Preference-based facility location for on-demand logistics

Paulusse, J.P.M.; Atasoy, B.; Maknoon, M.Y.; Rezaei, J.

Publication date
2019

Document Version
Accepted author manuscript

Citation (APA)
Paulusse, J. P. M., Atasoy, B., Maknoon, M. Y., & Rezaei, J. (2019). *Preference-based facility location for on-demand logistics*. Paper presented at 98th Annual Meeting of the Transportation Research Board (TRB), Washington D.C., District of Columbia, United States.

PREFERENCE-BASED FACILITY LOCATION FOR ON-DEMAND LOGISTICS

Joep Paulusse
Delft University of Technology
Delft, 2628 CD, The Netherlands
E-mail: paulusse.joep@gmail.com

Bilge Atasoy, Corresponding Author
Department of Maritime and Transport Technology
Faculty of Mechanical, Maritime and Materials Engineering
Delft University of Technology
Delft, 2628 CD, The Netherlands
Email: B.Atasoy@tudelft.nl

Yousef Maknoon
Faculty of Technology, Policy and Management
Delft University of Technology
Delft, 2628 CD, The Netherlands
Email: M.Y.Maknoon@tudelft.nl

Jafar Rezaei
Faculty of Technology, Policy and Management
Delft University of Technology
Delft, 2628 CD, The Netherlands
Email: J.Rezaei@tudelft.nl

Word Count: 6,142 words + 4 tables = 7142 words

Submitted [August 1, 2019]

ABSTRACT

In the context of on-demand logistics systems, the facility location is becoming even more critical as demand characteristics and customer preferences are changing with respect to the location, time, customer segments etc. Classic facility location models do not take into account customer preferences when the set of facility locations are optimized and therefore the expected profit of the classic facility location models is not an accurate representation of reality. This paper develops a preference-based facility location model which incorporates customer preferences while maximizing the system-wide expected profit in the context of an on-demand logistics provider. The customers are first segmented based on historical data and segment specific preferences are estimated by logit mixture where we take into account heterogeneity within the segment. The performance of the preference-based facility location model is measured by total expected profit and consumer surplus. It is found that, the preference-based facility location model is not only a more accurate representation of reality but also has the potential to increase the expected profit compared to typical facility location models.

Keywords: Customer preferences, customer segmentation, facility location, preference-based facility location, on-demand logistics.

INTRODUCTION

The logistics provider business is developing as a result of the emerging demand of advanced logistics services. As the market is changing, logistics companies have to adapt proactively to the industry. Customer expectations are increasing greatly as both individuals and businesses expect to get goods faster, more flexibly, with more transparency and at low or no delivery costs. Globalization, lead time reductions, customer orientation, and outsourcing are more major changes contributing to this interest in logistics *(1)*. Together with a growing demand for customized manufacturing, this complicates transportation processes and forces logistics providers to change their delivery strategy and tactics. An increasingly competitive environment is another big factor in the mix as high demand for logistics services through digitization attracts more logistics service providers to join the logistics market. Throughout the whole transport and logistics industry, from large freighters to last-mile delivery services, service providers are adjusting their processes to adapt to the competitive environment.

Although the quality of on-demand logistics depends on the complete supply chain, the last-mile delivery is crucial for on-demand logistics. To keep or gain a competitive advantage, on-demand logistics service providers need to reconsider their distribution network. The distribution network consists of a set of demand and depot locations where customers generate demand and are supplied by the logistics service providers from the set of depot locations. McKinnon (2009) finds that the imposition of environmental taxes, worsening congestion and congestion charges are factors that cause distribution managers to decentralize distribution centers, whereas the development of the motorway network stimulates the decision-makers to centralize their distribution structures *(2)*. Logistics companies may use location data and mathematical models to decide on where and how many depots they should build.

Given a centralized or decentralized structure, logistics managers should define the locations of their warehouses, depots, or other forms of facilities from where goods are distributed. The fixed-charge facility location (FLM) problems are one of the core problems in location science *(3)*. The problem is built on a finite set of users with demand of service and a finite set of potential facility locations that can offer service to users. The decision to be made in this problem is twofold: the location decisions determine the locations of new facilities and the allocation decisions determine which opened facilities will supply which users. The FLM, as described by Fernandez and Landete (2015), helps to find the optimal locations to open a facility by minimizing the total costs of the system, where the total costs contain transportation costs and fixed facility opening costs and where the system is represented by the complete set of facilities in the area of operations *(3)*. The result of the model is a set of locations with opened facilities from where demand is fulfilled. As the model is uncapacitated and minimizes total costs, and transport costs are based on distance, each demand location is assigned to and will place all its demand at the nearest facility.

