Choice-based service region assortment problem: equitable design with statewide synthetic data

Xiyuan Ren, Joseph Y. J. Chow*

***Abstract*— Incorporating individual user preferences in statewide transportation planning is of great importance regarding revenue management and behavioral equity. However, an enduring challenge is that consistent population travel data remains scarce, particularly in underserved and rural areas. Moreover, large-scale optimization models are computationally demanding when considering stochastic travel demands in a discrete choice model (DCM) framework. These can be addressed with a combination of synthetic population data and deterministic taste coefficients. We formulate a choice-based optimization model, in which the mode share in each block group-level trip origin-destination (OD) is determined by a set of deterministic coefficients reflecting user preferences. In that case, statewide service region design becomes an assortment optimization problem with known parameters and linear constraints, which can be efficiently solved through linear or quadratic programming (depending on variant). We test the method using a hypothetical new mobility service considered for New York State. The proposed model is applied to optimize its service region with one of the three objectives: (1) maximizing the total revenue; (2) maximizing the total change of consumer surplus; (3) minimizing the disparity between disadvantaged and non-disadvantaged communities.**

I. Introduction

Transportation policies, plans, and projects require the support of mathematical models due to the substantial cost of infrastructure and the need to assess system performance [1]. There are many studies that formulated optimization models to support transportation planning, which aims to improve service efficiency [2], collect revenue [3], or capture behavioral equity [4]. At the large scale, statewide transportation models are crucial to analyze the impact of policies and trends that are implemented or addressed by state governments, but not captured at a local city or community level [5].

However, an enduring challenge in incorporating user preferences into statewide models is that consistent population travel data remains scarce, particularly for underserved and rural communities. Models are estimated using survey data collected by metropolitan planning organizations (MPOs) for urban areas, not for rural communities. The lack of representative data at the state level can lead to the ignorance of specific user groups and exacerbations of social inequity [6]. Moreover, though user preferences can be captured by the DCMs, their stochastic properties result in nonlinear or nonconvex demand functions, which are difficult to embed in large-scale optimization models governing the supply-related decisions [7-9]. Whereas Pacheco et al. [10] have presented the feasibility of integrating mixed logit models into mixed-integer linear programming (MILP) models via a simulation-based linearization approach, longer computational time compared with conventional MILP still hinders its application to large transportation network.

These issues can be addressed with a combination of synthetic population data and deterministic user preferences estimated within a DCM framework. On the one side, a growing number of companies and institutions have synthetized trip details for total population by integrating large-scale information and communication technology (ICT) data [11-12]. For instance, Replica Inc. (2021) has developed a nationwide synthetic population dataset that includes both sociodemographic information and trip details [13]. With this unique data opportunity, it is now possible to develop choice models that can account for travelers in underserved areas. On the other side, the availability of large datasets has induced the development of individual parameter logit (IPL) models, which estimate unique sets of taste coefficients per individual or agent [14-15]. In that case, the derivation of travel demands can be deterministic (a summation of individual/agent choices) instead of stochastic (an integral of parametric distribution).

This paper formulates a choice-based optimization model for statewide mobility service region design. The mode share in each block group-level trip OD pair is determined by a set of coefficients deterministically estimated by an agent-based mixed logit (AMXL) model [9]. We show that given one or two new mobility services, the statewide service region design can be formulated as an assortment optimization problem in which the mobility providers pick regions and OD pairs to serve according to the mode choice decisions made by travelers. In an empirical study, we apply the proposed model to New York State. The synthetic population data is provided by Replica Inc. The block group-level mode choice coefficients are retrieved from a public dataset owned by C2SMART center [16]. We illustrate the method with a hypothetical new mobility service in New York State to show how we would optimize its service region regarding the total revenue and equity impacts.

The remainder of the paper is organized as follows: Section II introduces the proposed model in detail, including its theoretical structure and programming formulation. Section III gives the results of the empirical study and the analysis of

Research supported by C2SMART Center (USDOT #69A3551747124)

Joseph Y. J. Chow, corresponding author, Associate Professor, C2SMART Center, Department of Civil & Urban Engineering, New York University Tandon School of Engineering, NY, USA; (e-mail: joseph.chow@nyu.edu).

Xiyuan Ren, PhD candidate, C2SMART Center, Department of Civil & Urban Engineering, New York University Tandon School of Engineering, NY, USA (e-mail: xr2006@nyu.edu).

optimization results. Section IV draws some conclusions and makes some discussions for future work.

