobility flow do not rely on this assumption.) Additionally, the gravity, IO, and extended radiation models yield trips with similar distance distributions to the empirical distributions. However, consistent with other studies on small spatial scales (which generally have been in intra-urban settings) [9,29,37], the basic radiation model was not successful at reproducing features of the data.

The gravity, IO, and extended radiation models each have one parameter, and their performance depends on the value of their parameter. In our investigation, we found that it is sufficient to use the "optimal" model parameters that we calibrated on a single store to give good estimates of the mobility flows of all other stores. The only additional information that we needed for the other stores is the number of trips from and to each node; one can estimate these quantities from the purchase data of these stores. For a given store, we were also successful at using the models to estimate mobility flow during a time period using a parameter value from fitting to data from another time period of the same store. Given our success at translating optimal parameter values across both stores and time periods, our approach provides a potentially valuable test bed for experimentation by supermarket companies using sales data from existing stores before trying out new store layouts.

Finally, we showed how to use the gravity model in conjunction with a congestion model—with tests using congestion that we measured using either the maximum number

$\lambda_{\max}$ of visits or the total mean queue size Q—and an optimization algorithm to reduce congestion in supermarkets. We considered a congestion model in which each node acts as a queue with service rate μ, assumed that customers traverse a shortest path between two nodes, and explored the space of store layouts in which one can permute aisles (but one cannot permute individual store zones, except within the same aisle). We then used the gravity model to estimate how mobility flow changes from permutations of a store layout. In the layouts that we obtained by minimizing $\lambda_{\max}$ or minimizing Q with low service rate μ, popular nodes (as measured by the number of trips from and to a node) move from the center of a store to the left and upper perimeters. By contrast, in the layouts that we obtained by minimizing Q with a high service rate μ, some popular nodes move to a previously unpopular corner of a store.

There are several ways to build on our work. Possibilities include further development of mobility models and congestion models, analyzing seasonal effects and customer heterogeneity, allowing service rates to be heterogeneous, exploring the effects of our choice of space discretization, and applying our approach to situations other than supermarkets. We discuss several of these items in the following paragraphs.

In our investigation, we inferred empirical mobility flow from anonymized, ordered basket data of the mobility of a relatively small sample of customers (approximately 7%) from 17 supermarkets. Naturally, this sample also has certain biases, as our data consist primarily of baskets from regular customers. It is likely that these customers possess better knowledge than other customers of the stores in which they shop (given that they do so regularly), so their mobility patterns may not be representative of all customers of a given store.

We have also neglected temporal information and seasonal effects in our data by aggregating the mobility flow over $\tau = 91$ days. However, we expect mobility flow to be different at different times of a day (and on weekdays versus weekends) and at different times of a year (e.g., during certain holidays). We also expect different zones of a store to be the most congested ones at different times. Given sufficient data, one can apply mobility models to data that is segmented by the time of day or by the day of a year and then compare the parameter values from independent fitting to data of different time periods.

To study congestion, we used a simple routing and congestion model (using queues in each zone of a store). We assumed shortest-path routing, but some researchers have noted that customers deviate from shortest paths between purchases [44,56]. It is important to improve understanding of the routes that customers take between purchases. One possible approach is to use anonymized customer-trajectory data to develop and calibrate a stochastic routing model (e.g., using a variant of a random walk, perhaps with probabilities that depend on heterogeneous fitness values for different zones of a store). One can incorporate such a routing model in a straightforward way into our framework to better estimate the number v_k of visits to each zone k. When we estimated the mobility flow of different store layouts, we assumed that the number of customer shopping visits at node k is fixed and does not depend on the location of k. However, the location of a zone that contains items that are typically bought in an unplanned way likely affects the number of shopping visits to that zone. One can incorporate increasingly accurate models of the number and zone distribution of shopping visits into our framework to improve estimates of the mobility flow from different store layouts. Additionally, more empirical research is necessary to attain a detailed mechanistic understanding of the causes of congestion in supermarkets. We modeled congestion as queues in a zone; if such a model is conceptually accurate, we can incorporate more realistic types of queues (e.g., with variable service rates or with customers who do not enter a queue if it is too long). One can infer service rates using a method that is analogous to what we described in Sec. VI, provided one possesses data on customer dwell time (or can somehow infer such times) for each zone of a store.