In reality, customers do not always choose the nearest facility. They may travel further to buy items at a lower price or with a higher quality. In other words, given a choice set of multiple alternatives, customer choices may depend on their preferences for other key purchasing attributes in addition to distance, such as price and quality. The customer preferences can be measured to find the probability that a customer chooses an alternative among others. This can be done through random utility models based on historical choices so to measure the importance of the attributes as perceived by the customers *(4)*.

In this paper, we present a preference-based facility location model by incorporating customer preferences into the classic facility location model. The preference-based facility location model integrates customer preferences to predict customer choices and improve the set of selected facility locations. The need for a preference-based facility location model is also addressed by Benati (1999) who incorporates random utility functions in the location problem by solving a maximum capture problem that maximizes the market share *(5)*. Benati and Hansen (2002) elaborate on that model but make restrictive assumptions about customers utility as no variability is taken into account such that the customers' choice probabilities for each facility location are given by a simple logit model *(6)*. There are other attempts to include the attractiveness of facilities when optimizing the location *(7)*. However, they are based on aggregate level information (e.g., distance) and also typically they do not consider heterogeneity across users. Haase and Muller (2013) develop a non-linear FLM model which incorporates customer behavior to estimate demand per facility in the context of public school choice *(8)*. They adopt a simulation-based approach for representing the utility. Muller and Haase (2014) improve the assumptions of the simple logit model by introducing customer segmentation in the context of retail facility location planning *(9)*. There are also various other contexts where the choice behavior was in different ways incorporated in facility location models, e.g., Zhang et al. (2012) in preventive healthcare network design *(10),* Zheng et al. (2017) for the location of charging stations for EVs *(11)*.

It is clear that there is more and more interest in the literature to incorporate preferences of customers through choice models in optimization models for facility location problems. Yet there is still not much attention to the heterogeneity of customers. In this paper, we develop a preference-based facility location model where we represent the choice through a logit mixture model for two segments of customers. We represent the model by a mixed integer linear programming problem where the decision maker maximizes the system-wide expected profit. Moreover, the proposed model is applied for a real-life case study of an on-demand logistics supplier with a large dataset.

METHODOLOGY

As the first step of the proposed methodology, we segment customers based on their characteristics. For each of the segments we develop logit mixture models where we have random sensitivities of customers across the segments towards attributes. These choice models are then integrated into the facility location model for the optimization of the locations with maximum expected system-wide profit. In this section we provide the details on these different methodologies and in the case study section we will present the results for them accordingly.

Customer Segmentation

The customers are segmented according to the K-means clustering technique, which is a generally utilized technique meant for creating groupings by optimizing the qualifying criterion function, defined either globally or locally *(12)*. Clustering is usually performed based on the standardized attribute values as it is easier to analyze customer preferences for high or low attribute values and its magnitude of importance. The data used for segmentation finally is a customer-level data set, where each row represents a customer with the average of the standardized attribute values of the chosen alternatives.

Facility Choice Model

One of the simplest and most widely known discrete choice models that handles cross-sectional data is the logit model where it is assumed that random error terms are identical and independently distributed across alternatives *(13)*. This leads to the property of independence of irrelevant alternatives (IIA) which states that the relative probability of choosing any pair of alternatives is independent of the presence of attributes of any other alternatives *(4)*.

When an individual makes a set of consecutive choices, the correlation between those choices should be considered. The data can be divided in several groups and is called panel data *(14)*. Panel data is essentially multi-day, multi-group data where repeated measurements on the same sample at different points in time are gathered. Panel data offer major advantages over the cross-sectional data as repeated observations from the same individual generally give more precise measurements of individuals' tastes. Furthermore, logit mixture models provide flexibility in representing random terms for choice model parameters as well as variances and correlation across alternatives and individuals *(15)*. In this paper, in order to represent heterogeneity we consider logit mixture models for each customer segment.