II. Service Region Assortment Model

A. Choice-based optimization with deterministic coefficients

Choice-based optimization typically anticipates individuals' (or agents') choice behavior with a DCM following random utility theory. We refer interested readers to [17-18]. The basic assumption is that individuals or agents make choices by maximizing their overall utility that consists of a deterministic part and a random part. McFadden and Train (2000) defined a general framework that includes any DCM with the assumption of Gumbel distributed random utility [19]. They called this mixed logit (MXL), which is a multinomial logit (MNL) model with stochastic coefficients θ drawn from a cumulative distribution function.

In the context of this paper, the utility function of choosing mode k to travel from node u to w is defined in Eq. (1):

$$U_{uw}^{k} = V_{uw}^{k} + \varepsilon_{uw}^{k} = \theta_{uw} X_{uw}^{k} + \varepsilon_{uw}^{k} \tag{1}$$

where V_{uw}^{k} is the deterministic utility that is determined by a vector of trip attributes X_{uw}^{k} and a vector of taste coefficients θ_{uw} ; ε_{uw}^{k} is the random utility usually assumed to be independent and identically distributed (i.i.d.). According to [19], the total demand for trip mode k is defined in Eqs. (2)-(4):

$$D^{k} = \sum_{u \in N} \sum_{w \in N} d_{uw}^{k}, \qquad \forall k \in K^{-} \ or \ K^{+} \tag{2}$$

$$d_{uw}^{k} = d_{uw} . P_{uw}(k | X_{uw}^{k}, \theta), \qquad \forall k \in K^{-} \ or \ K^{+} \tag{3}$$

$$P_{uw}(k | X_{uw}^{k}, \theta) = \int \frac{e^{X_{uw}^{k} . \theta}}{\sum_{k' \in K} e^{X_{uw}^{k'} . \theta}} . g(\theta | \Omega) \, d\theta \tag{4}$$

where K is the choice set of trip modes; d_{uw}^{k} is the demand for mode k for OD uw; d_{uw} is the total travel demand for OD uw assumed to be fixed (won't change with the choice set) and can be observed from travel data; $P_{uw}(k|X_{uw}^{k}, \theta)$ is the probability of choosing mode k given trip attributes X_{uw}^{k} and taste coefficients θ. The taste coefficients θ vary across individuals or agents (in MXL) according to a probability distribution function $g(.)$ with distribution parameters included in Ω (e.g., means and covariance of Gaussian distribution).

Incorporating Equation (4) into optimization models is computationally demanding. A more efficient way is to estimate a set of coefficients, θ_{uw}, for each OD pair within a DCM framework. In that case, the integral of the parametric distribution can be replaced by the summation of OD pair-level mode shares, as shown in Eq. (5):

$$D^{k} = \sum_{u \in N} \sum_{w \in N} d_{uw} . \frac{e^{X_{uw}^{k} . \theta_{uw}}}{\sum_{k' \in K} e^{X_{uw}^{k'} . \theta_{uw}}}, \forall k \in K \tag{5}$$

where θ_{uw} can be estimated through various approaches, including linear regression, evolutionary plus gradient-based algorithm, and inverse optimization. We refer interested readers to [9,14,15].

In this paper, X_{uw} , d_{uw} and θ_{uw} are assumed to be deterministic values (treated as inputs in the optimization model), where θ_{uw} varies for each OD pair. Since the demand can be directly calculated, determining whether mode service k should be available for OD uw can be formulated as a linear knapsack problem or assortment problem with efficient algorithms to solve on a large scale [20].

B. Parameters and Decision Variables

The proposed model for service region optimization assumes that there will be new mobility services selecting operating zones and OD pairs, in which the fleet size should be decided to provide bi-direction trip services to meet the demands on the operated OD pairs. Each vehicle has a maximum service distance and a maximum number of trips per day. Under this setting, parameters in the proposed model consist of three parts:

- Trip attributes observed from synthetic population data.
- Deterministic taste coefficients estimated within the DCM framework.
- Variables defining the budget and performance of the new mobility services.

For a single mobility service, the single-service region design problem can be formulated as a linear programming (LP) problem. Table I lists the parameters and decision variables used in our model. It is noted that our model can also deal with multi-service region design. In that case, the problem is formulated as a quadratic programming (QP) problem with modifications on the decision variables. We refer interested readers to the Appendix.

