Inferring OD-Pairs and Utility-Based Travel Preferences of Shared Mobility System Users in a Multi-Modal Environment

Anshuman Anjani Kumar[1], Jee Eun Kang[1], Changhyun Kwon[2], and Alexander Nikolaev[1]

[1]Department of Industrial and Systems Engineering, University at Buffalo, SUNY
[2]Department of Industrial and Management Systems Engineering, University of South Florida

January 24, 2016

Abstract

This paper presents a methodological framework to identify population-wide traveler type distribution and simultaneously infer individual travelers' Origin-Destination (OD) pairs, based on the individual records of a shared mobility (bike) system use in a multimodal travel environment. Given the information about the travelers' outbound and inbound bike stations under varied price settings, the developed Selective Set Expectation Maximization (SSEM) algorithm infers an underlying distribution of travelers over the given traveler "types," or "classes," treating each traveler's OD pair as a latent variable; the inferred most likely traveler type for each traveler then informs their most likely OD pair. The experimental results based on simulated data demonstrate high SSEM learning accuracy both on the aggregate and dissagregate levels.

Keywords Origin-Destination estimation • Traveler preferences • Expectation Maximization • Probabilistic inference • Multimodal route choice • Bike Sharing Systems • Shared Mobility Systems

1 Introduction

Bike sharing systems are gaining prominence in the United States and world-wide as a viable shared mobility option. In driving green transportation initiatives, such systems are expected to alleviate the congestion in urban areas and provide commuters with additional travel utility as well as health and socio-demographic benefits. Researchers position bike sharing systems as a solution to the "first and last mile problem," stimulating users to switch to public transit modes and avoid relying on personal vehicles for reaching transit stations (**?**). The main challenge of shared mobility system operation is that, as many travelers tend to follow similar routes, the decreasing vehicle counts in trip origin areas (and parking spot counts in destination areas) cause vehicle imbalance across multiple stations in these areas. These operational issues are currently handled by system managers in a reactive manner; however, recent research (Haider et al., 2014) suggests a more pro-active solution—a strategic offering of incentives to travelers so as to reduce the imbalance build-up. The challenge is that in such efforts, the knowledge of travel demand and traveler preferences is critical for calculated incentive (pricing) planning.

Extensive analyses of network data and customer surveys have been conducted to understand traveler needs on an aggregate level (Vogel and Mattfeld, 2010). However, analytical and mathematical optimization models found in the literature have had limited success inferring *individual* traveler behavior (Vogel and Mattfeld, 2010). Descriptive analyses of shared vehicle usage patterns, reported in the recent past (Froehlich et al., 2009; Borgnat et al., 2009; Vogel and Mattfeld, 2010), may help this cause.

In order to parametrize a Mixed Multinomial Logit model (Hensher and Greene, 2003), often used to describe traveler routing decisions (Ben-Akiva and Lerman, 1985; Bovy and Hoogendoorn-Lanser, 2005; Wardman, 2004), true OD pair information is required. The main deficiency of OD pair estimation approaches that neglect the multimodal nature of transit is that they lose the information of demand elasticity and flexibility. Prior research has examined the problem of disaggregate multimodal (bus and metro) OD matrix estimation at the stop level (Munizaga and Palma, 2012), using automatic fare collection system records of boarding counts for two different transit mode systems. Stop-level OD pair estimation based on system data (e.g., traffic count, passenger count) has also been carried out (Abrahamsson, 1998; Lam et al., 2003; Wong and Tong, 1998; Li and Cassidy, 2007), including a study that exploited fare card transaction data (Lee and Hickman, 2014). However, no method exists that does disaggregate (i.e., individual-based) inference of traveler preferences from stop-level OD pair information or automated fare collection, and then uses these inferred preferences to distill "true" (i.e., not stop-level) OD pairs.

The OD estimation problem is paid much attention in transportation modeling and planning research. This problem is often referred to as the trip demand estimation problem, where an estimate of OD trip demand matrix is to be computed using traffic flow data and other available information (Cascetta and Nguyen, 1988). Several model formulations and heuristic methods have been proposed so far: they employ diverse theoretical approaches including the minimization of the sum of squares of the predicted and observed OD matrix value differences (Cascetta and Nguyen, 1988; Bell, 1991), column generation (Garcia-Rodenas and Verastegui-Rayo, 2008; Sherali and Park, 2001), bi-level formulations (Yang, 1995; Lundgren and Peterson, 2008), entropy/information based inference (Xie et al., 2010; van Zuylen, 1978; Xie et al., 2011), and path flow estimation (Nie et al., 2005; Chen et al., 2010; Ryu et al., 2014). Bayesian methods, exploiting the properties of certain families of parametrized distributions (Maher, 1983; Tebaldi and West, 1998; Mahmassani and Sinha, 1981; Hazelton, 2008; Castillo et al., 2008), have found use for updating the trip generation parameters and generating the trip matrix. Notably, recent work using Bayesian inference combines OD estimation with route choice behaviour (Sun et al., 2015) to assign passenger flows in networks. There, a route choice model is first developed, and then, using observable passenger data, the model parameters for the route choice model are calibrated.

The problem attacked in the present paper is even more complex, as it involves two unknowns: the trip ODs and the traveler preferences expressed via a set of feasible traveler types. In addition, the existing approaches are based on *aggregate* system (zonal/station/stop) level OD estimation. Meanwhile, for more effective pricing/incentive program implementation, the information of *individual*, i.e., *disaggregate*, traveler ODs and preferences is crucial, particularly for bike sharing systems and more broadly shared-mobility systems. The presented framework is developed to estimate both the individual trip OD and traveler types, using Bayesian logic, which is enabled by collecting multiple responses of same individuals (through user id or cards). It allows one to infer the OD-demand not at the station level but at the more granular traveler ("true" OD) level, exploiting a data driven approach, coupled with Bayesian learning, that enables the mining of trip details for each individual traveler.

This paper presents a methodological framework for traveler preference (represented as a traveler type) and OD pair inference in complex multimodal transportation systems. The developed Selective Set Expectation Maximization (SSEM) algorithm allows for estimating the unknown traveler type distribution by treating the OD pairs as latent variables, using the information about the changes in traveler route choices under varied circumstances. Due to the flexibility of system operation and the dynamic nature of the bike sharing systems, such systems offer a convenient test bed for implementing the presented estimation method. The estimation framework relies on the observed traveler responses to pricing incentives, road closures or extreme weather events. Such partial route information pieces can be collected from automatic fare collection system-type data (or from GPS or user pass-card data) as travelers respond to system perturbations. Using the partial information about a traveler's route under multiple price settings, it becomes possible to identify eligible OD points from a set of points in a geographical zone with certain belief/probability. We thereby manage to gauge the sensitivity of travelers to incentives by learning the travel utility - traveler type - distribution over the population of travelers. This distribution indicates how the travelers valuate travel options by time, price and convenience: a combination of the disutility weights for these respective measures defines a traveler type; the feasible range of the traveler types is assumed to be discrete, finite and given. Then, one can infer the bikers' origins and destinations, by first inferring the distribution of the traveler types in the entire population. A Bayesian model is used in the SSEM algorithm to quantify the likelihood of the observed data and the model parameters are adjusted in an iterative manner to maximize this likelihood.

The SSEM algorithm works to assess the sensitivity of bike-using travelers to pricing incentives, offered at outbound and inbound bike stations and varied over multiple scenarios. The algorithm can be viewed as an extension of the conventional Expectation Maximization (EM) algorithm. The EM algorithm has an issue of favoring the extreme traveler types: from each pricing scenario, it tends to conclude that travelers are either too sensitive or not-at-all sensitive to pricing incentives. The SSEM algorithm, on the other hand, infers *a range* of suitable traveler types from each scenario; an intersection of these ranges for a given traveler over all the scenarios is taken as the most likely traveler type for this traveler. The SSEM algorithm is also enhanced by the data pre-processing stage that effectively screens out infeasible route choice solutions for each traveler. In the reported computational studies, the SSEM algorithm is able to infer the true OD pairs for the travelers with the accuracy of about 75 % by correctly learning the population's sensitivity to pricing incentives.

The remainder of the paper is structured as follows. Section 2 describes the estimation framework, its assumptions and different components for inference. Section 3 dives into the details of the SSEM algorithm design. Section 4 explains the creation of the multimodal environment and the agent based simulation model for generating the data to testing and evaluation. Section 5 provides the results of the SSEM inference along with comparison of the estimated travelers' OD pairs against the true OD pairs. The paper is concluded by Section 6 which discusses the future research directions. A parallelized implementation of the SSEM algorithm is provided in the Appendix section—this implementation improves the method's scalability and speed with large data sets.

2 Model Description

The presented methodology relies on the accepted models of traveler routing choice behavior. It is assumed that (a) a population of travelers uses a transit system that is multimodal, i.e., the transportation infrastructure is a combination of multiple public transit modes such as bus, metro, bike and pedestrian walkways, (b) the travelers have different preferences for what they consider a "good" route, and (c) the travelers are rational decision-makers who readily adapt to the changing travel environment.

Figure 1 outlines the steps taken to develop and test the presented estimation framework: the partial traveler routes are assumed to be first observed, then processed and fed into the SSEM algorithm; the following sections give the detailed description of each step. For the convenience of keeping track of the notations used, the reader is welcome to refer to Table 1: it appears later in this section, after most of the notations have been introduced in the text. We now begin describing the assumed traveler routing choice model, captured in the *traveler type distribution*, and proceed with explaining the logic of the OD inference approach leading to the SSEM algorithm.

Table 1 presents the notation required to describe our estimation framework.

2.1 Multinomial Route Choice in Multimodal Travel

This section describes the Multinomial Logit (ML) model of traveler decision-making in multimodal transportation systems: this model expresses the probabilities for the route alternatives to be selected by a traveler, as a function of the traveler's preferences. More specifically, these choice probabilities are a function of the disutility values of the routes, weighted so as to be expressed in terms of cost in dollars as described by Wardman (2001).

Let w be an OD pair for a traveler with a given set of feasible route choices $\boldsymbol{R}(w)$, for which the disutility values are all pre-computed. As per the ML model, the probability that a traveler chooses a route $r \in \boldsymbol{R}(w)$ for a given disutility vector $d_r(p)$ under a certain pricing setting p can be expressed as

$$P(r(p)) = \frac{e^{-(d_r(p))}}{\sum_{r' \in \boldsymbol{R}(w)} e^{-(d_{r'}(p))}} , \tag{1}$$

where $d_r(p)$ is the disutility calculated for using a particular route computed by adding the measurable travel components of the route segments of different transit

Table 1: Mathematical Notation

t	traveler index
$\boldsymbol{T}$	Set of all travelers
p	price setting index
$\boldsymbol{P}$	Set of all price settings
obs	one bike segment observation
obs	Set of all bike segment observations
(i, j)	bike check-in and check-out stations (i.e., bike route segment)
$(i, j)_t^{\text{obs}}$	bike check-in and check-out stations (i.e., bike route segment) of traveler t for a given day observation, *obs*
$\beta_{\text{time}}^{\text{mode}}, \beta_{\text{price}}^{\text{mode}}$	weights of travel cost and time for each mode
$T^{\text{mode}}, P^{\text{mode}}$	Travel time and cost for a particular mode
$\mu = [\beta_{\text{time}}^{\text{mode}}, \beta_{\text{price}}^{\text{mode}}]$	A particular traveler type vector of weights multimodal travelers
$\boldsymbol{M}$	Set of different traveler types μ
$\boldsymbol{M}_t$	traveler type set for traveler t
$\boldsymbol{M}_t(p)$	traveler type set for traveler t under price setting p
Θ	Traveler type distribution, $\mu \sim \Theta$
$\Theta^{(n)}$	Current best estimate of traveler type distribution at iteration n
w	true OD pair for a multimodal traveler
w_o, w_d	true origin and destination of OD pair w
$\boldsymbol{\Omega}_t$	Set of eligible OD pairs for multimodal traveler t
r	a multimodal route
$d_r(p)$	Disutility of route r under price setting/scenario p
$\boldsymbol{R}(w)$	Set of route alternatives for an OD pair w
$\delta_{\mu}^{w,(i,j)}(p)$	OD-bike segment incident indicator under pricing p for traveler type μ
$\Delta_{\mu}(p)$	OD-bike segment incidence matrix under pricing p for traveler type μ
$\Delta(p)$	OD-bike segment incidence matrix for all traveler types μ
$P_{\Theta}(\mu)$	Probability value for traveler type μ under the unknown traveler type distribution Θ
$P(w)$	Prior probabilities for eligible OD pair w
$P(i, j)_{\mu}$	Logit probability of using bike segment (i, j) for traveler type μ
$P_{\text{logit}}(w\|\mu)_t^p$	Logit probability for OD pair w under the given traveler type μ for multimodal traveler t under price setting p
ϖ_r	Path correction factor for route r
$B_t^p(w)$	Bayes probability for eligible OD pair w for traveler t under price setting p
Z_{μ}	Normalizing constant for logit probabilities under preference category μ
Z'	Normalizing constant for Bayes probability
$Q(\Theta, \Theta^{(n)})$	Lower bound or expected log likelihood to be maximized at iteration n
B^p	Multi-dimensional array for Bayes probability for all travelers under price setting p
L^p	Multi-dimensional array for Logit probability for all travelers under price setting p