Preference-Based Facility Location Model

Based on the choice model, we develop the preference-based facility location model. The set of customers is represented by I and the set of potential facility locations are given by J. Not all locations are suitable to open a facility as facilities may need to comply with industry specific regulations obliged by the authorities such as safety regulations. Moreover, when the facility is in a large distribution network it may serve as a transshipment node. Therefore, the facility should have storage capacity and be easily accessible by large trucks to load/unload products. Based on all these considerations the set of potential locations are generated.

The mathematical model for the preference-based facility location problem is provided in (1)-(4).

$$\text{Max} \sum_{i\in I} \sum_{j\in J} \mathcal{P}_{ij} D_i \left(R_j - C_{ij}\right) \tag{1}$$

Subject to:

$$\mathcal{P}_{ij} = \frac{\exp(V_{ij}) A_{ij} Y_j}{\sum_{j'\in J} \exp\left(V_{ij'}\right) A_{ij'} Y_{j'} + \exp(V_0)} \qquad \forall\, i \in I, j \in J \tag{2}$$

$$\sum_{j\in J} Y_j = P \tag{3}$$

$$Y_j \in \{0,1\} \qquad \forall\, j \in J \tag{4}$$

The objective function (1) is the maximization of the total expected profit, where $\mathcal{P}_{ij}$ is the probability that customer i chooses facility j, D_i represents the demand for customer i, R_j is the gross margin per item at facility j, and C_{ij}represents the transportation costs between customer i and facility j. Note that, depending on how the routing is performed to serve customer i, the transportation cost may be directly linked to the distance between locations i and j in the case of direct shipments. However, the cost may be lower if several customers are visited in the same tour. In order to get a more accurate representation of transportation costs, we analyzed historical data and between each pair of locations we came up with measures that map the distance to the costs.

The binary decision variable Y_j is 1 if there is a facility located in location j, and 0 otherwise. Constraints (2) represent the choice probability where V_{ij} is the utility of customer i for facility j. We also included an opt-out alternative that represents the case when the customer does not work with the on-demand logistics provider at all, i.e., works with another provider. The utility of this opt-out alternative is given by V_0. We have parameters, A_{ij}, that are binary input parameters indicating whether it is possible to serve customer i from facility j. These are typically provided by the context of the on-demand logistics provider. Notice that, depending on which facility locations are opened, the choice set provided to each customer will vary and therefore the model in its current form is not a linear representation of the problem at hand. Constraints (3) limits the number of opened facilities to P as in the case of P-median problems *(16)*. This was a decision in order to reduce the complexity of the problem and in the case study we experiment with different values of P.

Mixed integer linear representation of the model

As mentioned above, the model in (1)-(4) is nonlinear due to the choice probability representation in constraints (2). In order to work with a computationally easier problem, we transformed the model into a mixed integer linear programming problem similar to the idea by Davis et al. (2013) as follows *(17)*:

$$\text{Max} \sum_{i\in I} \sum_{j\in J} \mathcal{P}_{ij} D_i \left(R_j - C_{ij}\right) \tag{5}$$

Subject to:

$$0 \le \frac{\mathcal{P}_{ij}}{\exp\left(V_{ij}\right)} \le \frac{\mathcal{P}_{i0}}{\exp\,(V_{i0})} \qquad \forall\, i \in I, \forall\, j \in J \tag{6}$$

$$\mathcal{P}_{i0} + \sum_{j\in J} \mathcal{P}_{ij} = 1 \qquad \forall\, j \in J \tag{7}$$

$$\mathcal{P}_{ij} \le A_{ij} Y_j \qquad \forall\, i \in I, \forall\, j \in J \tag{8}$$

$$\sum_{j\in J} Y_j = P \tag{9}$$

$$Y_j \in \{0,1\} \qquad \forall\, j \in J \tag{10}$$

The objective function (5) is the same as (1). The main change is the representation of the choice probability. In this version of the problem, the choice probability is not explicitly given with the logit model. Instead it is defined as a continuous decision variable. The relative values of the choice probabilities are maintained based on their utility values in reference to the opt-out alternative as ensured by constraints (6). Constraints (7) maintain that the sum of the choice probabilities across all facilities and the opt-out alternative sums up to one. Constraints (8) ensure that the probability for customer i is zero towards location j if this location is not serving the customer or if there is not a facility opened in that location.