TABLE I. Notations Used in the Proposed Model

Trip attributes observed from synthetic population data	
Z	The set of counties in New York state
N	The set of block groups in New York state
N_i	The set of block groups in county $i \in Z$
K^{-}	The mode choice set without the new mobility service
K^{+}	The mode choice set with the new mobility service
X_{uw}^{k}	A vector of trip attributes (including average travel time, average monetary cost, and mode constant) from block group $u \in N$ to $w \in N$ using mode k
d_{uw}	Travel demand (trips/day) on the OD pair from block group $u \in N$ to $w \in N$, $u \neq w$
l_{uw}	Trip length (km) from block group $u \in N$ and $w \in N$, $u \neq w$
Deterministic taste coefficients estimated within the DCM framework	
θ_{uw}	A vector of mode choice coefficients for trips from block group $u \in N$ to $w \in N$, $u \neq w$
V_{uw}^{k}	The deterministic utility of traveling from block group $u \in N$ to $w \in N$ using mode k
d_{uw}^{k}	The estimated demand (trips/day) of mode k on the OD pair from block group $u \in N$ to $w \in N$, $u \neq w$
s_{uw}^{K}	Consumer surplus of traveling from block group $u \in N$ to $w \in N$, $u \neq w$, given the mode choice set K
Variables defining the new mobility services	
O	The maximum number of operating zones
$\mathcal{F}_{max}$	The maximum fleet size in total (vehicles/day)

F_{max}, F_{min}	The maximum and minimum fleet size in each operating zone (vehicles/day)
$t_{uw}^{\hat{k}}$	Trip duration (minutes) of the new mobility service on the OD pair from block group $u \in N$ to $w \in N, u \neq w$
$c_{uw}^{\hat{k}}$	Trip fee ($/trip) of the new mobility service on the OD pair from block group $u \in N$ to $w \in N, u \neq w$
L	The maximum distance (km) a vehicle can serve per day
T	The maximum number of trips a vehicle can serve per day
Decision variables (single-service region design)	
y_i	A binary variable that indicates whether county $i \in Z$ is included into the service region
x_{uw}	A binary variable that indicates whether the OD pair from block group $u \in N$ to $w \in N$ is operated
f_{uw}	The fleet size (vehicles/day) on the OD pair from block group $u \in N$ to $w \in N, u \neq w$

C. *Objective function*

We considered three objectives of the new mobility service from the perspective of different stakeholder groups: (1) maximizing total revenue, which is a typical objective of Transportation Network Companies (TNCs); (2) maximizing the total change of consumer surplus, which is a typical objective of state governments aiming to improve the overall social welfare; (3) minimizing the disparities between disadvantaged and non-disadvantaged communities, which is a typical objective of NGOs focusing on disadvantaged communities. We set these objectives to provide optimal strategies for these stakeholders, as well as compare the difference of their aims in space.

The consumer surplus of traveling of OD uw given mode choice set K, s_{uw}^K, is defined as a log-sum of the utilities that is shown in Eq. (6):

$$s_{uw}^K = \ln\left(\sum_{k \in K} e^{X_{uw}^k \cdot \theta_{uw}}\right), \forall u, w \in N, u \neq w \tag{6}$$

For the single-service region design problem, three objective functions according to the objectives above is defined in Eqs. (7)-(9):

$$\max_{y_i, x_{uw}, f_{uw}} \sum_{u \in N} \sum_{w \in N} c_{uw}^{\hat{k}} d_{uw}^{\hat{k}} x_{uw} \tag{7}$$

$$\max_{y_i, x_{uw}, f_{uw}} \sum_{u \in N} \sum_{w \in N} \left(s_{uw}^{K^+} - s_{uw}^{K^-}\right) d_{uw} x_{uw} \tag{8}$$

$$\min_{y_i, x_{uw}, f_{uw}} \sum_{u \in N^{non\_dis}} \sum_{w \in N} \left(s_{uw}^{K^+} - s_{uw}^{K^-}\right) d_{uw} x_{uw} - \sum_{u \in N^{dis}} \sum_{w \in N} \left(s_{uw}^{K^+} - s_{uw}^{K^-}\right) d_{uw} x_{uw} \tag{9}$$