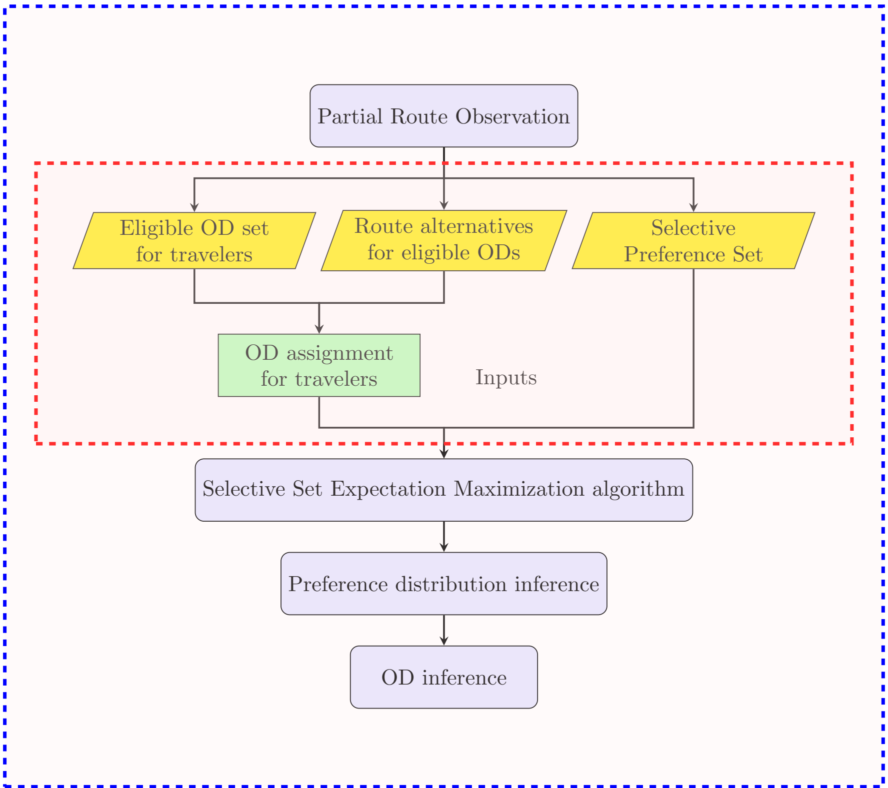

Figure 1: Framework for inferring traveler type distribution and OD pairs

system modes multiplied by the travelers' preference of each travel measurement. The negative sign accounts for the disutility term and (1) gives the probability of choosing the route with maximum utility (and minimum disutility). The disutility of each route in the route choice set of a traveler, under a given price setting p, is expressed by the following linear disutility function

$$d_r(p) = \sum_{\text{mode}} \left(\beta_{\text{time}}^{\text{mode}} \, T^{\text{mode}} + \beta_{\text{price}}^{\text{bike}} \, (P_{\text{entry}}^{\text{mode}} + P_{\text{exit}}^{\text{mode}}) \right) , \tag{2}$$

where $\beta_{\text{time}}^{\text{mode}}$ and $\beta_{\text{price}}^{\text{mode}}$ are the factors (weights) for travel time $T_{\text{time}}^{\text{mode}}$ on any mode (bus, metro, bike and walking). The price variables in the linear disutility model, P_{entry} and P_{exit}, are one-time charges a bike user incurs while picking up or dropping off a bike at a hub station. Different pricing strategies can be incorporated based on the travel length and time for which the bike is used. In the current disutility model, P_{entry} and P_{exit} are independent of the distance or time traveled. The disutility of route r under the scenario with price setting p is expressed as $d_r(p)$. Each traveler thus solves a utility-maximization problem, or equivalently, the disutility-minimization problem. A notable advantage of using the disutility approach is the ease of using shortest path algorithms, e.g., Dijkstra's or Floyd's, to compute the best route(s) under a given price setting for route alternative generation.

Given the logic behind the multi-modal route/mode choice behavior, it is acknowledged that travelers are heterogeneous decision makers. That is, travelers can be of different types, with each type characterized by its own unique combination of weights (parameters of the disutility model), $\mu = [\beta^{\text{mode}}_{\text{time}}, \beta^{\text{bike}}_{\text{price}}]$. All the feasible traveler types are assumed to be contained in a discrete finite set $\boldsymbol{M}$; the distribution of the traveler population over set $\boldsymbol{M}$ is assumed to be parametric, with parameter vector Θ.

Indeed, it is natural to assume that travelers differ from each other, and that it is not immediately known how each individual traveler valuates travel disutility or even what percentage of the traveler population shares the same views in disutility valuation. The introduction of the traveler type distribution that parametrically specifies $P_\Theta(\mu_t = \mu)$, $\mu \in \boldsymbol{M}$, for a random traveler t, allows us to express this uncertainty in a most general, realistic and useful way. As will be shown, the whole distribution $P_\Theta(\mu)$ and the weight combination for any particular traveler can be inferred, given the observations of routing choices of a large number of travelers. The inferred type for each traveler can directly inform the individual OD estimation and population-wide incentive/pricing planning: the inferred ODs and types will be sufficient for identifying bike station locations where certain pricing/incentive strategies can be expected to have high impact on the nearby travelers. Moreover, the developed theoretical framework can be further used for the specification of any particular type of utility function (of the researcher's choice); such parametric functions can involve any complex measures adopted in multi-modal transit analysis, e.g., level of service, time spent walking, time spent waiting, time spent in transfer, etc., as well as components that are crucial for understanding bike user behavior, e.g., elevation gain, weather, queueing at stations, etc.

Proceeding with the discussion of modeling assumptions, we posit that all travelers differ by their traveler types but choose routes deterministically. Specifically, it is assumed that a traveler always takes the minimum disutility-cost route, given their traveler type. The factors such as in-vehicle travel time and pricing are considered as parts of the traveler type $\mu \in \boldsymbol{M}$. The traveler population is heterogeneous, with a distribution over the traveler type set μ describing the whole population. We now define route set $\boldsymbol{R}(w)$ as the set of routes for OD pair w, generated under different pricing settings for a traveler under the assumption that the traveler type μ remains the same across these settings for a particular traveler. This assumption is quintessential for distilling the OD pair of a traveler based on their behavior under different (pricing) scenarios.

Note that in modeling the consumer's preferences over different route alternatives in a choice set, models based on Random Utility Theory (RUT) are extensively used by researchers in social sciences and economics. In the developed estimation framework, travelers are assumed to make routing decisions in the same deterministic manner. The Random Utility Theory (RUT) posits that a consumer is a rational decision maker who chooses an alternative that provides them with the maximum utility (Li et al., 2013); its advantages are highlighted in a comparative study of different decision theories used to model route choice behavior (de Moraes Ramos et al., 2011). We resort to the RUT model due to its intuitive logic and straightforward implementation for route choice decision-making modeling.

We adopt a basic utility function that effectively captures the price sensitivity, and allows for learning the traveler type distribution. Observe that a traveler with a more price-favoring traveler type is more likely to change routes as compared to a traveler with a more time-favoring traveler type; the *price vs. time* sensitivity becomes a defining number for the traveler, reflected in their behavior across multiple scenarios (recall that the set of parameters $\left[\beta^{\text{mode}}_{\text{time}}, \beta^{\text{mode}}_{\text{price}}\right]$ comprises a traveler type, μ). Note that other factors that could affect the commuters utility valuation may include mode preference, weather conditions, waiting time, egress time, convenience based on trip purpose (leisure, business, commute etc.), economic and social variables. For example, the generalized distutility cost of the following can be used,

$$d_r(p) = \beta^{\text{mode}}_{\text{wait}} \times W^{\text{mode}}(p) + \beta^{\text{mode}}_{\text{time}} \times T^{\text{mode}}(p) + \beta^{\text{mode}}_{\text{price}} \times P^{\text{mode}}(p) \\ + \beta^{\text{mode}}_{\text{LOS}} \times L^{\text{mode}}(p) + \beta_{\text{num}} \times N,$$

where $\beta^{\text{mode}}_{\text{wait}}$, $\beta^{\text{mode}}_{\text{time}}$, $\beta^{\text{mode}}_{\text{price}}$, and $\beta^{\text{mode}}_{\text{LOS}}$ are the weights for the waiting time $W^{\text{mode}}(p)$, in-transit travel time $T^{\text{mode}}(p)$, price $P^{\text{mode}}(p)$, and level of service $L^{\text{mode}}(p)$, for each mode respectively, where p is the price; finally, β_{num} is the weight for the number transfers N.

Note also, that the increase in the number of variables does not affect the applicability of the proposed framework, since the number and the weight combination of traveler types are set exogenously. Given more detailed real-world data, an analyst can estimate these parameters and construct a more intricate disutility model with variable dependencies to more closely model the decision making process of multimodal travelers—this indeed is the work focus for many travel economics researchers. The model parameters and assumptions in the present paper are set simple for the convenience of setting up and interpreting the results obtained with a synthetic testbed. Indeed, the objective of this paper is to highlight the theoretical aspects of the presented framework and test its inferential power; the synthetic data-based experimental investigation is more valuable for this purpose, since synthetic data here also serve as a reliable ground truth, against which the inference algorithm outputs can be compared.

Once the disutility values of the routes are obtained from (2), the probability that a particular route will be chosen by a traveler can be computed using the logit choice model function (1). Note that more complex choice probability models can also be adopted at this stage (e.g., see **?**), however, we use the logit model for the ease of presentation and without loss of generality. One issue with the basic logit model in (1), is that it assumes independence between the route alternatives. In other words, the choice set $\boldsymbol{R}(w)$ is assumed to be made of mutually exclusive routes, i.e., those that do not share any route segments. Hence, a correction factor must be incorporated while computing the disutility for each path. The Path Size (PS) logit model, developed by Ben-Akiva and Bierlaire (2003), accommodates for such a correction factor in the disutility values,

$$\varpi_r = \sum_{a \in \tau_r} \frac{L_a}{L_r} \frac{1}{N_a} = \sum_{a \in \tau_r} \frac{L_a}{L_r} \frac{1}{\sum\limits_{j \in \boldsymbol{R}(w)} \delta_{aj}} \ , \ \forall \ r \in \boldsymbol{R}(w) \ .$$

In the above formulation, ϖ_r is the path size correction factor for route r; the term L_a is the length of the link a overlapping the segments of route r denoted by τ_r; the length of route alternative $r \in \boldsymbol{R}(w)$ for the eligible OD pair w is expressed by L_r; the number of paths in the route set $\boldsymbol{R}(w)$ for the eligible OD w using the link a is denoted by N_a, where $N_a = \sum_{j \in \boldsymbol{R}(w)} \delta_{aj}$; the indicator variable δ_{aj} is set to one if link a is contained in the path j and to zero otherwise. With this correction factor incorporated into the basic logit model (1), the PS logit model becomes

$$P(r(p)) = \frac{\varpi_r \ e^{-(d_r(p))}}{\sum_{r' \in \boldsymbol{R}(w)} \ \varpi_{r'} \ e^{-(d_{r'}(p))}} \ .$$

2.2 Bayesian Approach to OD-Pair Inference

We now describe the Bayesian approach to inferring the true OD pairs that employs the ML model of Section 2.2. The Bayesian approach makes use of the fact that bike route data can be collected for each traveler individually (i.e., at the disaggregate level), by matching their trip records over multiple system usage instances (e.g., by the member pass-card number or by tracking bikes by GPS for a specific member). Once the bike segment information is observed, Bayes rule can be used to describe a mathematical structure, which gives the probability of an OD pair being the true OD from the eligible OD set under the information of the observed bike segment and traveler type distribution.