Note that in the case of Davis et al. (2013), they guarantee a totally unimodular constraint matrix for the assortment optimization problem since they serve the same size choice set across individuals *(17)*. However, in our case we do not have this structure and therefore we cannot prove that this is an exact linear transformation of the model and instead it serves as an approximation.

Benchmark facility location model

In order to assess the proposed preference-based facility location model, we use a state-of-the-art facility location model to serve as a benchmark. As most FLMs aim to minimize system-wide costs, distance is the only attribute that is considered in these models. However, we cannot make a fair comparison between the performance of a profit maximization model and a transport cost minimization model. Therefore, the FLM *(3)*, should be adjusted to a profit maximization problem. Moreover, their model incorporates the fixed facility opening costs, however, as stated before, we do not include the facility opening costs and design a P-facility location model. Moreover, the classic distance minimization models force the model to satisfy all demand, otherwise the model will not assign any demand to the facilities as this generates no costs at all. In the profit maximization model, we do not force the model to satisfy all demand as satisfying all demand may incur high transport costs and lowers the objective value.

Similar to Fernandez and Landete (2015), the benchmark model is an uncapacitated single allocation model which means that there is no limit on the demand assigned to a facility by an individual and each individual is assigned to only one facility *(3)*. We replace the minimization objective with a profit maximization objective and provide the benchmark model as follows:

$$\text{Maximize } \sum_{i \in I} \sum_{j \in J} X_{ij} D_i \left(R_j - C_{ij} \right) \quad (11)$$

Subject to:

$$\sum_{j \in J} X_{ij} \leq 1 \qquad \forall\, i \in I \qquad (12)$$

$$X_{ij} \leq A_{ij} Y_j \qquad \forall\, i \in I, \forall\, j \in J \qquad (13)$$

$$\sum_{j \in J} A_{ij} Y_j \leq M \sum_{j \in J} X_{ij} \qquad \forall\, i \in I \qquad (14)$$

$$\sum_{j \in J} Y_j = P \qquad (15)$$

$$X_{ij}, Y_j \in \{0,1\} \qquad \forall\, i \in I, j \in J \qquad (16)$$

In addition to the facility location decision variables, we have also X_{ij}, that represent the decision of the allocation of each customer to each potential facility. Constraints (12) ensure that each customer is assigned to at most one facility. Constraints (13)-(14) link the two decision variables such that X_{ij} is zero when there is no facility opened at location j or if this location cannot serve the customer. Constraint (15) is as before maintaining a total of P facilities to be opened.

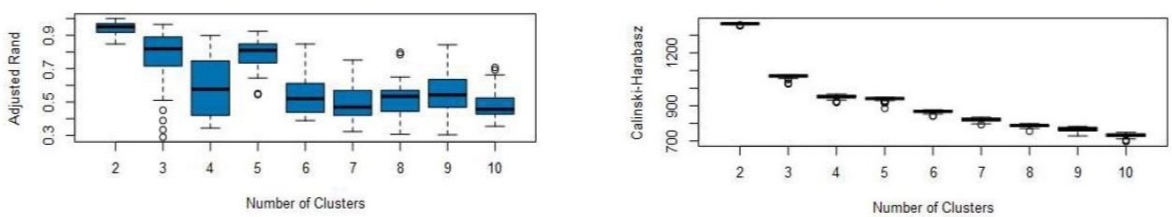
NUMERICAL EXPERIMENTS

The proposed methodology is applied in a set of experiments representing a general on-demand logistics provider. We assume that the on-demand provider is providing a set of products that are offered by different suppliers through a smartphone app. The customers are using the app to see the alternative products available to them. Let us assume that the products are listed on the app with typical attributes: price, estimated arrival time (ETA) and customer rating (that represents perceived quality of the product). We work with a dataset that represents previous product choices of customers together with the attributes for all the available alternatives in the choice set of the customer. We consider customers that have at least three alternatives in their choice set. This results with around 2,500 individuals (I) with a total of 25,000 choices.