where y_i, x_{uw}, f_{uw} are decision variables indicating whether county i is included into the service region, whether OD uw is operated, and the fleet size serving OD uw. d_{uw} and $d_{uw}^{\hat{k}}$ denote the total demand and demand for the new mobility service for OD uw. $c_{uw}^{\hat{k}}$ denotes the trip fare of the new mobility service for OD uw. N^{dis} denotes the set of block groups that are identified as disadvantaged communities by NYSERDA (2021) [21]. N^{non_dis} denotes the set of block groups that are identified as non-disadvantaged communities. $s_{uw}^{K^+}$ and $s_{uw}^{K^-}$ denote the social welfare (or consumer surplus) with and without the new mobility service, as defined in Eq. (6), in which V_{uw}^k denotes the utility of traveling from block group $u \in N$ to $w \in N$ using mode k. The objective functions for multi-service region assortment problem can be found in the Appendix.

D. *Constraints*

Eqs. (10) - (20) are constraints of the single-service region design problem regarding total budget and network characteristics. Eqs. (10) - (11) ensure the number of service zones and vehicle fleet size are restricted by the total budget. Eqs. (12) - (13) ensure only OD pairs within the service zones can be operated and vehicles can only be assigned to operating OD pairs, in which M is a large positive integer. Eqs. (14) - (15) ensure a maximum and minimum fleet size in each service zone. Eq. (16) ensures a bi-direction trip of the new mobility service. Eqs. (17) - (18) ensure that the travel demand for an operating OD pair should be met within the maximum distance and number of trips per vehicle. Eqs. (19) - (20) define the types of decision variables.

$$\sum_{i \in Z} y_i \leq O \tag{10}$$

$$\sum_{u \in N} \sum_{w \in N} f_{uw} \leq \mathcal{F}_{max}, \quad u \neq w \tag{11}$$

$$\sum_{u \in N_i} \sum_{w \in N_i} x_{uw} \leq M y_i, \quad \forall i \in Z, u \neq w \tag{12}$$

$$f_{uw} \leq M x_{uw}, \quad \forall u, w \in N, u \neq w \tag{13}$$

$$\sum_{u \in N_i} \sum_{w \in N_i} f_{uw} \leq F_{max} y_i, \quad \forall i \in Z, u \neq w \tag{14}$$

$$\sum_{u \in N_i} \sum_{w \in N_i} f_{uw} \geq F_{min} y_i, \quad \forall i \in Z, u \neq w \tag{15}$$

$$f_{uw} = f_{wu}, \quad \forall u, w \in N, u \neq w \tag{16}$$

$$L f_{uw} \geq d_{uw}^{\hat{k}} l_{uw} x_{uw}, \quad \forall u, w \in N, u \neq w \tag{17}$$

$$T f_{uw} \geq d_{uw}^{\hat{k}} x_{uw}, \quad \forall u, w \in N, u \neq w \tag{18}$$

$$y_i, x_{uw} \in \{0,1\} \tag{19}$$

$$f_{uw} \in Z^+ \tag{20}$$

III. Case Study

In this section, we apply the proposed model to New York State. We assume there will be a new mobility service selecting counties and block-group OD pairs as service regions, as well as deciding the fleet size for each OD pair. X_{uw}^k, d_{uw} are retrieved from Replica's synthetic population data. θ_{uw} are retrieved from a public dataset owned by C2SMART center. The model is solved with a Gurobi package in Python, which takes 5-8 min for each service region optimization on a local machine with Intel (R) Core (TM) i7-10875H CPU and 32GB installed RAM.

A. *Data Collection*

Replica Inc. developed a nationwide synthetic population dataset that includes both sociodemographic information and trip details (see Fig. 1). Since it is infeasible to build DCMs with such a large dataset (e.g., the maximum sample size for a mixed logit model to converge on a local machine is round 100 thousand choice observations), we retrieved an aggregated dataset for New York state that contains 120,740 rows in total. Each row contains the mode choice information of a block-group level trip OD, including the block group ID of the origin and destination, number of trips per day along an OD pair, average travel time by each mode, average monetary cost by each mode, and the current mode share.

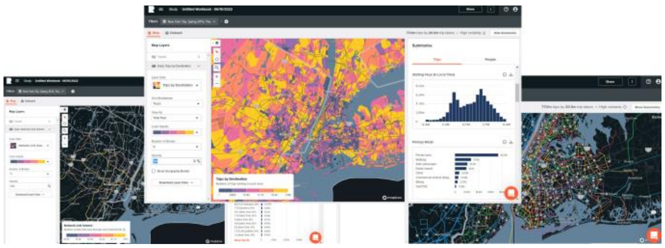


Figure 1. Replica's synthetic population data platfrom source: https://www.replicahq.com/

Taste coefficients, θ_{uw}, are retrieved from a public dataset generated by one of the research projects in C2SMART center. These coefficients were estimated within the group level agent-based mixed (GLAM) logit framework using Replica's data. These coefficients follow an empirical distribution revealing to be neither Gumber nor Gaussian, which captures taste heterogeneity to a great extent. We refer interested readers to [22].

