



Transportation Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Individual Path Recommendation Under Public Transit Service Disruptions Considering Behavior Uncertainty

Baichuan Mo, Haris N. Koutsopoulos, Zuo-Jun Max Shen, Jinhua Zhao

To cite this article:

Baichuan Mo, Haris N. Koutsopoulos, Zuo-Jun Max Shen, Jinhua Zhao (2025) Individual Path Recommendation Under Public Transit Service Disruptions Considering Behavior Uncertainty. Transportation Science

Published online in Articles in Advance 24 Oct 2025

. <https://doi.org/10.1287/trsc.2025.0042>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2025, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Individual Path Recommendation Under Public Transit Service Disruptions Considering Behavior Uncertainty

Baichuan Mo,^{a,b,*} Haris N. Koutsopoulos,^c Zuo-Jun Max Shen,^d Jinhua Zhao^e

^aDepartment of Civil Engineering, Tsinghua University, Beijing 100084, China; ^bDepartment of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ^cDepartment of Civil and Environmental Engineering, Northeastern University, Boston, Massachusetts 02115; ^dDepartment of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720; ^eDepartment of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

*Corresponding author

Contact: baichuan@mit.edu,  <https://orcid.org/0000-0003-0323-0066> (BM); h.koutsopoulos@northeastern.edu (HNK); maxshen@berkeley.edu (Z-JMS); jinhua@mit.edu,  <https://orcid.org/0000-0002-1929-7583> (JZ)

Received: May 27, 2023

Revised: February 1, 2025; June 21, 2025; September 2, 2025

Accepted: September 14, 2025

Published Online in Articles in Advance: October 24, 2025

Area of Review: Emerging and Cross-functional Topics

<https://doi.org/10.1287/trsc.2025.0042>

Copyright: © 2025 INFORMS

Abstract. Public transit passengers need guidance during service disruptions. This study proposes an individual-based path recommendation (IPR) model. The model decides which paths to recommend for each passenger with the objective of minimizing system travel time and respecting passengers' path choice preferences. We assume the recommendations could affect passengers' path choice probabilities, but their actual choices are uncertain. This behavior uncertainty makes the problem a stochastic optimization with decision-dependent distributions. We propose a single-point approximation method to eliminate the expectation operator by introducing two new concepts: ϵ -feasibility and Γ -concentration, which control the mean and variance of path flows in the optimization problem. The approximation yields a tractable single-stage mixed integer linear formulation, which can be solved efficiently with Benders decomposition. The approximation gap is proved to be bounded from above. Additional theoretical analysis shows that ϵ -feasibility and Γ -concentration are strongly connected to expectation and chance constraints in a typical stochastic optimization formulation, respectively. The model is implemented in a real-world case study using data from an urban rail disruption in the Chicago Transit Authority system and a synthetic case study with varied network sizes and incident locations. In the real-world case study, results show that the proposed IPR model reduces the average travel times in the system by 6.6% compared with the status quo and by 4.2% compared with a capacity-based benchmark model. In the synthetic case study, the proposed model shows 15.0%–1.8% lower system travel time compared with the capacity-based benchmark, depending on the network sizes and demand situations.

Funding: This work was supported by the Chicago Transit Authority.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/trsc.2025.0042>.

Keywords: individualized recommendation • behavior uncertainty • stochastic optimization

1. Introduction

With aging systems and near-capacity operations, service disruptions often occur in urban public transit (PT) systems. For example, the Chicago Transit Authority (CTA) reported an average of 75.4 abnormal events per day in 2019, with at least 1.1 events each day resulting in delays of over 20 minutes (Mo 2022). These incidents may result in passenger delays, cancellation of trips, and economic losses (Cox, Prager, and Rose 2011). During a significant disruption where the service is interrupted for a relatively long period of time (e.g., one hour), affected passengers usually need to find an alternative path or use other travel modes (such as transfer to another bus route). However, because of a lack of knowledge of the system (especially during incidents), the routes chosen by passengers may not be optimal or

even cause more congestion (Mo et al. 2022). For example, during a rail disruption, most of the passengers may choose bus routes that are parallel to the interrupted rail line as an alternative. However, given limited bus capacity, parallel bus lines may become oversaturated and passengers may have to wait for a long time to board because of being denied boarding (or left behind).

One of the strategies to better guide passengers is to provide path recommendations so that passenger flows are redistributed in a better way and the system travel times (STTs) are minimized. This can be seen as solving an optimal passenger flow distribution (or assignment) problem over a public transit network. However, different from the typical flow redistribution problem, there are several unique characteristics and challenges for the

path recommendation problem under PT service disruptions.

- Passengers may have different preferences on different alternative paths. This heterogeneity suggests that we cannot treat a group of passengers simply as flows. Individualization is needed in the path recommendation design.

- Passengers may not follow the recommendation. When a specific path recommendation is provided to a passenger, their actual path choices are uncertain (though the recommendation may change their preferences). This behavior uncertainty (BU) brings challenges to the recommendation system design and has not been considered in the path recommendation literature. In the context of individualization, the behavior uncertainty is also individual-specific, which requires a more granular modeling approach.