Another consideration is the choice of space discretization and spatial resolution, and it is necessary to examine how such choices affect qualitative results of both mobility models and congestion models. (In our work, we divided each store into zones of approximately similar size, with zone lengths of about 7 m.)

One can apply our approach for modeling mobility flow and congestion at any spatial scale, and we expect that one can implement our methodology in practice in systems in which one can modify the underlying spatial structure. Many such applications have small spatial scales, as rewiring a small system is often a lot less costly than rewiring a large one. For example, when considering commuting flow, one cannot change the locations of countries or buildings. However, one can apply our tools for modeling mobility flow and congestion in a museum and use our optimization procedure to suggest better locations for the exhibits. Other examples include poster sessions in academic conferences and food stations in buffet restaurants. Applying our approach to these settings will help reveal which of our findings are specific to mobility flow in supermarkets and which ones apply more generally to human mobility on small spatial scales.

ACKNOWLEDGMENTS

F.Y. acknowledges support from the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1) in collaboration and Tesco PLC. M.B.D. acknowledges support from the Oxford–Emirates Data Science Lab. We also thank David Allwright, Robert Armstrong, Chico Camargo, Albert Díaz-Guilera, Scott Hale, Renaud Lambiotte, Neave O'Clery, Trevor Sidery, and an anonymous referee for helpful comments.

APPENDIX A: DEFINING ZONES IN A STORE

In this appendix, we discuss how we define the zones in a store.

First, we manually identify the aisles in a store. Each aisle is a long rectangle between two lines of shelves. The edges of each aisle are either parallel to the horizontal axis or parallel to the vertical axis on a floor plan. (In other words, there are no tilted aisles.)

We then subdivide each aisle into Z rectangular zones of equal dimensions, where we choose the number Z of zones

such that the length of the zones in the aisle is as close as possible to 7 m (our targeted mean zone length). We define the "length" of a zone as the longer side of the rectangle that encloses the zone. The number Z of zones can be different in different aisles; it is typically either two or three. The width of each aisle is smaller than 7 m, so if an aisle is horizontal (i.e., the longer edges of the aisle are parallel to the horizontal axis), then all zones in the aisle are also horizontal. Similarly, if an aisle is vertical (i.e., the longer edges of the aisle are parallel to the vertical axis), all zones are also vertical.

We then manually place the entrance zone and the till zone. We fill the remaining floor area of a store with long, large rectangles whose widths are similar to those of the store aisles. We divide each of these large rectangles into zones of equal size (where zones for different large rectangles may have different sizes). See Ref. [57] for a lengthier discussion of how we define zones.

APPENDIX B: ITERATIVE PROPORTIONAL-FITTING PROCEDURE FOR DETERMINING A_i AND B_j

In the doubly-constrained mobility models, we are given O_i, D_j, and f_{ij} for all nodes i and j. We seek to determine A_i and B_j (the balancing factors) that satisfy

$$O_i = \sum_j T_{ij}^{\text{model}} = \sum_j A_i O_i B_j D_j f_{ij} \,, \tag{B1}$$

$$D_j = \sum_i T_{ij}^{\text{model}} = \sum_j A_i O_i B_j D_j f_{ij} \,. \tag{B2}$$

Rearranging Eqs. (B1) and (B2) yields

$$A_i = \left(\sum_j B_j D_j f_{ij} \right)^{-1} , \tag{B3}$$

$$B_j = \left(\sum_i A_i O_i f_{ij} \right)^{-1} . \tag{B4}$$

In our iterative proportional-fitting procedure, we initialize $A_i = 1$ for all i. We then calculate B_j using Eq. (B4) from A_i, followed by an update of A_i using Eq. (B3). We repeat this procedure until the values on the right-hand sides of Eqs. (B1) and (B2) are close (specifically, within 1%) of the values on the left-hand sides. In our computations, this procedure converged within 1000 iterations for all of the employed models.