Six trip modes are considered, including private auto, public transit, on-demand auto, biking, walking, and carpool. For each block-group level OD as an agent, the vector θ_{uw} contains ten values, including the coefficients of travel time for auto ($\theta_{uw}^{auto\_tt}$), in-vehicle-time for public transit ($\theta_{uw}^{trans\_tt}$), access time for public transit ($\theta_{uw}^{trans\_at}$), egress time for public transit ($\theta_{uw}^{trans\_et}$), number of transfers for public transit ($\theta_{uw}^{trans\_n}$), travel time for non-vehicle ($\theta_{uw}^{non\_vehicle\_tt}$), trip monetary cost ($\theta_{uw}^{cost}$), mode constant for auto ($\theta_{uw}^{asc\_auto}$), mode constant for public transit ($\theta_{uw}^{asc\_trans}$), and mode constant for non-vehicle ($\theta_{uw}^{asc\_non\_vehicle}$). Figure 2 shows the distribution of travelers' value of time (VOT), from which we find that the VOT in NYC is generally higher than other areas in NY state, and that within NYC, trips related to Manhattan and trips pointing to JFK airport have relatively higher value of time. These are consistent with our empirical knowledge.

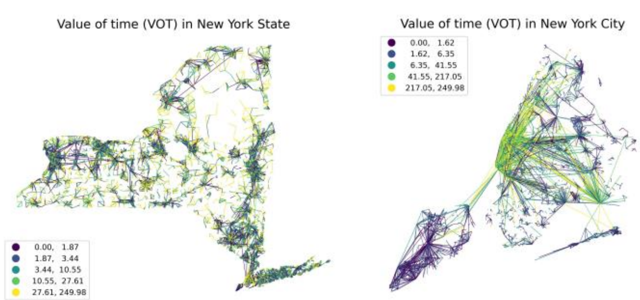


Figure 2. Travelers' value of time (VOT) at block-group OD level

B. *Pre-settings and Performance Metrics*

The budget level and service characteristics should be defined before running the optimization model. The values of these parameters are listed in Table II. The trip fare of the new mobility service is set to half of the on-demand mode, and the trip duration of the mobility service is set to the trip duration of on-demand mode plus a five-minute waiting time. We use relative trip fare and duration for simplicity, while price level and performance can also be defined using functions. Three budget levels are considered, including: (1) $O = 5, \mathcal{F}_{max} = 2{,}000$; (2) $O = 10, \mathcal{F}_{max} = 5{,}000$, and; (3) $O = 10, \mathcal{F}_{max} = 10{,}000$.

TABLE II. VARIABLES DEFINING THE NEW MOBILITY SERVICE

Variable	Explanation	Value in this paper
O	The maximum number of operating zones	$[5, 10]$
$\mathcal{F}_{max}$	The maximum fleet size in total (vehicles/day)	$[2000, 5000, 10000]$
$F_{max'}$	The maximum fleet size in each operating zone	Default to $2\mathcal{F}_{max}/O$
F_{min}	The minimum fleet size in each operating zone	Default to $\mathcal{F}_{max}/2O$
$t_{uw}^{\hat{k}}$	Trip duration (min) of the new mobility service on the OD pair	Default to on-demand travel time plus a five- minute waiting time
$c_{uw}^{\hat{k}}$	Trip fee ($/trip) of the new mobility service on the OD pair	Default to half of the on-demand mode
L	The maximum distance (km/day) a vehicle can serve	Default to 200 km per day
T	The maximum number of trips a vehicle can serve per day	Default to 10 trips per day

We consider several metrics when comparing the optimization results, including the number of operating OD pairs, vehicle kilometer traveled per vehicle (km/day), total revenue (objective 1, $/day), average welfare (objective 2, measured as total change of consumer surplus), and welfare disparity (objective 3, measured as change in consumer surplus in disadvantaged communities minus that in non-disadvantaged communities).