To tackle these challenges, this study proposes an individual-based path recommendation (IPR) model to reduce system congestion during public transit disruptions considering passenger behavior uncertainty. The decision variable in this study is $x_{p,r}$, a binary variable indicating whether path r is recommended to passenger p or not. However, passengers may not follow the recommendation. Their behavior uncertainty is modeled using a conditional path choice probability distribution given their received path recommendation. The original IPR formulation yields a stochastic optimization with decision-dependent distributions (Goel and Grossmann 2006, Drusvyatskiy and Xiao 2023). We propose a single-point approximation method to eliminate the expectation operator by introducing two new concepts: ϵ -feasibility and Γ -concentration. The former describes the relationship between path flows and recommendations from the “expectation” perspective, whereas the latter constrains the variance of the path flows under recommendations. The approximation yields a tractable single-stage mixed integer linear formulation, which can be solved efficiently with Benders decomposition (BD) in large-scale scenarios. The approximation gap is proved to be bounded from above. Additional theoretical analysis shows that ϵ -feasibility and Γ -concentration are strongly connected to expectation and chance constraints in a typical stochastic optimization formulation, respectively. The proposed framework is implemented in a real-world case study using data from an urban rail disruption in the CTA system and a synthetic case study with varied network sizes and incident locations.

The main contributions of the paper are twofold:

- The paper proposes a framework with prior path utility and posterior path choice distribution given recommendations to model behavior uncertainty, yielding a stochastic optimization problem with decision-dependent distributions. We propose a single-point approximation method to eliminate the expectation

operator by introducing two new concepts: ϵ -feasibility and Γ -concentration, which control the mean and variance of path flows in the optimization problem.

- The proposed single-point approximation yields a tractable single-stage mixed integer linear formulation, which can be solved efficiently with Benders decomposition. We prove that, with ϵ -feasibility and Γ -concentration, the gap of the proposed approximation is bounded from above. Additional theoretical analysis shows that ϵ -feasibility and Γ -concentration are strongly connected to expectation and chance constraints in a typical stochastic optimization formulation, respectively.

The remainder of the paper is organized as follows. A literature review is discussed in Section 2. In Section 3, we describe the problem conceptually and analytically. Section 4 develops the formulation for the individual path recommendation problem and the modeling of the behavior uncertainty. Section 5 provides all theoretical analysis of the proposed method. In Section 6, we apply the proposed model on the CTA system as a case study and analyze the results. Finally, we conclude the paper and summarize the main findings in Section 8.

2. Literature Review

In this section, we go through previous studies on path recommendations and behavior uncertainties. Specifically, we show how our paper is connected with stochastic optimization with decision-dependent distributions (also known as performative predictions).

2.1. Personalized Recommendations in the Transportation Field

Personalized recommendation systems in transportation have garnered significant attention in recent years, aiming to enhance user experience by tailoring travel options to individual preferences (Thiengburanatham, Cang, and Yu 2016; Quijano-Sánchez et al. 2020). For example, Zhu et al. (2020) explored the use of personalized incentives to promote sustainable travel behaviors, demonstrating that targeted rewards can effectively encourage eco-friendly transportation choices. Song et al. (2018) introduced a personalized menu optimization approach with a preference updater, applying it to a Boston case study to improve user satisfaction in transportation services. Wu, Lyu, and Liu (2022) proposed a conceptual framework for personalized travel mode recommendation in a multimodal system. Similarly, Liu et al. (2019) developed a context-aware, multimodal transportation recommendation system that integrates various travel modes, considering user context to provide optimal travel suggestions. Danaf et al. (2019) explored online discrete choice models, applying

them to personalized transportation recommendations to predict user preferences more accurately. Additionally, Lim et al. (2021) presented an origin-aware next-destination recommendation model utilizing personalized preference attention, enhancing the accuracy of destination predictions based on user history. Collectively, these studies underscore the importance of personalization in transportation systems, leveraging user data and advanced modeling techniques to offer tailored travel recommendations.

2.2. Path Recommendations During Disruptions

Path recommendation is a popular topic in the transportation field (Zhang et al. 2024). The main goal is to provide a suitable route for a user based on their travel constraints and preferences, with applications in navigation, deliveries (Liu et al. 2020), travel planning, etc. Methods for path recommendation include search-based (Stentz 1994, Geisberger et al. 2008, Dellinger et al. 2011, Wang et al. 2014, Dai et al. 2015), probability-based (Karaman and Frazzoli 2011; Chen, Ong, and Xie 2016; Qu et al. 2019), constraint-based (Qu et al. 2014; Reza, Ali, and Cheema 2017; Lai et al. 2018; Cheng, Zhou, and Lai 2019), deep learning-based (Huang, Xu, and Weng 2020; Wang, Wu, and Zhao 2021; Bhumika and Das 2022; Wang et al. 2023), and reinforcement learning-based (Ji et al. 2020, Chen et al. 2022, Xia et al. 2022) methods.

Most previous studies on path recommendation under incidents are like designing a “trip planner.” That is, the main objective is to find available routes or the shortest path given an origin-destination (OD) pair when the network is interrupted by incidents. For example, Bruglieri et al. (2015) designed a trip planner to find the fastest path in the public transit network during service disruptions based on real-time mobility information. Böhmová et al. (2013) developed a routing algorithm in urban public transportation to find reliable journeys that are robust for system delays. Roelofsen et al. (2018) provided a framework for generating and assessing alternative routes in case of disruptions in urban public transport systems. Fang et al. (2024) use large language models to recommend alternative routes in public transit during network disruptions. To the best of the authors’ knowledge, most of the previous studies did not consider minimizing STT as the goal for path recommendations. The only exceptions are Mo et al. (2023a) and Dai, Liu, and Xu (2024), where they aim to provide station-based path recommendations to guide passengers in urban rail systems in order to minimize system travel time. Compared with typical path recommendation studies where the routes are selected independently for each person, the recommendation with system travel time minimization needs to codesign the strategy of all individuals simultaneously, which is more challenging.