APPENDIX C: PERFORMANCE OF THE DOUBLY-CONSTRAINED GRAVITY MODEL WITH AN EXPONENTIAL DETERRENCE FUNCTION

In the doubly-constrained gravity model with an exponential deterrence function, the OD matrix $\boldsymbol{T}^{\text{model}}$ is given by Eq. (3) with

$$f_{ij} = f_{\text{g}}(O_i, D_j, d_{ij}) = O_i D_j e^{-\gamma d_{ij}/l} . \tag{C1}$$

We find that both the CPC score and NRMSE$_v$ are slightly worse on average than for the gravity model with a power-law deterrence function (see Table VIII).

TABLE VIII. Mean CPC scores and NRMSE$_v$ when fitting the doubly-constrained gravity model with an exponential deterrence function versus the doubly-constrained gravity model with a power-law deterrence function.

Model	Mean CPC	Mean NRMSE$_v$
Gravity (exponential)	0.677	0.049
Gravity (power law)	0.686	0.045

APPENDIX D: ESTIMATING $\{O_k\}_{k=1}^n$ AND D_1^{data} FROM PURCHASE DATA

In our models, we used $\{O_k\}_{k=1}^n$ and D_1^{data} to calculate mobility flow. In the main text, we assumed that we know these values. In this section, we show how to estimate $\{O_k\}_{k=1}^n$ and D_1^{data} from customer-level purchase data.

For each customer c, our data include a list of items that were purchased during a shopping journey. Using item-location data, we can identify the possible zones in which a customer could have picked up each item. We assume that all items in the same zone were picked up in one visit by a customer, so customers do not visit a zone more than once. The main challenge is how to account for multilocated items. We calculate the number $O_k^{\text{data},c}$ of trips that start from nodes $k = 1, \ldots, n$ for each customer c as follows. If a customer buys an item that is located only in zone k, we set $O_k^{\text{data},c} = 1$. Otherwise, we check whether a customer buys an item that is located both in zone k and in other zones. (In other words, there are multiple possible zone locations for that item.) If this is the case, let M_c be the number of such purchased items, and let $N_1, \ldots, N_{M_c}$ be the numbers of possible zone locations for the items. We set $O_k^{\text{data},c} = \min\{\sum_{l=1}^{M_c} 1/N_l, 1\}$. Therefore, each item that is located in zone k and in $N-1$ other zones counts as a fraction $1/N$ of a visit, with $O_k^{\text{data},c}$ capped at 1.

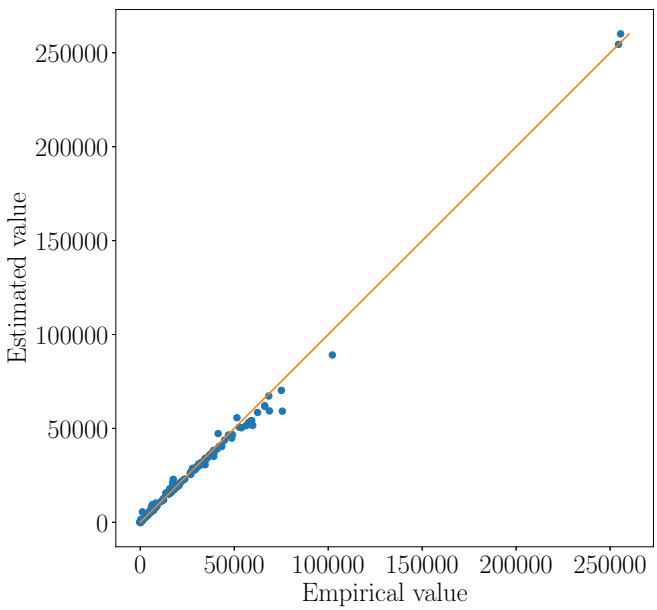

FIG. 8. Comparison of our estimates and empirical values of O_k, D_1^{data}, and D_n^{data} for Store A.

If a customer has not purchased any items that are located in zone k, then $O_k^{\text{data},c} = 0$. By construction, the value of $O_k^{\text{data},c}$ is at most 1, and it can take a fractional value when a customer buys (one or more) items that are located both in k and in other zones.