Given an observation of an individual using bike segment $(i,j)_t^{\text{obs}}$ as part of at least one of their feasible routes, let $\boldsymbol{\Omega}_t = \{w : \exists r \in \boldsymbol{R}(w) \text{ s.t. } (i,j)_t^{\text{obs}} \in r\}$ denote the set of eligible OD pairs for traveler t. By Bayes Rule, for any OD-pair $w \in \boldsymbol{\Omega}_t$, the conditional probability that it is the individual's true OD-pair (as opposed to any other one), given a vector of choice model parameters Θ for the traveler type distribution and the fact he/she chooses bike segment $(i,j)_t^{\text{obs}}$, can be expressed as

$$P(w|\Theta,(i,j)_t^{\text{obs}}) = \frac{P(w)\sum_{\mu \in \boldsymbol{M}} P((i,j)_t^{\text{obs}}|\mu, w) P_\Theta(\mu)}{\sum_{w \in \boldsymbol{\Omega}_t} P(w) \sum_{\mu \in \boldsymbol{M}} P((i,j)_t^{\text{obs}}|\mu, w) P_\Theta(\mu)} \ , \tag{3}$$

where $P(w)$, $w \in \boldsymbol{\Omega}_t$, are prior probabilities that are assumed equal, i.e., meaning that each eligible OD pair is equally likely if no information is available that can help differentiate them (note that such information could come, for example, from census data, allowing one to scale the priors by the total population around their route origins and the number of available workplaces or services around their route destinations). Probabilities $P((i,j)_t^{\text{obs}}|\mu, w)$ are the route choice probabilities given by (1) for the observed bike segment (i,j) and eligible OD-pair w under a given traveler type $\mu \in \boldsymbol{M}$. The probability associated with choosing a traveler type from the traveler type distribution is scaled into (2) as $P_\Theta(\mu)$, where Θ is the vector of parameters for the distribution. The traveler type distribution weights can be discrete or continuous functions of Θ (while the range of this distribution itself is discrete and finite). Since the input traveler type distribution is unknown, we take the expected value for observing the bike segment $(i,j)_t^{\text{obs}}$ over the traveler type distribution Θ.

The described probability structure will be exploited in the maximization step of the SSEM algorithm. The route choices of the same individual are observed in multiple scenarios—over multiple days and under multiple price settings. Assume that the observations are collected over the same periods (e.g., on the same weekdays), to ensure that the travel conditions as well as the travelers' OD pairs remain the same over all the observations.

We proceed with explaining the Bayesian logic exploited for OD pair inference. Figure 2 depicts some eligible origins and destinations around the bike station

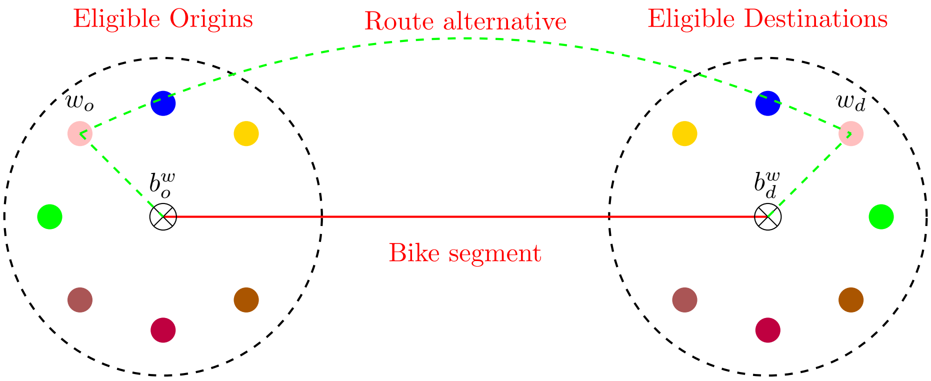

Figure 2: Eligible OD-pairs and route alternative setting for Bayesian inference

entry and exit points, together with the respective route alternatives, out of which at least one route includes the bike segment in red. An origin point which is closer to the observed bike segments from where the traveler enters the bike sharing system and the destination point which is closer to the bike segments where the traveler leaves the bike sharing system have a higher chance of being the true origin and destination locations. The level of inference that we can make is only restricted to the locations around bike stations, since we are observing only the partial route data (bike segments traversed) for the traveler. Given that any complete route may be multi-modal, one can only infer where a traveler arrives from to pick up a bike and where they depart to after they surrender it. Eligible OD locations are seeded inside a geographical zone spanned by the transportation system of interest, i.e., each residential building (or block centroid) and each alternative transportation mode stop (metro, bus station, etc.) that is within the walking distance of any bike station is eligible.

The subsequent subsections discuss the eligible OD set creation and route choice generation for these ODs which serve as the input for the estimation framework.

2.3 Eligible OD Set Generation

The learning algorithm for demand estimation requires a set of eligible or candidate origin-destination pairs for each bike traveler from which the true OD pairs for a given traveler will be distilled with certain belief or probability. In this subsection, we define a strategy to generate an eligible set of OD pairs for each traveler based on the bike segment observation of travelers.

One approach to generating an eligible OD set is to directly sample the observations of travelers. Prior knowledge about an eligible set of OD pairs can be obtained using passenger surveys or travel forecasting data. However, such survey data is difficult to acquire and organize because of the enormity of the bike sharing system. The second methodology for defining an OD pair set relies on observing the routing choices of travelers. The all pair shortest paths (APSP), with a feasible route alternative for each such path for every OD pair, are found in the network, for multiple price settings. Based on the traveler's observed bike segment, those OD pairs are chosen which include the observed bike segment in at least one of its route alternatives. This method may be able to capture the set of eligible OD pairs completely, but suffers from being too computationally expensive and complex to handle.

In order to avoid enumerating all the point on the map as the potential endpoints of a traveler's true route, a zone of eligible origins and destinations can be defined as shown in Figure 2: the circles of a pre-set radius around the entry and exit bike stations serve this purpose well. Once the OD pairs have been generated for each bike segment, we assign the ODs for each traveler by following the steps below.

In the assignment process, the OD pair which includes a traveler's observed bike segment as part of at least one of its route alternatives, across all the scenarios, is called "eligible" for the traveler. The traveler's true OD is expected and assumed to be included in the eligible OD set. Note that ensuring this in practice might turn out rather time consuming; then, further steps might be required to reduce the size of the resulting set: prior knowledge of population's habits, zonal attraction characteristics, and forecasting analyses can be used to this end. Next, aggregated over all the scenarios, an OD-Bike segment matrix is constructed, for each of the distinct traveler type values. Let $\delta_{\mu}^{w,(i,j)}(p)$ be the incidence indicator of the event "Route $r(w) \in \boldsymbol{R}(w)$ for OD w contains bike segment (i,j)," under the pricing strategy p for traveler type μ, and $\Delta_{\mu}(p)$ be a binary matrix whose elements are one if a specific bike segment is a part of the OD pair's route alternative and zero otherwise,

$$\Delta_{\mu}(p)[w,(i,j)] = \begin{cases} 1 & \text{if } (i,j) \in r(w) \\ 0 & \text{if } (i,j) \notin r(w) \end{cases} \qquad \forall\ \mu\ \in \boldsymbol{M}\ .$$

Once the matrices for each traveler type and price combinations are constructed for a given traveler, the traveler's route changes are observed under multiple price settings. The union of the observed bike segment column vectors from the different pricing settings gives the eligible set of OD pairs for the traveler. Let $\Delta(1), ..., \Delta(p), ..., \Delta(|P|)$ denote the block matrices representing the OD-Bike segment matrix for price setting p under a given traveler type for $|\boldsymbol{P}|$ scenarios. The structure of such a block matrix under a given price setting is given,

$$\Delta(p) = \begin{array}{c} \begin{array}{cccccccc} \quad\mu_1 & \qquad\qquad\mu_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \mu_{|\boldsymbol{M}|} \end{array} \\ \left[\begin{array}{ccc|ccc|ccc|ccc} 1 & \cdots & 0 & 1 & \cdots & 0 & & \cdots & & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 & & \cdots & & 1 & \cdots & 0 \end{array}\right] \end{array} .$$

With the observed traveler, t, of traveler type μ, the eligible OD set can be found as an intersection of the column vectors from each observation $(i,j)_t^{\text{obs}}$ as $\Delta(p_1)[(i,j)_t^{\text{obs}_1}] \cap \Delta(p_2)[(i,j)_t^{\text{obs}_2}] \cap ... \cap \Delta(p_{|P|})[(i,j)_t^{\text{obs}_{|P|}}]$, where $\Delta(p)[(i,j)_t^{\text{obs}}]$ is the column of $\Delta(p)$ that corresponds to bike segment $(i,j)_t^{\text{obs}}$ given the day's pricing strategy p. In order to identify the eligible OD set, we take the minimum across all the scenarios,

$$\min\left(\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} ... \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} \right) \equiv \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix} .$$

The resultant vector obtained as a union of the column vectors over all the scenarios contains all the eligible OD pairs for a traveler. This operation is carried out for every traveler in the observed population.

The eligible OD sets generated for each traveler described in this section is then fed into the learning algorithm from which the most likely OD pairs for the passengers are found. The eligible OD set is followed by route alternatives generation for each and every OD pair explained in the next section, out of which at least one of the alternatives may contain the observed bike segment. The route generation methodology is explained in greater detail in the next subsection.

2.4 Route Choice Set Generation

This section covers a computationally tractable method adopted for the generation of route alternatives that a traveler has under the multinomial discrete choice model based on RUT. These alternatives form a discrete finite choice set. Recall that a route can consist of any combination of walking, biking and other public transit mode-based travel segments. The demand estimation model requires a set of reasonable routes, $\boldsymbol{R}(w)$, for each OD pair w, generated as described in the previous subsection.

Path choice generation can be approached deterministically or stochastically. Deterministic algorithms including the labeling approach by **?**, link elimination by J.A. et al. (1993), link penalty by de la Barra et al. (1993), constrained k-shortest path approach by Van der Zijpp and Catalano (2005) and heuristics based on branch and bound by Friedrich et al. (2001); Lanser (2005); Prato and Bekhor (2006) have been used extensively by the researchers in the domain of route choice set generation. Frejinger and Bierlaire (2007) suggest a recently developed random sampling methodology to generate route alternatives and provide a stochastic path generation algorithm. **?** provides a simulation approach to route choice alternative identification, with the routes constructed with links drawn from certain probability distributions. Another stochastic approach suggested by Bovy and Fiorenzo-Catalano (2007) provides a more generalized simulation-based method for route choice generation by considering generalized cost functions in terms of utility values. The route choice generation algorithms based on both the deterministic and stochastic approaches can be computationally expensive to be used for the present work, since our route alternatives have to be generated for each and every individual in the travel population, and then, lumped together for further analysis. Also, our feasible route alternatives must contain bike segments as part of all the routes for the estimation framework to work.

A computationally tractable methodology is now presented for route choice generation. For each eligible OD pair w, generated for the respective bike segments as explained in the above section, let w_o denote the origin node and w_d denote the destination node. For every pair w_o and w_d, the bike stations close to these nodes are identified, with a limit set for the number of bike stations to be considered around each node. After identifying the respective bike stations, corresponding to the origin and destination nodes, a route is generated as a combination of the four nodes w_o, w_d, i and j, where i and j represent the bike stations near the origin and destination nodes, respectively, for the OD pair w.

A route alternative constructed in this manner can be expressed as

$$r(w) \ \rightarrow SP(w_o, i) + SP(i, j) + SP(j, w_d) \ ,$$

where $SP(\cdot\,,\,\cdot)$ stands for the shortest path between a pair of nodes. Since the shortest paths found based on disutility valuation may differ across scenarios, the routes should be generated separately for each scenario, for every combination of the candidate bike stations near w_o and w_d. Figure 3 illustrates the route

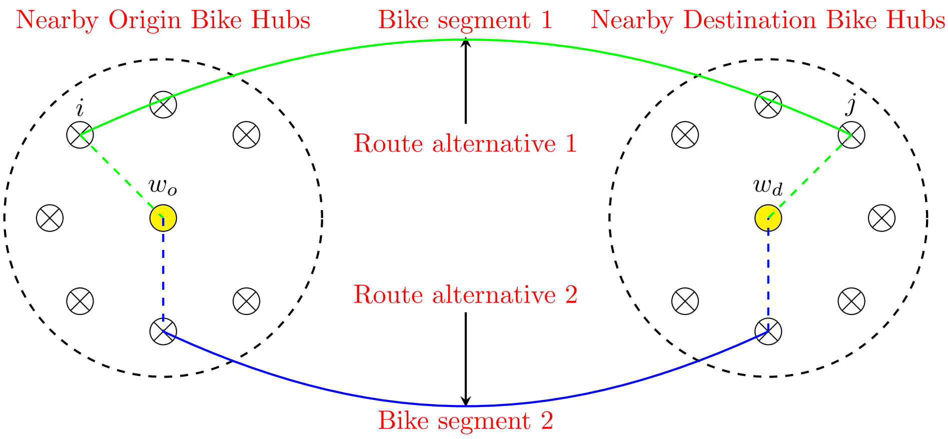

Figure 3: Route choice set generation for an eligible OD-pair

choice set formation process, with two route alternatives shown for an OD-pair (w_o, w_d). The size of the route choice set must be carefully selected for any given navigation or transportation system analysis: for computational tractability reasons, the number of alternatives in the route choice set cannot be large - we set it equal to two. Note, however, that this restriction allows the inference algorithm to consider around 50 distinct alternative OD pairs for every observed traveler.