Customer Segmentation Results

Before we quantify customer preferences, we first studied individuals in order to generate groups with similar preferences. The Adjusted Rand index and the Calinski-Harabasz index are used to find the optimal number of clusters of which the results are shown in Figure 1. The Adjusted Rand index generates relatively high values for 2, 3, and 5 clusters and the Calinski-Harabasz index scores highest at 2 clusters. This means that according to these indices, we can have 2 distinct groups of individuals where individuals in each group show similar behavior. The clusters have around 550 and 1,950 individuals, respectively.

The standardized average attribute values per cluster are shown in Figure 2. We directly observe that individuals in cluster 1 are less sensitive to price and care more about ETA and rating. On the other hand, the individuals in cluster 2 are very sensitive to price and rather insensitive to ETA and rating.



(a) Adjusted Rand index (b) Calinski-Harabasz index

Figure 1 Indices for clustering

Figure 2 Resulting two clusters with average standardized attributes

Choice Model Estimation Results

For the facility location choice model, we considered using a simple logit model as well as a logit mixture model. Even though we have customer segments, still we believe there should be random heterogeneity in each segment and we investigated this through logit mixture models. Based on the analysis of the historical data, we concluded that highest variability may lie in the ETA preferences of customers. Therefore, in the logit mixture specification we assumed a randomly distributed ETA parameter and estimated a mean and standard deviation. In order to ensure a negative parameter for ETA, we introduced it as a lognormally distributed random term.

The estimation results for both of the clusters are presented with simple logit and logit mixture specifications in Table 1. It is observed that, logit mixture provides a significantly better likelihood for both of the clusters considering only one additional parameter. All the parameters have the expected sign for all estimations. As expected, cluster 1 individuals are less sensitive to price and more sensitive to rating and ETA compared to cluster 2. For the logit mixture model, the presented mean and standard deviation values in Table 1 are from the corresponding normal distribution. After transformation, we get a mean value of -0.120 and -0.039 for clusters 1 and 2, respectively. Note that, the mean values of the ETA parameter with logit mixture model are larger in absolute value. This is indeed a clear sign that, when heterogeneity is not modeled the estimation results may have bias. In this case, the simple logit model gives a wrong indication that individuals are not that sensitive to ETA. However, when heterogeneity is explained, it is clear that part of the population is more sensitive.

For the integration of the preferences in the facility location model we used the logit mixture estimation results and for the ETA parameter we used the estimated mean value to keep the optimization procedure simpler.

TABLE 1 Choice model estimation results

				Estimated parameters			
Cluster	Model	Initial log likelihood	Final log likelihood	Price	ETA	Rating	σ_{ETA}
1	Simple Logit	-11,208.8	-9,583.4	-0.093	-0.073	2.02	
	Logit Mixture	-11,023.5	-9,116.2	-0.953	-3.03	2.16	1.35
2	Simple Logit	-36,729.1	-24,461.6	-0.459	-0.014	1.27	
	Logit Mixture	-47,136.6	-23,902.6	-0.48	-5.12	1.2	1.94

Facility Location Optimization Results

Given the clusters and the estimated choice model parameters we can optimize the facility locations with the proposed preference-based facility location model in reference to benchmark models. Note that, the performance of the models is measured by the total expected profit and consumer surplus, where the consumer surplus is the expected maximum utility an individual receives by the provided choice set.

The benchmark model provides a single location for each customer based on maximum profit. The output metrics of the benchmark model are first computed where each individual is fully assigned to the most profitable facility suggested by the optimization. However, in reality if customers find a facility in their area that matches their preferences better, they will reach out to those facilities. Therefore, in order to represent the reality better, we consider the benchmark model with multiple allocation where each individual is assigned to all available open facilities. The set of open facility locations is given by applying the single allocation model where after only the allocation is done based on the choice probabilities as a function of the attributes of the facilities. As we have the estimated choice model, we can use it for the recomputation of the metrics. Single and multiple allocations are illustrated in Figure 3. Figure 3a shows the behavior of individuals according the benchmark model where individuals are only assigned to the most profitable available and opened facility. In Figure 3b, the choice probabilities come into play and less profitable facilities will capture some demand of the more profitable facilities.

It is expected that recomputing the expected profit with multiple allocation in the benchmark model leads to a lower total profit value as we extend the choice sets of the individuals and apply choice probabilities when we open the same set of facilities. They will not be solely served by the most profitable location given by the benchmark model with single allocation.