We calculate O_k (with $k = 1, \ldots, n$) with the formula

$$O_k = \sum_c O_k^{\text{data},c} + \delta_{k1} C \,, \tag{D1}$$

where C is the number of journeys in the data set and δ_{k1} is the Kronecker delta. The term $\delta_{k1}C$ accounts for the first trips of each shopping journey, as these start at the entrance of a store. We calculate D_1^{data} with the formula

$$D_1^{\text{data}} = \sum_c O_1^{\text{data},c} . \tag{D2}$$

We rescale $\{O_k\}_{k=1}^n$ and D_1^{data} by dividing by ρ (i.e., by the number of baskets in the data set during the time period divided by the total number of baskets during the time period).

In Fig. 8, we compare the empirical values of $\{O_k\}_{k=1}^n$ and D_1^{data} with ones that we estimate for Store A. In our calculations, the estimated values are close to the empirical ones, so we conclude that we can estimate O_k and D_k using only purchase data.

APPENDIX E: DERIVATION OF THE ATTRACTION FACTOR f_{ij} IN SCHNEIDER'S IO MODEL

In Schneider's IO model, each customer who leaves node i considers each opportunity in nondecreasing order of distance from i (i.e., from the nearest node to the farthest one). For simplicity, we assume that there are no equidistant nodes. A customer accepts the current opportunity with probability L/N, where $L \in [0, N]$ is a dimensionless fitting parameter and $N = \sum_i O_i^{\text{data}}$ is the total number of trips [58]. Upon accepting an opportunity at j, a customer takes a trip from node i to node j and thus does not consider any other opportunities. A customer who has not accepted any opportunity restarts the process and considers all opportunities again in nondecreasing distance from i. This process continues until the customer accepts an opportunity.

We calculate the probability of a customer accepting an opportunity at node j as follows. Let M be the number of rejected opportunities in the iteration when an opportunity is accepted (and thus the final iteration). The random variable M has a truncated geometric distribution. A customer who accepts an opportunity at node j must reject all S_{ij} opportunities at closer nodes and accept one of the D_j opportunities at j. In other words, a customer in zone i takes a trip to zone j if $S_{ij} + D_j > M \geqslant S_{ij}$. The probability of rejecting at least k opportunities is

$$\mathbb{P}(M \geqslant k) \propto \left(1 - \frac{L}{N}\right)^k \approx \exp\left(-\frac{kL}{N}\right), \quad k < N \,, \tag{E1}$$

where the approximation becomes exact as N and k tend to infinity with k/N constant. Therefore, the probability that a customer accepts an opportunity at node j is

$$\begin{aligned}\mathbb{P}(S_{ij} + D_j > M \geqslant S_{ij}) &= \mathbb{P}(M \geqslant S_{ij}) - \mathbb{P}(M \geqslant S_{ij} + D_j) \\ &\approx e^{-\frac{L}{N} S_{ij}} - e^{-\frac{L}{N}(S_{ij}+D_j)} = f_{ij} \,. \end{aligned} \tag{E2}$$

In Eq. (E2), the quantity f_{ij} equals the number of customers who make a trip from i to j divided by the number of customers who leave i. However, because we use a doubly-constrained model, f_{ij} gives only the attraction value of zone j to a customer in zone i. In a doubly-constrained model, the quantity $T_{ij}^{\text{model}}/O_i = A_i B_j D_j f_{ij}$ equals the actual number of customers who make a trip from i to j divided by the total number of trips that originate at i.

APPENDIX F: PARAMETERS OF THE SIMULATED-ANNEALING ALGORITHM

We set the initial computational temperature to 200 when minimizing $\lambda_{\max}$ and to 20 when minimizing Q. We use a cooling schedule in which we reduce the temperature by 0.18% of the current temperature at each step, and we use 5000 steps in total. We use the standard acceptance probability function $\exp(-\Delta E/T)$, where ΔE is the change in the objective-function value of the current step and T is the computational temperature.