C. *Optimization Results*

The model is solved with Gurobi package in Python, which takes 5-8 min for each service region optimization on a local machine with Intel (R) Core (TM) i7-10875H CPU and 32GB installed RAM. Table III compares the performance metrics of single-service region optimization under different objectives and budge levels, in which the baseline refers to the current states observed from Replica's data. Each entry represents the value of a metric, and the number in the parenthesis is the percentage change compared to baseline (without any new mobility services). Several interesting points were found.

- Entries in bold font indicate the extreme values found for each metric across the three different objectives.
- The percentage change of the metrics is relatively small, though most of them have the expected signs. This is because the new mobility service can only impact a small part of the total trips (only 2.5% given budget level C).
- Maximizing total revenue and maximizing total welfare will increase welfare disparity by up to 0.59%, and

revenue per vehicle decreases with the increase of maximum fleet size (particularly from 5,000 to 10,000).

- Minimizing welfare disparity helps to decrease transportation inequities by up to 7.37%, though this is at the cost of losing total revenue. Moreover, its service region includes more OD pairs and smaller VKT per vehicle (targeting at underserved or rural block groups).

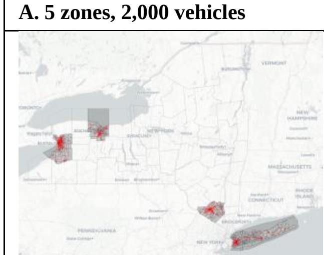
TABLE III. METRICS OF THE OPTIMIZATION RESULTS

	Baesline	**Obj. 1**	**Obj. 2**	**Obj. 3**
A. 5 zones, 2,000 vehicles				
Num. OD pairs	--	694	1,387	1,492
VKT/vehicle (km)	--	63	57	33
Total revenue ($)	--	**148,635**	136,442	73,223
Average welfare	5.443	5.452 (+0.16%)	**5.454 (+0.19%)**	5.446 (+0.04%)
Welfare disparity	0.482	0.483 (+0.02%)	0.484 (+0.37%)	**0.471 (-2.48%)**
B. 10 zones, 5,000 vehicles				
Num. OD pairs	--	2,503	3,445	3,439
VKT/vehicle (km)	--	62	53	33
Total revenue ($)	--	**350,623**	308,357	171,805
Average welfare	5.443	5.460 (+0.29%)	**5.462 (+0.33%)**	5.449 (+0.10%)
Welfare disparity	0.482	0.483 (+0.06%)	0.484 (+0.38%)	**0.459 (-4.84%)**
C. 10 zones, 10,000 vehicles				
Num. OD pairs	--	6,151	7,050	5,202
VKT/vehicle (km)	--	55	46	28
Total revenue ($)	--	**593,845**	538,666	237,878
Average welfare	5.443	5.470 (+0.49%)	**5.476 (+0.59%)**	5.452 (+0.14%)
Welfare disparity	0.482	0.485 (+0.40%)	0.486 (+0.59%)	**0.447 (-7.37%)**

Table IV lists the optimal service region given different scenarios, in which each figure is a snapshot of the online maps visualizing operated OD pairs and served counties (for larger images please click the OD pairs under each figure). We find that optimal service region under objective 3 is different from objective 1&2, by covering more OD pairs and counties in rural areas.

TABLE IV. VISUALIATION OF OPTIMAL SERVICE REGION

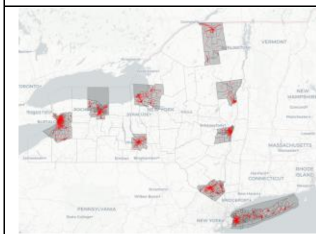
A. 5 zones, 2,000 vehicles

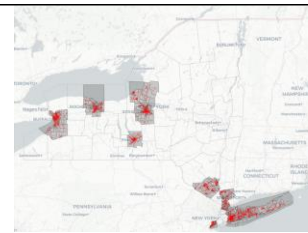

Objective 1 (online map)

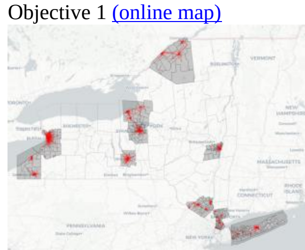

Objective 2 (online map)


Objective3 (online map)