In the remainder of this section we present results over three models: (1) benchmark model with single allocation, (2) benchmark model with multiple allocation (where preferences are not considered while optimizing the facility locations) and (3) preference-based facility location model.

(a) Single allocation (b) Multiple allocation

Figure 3 Benchmark facility location model illustration with single and multiple allocation

Table 2 presents the total profit and consumer surplus across the three models for different values of P where there are assumed to be 90 possible facility locations in total. It is seen that the profit increases for relatively lower values of P, where after it stabilizes and drops at the end when we select all potential locations. This trend is intuitive as the model first selects the most profitable facilities. However, at some point, additional facilities are located in the same service areas as other open facilities since locating them outside these already covered service areas would be unprofitable by either low revenues or high transport costs. Opening facilities in areas that are already mostly covered by more profitable facilities does not satisfy much more demand and therefore the profit stabilizes in the mid-range values of P. However, when we force the model to open more facilities it needs to open even these unprofitable facilities. These facilities do not have overlapping service areas with other facilities and therefore satisfy new demand, increase the total revenue, but also increase the total transport costs. As the profit decreases for high values of P, the increase in revenue by satisfying more demand does not outweigh the increase in transport costs.

TABLE 2 Results with the three models across different number of facilities

	Number of opened facilities (P)								
Profit	10	20	30	40	50	60	70	80	90
Benchmark: single allocation	*base*	17%	6%	2%	0%	0%	0%	-2%	-11%
Benchmark: multiple allocation	*base*	12%	-2%	-2%	-4%	-4%	-2%	-4%	-22%
Preference-based	*base*	17%	5%	-2%	0%	0%	-3%	-6%	-17%
Consumer surplus	10	20	30	40	50	60	70	80	90
Benchmark: single allocation	*base*	68%	20%	10%	1%	0%	7%	0%	15%
Benchmark: multiple allocation	*base*	87%	39%	20%	4%	0%	8%	4%	12%
Preference-based	*base*	87%	26%	28%	1%	3%	4%	11%	12%

Remember that, in the multiple allocation version of the benchmark model we use the same location decisions as the single allocation version of the model. However, we recompute the output metrics with respect to the choice probabilities after allowing the customers to be served by the opened facilities around them. Therefore, the trends turn down quicker and keep decreasing as facilities have overlapping service areas and the most profitable facilities loose demand to less profitable facilities. Individuals that were fully allocated to a very profitable facility in the benchmark model with single allocation, are now partly assigned to another available and open facility which is less profitable. This causes the profit to decrease faster as more facilities are opened. The largest drops are at the end where we force the model to open the most unprofitable facilities.

In the preference-based model, the set of facility locations is optimized considering that individuals choose from the available and open facilities according their choice probabilities. We expect that the total expected profit is lower than the profit in the benchmark model with single allocation as in the preference-based model not all demand will be assigned to the most profitable facilities. However, the profit is expected to be higher than the profit in the benchmark model with multiple allocation as facilities are optimized considering upfront that individuals will be assigned to all open and available alternatives according the choice probabilities. The results in Table 3 confirm this as we see that the preference-based model outperforms the benchmark model with multiple allocation by 12%. Note that the results in Table 3 are in reference to the benchmark model with single allocation.

Looking at the results in Table 2, one can conclude that, when a logistics provider has a set of 90 potential facility locations and uses the preference-based facility location model to maximize the expected profit, 30 facilities should be opened as this generates the highest expected profit. However, when the logistics provider cares about the consumer surplus, 40 facilities may be opened as this decreases the profit only by 2% but increases the consumer surplus by 28%. These indicate the clear trade-off between the consumer surplus and expected profit.

As the consumer surplus represents the expected maximum utility that a person receives with the provided choice set, it keeps increasing as we open more facilities. Notice that at the very end, facilities with preferred attribute values are selected and boost the consumer surplus since the model is forced to open even unprofitable facilities which may have low prices and therefore bring more utility to the individuals. Note that our largest cluster is price sensitive and therefore receive higher utility for alternatives with low prices. For the case of the multiple allocation and preference-based models, the increase in consumer surplus is larger than the single allocation model as expected since the customers have a choice set in both cases.