APPENDIX G: RESULTS OF THE SIMULATED-ANNEALING ALGORITHM WITHOUT THE AISLE CONSTRAINT

We now apply the SA algorithm without the aisle constraint. Specifically, in one swapping step, we choose an edge uniformly at random among all edges, except for those that are incident to the entrance or till nodes; if we accept the step, we swap the two nodes that are incident to chosen edge. We use the same parameters that we described in Appendix F. The SA algorithm without the aisle constraint finds store layouts

TABLE IX. Minimum and mean values of objective functions of the final store layouts from 20 runs without the aisle constraint and 20 runs with the aisle constraint of the SA algorithm for optimizing Store A. For each objective function, we show the original value of the objective function, its minimum final value across 20 runs of the optimization algorithm, and its mean final value across the 20 runs. In parentheses, we indicate the amount of improvement as a percentage.

		Without the aisle constraint		With the aisle constraint	
Objective function	Original	Minimum value	Mean value	Minimum value	Mean value
$\lambda_{\max}$	6575	5040 (−23.3%)	5150 (−21.7%)	5009 (−23.8%)	5042 (−23.3%)
Q (with $\mu = 7500$)	38.10	27.21 (−28.6%)	28.02 (−26.5%)	29.10 (−23.6%)	29.28 (−23.1%)
Q (with $\mu = 15\,000$)	12.78	11.06 (−13.5%)	11.23 (−12.1%)	11.86 (−7.2%)	11.93 (−6.7%)

with smaller values of Q on average than when we run the algorithm with the aisle constraint (see Table IX). We find that the relative decrease in Q is smaller for $\mu = 7500$ than for $\mu = 15\,000$. Interestingly, the store layouts that we obtain when minimizing $\lambda_{\max}$ have larger values of $\lambda_{\max}$ without the aisle constraint than with the constraint. This suggests that the SA algorithm, which is a heuristic algorithm, is not very efficient at exploring the state space of store layouts, as it is unable to find the store layouts (with small values of $\lambda_{\max}$) that we obtained when including the aisle constraints.