This section described the details of the inputs and intermediary processes required for generating the inference data to be fed into the SSEM algorithm. The selective preference set, which is also an input to the SSEM algorithm, will be discussed next. To estimate the most likely value of the parameters Θ for the traveler type distribution, a set of eligible ODs with route alternatives for each OD pair needs to be supplied to the SSEM algorithm which will infer the traveler type distribution and use this inferred distribution to learn the true travel OD-pairs by utilizing the observations across the entire travel population under different price settings.

3 Selective Set Expectation Maximization (SSEM) Algorithm Design

This section presents the details of the SSEM algorithm which iteratively estimates the unknown parameters of the traveler type distribution by recovering (learning) the hidden OD pairs the observed travelers. SSEM is a development of the conventional Expectation Maximization (EM) idea first described by

Dempster et al. (1977). The traditional EM algorithm cannot be successfully applied to the presented problem. This is because in each pricing scenario, there exist multiple traveler types that support an observed traveler's route choice; the EM makes a pick from this vector set and then cannot correctly fuse the inference results over all the scenarios, favoring the extreme traveler types (extreme in terms of sensitivity to cost incentives). SSEM, on the other hand, stores the likely solutions to all the scenarios as sets, and makes inference based on the intersection of these sets.

The Maximum Likelihood Estimator (MLE) maximizes the likelihood of the data as a function of model parameters. The maximization of the likelihood can be achieved by the first derivative principle. For many problems, however, the first derivative is difficult to evaluate under latent variables, and hence, one has to resort to more elaborate techniques. The SSEM algorithm is used in the case where the likelihood function is intractable or incomplete or missing data are present. The formal steps leading to the design of the SSEM algorithm are presented next.

Let the likelihood function for the OD-pair inference of a single traveler t, under multiple price settings, be expressed as

$$\prod_{\text{obs}} P((i,j)_t^{\text{obs}})|\Theta) \ . \tag{4}$$

The MLE, i.e., the estimate of the best-fit parameter vector Θ of the traveler type distribution, is found using (4),

$$\widehat{\Theta}_{\text{MLE}} = \prod_t \prod_{\text{obs}} P((i,j)_t^{\text{obs}})|\Theta) \ . \tag{5}$$

Taking the logarithm in (5), the likelihood function is defined as

$$\begin{aligned} L(\Theta)_{\widehat{\Theta}_{\text{MLE}}} = \arg\max_{\Theta} L(\Theta) &= \arg\max_{\Theta} \ \log \ \prod_t \prod_{\text{obs}} P((i,j)_t^{\text{obs}})|\Theta) \\ &= \arg\max \sum_t \sum_{\text{obs}} \log P((i,j)_t^{\text{obs}})|\Theta) \ , \end{aligned} \tag{6}$$

where $P((i,j)_t^{\text{obs}}|\Theta)$ is the marginal density function found by summing over all the latent variables w_t $\forall t$. The true OD pair w_t for traveler t is unknown and treated as a hidden/latent variable. The set of the bike segments $(i,j)_t^{\text{obs}}$, along which the traveler was observed to bike, under each price setting, are termed available data; the traveler's unobserved OD is termed missing data. A new likelihood function can now be defined by working with the joint distribution of the observed bike segments and the unobserved or unknown OD pairs; this step gives the log-likelihood function that allows one to obtain an MLE based only on the available data,

$$L(\Theta) = \arg\max_{\Theta} \sum_t \sum_{\text{obs}} \log P((i,j)_t^{\text{obs}}, w_t|\Theta) \ . \tag{7}$$

By using the multiplicative rule for conditional probability we marginalize the observed bike segments in (7) over the eligible OD pairs to obtain

$$L(\Theta) = \sum_t \sum_{\text{obs}} \log \sum_{w_t \in \mathbf{\Omega}} P((i,j)_t^{\text{obs}} | w_t, \Theta) P(w_t|\Theta) \ , \tag{8}$$

where the set of eligible OD pairs $\mathbf{\Omega}$ is built for traveler t. Suppose the best incumbent found so far for Θ is $\Theta^{(n)}$, and one seeks to improve this incumbent. Transforming (8) by simple manipulations, one obtains

$$\begin{aligned}
L(\Theta) - L(\Theta^{(n)}) &= \sum_t \sum_{\text{obs}} \log \sum_{w_t \in \mathbf{\Omega}} P((i,j)_t^{\text{obs}} | w_t, \Theta) P(w_t|\Theta) - L(\Theta^{(n)}) \\
&= \sum_t \sum_{\text{obs}} \log \sum_{w_t \in \mathbf{\Omega}} \left[\frac{P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) P((i,j)_t^{\text{obs}}|w_t, \Theta) P(w_t|\Theta)}{P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)})} \right] \\
&\qquad - \sum_t \sum_{\text{obs}} \log P((i,j)_t^{\text{obs}} | \Theta^{(n)}) \ .
\end{aligned}$$

Since the log term in above expression is convex, by using the Jensen's inequality for log-concave functions, one can replace the logarithm of sums by the sum of logarithms to obtain the inequality,

$$\begin{aligned}
L(\Theta) - L(\Theta^{(n)}) \geq \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} & P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \times \\
& \log \left[\frac{P((i,j)_t^{\text{obs}}|w_t, \Theta) P(w_t|\Theta)}{P((i,j)_t^{\text{obs}}|\Theta^{(n)}) P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)})} \right] \ ,
\end{aligned}$$

and hence,

$$\begin{aligned}
L(\Theta) - L(\Theta^{(n)}) \geq \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} & P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \times \\
& \log \left[\frac{P((i,j)_t^{\text{obs}}, w_t|\Theta)}{P(w_t, (i,j)_t^{\text{obs}}|\Theta^{(n)})} \right] \ ,
\end{aligned} \tag{9}$$

where $P((i,j)_t^{\text{obs}}, w_t|\Theta)$ and $P(w_t, (i,j)_t^{\text{obs}}|\Theta^{(n)})$ are the joint probability distributions under Θ and $\Theta^{(n)}$, respectively. Denoting the right-hand side of the above expression by $B(\Theta, \Theta^{(n)})$,

$$\begin{aligned}
B(\Theta, \Theta^{(n)}) \equiv \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} & P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \times \\
& \log \left[\frac{P((i,j)_t^{\text{obs}}, w_t|\Theta)}{P(w_t, (i,j)_t^{\text{obs}}|\Theta^{(n)})} \right] \ ,
\end{aligned} \tag{10}$$

and using (10), the inequality in (9) becomes

$$L(\Theta) \geq B(\Theta, \Theta^{(n)}) + L(\Theta^{(n)}) \ .$$

If $\Theta = \Theta^{(n)}$, then $\log \left[\frac{P((i,j)_t^{\text{obs}}, w_t|\Theta)}{P(w_t, (i,j)_t^{\text{obs}}|\Theta^{(n)})} \right] = 0 \implies L(\Theta) = L(\Theta^{(n)})$. By maximizing the lower bound $B(\Theta, \Theta^{(n)})$ over all possible values of Θ, one guarantees that at $\Theta^{(t+1)}$ a higher value of the likelihood is achieved, $L(\Theta^{(n+1)}) \geq$

$L(\Theta^{(n)})$. Since $P(w_t, (i,j)_t^{\text{obs}}|\Theta^{(n)})$ is not a function of Θ, it is convenient to express this lower bound (10) as

$$
\begin{aligned}
B(\Theta, \Theta^{(n)}) &= \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \ \times \log P((i,j)_t^{\text{obs}}, w_t|\Theta) \\
&= \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \ \times \log[P((i,j)_t^{\text{obs}}|w_t, \Theta) P(w_t|\Theta)] \ ,
\end{aligned} \tag{11}
$$

where $B(\Theta, \Theta^{(n)}) \iff E_{w|(i,j),\Theta^{(n)}}[\log P((i,j)_t^{\text{obs}}, w_t|\Theta)]$, i.e., the expected complete log-likelihood with respect to the probability distribution function $P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)})$ in (2). The likelihood of an eligible OD pair for a traveler, across all the observations available for this traveler, is independent of the traveler type Θ and only depends on the bike segment along which trips were made. This implies that w_t is independent of Θ and it suffices to say that $P(w_t|\Theta) \iff P(w_t)$ for a certain observation of traveler t under the current traveler type. Equivalently, one can re-write (10) as

$$
\begin{aligned}
B(\Theta, \Theta^{(n)}) = \sum_t \sum_{\text{obs}} \sum_{w_t \in \mathbf{\Omega}} & P(w_t|(i,j)_t^{\text{obs}}, \Theta^{(n)}) \ \times \\
& \log[P((i,j)_t^{\text{obs}}|w_t, \Theta) P(w_t)] \ .
\end{aligned} \tag{12}
$$

Gradient ascent-based methods (e.g., the Newton-Raphson algorithm) perform continuous optimization iteratively, using an approximation of the objective function to determine how to move to an improved solution at each iteration. These methods, however, generally suffer from the quality of the employed approximations and the difficulty in determining the appropriate step size (i.e., in finding how far to move along an identified improving direction). The EM algorithm computes a local approximation of the likelihood function (at a given, current best solution) as a lower bound to the objective function, and at every iteration, re-evaluates this lower bound (see, e.g., (Minka, 1998)). Per the Jensen's inequality, the algorithm will find an improving move as long as the gradient at a current solution is not zero; in practice, algorithm stopping criteria typically prescribe to stop it when the gradient gets "small enough."

Once the lower bound in computed during the E-Step of the SSEM algorithm, one can restrict the set of travelers types to be searched over by the M-Step of the algorithm, for each traveler in the population. This rule-based approach is developed to capture the semantics of the true travel population and tie this information to the traditional EM algorithm. This is achieved by generating a set of traveler types for which the logit probability for the observed bike segment is greater than that for any alternative path for any eligible OD $w \in \mathbf{\Omega}$, under a given scenario. The finalized traveler type set is then obtained as the intersection of such sets obtained for all the scenarios. These sets are then fed to the E-Step while evaluating the lower bound $B(\Theta, \Theta^{(n)})$.

More specifically, let $\boldsymbol{M}_t(p)$ be the traveler type set for traveler t under scenario p. Then $\boldsymbol{M}_t(p) = \{\mu : P(i,j)_\mu > P(\text{alt})_\mu\}$, where $P(i,j)_\mu$ and $P(\text{alt})_\mu$ are the logit probabilities for the observed bike segment and the alternative bike segment under a traveler type $\mu \in \boldsymbol{M}$. The intersection across the different scenarios gives the traveler type set for traveler t, $\boldsymbol{M}_t = \boldsymbol{M_t(p_1)} \cap \boldsymbol{M_t(p_2)} \cap ... \boldsymbol{M_t(p_{|P|})}$. The resulting effect of this is that instead of summing $P((i,j)_t^{\text{obs}}|w_t, \mu) \times P_\Theta(\mu)$ over all traveler types, it is summed across only those traveler types that are contained in the traveler type set of traveler t.

We now outline the full procedure of the SSEM algorithm for true OD-pair inference. At each iteration of SSEM, a lower bound is found by calculating the expected value of the log-likelihood function. This expected value is conditioned on the distribution over the unknown OD-pairs, given the information about the bike segments under the current estimate of the traveler type probabilities. A selective set of traveler types is also used, while evaluating this expected value: this step is called the expectation step (E-Step). In the maximization step (M-Step) we find such an estimate for the traveler type probabilities that maximizes our lower bound. These E-Steps and M-Steps are performed iteratively, for the entire traveler population, to find the best-fit estimate of the traveler type distribution in the population and infer each traveler's most likely OD-pair. See Algorithm 1 steps, summarizing the SSEM procedure.