Overall, it is clear and expected that the benchmark model with single allocation generates the highest profit for each value of P as it assigns the customers to the most profitable facility. However, as discussed before the individuals should be allocated to all open facilities that operate in their area. Therefore, we believe that it is an overestimation of the potential profit. A better representation of a realistic benchmark model is with multiple allocation, where the set of facilities that were found by the case with single allocation are used and the output metrics are computed differently. It is observed that the multiple allocation profit outputs are lower than the single allocation outputs. This is because the demand of individuals is shared among the facilities according to the choice probabilities instead of fully allocating demand to the most profitable facility. The preference-based model considers upfront that individuals will be allocated to all open and available facilities when it optimizes the set of facility locations. The profit of the preference-based model cannot be higher than the benchmark model with single allocation as it would be most profitable to assign all demand to the facilities with the highest profit. The preference-based model does take into account the choice probabilities and finds a different set of facility locations which generates a higher profit than the benchmark model with multiple allocation, where preferences are not taken into account when the facilities are optimized.

TABLE 3 Results in reference to the benchmark model with single allocation

	Number of opened facilities (P)									
Profit	10	20	30	40	50	60	70	80	90	Average
Benchmark: multiple allocation	0%	-5%	-12%	-16%	-20%	-23%	-24%	-25%	-34%	-18%
Preference-based	0%	0%	-1%	-6%	-6%	-6%	-8%	-12%	-18%	-6%
Consumer surplus	10	20	30	40	50	60	70	80	90	Average
Benchmark: multiple allocation	2%	14%	33%	45%	49%	49%	51%	56%	51%	39%
Preference-based	2%	14%	20%	40%	40%	45%	41%	56%	51%	34%

Finally, we would like to perform sensitivity analysis on the utility of the opt-out alternative. The value we used for this utility was very low so that practically almost no customers were opting out. Our main aim to use the opt-out alternative was rather to be able to transform the model as it gave a reference choice probability. In order to see the effect of this assumption on our results, we worked with two additional utility values: (i) min – that is the minimum utility value across the whole data set, (ii) mean – mean utility value across the data set.

Table 4 presents the profit across different opt-out utility values and the reference is the benchmark model with single allocation. We observe that the preference-based facility location model outperforms the benchmark model with multiple allocation by 12% when we use an extremely low opt-out utility value, as we originally did. The preference-based model outperforms the benchmark model by 11% when we set the utility of the opt-out option equal to the minimum utility value in the data set. Lastly, the preference-based model outperforms the benchmark model by 54% when we set the opt-out option equal to the mean of the utility values in our data set. This means that the improved performance of the preference-based facility location model is consistent across different values of the opt-out utility and it is even more evident when we have a strong opt-out alternative. Moreover, as the opt-out utility increases the profit decreases as expected.

TABLE 4 Sensitivity analysis on the opt-out alternative (expected profit)

		Number of opened facilities (P)								
Model	V_0	10	20	30	40	50	60	70	80	90	Average
Benchmark: multiple allocation	Low	0%	-5%	-12%	-16%	-20%	-23%	-24%	-25%	-34%	-18%
	Min	-1%	-5%	-12%	-16%	-20%	-23%	-24%	-25%	-34%	-18%
	Mean	-78%	-77%	-78%	-79%	-79%	-80%	-80%	-82%	-78%	-79%
Preference-based	Low	0%	0%	-1%	-6%	-6%	-6%	8-%	-12%	-18%	-6%
	Min	0%	-3%	-5%	-6%	-7%	-7%	-9%	-12%	-18%	-7%
	Mean	-20%	-21%	-23%	-24%	-24%	-24%	-26%	-28%	-33%	-25%

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we developed a preference-based facility location model in the context of an on-demand logistics provider, which gives a better representation of reality and therefore improves the facility selection results. The preference-based facility location model incorporates choice probabilities which are used to estimate individuals' demand allocation and to select the optimal set of opened facility locations. We measured its performance by total expected profit and consumer surplus to see if it outperforms the benchmark which is a basic facility location model.

The expected profit in the preference-based model could not outperform the single allocated benchmark model as choice sets with more than one alternative include less profitable facilities which take a share of the most profitable facilities. However, the preference-based model outperforms the benchmark model that is adapted to a realistic context where consumers can go to all open facilities in their area. In the preference-based model, the facilities are selected such that the total profit is maximized taking into account that the individuals are assigned to all the facilities in their choice set and allocate demand according to a choice probability. As this model accounts for the choice probabilities upfront when the facilities are selected, it generates a higher total expected profit compared to the benchmark model with multiple allocation. This means that incorporating customer preferences in the facility location model improves the selection of facility locations to maximize the system-wide profit. The consumer surplus of the preference-based model (with expected profit maximization as the objective) is slightly lower compared to the benchmark model with multiple allocation and much better than the benchmark model with single allocation.