Providing path recommendations during disruptions is similar to the topic of passenger evacuation under emergencies. The objective of evacuation is usually to minimize the total evacuation time, where the decisions need to consider interaction among individual choices. For example, Abdelgawad and Abdulhai (2012) developed an evacuation model with routing and scheduling of subway and bus transit to alleviate congestion during the evacuation of busy urban areas. Wang, Yuan, and Yin (2019) proposed an optimal bus bridging design method under operational disruptions on a single metro line. Tan et al. (2020) propose an evacuation model with urban bus networks as alternatives in the case of common metro service disruptions by jointly designing the bus lines and frequencies. However, although these passenger evacuation papers focus on minimizing the system travel time, there are several differences from this paper. First, in our paper, the service disruption is not as severe as the emergency situation. We assume the service will recover after a period of time and passengers are allowed to wait. They do not necessarily need to cancel trips or follow the evacuation plan, as assumed in previous evacuation studies. Second, in this article, we assume the supply-side adjustments (e.g., bus rerouting) are given and are not considered decision variables in the optimization. We focus on providing information to the passengers to better utilize the existing resources/capacities of the system, whereas many evacuation studies focus on supply-side adjustment (e.g., rerouting and rescheduling). Third, this paper considers passenger heterogeneity and focuses on individual-level path recommendations, whereas previous evacuation papers simply model passengers as flows. Besides, we also assume that passengers may not follow the recommendation (i.e., behavior uncertainty), which has not been considered in any evacuation paper before.

2.3. Behavior Uncertainty and Compliance

Behavior uncertainty is a well-known challenge in transportation modeling (Mahmassani 1984). It is one of the major reasons for traffic flow instability and the difficulties in predicting traffic conditions (Han et al. 2024). In the context of path recommendation, behavior uncertainties cause noncompliance (i.e., passengers may not follow the recommendations), which brings more challenges in evaluating system travel time and other objective functions that depend on passenger flows. For example, Wong et al. (2023), using survey data, revealed there is an evacuation-reluctant group of people who are not willing to follow mandatory evacuation orders in California wildfires. Wang and Wallace (2022) considered passengers’ noncompliance behavior in a transit-based evacuation pick-up point assignment problem.

Typically, passenger behavior is modeled using various econometrics approaches (Ben-Akiva and Lerman 1985, Train 2009, Mo et al. 2021) or machine learning models (Mirchevska 2013; Wang, Mo, and Zhao 2020). These models output the probability distribution for the passenger’s possible behavior. At the aggregate level, numerous studies have taken demand uncertainty into consideration for decision making. The applications include ride-sharing (Guo, Caros, and Zhao 2021), bus scheduling (Guo et al. 2024), transit route planning (Yoon and Chow 2020), and supply chain management (Jung et al. 2004).

However, at the individual level, the number of studies is limited. The main reason is that individual-level decision making is usually discrete. So it is challenging to use typical robust optimization to address discrete uncertain variables (Subramanyam et al. 2021). In terms of stochastic optimization, the number of possible scenarios increases exponentially with the number of individuals in the system. Typical ways to address a large number of scenario combinations include (1) Monte Carlo simulation (Niederreiter and Winterhof 1992, Homem-de Mello and Bayraksan 2014, Bartl and Mendelson 2022), (2) scenario reduction by clustering (Hu and Li 2019; Bertsimas and Mundru 2023; Keutchayan, Ortmann, and Rei 2023), and (3) better scenario generation methods, such as optimal quantization (Bardou, Bouthemy, and Pagès 2009) and moment matching (Høyland, Kaut, and Wallace 2003; Mehrotra and Papp 2013). Though these methods help to deal with a large number of combinations of scenarios, most of them only apply to “exogenous” uncertainties (i.e., the distribution of the uncertain parameters is given). In this study, we assume an individual’s path choice probabilities are affected by our path recommendation decisions (e.g., they are more likely to choose paths that we recommend). This is known as stochastic optimization with decision-dependent distributions (Goel and Grossmann 2006; Hellemo, Barton, and Tomasgard 2018; Drusvyatskiy and Xiao 2023), or performative predictions in the machine learning field (Mendler-Dünner et al. 2020, Perdomo et al. 2020). Specifically, these problems can be formulated as

$$\min_{x \in \mathcal{X}} \mathbb{E}_{\zeta \sim \mathbb{P}(x)} [f(x, \zeta)], \quad (1)$$

where ζ is the random parameter with distribution \mathbb{P} depending on x and $f(x, \zeta)$ is the objective function. Typical methods for solving these problems include (1) performative stochastic gradient descent (for continuous problems) and (2) applying sample average approximation (SAA) in branch-and-cut (for integer problems). However, $f(x, \zeta)$ in the IPR problem represents the system travel time, which usually has no analytical formulations because of the hard capacity constraints in transit network loading processes

(Mo et al. 2023a, b). Therefore, the traditional ways are hard to adopt.

3. Problem Description

In this section, we describe how the individual path recommendation problem is defined, including network and terminology, decision variables and objectives, as well as the scope and time point for decision making.