3.1 Inferring traveler type distribution via constrained non-Linear optimization

In the maximization step of the SSEM algorithm, a constrained nonlinear optimization problem is solved. This section details the solution approach to this problem. The unknown model parameters for the traveler type distribution are just the probability values for each traveler type: these values must be real numbers between zero and one that sum up to one. The decision variables in the formulation are the unknown probability values of the discrete traveler type distribution. The constrained nonlinear formulation maximizes the lower bound $B(\Theta, \Theta^{(n)})$ computed in the E-Step using the partial observed data and the reconstruction of the hidden OD-pair structure for the travelers under the current estimate of the unknown distribution. The formulation for the maximization step is given by model (P)

$$
\text{(P)} \qquad
\begin{aligned}
\max \sum_{t} \sum_{\text{obs}} \sum_{w_t \in \boldsymbol{\Omega}} C \ \log[x_1^{\text{obs}}\ \Theta_1 + x_2^{\text{obs}}\ \Theta_2 + ... + x_9^{\text{obs}}\ \Theta_9] \\
\text{s.t.} \ \sum \Theta_i = 1 \ , \ i = 1, 2, ..., 9 \ , \qquad (P1) \\
0 \leq \Theta_i \leq 1 \ , \ i = 1, 2, ..., 9 \ . \qquad (P2)
\end{aligned}
$$

In (P), constant C stands for the probability $P(w_t|(i,j)_t^{\text{obs}})$ and x_i^{obs}, $i = 1, 2, ..., 9$, is the product of the logit and prior probabilities $P((i,j)_t^{\text{obs}}|w_t, \mu) \times P(w_t)$. Since the traveler type space is discrete, the values Θ_i, $i = 1, 2, ..., 9$, are the probability (mass) values of the distribution $P_\Theta(\mu)$. Problem (P) can be solved, e.g., using the `fmincon` function in Matlab which implements an interior point method. The summation across Θ_i values in constraint (P1) represents the normalization probability axiom. The second constraint ensures that each Θ_i is non-negative to follow the non-negativity axiom from the law of probabilities. At every iteration at the maximization stage, (P) is formulated and solved until convergence to the input traveler type distribution.

The time complexity of an E-Step is $\mathcal{O}(NMKR)$, where N is the number of travelers, M is the number of scenarios/price settings and K and R are the number of OD pairs in the eligible OD set and the number of route alternatives for each OD pair w, respectively. If the counts of alternatives and eligible OD pairs are restricted, the complexity is $\mathcal{O}(NM)$. The complexity of the M-Step, and hence, of the whole SSEM algorithm is also $\mathcal{O}(NM)$. To improve the runtime

Algorithm 1 SSEM algorithm for MLE estimator

function E STEP(μ , $\Theta^{(n)}$) $\triangleright$ The expectation step
    Inputs:
    $\mu \leftarrow \mu \quad , \forall \ \mu \in \boldsymbol{M}$
    $\Theta^{(n)} \leftarrow P_{\Theta}(\mu) \quad , \forall \ \mu \in \boldsymbol{M}$
    for $t \in \boldsymbol{T}$ **do**
        for obs $\in$ **obs do**
            Initialize:
            $P(w) \leftarrow 1/|\boldsymbol{\Omega}_t| \quad , \forall \ w \in \boldsymbol{\Omega}_t$
            $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow 0 \quad , \forall \ w \in \boldsymbol{\Omega}_t, \forall \mu \in \boldsymbol{M}$
            $B_t^{\text{obs}}(w) \leftarrow 0 \quad , \forall \ w \in \boldsymbol{\Omega}_t$
            for $w \in \boldsymbol{\Omega}_t$ **do**
                $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow \frac{\varpi_r \exp(-d_r(p))}{|\boldsymbol{\Omega}|} \ , \forall \mu \in \boldsymbol{M} \ , \forall r \in \boldsymbol{R}(w)$
            $Z_\mu \leftarrow \sum_{r \in \boldsymbol{R}(w)} P_{\text{logit}}(w|\mu)_t^{\text{obs}} \ , \forall \mu \in \boldsymbol{M}$
            for $r \in \boldsymbol{R}(w)$ **do**
                $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow \frac{P_{\text{logit}}(w|\mu)_t^{\text{obs}}}{Z}$
            for $w \in \boldsymbol{\Omega}_t$ **do**
                $B_t^{\text{obs}}(w) \leftarrow \ P(w) * \sum_{\mu \in \boldsymbol{M}} P_{\text{logit}}(w|\mu)_t^{\text{obs}}$
            $Z' \leftarrow \sum_{w \in \boldsymbol{\Omega}_t} B_t^{\text{obs}}(w)$
            for $w \in \boldsymbol{\Omega}_t$ **do**
                $B_t^{\text{obs}}(w) \leftarrow \frac{B_t^{\text{obs}}(w)}{Z'}$
    return $B_t^{\text{obs}}(w) \ \ \forall$ obs $\in$ **obs**, $\ \forall \ \ t \in \boldsymbol{T} \ \ \forall w \in \boldsymbol{\Omega}_t$

function M STEP(B_t^{obs} , $\boldsymbol{M}_t$) $\triangleright$ The maximization step
    Inputs:
    $B_t^{\text{obs}}(w) \quad , \forall$ obs $\in$ **obs**, $\ , \forall \ \ t \in \boldsymbol{T} \ \ , \forall w \in \boldsymbol{\Omega}_t$
    Initialize:
    $Q(\Theta, \Theta^{(n)}) \leftarrow 0$
    $L_t^{\text{obs}}(w) \ \ \forall$ obs $\in$ **obs**, $\ \forall t \in \boldsymbol{T}, \ \ \forall w \in \boldsymbol{\Omega}_t$
    for $t \in \boldsymbol{t}$ **do**
        for obs $\in$ **obs do**
            $L_t^{\text{obs}}(w) \leftarrow \log[\sum_{\mu \in \boldsymbol{\Omega_p}} P((i,j)_t^{\text{obs}}|w,\mu) P_{\Theta}(\mu) P(w)]$
    Initialize:
    $B_{\text{obs} \leftarrow \left[[\ldots, B_1^{\text{obs}}[w], \ldots], [\ldots, B_1^{\text{obs}}[w], \ldots], \ldots, [\ldots, B_t^{\text{obs}}[w], \ldots]\right] \ \forall \text{obs} \in \textbf{obs}}$
    $L_{obs} \leftarrow \left[[\ldots, L_1^{\text{obs}}[w], \ldots], [\ldots, L_1^{\text{obs}}[w], \ldots], \ldots, [\ldots, L_t^{\text{obs}}[w], \ldots]\right] \ \forall \text{obs} \in \textbf{obs}$
    $Q(\Theta, \Theta^{(n)}) \leftarrow \sum_{\text{obs} \in \textbf{obs}} (B_{\text{obs}})^T \times L_{\text{obs}}$
    $\Theta^{(n+1)} \leftarrow \arg\max_{\Theta} Q(\Theta, \Theta^{(n)})$
    return $\Theta^{(n+1)}$

procedure SSEM( ) ▷ Main function implementing the SSEM algorithm
  Inputs:
  $\boldsymbol{M}_t \;\; \forall t \in \boldsymbol{T}$
  Initialize:
  $n \leftarrow 0$
  $\Theta^0 \leftarrow \text{rand}\big(P_\Theta(\mu)\big) \;\forall \boldsymbol{M}$
  repeat
    E STEP(μ , $\Theta^{(n)}$)
    M STEP(B_t^{obs} , $\boldsymbol{M}_t$) $\;\; \forall \text{obs} \in \textbf{obs}, \;\; \forall t \in \boldsymbol{T}$
    $n \leftarrow n + 1$
  until Convergence

performance, a parallel implementation of both E-Step and M-Step can be developed, exploiting multiprocessing. If the presented estimation methodology is used with the real-world, large scale multimodal network data, then a distributed computing approach should be considered for efficient scalability to accommodate the quartic nature of the expectation step, which takes most time. The Appendix section provides a parallelized version of the SSEM algorithm.

The output of SSEM is a traveler type distribution which is then used to infer the true OD-pairs, using the Bayesian Inference approach discussed in Section 2; it provides the most likely estimates for the true OD pairs, which is discussed in greater detail in the next section.

3.2 Inference approach for true OD-pairs

This subsection presents the final step of the inference algorithm, which infers the true traveler origin-destination pairs with certain belief/probability by utilizing the knowledge of the learned traveler type distribution.

The distribution of traveler types estimated by the SSEM algorithm is characteristic of the entire traveler population. In order to infer the OD pairs for all the individual travelers, one needs to assign to them such traveler types whose probability values are most probable and are in compliance with the inferred distribution. This can be done by observing what routing choices each traveler made across the different price settings and choosing the traveler types that support those choices. The selective set of the traveler types generated per the the approach explained in the previous section can be used here. Under any given scenario, one can compute the probability of selecting an eligible OD from the OD set $\boldsymbol{\Omega}_t$ as

$$P(w|\Theta,(i,j)_t^{\text{obs}}) = \frac{P(w)\sum_{\mu\in\boldsymbol{M}_t} P((i,j)_t^{\text{obs}}|\mu,w)P_\Theta(\mu)}{\sum_{w\in\boldsymbol{\Omega}_t} P(w)\sum_{\mu\in\boldsymbol{M}_t} P((i,j)_t^{\text{obs}}|\mu,w)P_\Theta(\mu)} \;. \tag{13}$$

This probability value for each eligible OD $w \in \boldsymbol{\Omega}_t$ must be computed for the different price settings (scenarios). Then, the most likely OD-pair for this individual (t) is identified as

$$w_t^* = \underset{w_t\in\boldsymbol{\Omega}_t}{\arg\max} \prod_{\text{obs}} P(w_t|\Theta,(i,j)_t^{\text{obs}}) \;. \tag{14}$$

4 Testing & Evaluation: Test bed specification

To test and validate the estimation model, two data sets are used. Simulated data sets are used for evaluating the performance of the proposed estimation framework. In this section, we first discuss the simulation environment developed for the case studies along with presenting the model characteristics and assumptions made in creating the simulation environment to generate the synthetic data for estimation purpose.

An agent based simulation (ABS) modeling approach is used to generate the synthetic data. ABS is a flexible method for analyzing uncertainties in complex networks, and in particular, modeling individual traveler's behavior. The two major components of this ABS model are (1) the construction of a data set of travelers/agents with known trip (OD) demand, and (2) the simulation of the travelers' routing over the multi-modal network. The ABS model generates travelers based on demand points and fixes their traveler types according to a pre-set distribution. We capture the information of the bike segments the agents travel on, which are fed into the inference algorithm for estimating the assumed distribution while creating the ABS model.

The simulation framework for generating travelers, assigning a traveler type and capturing the observed bike segment is outlined in Figure 4. The ABS model requires the creation of a multimodal network spanning such modes as bike, bus, metro and walking. The multi-modal network can be represented by

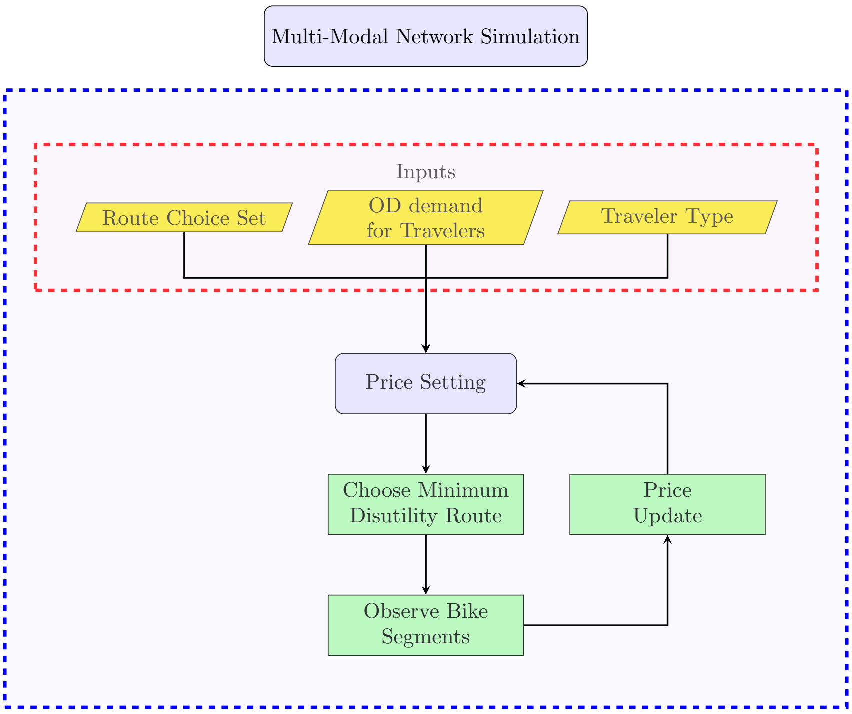

Figure 4: Framework for Agent Based Simualtion Model

Figure 5. The schematic representation shows the multimodal travel pattern of an individual agent in the network. The hexagonal symbols denote the origin