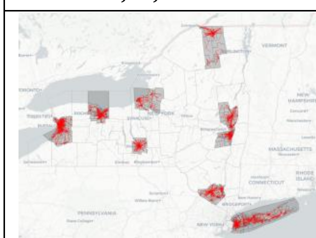
B. 10 zones, 5,000 vehicles

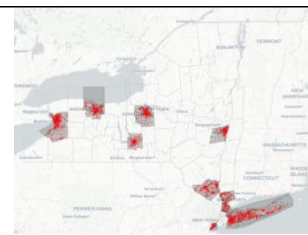

Objective 1 (online map)

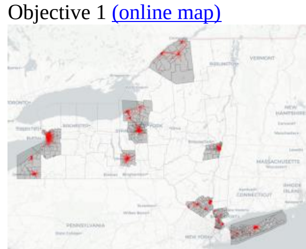

Objective 2 (online map)


Objective 3 (online map)

C. 10 zones, 10,000 vehicles


Objective 1 (online map)


Objective 2 (online map)


Objective 3 (online map)

IV. DISCUSSION AND CONCLUSION

A statewide transportation model considering user preferences can play a critical role in improving collected revenue and behavioral equity by providing state policymakers with a comprehensive, data-driven and transparent approach. This paper shows the feasibility of formulating a choice-based optimization model for mobility service region design. The challenges of incorporating user preferences in large-scale models are addressed by combining (1) synthetic population datasets that contain trips in underserved areas and (2) deterministic taste coefficients estimated within the AMXL framework.

The proposed optimization model takes X_{uw}^k, d_{uw}, and θ_{uw} as inputs, and outputs optimal service region including the service zones (y_i), operating OD pairs (x_{uw}), and fleet size per OD pair (f_{uw}). The single-service region design takes only several minutes to be solved on a local machine, which can serve as an efficient tool for statewide mobility service providers regarding total revenue or transportation equity.

There are many new research opportunities and use cases to be addressed. Though the performance and cost of the new mobility service is defined by a set of values ($X_{uw}^{\hat{k}}$), they can be replaced by cost or performance functions in response to travel demands. This won't increase the model complexity as long as travel demands can be directly computed using deterministic coefficients. However, an essential assumption in this study is that the total travel demand on each OD pair (d_{uw}) and performance of other modes ($X_{uw}^k, k \neq \hat{k}$) are fixed, i.e., insensitive to the new mobility service. This assumption is common in choice-based optimization studies but does not hold in every case. It requires further study in the future.

APPENDIX

TABLE V. NOTATIONS FOR MULTI-SERVICE REGION ASSORTMENT

Decision variables (multi-service region assortment)	
y_i	A binary variable that indicates whether county $i \in Z$ is included into the service region of service A and B
x_{uw}^A, x_{uw}^B	Binary variables that indicate whether the OD pair from block group $u \in N$ to $w \in N$ is operated by service A and B, respectively
f_{uw}^A, f_{uw}^A	The fleet size (vehicles/day) of service A and B on OD pair from block group $u \in N$ to $w \in N$, $u \neq w$

To showcase how to formulate the multi-service region assortment as a QP problem, Eqs. (A1)-(A4) are formulated to define the objective function of maximizing total revenue:

$$\max_{y_i,\dots,f_{uw}^B} \sum_{u \in N} \sum_{w \in N} c_{uw}^A d_{uw}^A + \sum_{u \in N} \sum_{w \in N} c_{uw}^B d_{uw}^B \quad \text{(A1)}$$

$$d_{uw}^A = d_{uw}^{A\_A} x_{uw}^A (1 - x_{uw}^B) + d_{uw}^{A\_AB} x_{uw}^A x_{uw}^B \quad \text{(A2)}$$

$$d_{uw}^{A\_A} = \frac{e^{V_{uw}^A}}{\sum_{k=1}^{|K^-|} e^{V_{uw}^k} + e^{V_{uw}^A}} \quad \text{(A3)}$$

$$d_{uw}^{A\_AB} = \frac{e^{V_{uw}^A}}{\sum_{k=1}^{|K^-|} e^{V_{uw}^k} + e^{V_{uw}^A} + e^{V_{uw}^B}} \quad \text{(A4)}$$

where the demand of service A on OD pair uw is defined as a combination of possible demands and the decision variables. Eq. (A2) ensures that if $x_{uw}^A = 0$, then $d_{uw}^A = 0$, and if $x_{uw}^A = 0$, $x_{uw}^B = 1$, then $d_{uw}^A = d_{uw}^{A\_AB}$. Eqs. (A3) - (A4) defines the demand of service A when only service A operates on OD pair uw ($d_{uw}^{A\_A}$) and when both service A and B operate on OD pair uw ($d_{uw}^{A\_AB}$), respectively.