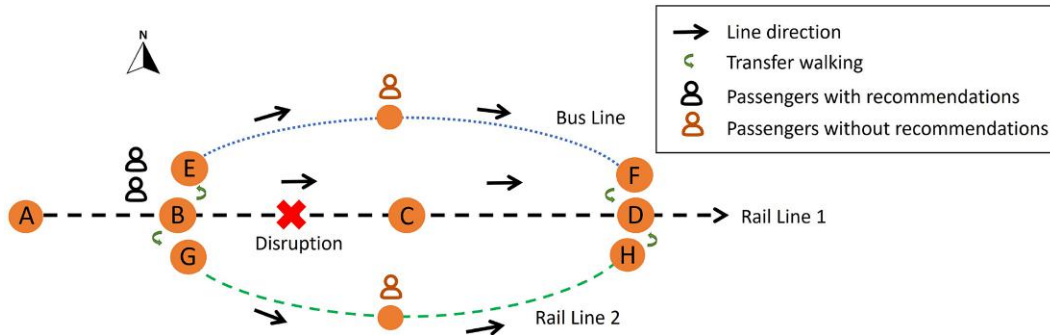
3.1. Conceptual Description

Consider a public transit system with both rail (or subway) and bus services. A public transit network consists of stations (bus or rail stations) and lines (bus or rail lines). A line is defined based on how vehicles (bus or train) are operated. A station is where passengers can board or alight. It is associated with three attributes: station identifier (ID), line ID, and line direction. Technically, a station in this study represents a boarding/alighting platform. Passengers can transfer from a station on line 1 to another station on line 2. We call the first station of a line where the vehicle departs from the garage the “terminal” station. We call a station with multiple lines crossing it a “transfer” station. Figure 1 shows an example of the transit network. In this example, station A is associated with rail line 1 in the east direction. It is also a terminal station. Station B is a transfer station, which can be connected to station E on the bus line or station G on rail line 2.

Consider a service disruption in a public transit system. During the disruption, some stations in the incident line (or the whole line) are blocked. Passengers in the blocked vehicles are usually offloaded to the nearest platforms. To respond to the incident, some operating changes are made, such as dispatching shuttle buses, rerouting existing services, short-turning in the incident line, headway adjustment, etc. Assume that all information about the operating changes is available. These changes define a new PT service network and available path sets. Our objective is to develop an individual-based path recommendation model that, when an incident happens, provides a recommended path to every passenger who uses their phones, websites, or electronic boards at stations to enter their origin, destination, and departure time. The recommendation considers the individual’s preferences and behavioral histories. Hence, passengers with the same origin, destination, and departure time may get different recommended paths. The overall system aims to minimize the total travel time for all passengers, including passengers on nearby rail or bus lines without incidents (note that these passengers may experience additional crowding because of transfer passengers from the incident line).

In this example of Figure 1, rail line 1 has an incident between stations B and C and cannot provide service for a period of time. Both of the two passengers at station B want to go to station D. Assuming that they

Figure 1. (Color online) Example of the Individual Path Recommendation Problem



request path recommendations, the alternative paths include using the bus line (blue dashed line), using rail line 2 (green dashed line), or waiting for the system to recover (i.e., still using rail line 1). Note that using either the bus line or rail line 2 will take away capacity from passengers who originally used these two services (i.e., the orange passengers in the figure). Hence, the model should consider the total travel time of all four passengers in the system to design recommendation strategies. Moreover, as mentioned in the introduction, behavior uncertainty needs to be considered. In this example, if we recommend that a passenger use a bus line, the person may not follow the recommendation and may choose rail line 2 instead.

3.2. Analytical Description

Let us divide the analysis period into several time intervals with equal-length τ (e.g., $\tau = 5$ minutes). Let t be the integer time index. $t = 1$ is the start of the incident, and $t \leq 0$ indicates the time before the incident. Let \mathcal{P} be the set of passengers who will receive path recommendations. We assume \mathcal{P} is known as we can obtain passengers' requests before running the model. Given the revised operation during the incident, let \mathcal{R}_p be the feasible path set for each passenger $p \in \mathcal{P}$. Note that \mathcal{R}_p includes all feasible services that are provided by the PT operator. A path $r \in \mathcal{R}_p$ may be waiting for the system to recover (i.e., using the incident line) or transferring to nearby bus lines, using shuttle services, etc. We do not consider non-PT modes such as transportation network companies (TNC) or driving for the following reasons: (1) This study aims to design a path recommendation system used by PT operators. The major audience should be all PT users. Consideration of non-PT modes needs the supply information of all other travel modes, and involves non-PT users (such as the impact of traffic congestion on drivers), which is beyond the scope of this study. Future research may consider a multimodal path recommendation system. (2) Passengers using non-PT modes can be simply treated as demand reduction for the PT system. So, their impact on the PT system can still be captured.

Given a passenger $p \in \mathcal{P}$, we aim to determine $x_{p,r}$ for each p , where $x_{p,r}$ indicates whether path $r \in \mathcal{R}_p$ is recommended to passenger p or not. Assuming only one path is recommended to each passenger, we have

$$\sum_{r \in \mathcal{R}_p} x_{p,r} = 1 \quad \forall p \in \mathcal{P}. \quad (2)$$

Note that we can relax this assumption by designing the recommendation as a "composition" including multiple paths or travel times. This generalization is discussed in Online Appendix A.1.

The set \mathcal{P} includes passengers with different origins, destinations, and departure times. If an incident ends at t^{end} , the recommendation should consider a time horizon after t^{end} because there is remaining congestion in the system. Hence, we provide recommendations until time $T^D > t^{\text{end}}$ (e.g., T^D can be one hour after t^{end}). Therefore, the departure times for passenger $p \in \mathcal{P}$ range from $[1, T^D]$ (T^D and t^{end} are both time indices).