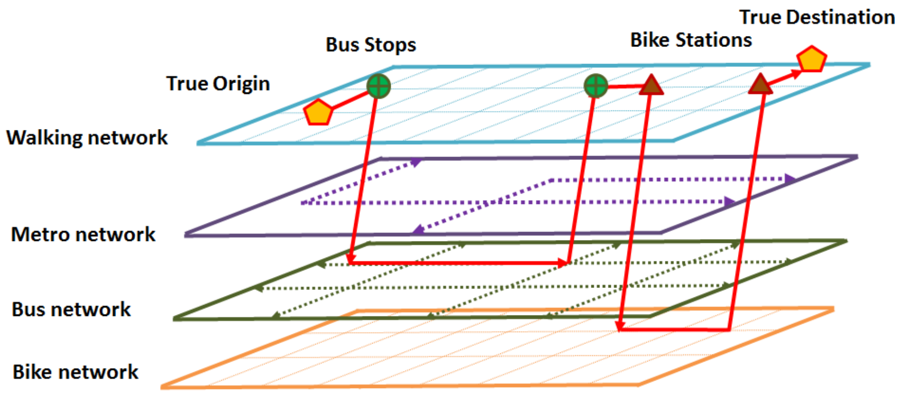

Figure 5: Multi-Modal Network for ABS framework

and destination of a traveler and is obtained from know demand data. The public transportation modes bus, metro and the shared mobility bike sharing system are all connected to the walking layer. A traveler can only enter and egress a mode (or switch between modes) by transitioning through the walking layer. The next step requires the creation of a distribution for assigning the traveler types to travelers. The traveler types are sorted in the increasing order of their price/time ratio and the traveler type with the lowest price/time ratio is assigned the maximum probability. We thus use an exponential distribution to assign the travelers, a traveler type. Once the multimodal-network is created, the price settings for the links between the walking layer and bike network transition are varied across scenarios in the Generate Scenario stage. After the scenario creation, the simulation model utilizes the data and distributes the demand over the network in the trip generation phase. Price settings, and traveler types assigned to each traveler are also provided as an input to the model. For each agent, a route choice set needs to be provided. There can be different strategies implemented here. Choice set generation algorithms discussed in the literature as mentioned in Section 2.4 can be used for generating routing alternatives for the agents. Another way is to compute the shortest path in each and every scenario for the OD assigned from the demand data and consider the set of unique paths across all scenarios as the choice set. However for simplicity and computational tractability, the choice set is generated by choosing the shortest path under the base case and an alternative meaningful route.

In the multimodal network, travelers' true origin and destination nodes need to be identified. Note that the origin-destination nodes connected to the multimodal network by the demand points are not the true OD pairs with respect to the bike sharing system. In fact, these locations cannot be learned and the inference algorithm can only infer true OD pairs around bike stations. So a true origin-destination node that the SSEM algorithm can successfully infer can be either a node on the walking layer, a bus station or a metro station. The inference of such nodes can aid an analyst in modeling external demand entering into the multimodal system. The route alternatives need to be generated based on these true OD pairs. Once the route alternatives are generated, an agent makes his/her routing decision my minimizing the disutility function. That is, every agent solves the problem of choosing an optimal route $r \in \boldsymbol{R}(w)$ represented by

$$\min_{r \in \boldsymbol{R}(w)} C_r(\mu, p) \ .$$

The subsequent subsections focus on implementing the ABS framework on two test cases.

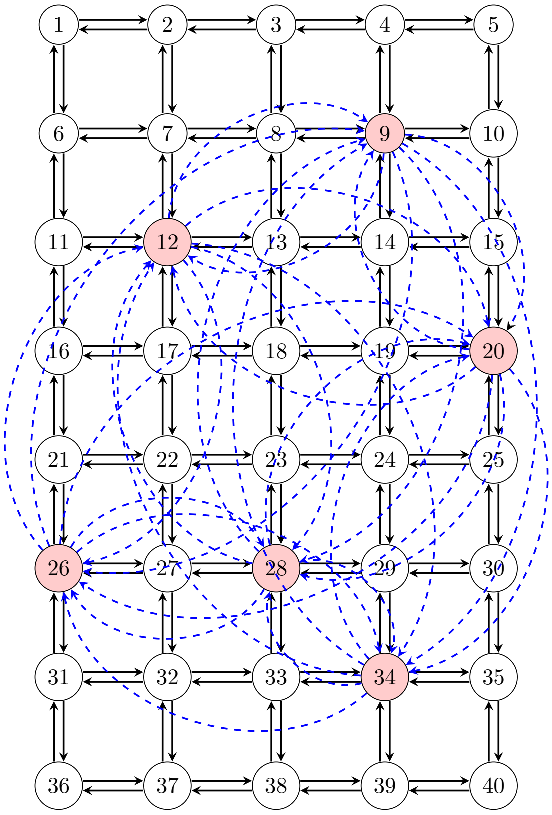

Figure 6: Weighted network representation for Case Study 1

4.1 Case Study 1

We first present a test example of doing inference on a small network with only two modes of transportation: a walking layer and a shared mobility system, i.e., a bike sharing system. The network shown in Figure 6 is composed of 40 nodes and 134 links in the walking layer and with 6 nodes as bike hubs/stations with 30 links as bike segments. All the links of the graph are weighted; travel time on walking and biking layer and entry/exit cost on the transition links between walking and biking. The network is a directed graph with no loops. Each of the bike hubs/stations represented by red nodes on the biking layer are connected to a corresponding node on the walking layer by two directed links; transitioning from walking to biking and vice-versa. The total number of travelers uploaded on this network are twenty. All 20 travelers are assumed to use the bike sharing system, with no traveler having the same origin-destination pair. Once the demand of 20 travelers is uploaded on the network, the traveler types are assigned to these 20 travelers using a distribution as shown in Figure 7. If all classes of travelers are considered, the overall distribution will be a

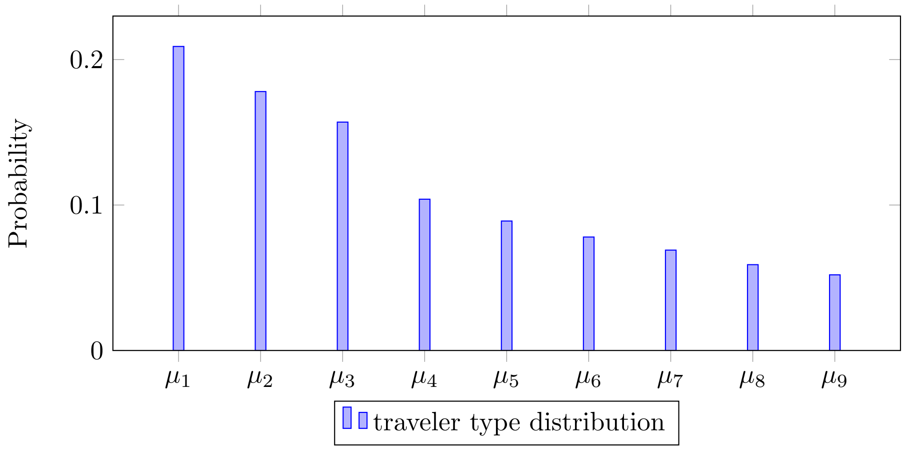

Figure 7: traveler type distribution.

traveler types ordered in increasing order of price/time ratio (sensitivity).

normal distribution with travelers having traveler types different than these nine values. The nine traveler types are assumed to be at the end or tail of the normal distribution. In the shown distribution, μ_1 has the highest probability since it is the least price sensitive and μ_9 has the lowest probability since it has the highest price sensitivity. Based on the count of the travelers using a specific traveler type, a new distribution is created which is the Traveler type distribution to be inferred using the estimation framework.

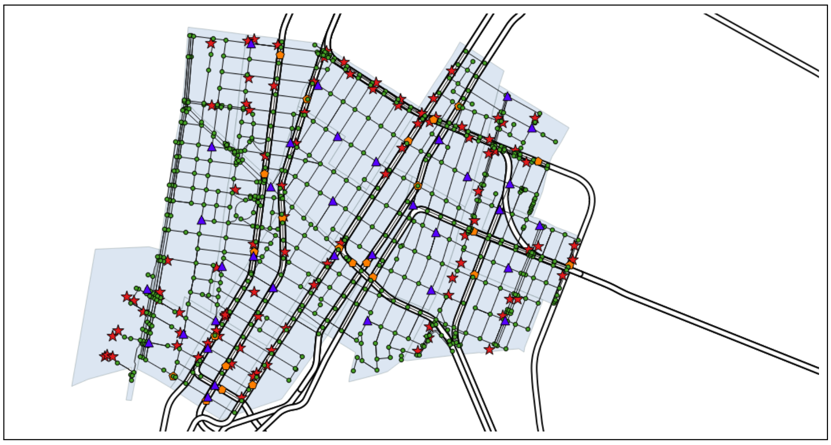

Figure 8: Multi-modal Network for Lower Manhattan, New York

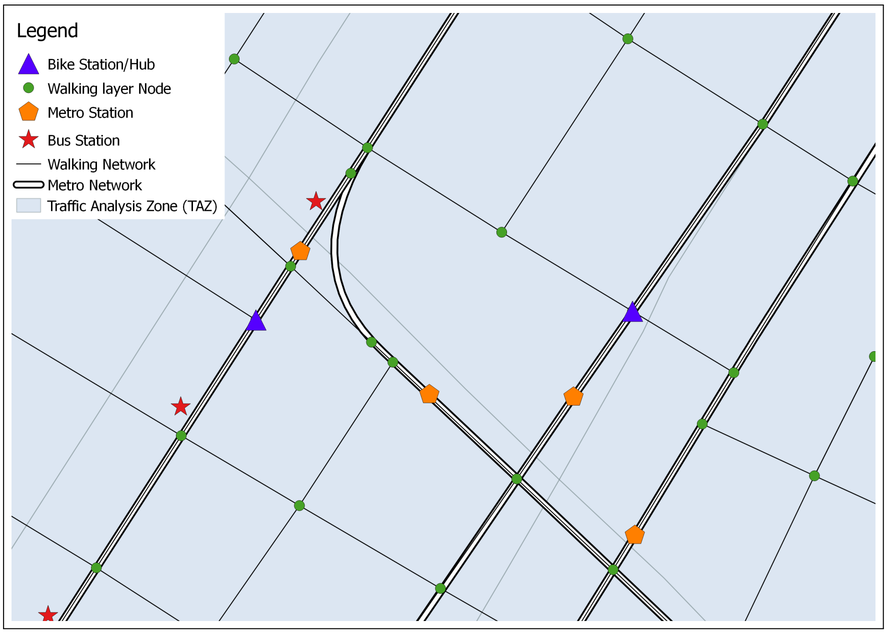

Figure 9: Zoomed snippet of the Multi-modal network for Manhattan

4.2 Case Study 2

In this section, the estimation model is implemented on a representative network of the Citi bike sharing system which includes a portion of Manhattan, New York. Figure 8 depicts the multimodal network for a portion of lower Manhattan, New York. The network has been modeled using real time schedules for the transit modes. The tool used to model the multimodal network is Q-GIS a Geographical Information System software. As per ESRI, 1995 GIS is a tool that organizes geographic data using a software platform for data management and integrates it with computer hardware to effectively display the different geographic features along with analyzing or manipulating the geospatial data. Q-GIS is a an open source version which is used to achieve these functionality. The data sets used for creating the multimodal network's walking layer have been acquired from the 'New York City, Department of City Planning'. The bike stations/hubs location coordinates have been identified by using the information available on Citi Bike's online maps. Geo reference data for constructing the networks for bus and metro modes have been obtained from the Metropolitan Transportation Authority (MTA). The link costs for the public modes; bus and metro are modeled using the transit schedules available by MTA. For the links between bike stations on the biking layer, we assume that the travelers bike on the roads with a few exceptions like freeways and underpasses; so the network for bike sharing system is similar to the walking layer factored by the speed of biking. To model transfers between different transit lines for bus and metro and waiting at the stations, headway calculated from the transit schedules are used. One major assumption made while creating the networks is that no congestion is taken into consideration for any mode, which can have a major effect on the travel disutility function. The period of study for this case is the peak morning period from 9:00 am to 10:00 am. This particular time frame is selected as there will be maximum load on the multimodal transit system during this period. In the zoomed snippet shown in 9, the green dots represent the nodes of the walking layer, the orange pentagons represent the metro stations, the blue triangles represent the bike stations/hubs and the red stars represent the bus stations. Demand data acquired from 'New York City, Department of City Planning'is then uploaded onto the network, after the network creation stage. The traffic flow from the demand points not included in the study area are considered as external travelers being born at the nearest metro or bus stations. Similarly for travelers flowing out of the study area, are considered to be leaving the study area through bus or metro stations located at the periphery of the network. The total number of travelers after uploading the demand is 45,860. Using a similar traveler type distribution, traveler types are assigned to the travelers and a 100 scenarios are generated. Those travelers that do not use the bike sharing system under any of the price settings are filtered out to make the problem more computationally tractable. The total number of travelers that actually end up using the bike sharing system are 2,852 which is around 6% of the travel population and is a reasonable number of agents for implementing the estimation model. Due to computational challenge of handling a large population along with their respective inputs, we test the estimation framework only on 100 travelers. In the next section we present the estimation results for the illustrative example and the case study discussed in the current section.