[1] G. K. Zipf, Am. Sociol. Rev. **11**, 677 (1946).
[2] S. A. Stouffer, Am. Sociol. Rev. **5**, 845 (1940).
[3] M. Schneider, Papers Regional Sci. **5**, 51 (1959).
[4] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, Phys. Rep. **734**, 1 (2018).
[5] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature (London) **484**, 96 (2012).
[6] J. H. Bergstrand, Rev. Econ. Stat. **67**, 474 (1985).
[7] D. Piovani, E. Arcaute, G. Uchoa, A. Wilson, and M. Batty, R. Soc. Open Sci. **5**, 171668 (2018).
[8] J. E. Anderson, Am. Econ. Rev. **69**, 106 (1979).
[9] A. P. Masucci, J. Serras, A. Johansson, and M. Batty, Phys. Rev. E **88**, 022812 (2013).
[10] In describing spatial scales as "large" or "small," we are considering sizes in the context of human mobility. In relative terms, our "small" spatial scales are much smaller for a human than for an ant.
[11] M. Gutiérrez-Roig, O. Sagarra, A. Oltra, F. Bartumeus, A. Díaz-Guilera, and J. Perelló, R. Soc. Open Sci. **3**, 160177 (2016).
[12] D. Helbing and P. Molnar, Phys. Rev. E **51**, 4282 (1995).
[13] However, as we show in Sec. IV, we can interpret population-level mobility models as flows that arise from random walks.
[14] J. U. Farley and L. W. Ring, Oper. Res. **14**, 555 (1966).
[15] A definition of edge length that is closer to the walking distance between two nodes is the length of a shortest path that does not cross any shelves between the centroids of zones i and j. We call this distance the "shelf-respecting shortest-path distance." However, the difference between the Euclidean distance and shelf-respecting shortest-path distance is small (in most cases, the straight line between the centroids of two zones only clips a corner of a shelf), so we use the former because it is considerably faster to compute.
[16] A. G. Wilson, Environ. Plan. A **3**, 1 (1971).
[17] W. E. Deming and F. F. Stephan, Ann. Math. Stat. **11**, 427 (1940).
[18] N. Masuda, M. A. Porter, and R. Lambiotte, Phys. Rep. **716–717**, 1 (2017).
[19] For sufficiently large τ, it would be interesting to generalize our analysis to incorporate seasonal shopping variations, such as differences in behavior during holidays.
[20] H. C. Carey, *Principles of Social Science*, Vol. 3 (J. B. Lippincott & Company, Philadelphia, 1867).
[21] A. G. Wilson, *Entropy in Urban and Regional Modelling* (Pion, London, 1970).
[22] S. Erlander and N. F. Stewart, *The Gravity Model in Transportation Analysis: Theory and Extensions*, Vol. 3 (CRC Press, Boca Raton, FL, 1990).
[23] W.-S. Jung, F. Wang, and H. E. Stanley, EPL (Europhys. Lett.) **81**, 48005 (2008).
[24] S. H. Lee, R. Ffrancon, D. M. Abrams, B. J. Kim, and M. A. Porter, Phys. Rev. X **4**, 041009 (2014).
[25] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, J. R. Soc. Interface **7**, 1093 (2010).
[26] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, J. Stat. Mech. (2009) L07003.
[27] W. Luo and F. Wang, Environ. Plan. B **30**, 865 (2003).
[28] M. Lenormand, A. Bassolas, and J. J. Ramasco, J. Transport Geogr. **51**, 158 (2016).
[29] X. Liang, J. Zhao, L. Dong, and K. Xu, Sci. Rep. **3**, 2983 (2013).
[30] E. R. Ruiter, Trans. Res. **1**, 47 (1967).
[31] T. R. Anderson, Am. Sociol. Rev. **20**, 287 (1955).
[32] S. Akwawua and J. A. Pooler, Geography Research Forum **20**, 33 (2000).
[33] S. Akwawua and J. A. Pooler, J. Geogr. Syst. **3**, 69 (2001).
[34] T. P. Davies, H. M. Fry, A. G. Wilson, and S. R. Bishop, Sci. Rep. **3**, 1303 (2013).
[35] A. Sim, S. N. Yaliraki, M. Barahona, and M. P. H. Stumpf, J. R. Soc. Interface **12**, 20150315 (2015).
[36] F. Simini, A. Maritan, and Z. Néda, PLoS ONE **8**, e60069 (2013).
[37] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant, PLoS ONE **7**, e45985 (2012).
[38] Y. Yang, C. Herrera, N. Eagle, and M. C. González, Sci. Rep. **4**, 5662 (2014).
[39] F. Gargiulo, M. Lenormand, S. Huet, and O. B. Espinosa, J. Artificial Soc. Social Sim. **15**, 6 (2012).
[40] G. McNeill, J. Bright, and S. A. Hale, EPJ Data Sci. **6**, 24 (2017).
[41] T. Sørensen, Biologiske Skrifter **5**, 1 (1948).
[42] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford, 2018).
[43] U. Brandes, J. Math. Sociol. **25**, 163 (2001).
[44] S. K. Hui, P. S. Fader, and E. T. Bradlow, Market. Sci. **28**, 566 (2009).
[45] B. Hilton, A. P. Sood, and T. S. Evans, arXiv:1909.07194.
[46] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[47] F. P. Kelly, *Reversibility and Stochastic Networks* (Cambridge University Press, Cambridge, 2011).
[48] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling* (Academic Press, San Diego, 2014).
[49] A. Arenas, A. Díaz-Guilera, and R. Guimerà, Phys. Rev. Lett. **86**, 3196 (2001).
[50] S. Chen, W. Huang, C. Cattani, and G. Altieri, Math. Probl. Eng. **2012**, 732698 (2012).
[51] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).
[52] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[53] G. Mukherjee and S. S. Manna, Phys. Rev. E **71**, 066108 (2005).
[54] J. D. Little, Oper. Res. **9**, 383 (1961).

[55] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[56] S. K. Hui, P. S. Fader, and E. T. Bradlow, Market. Sci. **28**, 320 (2009).
[57] F. Ying, Customer Mobility and Congestion in Supermarkets, D.Phil. thesis, University of Oxford, 2019.
[58] In the original formulation, L/N is replaced by L in Eq. (7) and L has dimensions of [(number of trips)$^{-1}$].