The recommendation model will be solved at $t = 1$ and will generate the recommendation strategies $x = (x_{p,r})_{p \in \mathcal{P}, r \in \mathcal{R}_p}$ for passengers who depart at time $t \in [1, T^D]$. In reality, the model can be implemented in a rolling horizon manner. Specifically, at each time interval $t \geq 1$, we first update the demand and supply information in the system, including new demand estimates, new recommendation-receiving passenger set \mathcal{P} , newly available path sets \mathcal{R}_p , new service routes and frequencies, new incident duration estimates, new onboard passenger estimates, etc. Based on this information, we solve the model to obtain recommendations for passengers with departure time in $[t, T^D]$. However, we only implement the recommendation strategies for passengers who depart at the current time t . More discussions can be found in Online Appendix A.2.

Therefore, in the following formulation, we only focus on solving the model at $t = 1$, which is the start of the incident. The whole analysis period includes warm-up and cool-down periods to better estimate the system states (e.g., vehicle loads, passenger travel times, etc.). Therefore, the analysis period is defined as $[t^{\text{min}}, T]$,

where $t^{\min} < 1$ (time before the incident) and $T > T^D$. For example, t^{\min} and T can be one hour before and after $t = 1$ and T^D , respectively, and we define all time intervals in the analysis period as $\mathcal{T} = \{t^{\min}, t^{\min} + 1, \dots, T\}$. The overall path recommendation framework can be summarized in Figure 2.

It is worth noting that the supply information is assumed to be known in this study. An ideal response strategy to disruptions should decide supply (e.g. rescheduling) and demand (e.g., path recommendation) strategies simultaneously. However, considering both in one decision-making problem is beyond the scope of this study and deserves a separate study. We admit this is a limitation of the paper.

In reality, only considering path recommendations also has practical implications. For example, some PT operators may manually decide the supply-side changes (such as dispatching shuttle buses and rerouting) by themselves, as they believe human decisions can incorporate better domain knowledge and have better interpretability. Our model could take their supply-side decision as input and only solve the path recommendation problem. Moreover, in the context of algorithm-driven supply change decisions, our model can be a follow-up subproblem to decide the demand-side strategies, thus making the supply-side and demand-supply strategy design two subproblems. Note that because the model can also be implemented with the rolling horizon method (Online Appendix A.2), the supply-side information may vary over time according to the newest updates from the operators.

4. Formulations

In this section, we elaborate on the detailed formulation of the individual path recommendation model. Section

4.1 describes how passengers' behavior uncertainties (i.e., noncompliance with recommendations) are modeled based on a random utility maximization framework. Section 4.2 provides the overall formulation of the individual path recommendation model as a stochastic optimization with decision-dependent distributions. Section 4.3 presents a linear programming (LP) model to calculate the system travel time. Section 4.4 elaborates on the proposed single-point approximation with ϵ -feasibility and Γ -concentration. Section 4.5 shows how the individual path recommendation model can be solved efficiently using Benders decomposition. The primary notations used in the paper are summarized in Online Appendix B.

4.1. Behavior Uncertainty

Consider a passenger p with a path set \mathcal{R}_p . Their inherent utility of using path r is denoted as V_p^r . If path r' was recommended, the impact of the recommendation on the utility of path r is denoted as $I_{p,r'}^r$. Hence, the passenger's overall utility of using path r can be represented as

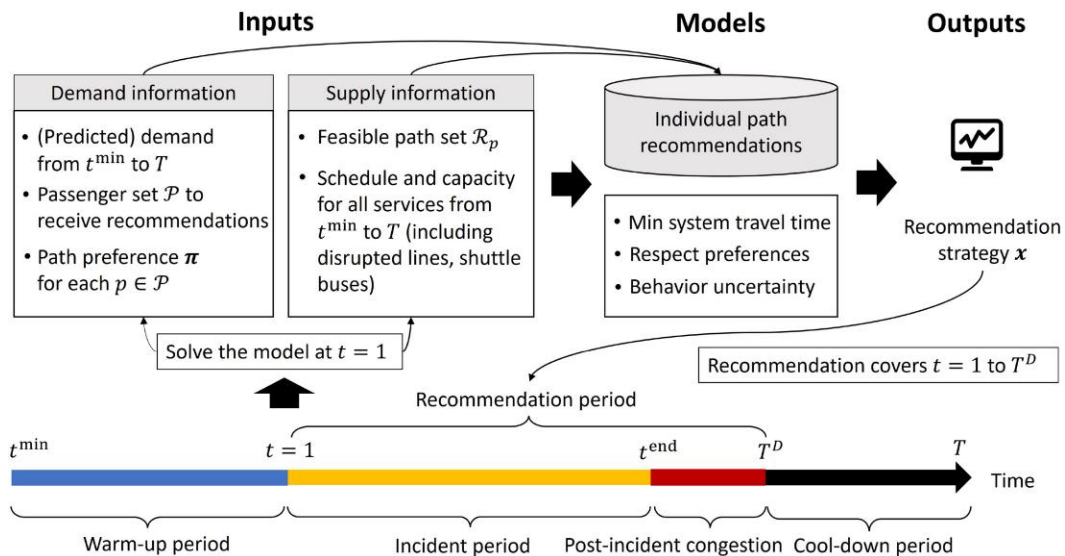
$$U_p^r = V_p^r + \sum_{r' \in \mathcal{R}_p} x_{p,r'} \cdot I_{p,r'}^r + \xi_p^r \quad \forall r \in \mathcal{R}_p, p \in \mathcal{P}, \quad (3)$$

where ξ_p^r is the random error, and $x_{p,r'} = 1$ if passenger p is recommended path r' , otherwise $x_{p,r'} = 0$. Let $\pi_{p,r'}^r$ be the conditional probability that passenger p chooses path r given that the recommended path is r' . Assuming a utility-maximizing behavior, we have

$$\pi_{p,r'}^r = \mathbb{P}(V_p^r + I_{p,r'}^r + \xi_p^r \geq V_p^{r''} + I_{p,r'}^{r''} + \xi_p^{r''}, \forall r'' \in \mathcal{R}_p). \quad (4)$$