5 Results and Discussion

In order to test the inference methodology, numerical studies were carried out over two sets of testing data described in the previous section. The first test set consists of a small network topology whereas the second is a network topology representing a part of Manhattan, New York City as described in Section 4.2. The simulation study period was chosen as the peak AM period from 9:00 AM to 10:00 AM, to have maximum traveler density in the test network. The inference algorithm is applied to both these networks and the inferred distributions are as shown in Figures 10 and 11.

Note that the inferred distribution overstates the probability values for certain traveler types and understates the values for others. More specifically, the estimation method is able to provide good estimates of the traveler types for the travelers that are fairly sensitive to pricing (see traveler types 7, 8 and 9) and for the travelers that are more time sensitive (see vectors 1, 2 and 3). However, the travelers whose traveler types are sensitive to both pricing and time are hard to distinguish. This issue can be resolved with an addition of more scenarios, as will be seen in case study 2.

Once the traveler type distribution is inferred, the true OD pairs for each traveler are deduced using the methodology presented in section 3.4. Table 2 provides the details of the estimated OD pairs on the illustrated example. If the traveler's Origin-Destination locations are found correctly it is represented by an indicator value of 1 else 0. The estimation framework fails to identify the true OD pair only in 5 cases. The closeness of the wrongly estimated origins and destinations can be computed to measure the true error in the estimation method. A closeness rank for origins and destinations around the wrongly estimated OD

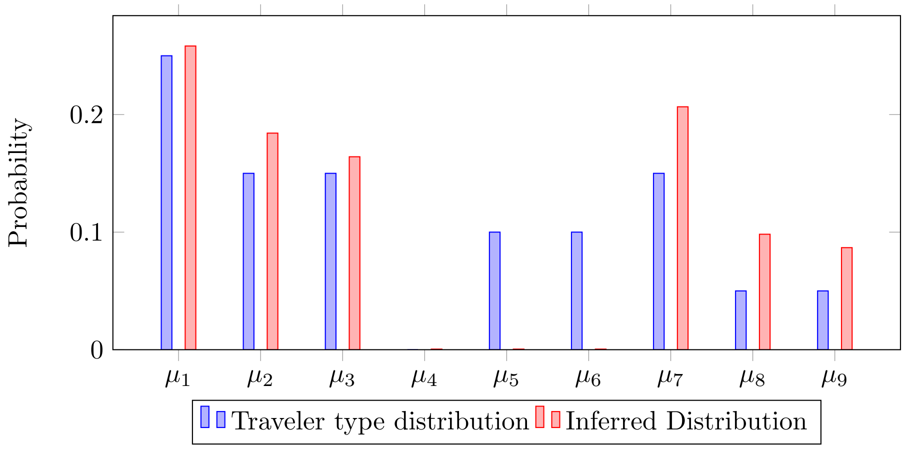

Figure 10: PDFs for Traveler Input and Learned distribution using Selective Set EM algorithm for MLE on small network topology

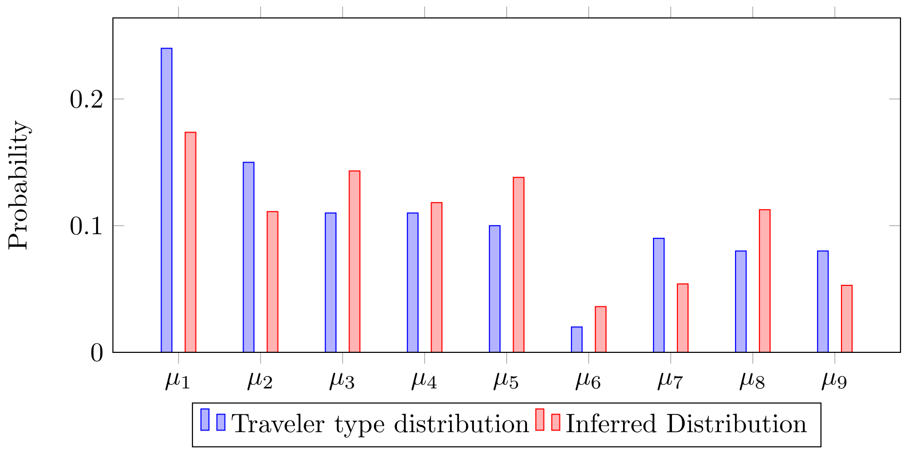

Figure 11: Learned distribution using Selective Search EM algorithm for MLE on network topology representing Manhattan, New York

pairs are computed. The eligible origins around the true origin and eligible destinations around the true destination node are ranked based on the closeness of the nodes to the true origin and destination nodes. From Table 2 it can be seen that for OD pair 30-16, the origin assigned as the true origin is the closest node whereas the destination assigned as the true destination is the second closest node from the set of eligible origins and destination points. The percentage of OD pairs correctly inferred for the small network is around 75 %, whereas for the larger network it is around 78.57 %. For the larger network we report the OD estimation results in terms of aggregate travel demand inferred from the inference algorithm and compare it with the true travel demands. Table 3 shows the true counts and inferred counts for each OD pair along which multiple individuals travel. The percentage of OD pairs correctly inferred is around 76.47 %.

We next present the sensitivity analysis for the SSEM inference algorithm

Table 2: Inferred OD pairs for each traveler.

Column1: Traveler ID, **Column2:** True Traveler OD from demand data, **Column3:** Traveler OD pair inferred using EM algorithm, **Column4:** Variable with value 1 if true and inferred OD are same else 0, **Column5:** Rank of the origin in the inferred OD in terms of distance from the true origin, **Column6:** Rank of the destination in the inferred OD in terms of distance from true destination.

Traveler ID	True OD	Inferred OD	Inference Indicator	Closeness Rank for Origin	Closeness Rank for Destination
T1	1-30	1-30	1	-	-
T2	36-5	36-5	1	-	-
T3	38-3	38-3	1	-	-
T4	30-16	24-8	0	1 ‖ 4	2 ‖ 4
T5	38-3	40-13	0	1 ‖ 3	2 ‖ 3
T6	8-24	8-24	1	-	-
T7	15-27	15-27	1	-	-
T8	16-35	16-35	1	-	-
T9	1-40	1-40	1	-	-
T10	13-37	13-37	1	-	-
T11	5-37	5-37	1	-	-
T12	1-15	1-15	0	1 ‖ 2	1 ‖ 2
T13	16-29	8-30	1	-	-
T14	15-32	15-32	1	-	-
T15	30-13	30-13	1	-	-
T16	2-32	19-32	0	3 ‖ 4	-
T17	22-5	36-13	0	1 ‖ 2	2 ‖ 5
T18	39-18	39-18	1	-	-
T19	10-21	10-21	1	-	-
T20	31-15	31-15	1	-	-

to the number of scenarios. Figure 12 shows the performance of the SSEM algorithm in terms of Mean Squared Error (MSE) for the inferred distribution and the percentage of OD-pairs correctly inferred across different number of scenarios. The sensitivity analysis is performed for both the test cases. It can be seen from Figure 12 that the MSE for the inferred distributions goes on decreasing as the number of scenarios increases, whereas the percentage of the correctly estimated OD-pairs goes on increasing as the number of scenarios increases. As a result, the SSEM algorithm is able to learn more about the traveler's preferences and true OD-pairs as the size of the data increases on increasing the number of scenarios.

Table 3: Aggregate travel demand inferred using the estimation framework

OD Pair	True Count	Inferred Count	OD Pair	True count	Inferred Count	OD Pair	True count	Inferred Count
35-66	2	2	66-243	5	2	251-51	1	1
35-51	2	2	51-42	1	0	14-66	2	1
61-66	1	1	689-66	1	0	23-273	1	1
99-51	1	1	40-40	1	1	92-273	1	1
40-69	1	1	582-98	1	1	251-273	1	1
655-14	2	1	722-66	5	5	92-51	2	2
117-1	1	0	720-66	2	2	23-388	1	1
298-202	1	1	127-217	1	1	66-40	2	2
66-308	3	2	722-217	6	6	23-122	1	1
66-14	3	2	244-243	1	1	23-112	1	1
66-19	1	1	127-66	2	2	99-99	1	1
582-306	1	1	720-69	1	1	66-547	1	1
309-314	1	1	722-319	1	1	66-127	1	1
190-298	1	1	51-251	1	1	66-720	1	1
202-112	1	1	720-217	1	1	217-720	1	1
66-57	1	1	720-319	1	1	66-35	1	1
217-127	1	0	40-273	1	1	66-722	1	1
66-684	1	1	40-582	1	1	558-112	1	1
14-217	1	0	200-51	1	1	558-69	2	0
23-582	2	1	40-66	4	1	558-388	1	1
251-582	1	1	40-51	1	1	12-112	1	1
14-69	1	0	51-243	1	1	1-112	1	1

Figure 12: Sensitivity of estimation framework to the number of scenarios

6 Conclusion and Future Research Directions

This paper presents a framework to acquire the knowledge about dissagregate traveler's OD pairs and their preferences by observing traveler responses to the pricing incentives. To this end, traveler type distribution is assumed to capture the heterogeneity of the travelers. The presented Selective Set Expectation Maximization (SEEM) algorithm gradually updates the beliefs about the travelers' OD location distribution; a system manager can exploit these results to introduce a new pricing scenario, with each such added scenario increasing the accuracy of OD pair inference. With the travel patterns changing over short time periods, the manager can react to these changes as they occur, offering pop-up incentives in addition to regular ones. With the flexibility in operations, bike sharing systems are an ideal test bed for the developed method, which can supply OD pair inference algorithms with rich data, resulting from perturbing price settings over various bike segments. The SSEM algorithm can also be used with real time transit data to learn travel population's taste variation and also for travel OD-estimation in real time. The SSEM algorithm can also be boosted to increase the accuracy of inference results for data sets with multiple observations by screening out infeasible choices in cases of similar estimation problems.

The methodology adopted in this paper can be extended as a future course of research with other modes of transportation, such as perturbations can arise as a result of road lane closures, as well as public events or extreme weather conditions. Further research interests can be invoked in the areas of eligible OD set generation and parameter estimation for pricing. Travelers value pricing differently similar to the way they value in-vehicle travel time or waiting time differently. The dissagregate travel OD estimation model developed in this paper can provide useful information for system operations, planning, incentives/toll policies, management, etc. In bike-sharing system, such an integrated solution for the shared vehicle re-balancing problem could alternate between optimizing pricing incentives based on the currently estimated parameters for the trip demand and traveler preferences, and updating those parameter values based on the additional data gathered after the system adjusts itself to a new pricing policy. This process can be thought of as a repeated probing of the system to better estimate its parameter values, which can then be effectively optimized.

7 Appendix

This section provides a parallel implementation for the Selective Set Expectation Maximization algorithm discussed in section 3. Multiprocessing or parallel processing on multiple cores is used to achieve higher scalability for the Expectation Step (E-Step) of the SSEM algorithm. A mapper and reducer functions are used to create a list of values with the traveler's id as the key for the list. For each value in the list (which has a unique traveler id); the $\langle key - value \rangle$ pair is sent to the parallelized functions E STEP() and LIKELIHOOD() to perform the expectation step of SSEM and compute the lower bound for maximization. As discussed in Section 3.2, the time complexity of E-Step is quartic in nature and a faster execution can be achieved using multiple cores on a high performance computing platform. For even higher scalability, this parallelized version of SSEM can be implemented on a distributed framework with multiple nodes each having multiple cores for processing.