The preference-based facility location model does not include facility opening costs as we did not have reliable data and were not able to make accurate estimations. Therefore, we worked with different scenarios where we changed the number of opened facilities and analyze the impact on the output metrics. We suggest to build on our preference-based facility location model and extend it by including the facility opening costs which allows the model to decide on the number of opened facilities instead of solving the model for different numbers of opened facility locations. Moreover, the proposed location model does not account for any capacity constraints at the facilities. The model could be improved by including a maximum capacity at the different type of facilities, where facilities with larger capacities have higher opening costs.

Lastly, the linearization of the preference-based model was inspired by Davis et. al (2013) who proved to design an exact transformation of the model into a linear model *(17)*. Although they offer the same number of alternatives to all individuals and guarantee a unimodular constraint matrix, in our problem the choice sets of individuals depend on their geographical locations and is therefore different. As the linearization is not proven to be an exact transformation, we can only state that our results are an approximation. Therefore, one other direction is the analysis of the bounds for the formulation to assess the performance.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: J. Paulusse, B. Atasoy, Y. Maknoon, J. Razaei; data collection: J. Paulusse; analysis and interpretation of results: J. Paulusse, B. Atasoy, Y. Maknoon, J. Razaei; draft manuscript preparation: J. Paulusse, B. Atasoy. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Hertz, S. and Alfredsson, M. Strategic development of third party logistics providers. *Industrial Marketing Management*, 2003. 32:139–149.

2. McKinnon, A. The present and future land requirements of logistical activities. *Land Use Policy*, 2009. 26S:S293–S301.

3. Fernandez, E. and Landete, M. Fixed-Charge Facility Location Problems. In *Location* Science (G. Laporte, S. Nickel. and F. Saldanha da Gama, eds.), Springer, Cham, 2015, pp. 44–74.

4. Ben-Akiva, M. and Lerman S.R. Discrete Choice Analysis. The MIT Press, Cambridge, MA, 1985.

5. Benati, S. The maximum capture problem with heterogeneous customers. *Computers and Operations Research*, 1999. 26:1351–1367.

6. Benati, S. and Hansen, P. The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 2002. 143:518–530.

7. Küçükaydın, H., Aras, N. and Altınel, İ. A discrete competitive facility location model with variable attractiveness. *Journal of the Operational Research Society*, 2011. 62(9): 1726-1741.

8. Haase, K. and Muller, S. Management of school locations allowing for free school choice. *Omega*, 2013. 41:847–855.

9. Müller, S. and Haase, K. Customer segmentation in retail facility location planning. *Business Research*, 2014. 7: 235-261.

10. Zhang, Y., Berman, O. and Verter, V. The impact of client choice on preventive healthcare facility network design. *OR Spectrum*, 2012. 34: 349-370.

11. Zheng, H., He, X., Li, Y. and Peeta, S. Traffic equilibrium and charging facility locations for electric vehicles. *Networks and Spatial Economics*, 2017. 17(2):435–457.

12. Vaishali, R. and Rupa, G. Impact of outlier removal and normalization approach in modified k-means clustering algorithm. *International Journal of Computer Science*, 2011. 8:331–336.

13. Ray, P. Independence of Irrelevant Alternatives. The Econometric Society, 1973. 41:987–991.

14. Cherchi, E. and Cirillo, C. A mixed logit mode choice model on panel data: accounting for systematic and random variations on responses and preferences. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.

15. Train, K. Discrete Choice Methods with Simulation. Cambridge University Press, 2009.

16. Tansel, B.C., Francis, R.L., and Lowe, T.J. State of the Art—Location on Networks: A Survey. Part I: The p-Center and p-Median Problems. *Management Science,* 1983. 29(4):482-497.

17. Davis, J., Gallego, G., and Topaloglu, H. Assortment Planning under the Multinomial Logit Model with Totally Unimodular Constraint Structures. 2013. Working paper