Different assumptions for the distribution of ξ_p^r can lead to different expressions. For example, if ξ_p^r are i.i.d. Gumbel distributed, the choice probability reduces to

Figure 2. (Color online) Problem Description and Model Framework



the multinomial logit model (Train 2009, Mo et al. 2025) and we have

$$\pi_{p,r}^r = \frac{\exp(V_p^r + I_{p,r}^r)}{\sum_{r'' \in \mathcal{R}_p} \exp(V_p^{r''} + I_{p,r''}^r)}. \quad (5)$$

The values of V_p^r and $I_{p,r}^r$ can be calibrated using data from individual-level surveys or smart card data, which deserve separate research. When developing the individual path recommendation model, we assume $\pi_{p,r}^r$ is known. Figure 3 shows an example of the conditional probability matrix. The specific values assume that paths with recommendations are more likely to be chosen.

The conditional probability $\pi_{p,r}^r$ captures the individual’s inherent preference for different paths as well as the response to the recommendation system. It varies across individuals and reflects their behavioral uncertainties. This study focuses on design path recommendation systems based on the value of $\pi_{p,r}^r$.

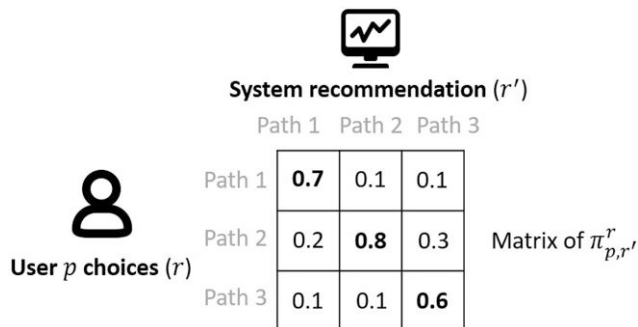
4.2. Individual Path Recommendation

Let $\mathbb{1}_{p,r}^r$ be the indicator random variable representing whether passenger p actually chooses path r or not given that the passenger is recommended path r' . By definition, $\mathbb{1}_{p,r}^r$ is a Bernoulli random variable with $\mathbb{E}[\mathbb{1}_{p,r}^r] = \pi_{p,r}^r$ and $\text{Var}[\mathbb{1}_{p,r}^r] = \pi_{p,r}^r \cdot (1 - \pi_{p,r}^r)$. Consider an OD pair (u, v) and departure time t , where u, v are two stations in the transit network. Let $\mathcal{R}^{u,v}$ be the set of feasible paths for OD pair (u, v) . Define $Q_t^{u,v,r}$ as the number of passengers in \mathcal{P} with OD pair (u, v) and departure time t , who use path $r \in \mathcal{R}^{u,v}$. We have

$$Q_t^{u,v,r}(\mathbf{x}) = \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \mathbb{1}_{p,r}^r \quad (6)$$

where $\mathcal{P}_t^{u,v} \subseteq \mathcal{P}$ is the set of passengers with OD pair (u, v) arriving at the system at time interval t who receive path recommendations. Note that $Q_t^{u,v,r}(\mathbf{x})$ is also a random variable and it is binomially distributed. Define $Q(\mathbf{x}) := (Q_t^{u,v,r})_{t \in \mathcal{T}, (u,v,r) \in \mathcal{F}}$, where \mathcal{F} is the set of all triplets (u, v, r) in the system. Denote its distribution

Figure 3. Example of Conditional Path Choice Probability



Note. The bold numbers in the diagonal represent people have higher probability of choosing the recommended paths.

as $\mathbb{P}_{Q(\mathbf{x})}$, which depends on the recommendation \mathbf{x} . The objective of this study is to provide a recommendation strategy that minimizes the STT, which can be formulated as

$$(S - \text{IPR}) \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{Q(\mathbf{x})}[STT(Q(\mathbf{x}))], \quad (7)$$

where $\mathcal{X} := \{\mathbf{x} : \sum_{r \in \mathcal{R}_p} x_{p,r} = 1, x_{p,r} \in \{0, 1\}, \forall p \in \mathcal{P}, r \in \mathcal{R}_p\}$ is the feasible recommendation set. $STT(Q(\mathbf{x}))$ is a pseudo-function that returns the system travel time of the public transit network given the path flow $Q(\mathbf{x})$ (e.g., a simulation or transit network loading model).