Algorithm 2 Parallel Implementation of SSEM algorithm for MLE estimator

function MAP(t) ▷ Mapper for creating Traveler-Scenario Keys
    for obs $\in$ **obs do**
        $K \leftarrow t$
        $V \leftarrow obs$
        $emit(K, V)$
function REDUCE($K, list[V]$) ▷ Reducer to Create Traveler-Scenario Pairs
    for $i \in 1...|list[V]|$ **do**
        $K \leftarrow t$
        $V \leftarrow list[i]$
        $emit(K, V)$

function E STEP($< t \;\; \text{obs} >$, μ , $\Theta^{(n)}$) ▷ The expectation step
    Inputs:
    $\mu \leftarrow \mu \;\; \forall \mu \in \boldsymbol{M}$
    $\Theta^{(n)} \leftarrow P_{\Theta}(\mu) \;\; \forall \mu \in \boldsymbol{M}$
    Initialize:
    $P(w) \leftarrow 1/|\boldsymbol{\Omega}_t| \;\; \forall \; w \in \boldsymbol{\Omega}_t$
    $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow 0 \;\; \forall \; w \in \boldsymbol{\Omega}_t, \; \forall \mu \in \boldsymbol{M}$
    $B_t^{\text{obs}}(w) \leftarrow 0 \;\; \forall \; w \in \boldsymbol{\Omega}_t$
    for $w \in \boldsymbol{\Omega}_t$ **do**
        $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow \frac{\varpi_r * exp(-\mu \; * \; d_r(p))}{|\boldsymbol{\Omega}|} \; \forall \mu \in \boldsymbol{M}, \; \forall r \in \boldsymbol{R}(w)$
    $Z_\mu \leftarrow \sum_{r \in \boldsymbol{R}(w)} P_{\text{logit}}(w|\mu)_t^{\text{obs}} \; \forall \mu \in \boldsymbol{M}$
    for $r \in \boldsymbol{R}(w)$ **do**
        $P_{\text{logit}}(w|\mu)_t^{\text{obs}} \leftarrow \frac{P_{\text{logit}}(w|\mu)_t^{\text{obs}}}{Z_\mu}$
    for $w \in \boldsymbol{\Omega}_t$ **do**
        $B_t^{\text{obs}}(w) \leftarrow \; P(w) * \sum_{\mu \in \boldsymbol{M}} P_{\text{logit}}(w|\mu)_t^{\text{obs}}$
    $Z' \leftarrow \sum_{w \in \boldsymbol{\Omega}_t} B_t^{\text{obs}}(w)$

for $w \in \boldsymbol{\Omega}_t$ **do**
$\quad B_t^{\text{obs}}(w) \leftarrow \frac{B_t^{\text{obs}}(w)}{Z'}$
return $B_t^{\text{obs}}(w) \quad \forall w \in \boldsymbol{\Omega}_t$

function LIKELIHOOD($< t \;\; \text{obs} >$, $\boldsymbol{M}_t$) $\triangleright$ Objective Function Terms
$\quad$ **Initialize:**
$\quad L_t^{\text{obs}}(w) \leftarrow \log[\sum_{\mu \in \boldsymbol{M}_t} P((i,j)_t)^{\text{obs}} | w, \mu) P_\Theta(\mu) P(w)]$
$\quad$ **return** $L_t^{\text{obs}}(w)$

function M STEP(B_t^{obs}) $\triangleright$ The maximization step
$\quad$ **Initialize:**
$\quad Q(\Theta, \Theta^{(n)}) \leftarrow 0$
$\quad B_{\text{obs}} \leftarrow \big[[..., B_1^{\text{obs}}[w], ...], [..., B_1^{\text{obs}}[w], ...], ..., [..., B_t^{\text{obs}}[w], ...]\big] \;\forall \text{obs} \in \textbf{obs}$
$\quad L_{\text{obs}} \leftarrow \big[[..., L_1^{\text{obs}}[w], ...], [..., L_1^{\text{obs}}[w], ...], ..., [..., L_t^{\text{obs}}[w], ...]\big] \;\forall \text{obs} \in \textbf{obs}$
$\quad Q(\Theta, \Theta^{(n)}) \leftarrow \sum_{\text{obs} \in \textbf{obs}} (B_{\text{obs}})^T \times L_{\text{obs}}$
$\quad \Theta^{(n+1)} \leftarrow \arg\max_{\Theta} Q(\Theta, \Theta^{(n)})$
$\quad$ **return** $\Theta^{(n+1)}$

procedure SSEM( ) $\triangleright$ Main function implementing the SSEM algorithm
$\quad$ **Inputs:**
$\quad \boldsymbol{M}_t \quad \forall t \in \boldsymbol{T}$
$\quad$ **Initialize:**
$\quad n \leftarrow 0$
$\quad \Theta^0 \leftarrow rand\big(P_\Theta(\mu)\big) \;\forall \mu \in \boldsymbol{M}$
$\quad TS \leftarrow list[\;\;]$
$\quad$ **for all** processors **do in parallel**
$\quad\quad$ MAP(t) $\quad t \in \boldsymbol{T}$
$\quad\quad TS \leftarrow$ REDUCE()
$\quad$ **repeat**
$\quad\quad$ **for all** processors **do in parallel**
$\quad\quad\quad$ E STEP($< t \;\; \text{obs} >$) $\;\forall\; < t \;\; \text{obs} > \;\in\; TS[< t \;\; \text{obs} >]$
$\quad\quad\quad$ LIKELIHOOD($< t \;\; \text{obs} >$) $\;\forall\; < t \;\; \text{obs} > \;\in\; TS[< t \;\; \text{obs} >]$
$\quad\quad$ M STEP( )
$\quad\quad t \leftarrow n + 1$
$\quad$ **until** Convergence

References

Abrahamsson, T. (1998). Estimation of origin-destination matrices using traffic counts–a literature survey. *IIASA Interim Report IR-98-021/May 27*, 76.

Bell, M. G. (1991). The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological 25*(1), 13–22.

Ben-Akiva, M. and M. Bierlaire (2003). Discrete choice models with applications to departure time and route choice. In *Handbook of transportation science*, pp. 7–37. Springer.

Ben-Akiva, M. E. and S. R. Lerman (1985). *Discrete choice analysis: theory and application to predict travel demand*. The MIT press.

Borgnat, P., P. Abry, P. Flandrin, and J. Rouquier (2009). Studying lyon's vélo'v: A statistical cyclic model. In *European Conference on Complex Systems 2009*.

Bovy, P. H. and S. Fiorenzo-Catalano (2007). Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica 3*(3), 173–189.

Bovy, P. H. L. and S. Hoogendoorn-Lanser (2005). Modelling route choice behavior in multi-modal transport networks. *Transportation 32*, 341–368.

Cascetta, E. and S. Nguyen (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological 22*(6), 437–455.

Castillo, E., J. M. Menéndez, and S. Sánchez-Cambronero (2008). Predicting traffic flow using bayesian networks. *Transportation Research Part B: Methodological 42*(5), 482–509.

Chen, A., S. Ryu, and P. Chootinan (2010). l_∞-norm path flow estimator for handling traffic count inconsistencies: Formulation and solution algorithm. *Journal of transportation engineering 136*(6), 565–575.

de la Barra, T., B. Perez, and J. Anez (1993). Multidimensional path search and assignment. In *PTRC Summer Annual Meeting, 21st, 1993, University of Manchester, United Kingdom*.

de Moraes Ramos, G., W. Daamen, and S. Hoogendoorn (2011). Expected utility theory, prospect theory, and regret theory compared for prediction of route choice behavior. *Transportation Research Record: Journal of the Transportation Research Board 2230.1*, 19–28.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Frejinger, E. and M. Bierlaire (2007). Random sampling of alternatives for route choice modeling. In *Swiss Transport Research Conference*, Number TRANSP-OR-TALK-2007-013.

Friedrich, M., I. Hofsäß, and S. Wekeck (2001). Timetable-based transit assignment using branch and bound techniques. *Transportation Research Record: Journal of the Transportation Research Board 1752*(1), 100–107.

Froehlich, J., J. Neumann, and N. Oliver (2009). Sensing and predicting the pulse of the city through shared bicycling. In *Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 1420–1426.

Garcia-Rodenas, R. and D. Verastegui-Rayo (2008). A column generation algorithm for the estimation of origin–destination matrices in congested traffic networks. *European Journal of Operational Research 184*(3), 860–878.

Haider, Z., A. Nikolaev, J. E. Kang, and C. Kwon (2014). Inventory rebalancing through pricing in public bike sharing systems. Working Paper 564, Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY, USA.

Hazelton, M. L. (2008). Statistical inference for time varying origin–destination matrices. *Transportation Research Part B: Methodological 42*(6), 542–552.

Hensher, D. A. and W. H. Greene (2003). The mixed logit model: the state of practice. *Transportation 30*(2), 133–176.

J.A., A., M. S. Costa, J. S. Madeira, and E. V. Martins (1993). An algorithm for the ranking of shortest paths. *Hxurshdq Mrxuqdo ri Rshudwlrqdo Uhvhdufk 69*, 97–106.

Lam, W. H. K., Z. X. Wu, and K. S. Chan (2003). Estimation of transit origin-destination matrices from passenger counts using a frequency-based approach. *Journal of Mathematical Modelling and Algorithms 2*, 329–348.

Lanser, S. (2005). *Modelling travel behaviour in multi-modal networks.* TU Delft, Delft University of Technology.

Lee, S. G. and M. Hickman (2014). Trip purpose inference using automated fare collection data. *Public Transport 6*(1-2), 1–20.

Li, D., T. Miwa, and T. Morikawa (2013). Dynamic route choice behavior analysis considering en route learning and choices. *Transportation Research Record: Journal of the Transportation Research Board 2383.1*, 1–9.

Li, Y. and M. J. Cassidy (2007). A generalized and efficient algorithm for estimating transit route ods from passenger counts. *Transportation Research Part B: Methodological 41*, 114–125.

Lundgren, J. T. and A. Peterson (2008). A heuristic for the bilevel origin–destination-matrix estimation problem. *Transportation Research Part B: Methodological 42*(4), 339–354.

Maher, M. (1983). Inferences on trip matrices from observations on link volumes: a bayesian statistical approach. *Transportation Research Part B: Methodological 17*(6), 435–447.

Mahmassani, H. S. and K. C. Sinha (1981). Baysian updating of trip generation parameters. *Journal of Transportation Engineering 107*(5).

Minka, T. (1998). Expectation-maximization as lower bound maximization. *Tutorial published on the web at http://www-white. media. mit. edu/tpminka/papers/em. html.*

Munizaga, M. A. and C. Palma (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies 24*, 9–18.

Nie, Y., H. Zhang, and W. Recker (2005). Inferring origin–destination trip matrices with a decoupled gls path flow estimator. *Transportation Research Part B: Methodological 39*(6), 497–518.

Prato, C. G. and S. Bekhor (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board 1985*(1), 19–28.

Ryu, S., A. Chen, H. M. Zhang, and W. Recker (2014). Path flow estimator for planning applications in small communities. *Transportation Research Part A: Policy and Practice 69*, 212–242.

Sherali, H. D. and T. Park (2001). Estimation of dynamic origin–destination trip tables for a general network. *Transportation Research Part B: Methodological 35*(3), 217–235.

Sun, L., Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen (2015). An integrated bayesian approach for passenger flow assignment in metro networks. *Transportation Research Part C: Emerging Technologies 52*, 116–131.

Tebaldi, C. and M. West (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association 93*(442), 557–573.

Van der Zijpp, N. and S. F. Catalano (2005). Path enumeration by finding the constrained k-shortest paths. *Transportation Research Part B:Methodological 39*(6), 545–563.

van Zuylen, H. (1978). The information minimization method: validity and applicability in transporation planning. In *Colloquium Vervoers-planologisch Speurwerk*.

Vogel, P. and D. Mattfeld (2010). Modeling of repositioning activities in bike-sharing systems. In *World Conference on Transport Research (WCTR)*.

Wardman, M. (2001). Public transport values of time. Working Paper 564, Institute for Transport Studies,University of Leeds.

Wardman, M. (2004). Public transport values of time. *Transport Policy 11*(4), 363–377.

Wong, S. C. and C. O. Tong (1998). Estimation of time-dependent origin–destination matrices for transit networks. *Transportation Research Part B: Methodological 32*, 35–48.

Xie, C., K. Kockelman, and S. Waller (2010). Maximum entropy method for subnetwork origin-destination trip matrix estimation. *Transportation Research Record: Journal of the Transportation Research Board* (2196), 111–119.

Xie, C., K. M. Kockelman, and S. T. Waller (2011). A maximum entropy-least squares estimator for elastic origin–destination trip matrix estimation. *Transportation Research Part B: Methodological 45*(9), 1465–1482.

Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research Part B: Methodological 29*(4), 231–242.