ACKNOWLEDGMENT

The researchers were supported by the C2SMART center. Data shared by Replica are gratefully acknowledged.

REFERENCES

[1] Bierlaire M. Mathematical models for transportation demand analysis[J]. *Transportation research. Part A, Policy and practice*, 1997, 31(1): 86-86.

[2] Singh L, Tripathi S, Arora H. Time optimization for traffic signal control using genetic algorithm[J]. *International Journal of Recent Trends in Engineering*, 2009, 2(2): 4.

[3] Ammirato S, Felicetti A M, Linzalone R, et al. A systematic literature review of revenue management in passenger transportation[J]. *Measuring Business Excellence*, 2020.

[4] Karsu Ö, Morton A. Inequity averse optimization in operational research[J]. *European journal of operational research*, 2015, 245(2): 343-359.

[5] Moeckel R, Donnelly R, Ji J. Statewide Transportation Models in the US: A Review of the State of Practice 2[J]. *Practice*, 2019, 2: 3.

[6] Karner A, London J, Rowangould D, et al. From transportation equity to transportation justice: within, through, and beyond the state[J]. *Journal of planning literature*, 2020, 35(4): 440-459.

[7] Robenek T, Azadeh S S, Maknoon Y, et al. Train timetable design under elastic passenger demand[J]. *Transportation research Part b: methodological*, 2018, 111: 19-38.

[8] Ljubić I, Moreno E. Outer approximation and submodular cuts for maximum capture facility location problems with random utilities[J]. *European Journal of Operational Research*, 2018, 266(1): 46-56.

[9] Ren X, Chow J Y J. A random-utility-consistent machine learning method to estimate agents' joint activity scheduling choice from a ubiquitous data set[J]. *Transportation Research Part B: Methodological*, 2022, 166: 396-418.

[10] Paneque M P, Bierlaire M, Gendron B, et al. Integrating advanced discrete choice models in mixed integer linear optimization[J]. *Transportation Research Part B: Methodological*, 2021, 146: 26-49.

[11] Hörl S, Balac M. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data[J]. *Transportation Research Part C: Emerging Technologies*, 2021, 130: 103291.

[12] He B Y, Zhou J, Ma Z, et al. Evaluation of city-scale built environment policies in New York City with an emerging-mobility-accessible synthetic population[J]. *Transportation Research Part A: Policy and Practice*, 2020, 141: 444-467.

[13] Replica Inc., 2023. Synthetic Population Demo. [online] Available at: <https://replicahq.com/> [Accessed 17 May 2023].

[14] Fox J T, Kim K I, Ryan S P, et al. A simple estimator for the distribution of random coefficients[J]. *Quantitative Economics*, 2011, 2(3): 381-418.

[15] Swait J. Distribution-free estimation of individual parameter logit (IPL) models using combined evolutionary and optimization algorithms[J]. *Journal of Choice Modelling*, 2022: 100396.

[16] C2SMART, 2023. Calibrated block-group level coefficients for NYS mode choice. [online] Available at: <https://zenodo.org/record/7718923#.ZGQ2d09BxPa> [Accessed 17 May 2023].

[17] Strauss A K, Klein R, Steinhardt C. A review of choice-based revenue management: Theory and methods[J]. *European journal of operational research*, 2018, 271(2): 375-387.

[18] Roemer N, Müller S, Voigt G. A choice-based optimization approach for contracting in supply chains[J]. *European Journal of Operational Research*, 2023, 305(1): 271-286.

[19] McFadden D, Train K. Mixed MNL models for discrete response[J]. *Journal of applied Econometrics*, 2000, 15(5): 447-470.

[20] Nace D, Orlin J B. Lexicographically minimum and maximum load linear programming problems[J]. *Operations research*, 2007, 55(1): 182-187.

[21] NYSERDA, 2021. Disadvantaged Communities. [online] Available at: <https://www.nyserda.ny.gov/ny/disadvantaged-communities> [Accessed 17 May 2023].

[22] C2SMART, 2023. NY Statewide Behavioral Equity Impact Decision Support Tool with Replica. [online] Available at: <https://c2smart.engineering.nyu.edu/ny-statewide-behavioral-equity-impact-decision-support-tool-with-replica/> [Accessed 17 May 2023].