Besides the total system travel time, many recommendation systems also aim to respect passengers’ preferences. That is, if possible, a path with high inherent utility V_p^r should be recommended. Hence, the following term is added to the objective function:

$$\max \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} x_{p,r} \cdot V_p^r \Leftrightarrow \min \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r} \cdot V_p^r. \quad (8)$$

The final S-IPR model considering individual preference can be extended as

$$(S\text{-IPR-P}) \min_{\mathbf{x} \in \mathcal{X}} \Psi \cdot \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r} \cdot V_p^r + \mathbb{E}_{Q(\mathbf{x})}[STT(Q(\mathbf{x}))], \quad (9)$$

where Ψ is a hyperparameter to adjust the scale and balance the trade-off between system efficiency and passenger preferences.

The S-IPR-P problem (9) is a stochastic optimization with decision-dependent uncertainties (Goel and Grossmann 2006, Drusvyatskiy and Xiao 2023), which is also known as “performative predictions.” This is because the expectation of the system travel time is taken over the distribution of the path flows $Q(\mathbf{x})$, and the distribution of $Q(\mathbf{x})$ depends on the decision variable \mathbf{x} . For normal stochastic optimization, the expectation of the objective function is usually taken over some exogenous parameters (with a fixed distribution that is invariant to decision variables).

The typical ways to solve it are performative stochastic gradient descent (for continuous problems) or applying sample average approximation in branch-and-cut (for integer problems). However, $STT(\cdot)$ usually has no analytical formulation because of the hard capacity constraints in transit systems (Mo et al. 2023a, 2023b). Therefore, the traditional ways are hard to adopt.

In this study, we propose a new idea to approximate the performative stochastic optimization using a deterministic formulation with two new constraints: ϵ -feasibility and Γ -concentration. The idea is that, because the distribution of $Q(\mathbf{x})$ is flexible and dependent on \mathbf{x} , we aim to restrict its distribution such that $\mathbb{E}_{Q(\mathbf{x})}[STT(Q(\mathbf{x}))]$ can be approximated by some deterministic values.

4.3. Solving System Travel Time as Linear Programming

One of the main challenges in solving the S-IPR-P problem (9) is that $STT(\cdot)$ has no analytical formulation. It is usually obtained by transit network loading or simulation models (Mo et al. 2020). Bertsimas, Sian Ng, and Yan (2020) show that the network loading process can be formulated as LP. In this section, we adapt their formulations for the transit network loading in a public transit system with service disruptions.

Consider an OD pair (u, v) and departure time t . Define $f_t^{u,v,r}$ as the number of passengers **not in** \mathcal{P} with OD pair (u, v) and departure time t , who use path $r \in \mathcal{R}^{u,v}$. Specifically, $Q_t^{u,v,r}(x)$ represents the passenger flows who receive recommendations whereas $f_t^{u,v,r}$ represents those who do not. Hence, the total path flow in $r \in \mathcal{R}^{u,v}$ is $Q_t^{u,v,r} + f_t^{u,v,r}$. Letting $d_t^{u,v}$ be the total demand of OD pair (u, v) at time t , we have

$$\sum_{r \in \mathcal{R}^{u,v}} Q_t^{u,v,r}(x) + f_t^{u,v,r} = d_t^{u,v} \quad \forall (u, v) \in \mathcal{W}, t \in \mathcal{T}, \quad (10)$$

where \mathcal{W} is the set of all OD pairs.

Consider a path r for OD pair (u, v) . A path may include multiple legs, where each leg is associated with the service on a rail or a bus line. For example, the path r in Figure 4 with origin at station A and destination at station D (indicated by green arrows) has two legs: the first one (A to B) on the rail line and the second on the bus line (C to D). Every leg has a boarding and an alighting station. For example, leg 1 (resp. 2) in this example has boarding station A (resp. C) and alighting station B (resp. D). Let $\mathcal{I}^{u,v,r} = \{1, \dots, |\mathcal{I}^{u,v,r}|\}$ be the set of legs for path r . We use a four-element tuple (u, v, r, i) to represent a leg i of path r for OD pair (u, v) , where $i \in \mathcal{I}^{u,v,r}$. In this example, leg 1 is $(A, D, r, 1)$ and leg 2 is $(A, D, r, 2)$.

Let $\Delta_t^{u,v,r,i}$ (resp. $\delta_t^{u,v,r,i}$) be the travel time between the **terminal** and the **boarding** (resp. **alighting**) station of leg (u, v, r, i) for a vehicle **departing** from the terminal at time t , where a terminal is the first departure station

of a line. As shown in the example (Figure 4), $\Delta_t^{A,D,r,1}$ (resp. $\delta_t^{A,D,r,1}$) represent the travel time from the terminal to station A (resp. station B) on the rail line. $\Delta_t^{A,D,r,2}$ (resp. $\delta_t^{A,D,r,2}$) represent the travel time from the terminal to station C (resp. station D) on the bus line.

Hence, if a vehicle departs at time t , its arrival time at the boarding (resp. alighting) station of leg (u, v, r, i) is $t + \Delta_t^{u,v,r,i}$ (resp. $t + \delta_t^{u,v,r,i}$). Then, $\delta_t^{u,v,r,i} - \Delta_t^{u,v,r,i}$ represents the total in-vehicle time of leg (u, v, r, i) for the vehicle.

Define $z_t^{u,v,r,i}$ (decision variable) as the total number of onboard passengers in leg (u, v, r, i) who board a vehicle that **had departed** from the terminal at time t . There are three types of constraints for the network flow description: (1) existing flow constraints, (2) vehicle capacity constraints, and (3) flow conservation constraints.

4.3.1. Existing Flow Constraints. Although the path recommendations start at time $t = 1$, there are passengers who have already boarded the vehicles. Ignoring these existing flows may lead to an overestimation of the system's available capacity. To capture the existing onboard flows at $t = 1$, we define the set of all onboard flow indices at time $t = 1$ as

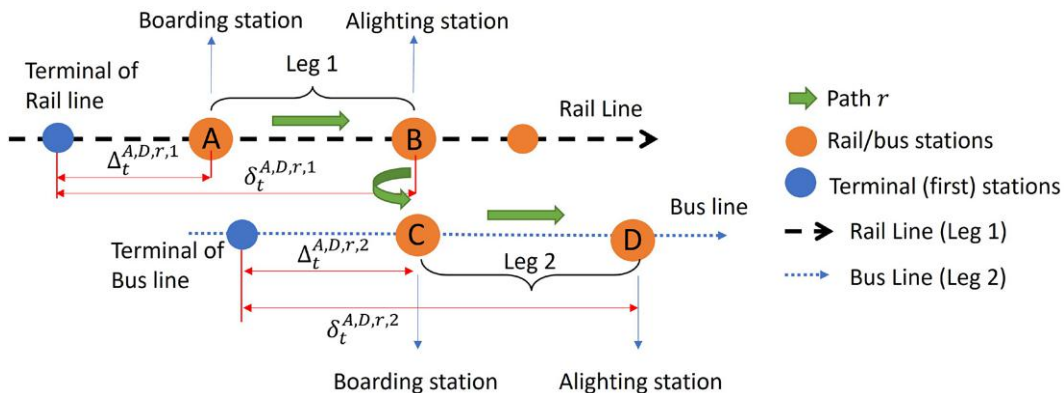
$$\Omega_1 = \{(u, v, r, i, t) : t + \Delta_t^{u,v,r,i} \leq 1 \leq t + \delta_t^{u,v,r,i}\}, \quad (11)$$

where Ω_1 represents the indices of passenger flows who enter the system before the incident starts and have not left the system at the time of the incident (i.e., $t = 1$). The existing flow constraints can be expressed as

$$z_t^{u,v,r,i} = \hat{z}_t^{u,v,r,i} \quad \forall (u, v, r, i, t) \in \Omega_1, \quad (12)$$

where $\hat{z}_t^{u,v,r,i}$ are constants that capture the existing onboard flows when the incident happens. These flows can be directly obtained from a simulation model or real-time passenger count data.

Figure 4. (Color online) Definition of Paths and Legs



4.3.2. Capacity Constraints. Transit vehicles have limited capacity. Consider a vehicle departing at time t on line l (referred to as vehicle (l, t)). We denote its total number of onboard passengers at time t' as $O_{l,t,t'}$. Specifically, $O_{l,t,t'}$ can be expressed as

$$O_{l,t,t'}(\mathbf{z}) = \sum_{\{(u,v,r,i) \in \text{OBLEgs}(l,t,t')\}} z_t^{u,v,r,i} \quad \forall l \in \mathcal{L}, \forall t \in \mathcal{T}, t' = t, t+1, \dots, T_{l,t} \quad (13)$$

where $T_{l,t}$ is the time index that vehicle (l, t) arrives at the last station of line l , \mathbf{z} is the decision variable vector defined as $\mathbf{z} = (z_t^{u,v,r,i})_{t \in \mathcal{T}, (u,v,r) \in \mathcal{F}, i \in \mathcal{I}^{u,v,r}}$, and $\text{OBLEgs}(l, t, t')$ is the set of legs for flows that could board vehicle (l, t) at time t' , defined as

$$\text{OBLEgs}(l, t, t') = \{(u, v, r, i) : \text{Leg}(u, v, r, i) \text{ on line } l, \text{ and } t + \Delta_t^{u,v,r,i} \leq t' \leq t + \delta_t^{u,v,r,i}\}. \quad (14)$$

Equation (14) implies that the legs can be selected based on their associated vehicle arrival time at the boarding and alighting stations. Then, the capacity constraint is

$$O_{l,t,t'}(\mathbf{z}) \leq K_{l,t} \quad \forall l \in \mathcal{L}, t \in \mathcal{T}, t' = t, t+1, \dots, T_{l,t} \quad (15)$$

where $K_{l,t}$ is the capacity of the vehicle (l, t) , and \mathcal{L} is the set of all lines.

4.3.3. Flow Conservation Constraint. There are two different flow conservation constraints: (1) flow conservation at origin stations and (2) that at transfer stations. To ensure the origin flow conservation, the cumulative number of arriving passengers should be larger than the cumulative number of boarding passengers at an origin at any time (note that we use “larger than” because passengers left behind because of capacity constraints are allowed). This indicates that not all arriving passengers can board because of potentially being left behind because of capacity constraints.

The number of arriving passengers (i.e., demand) for path (u, v, r) at time t is $d_t^{u,v}$, and the number of boarding passengers at the origin station (i.e., u) at time t is $z_t^{u,v,r,1}$ (i.e., the first leg) with $t' + \Delta_{t'}^{u,v,r,1} = t$. t' is the vehicle departure time from the terminal. $t' + \Delta_{t'}^{u,v,r,1}$ is the time when the vehicle arrives at the boarding station of leg $(u, v, r, 1)$. Therefore, the origin flow conservation constraint can be written as

$$\sum_{\{t': t_{\min} \leq t' + \Delta_{t'}^{u,v,r,1} \leq t\}} z_{t'}^{u,v,r,1} \leq \sum_{t'=t_{\min}}^t Q_t^{u,v,r}(\mathbf{x}) + f_t^{u,v,r} \quad \forall (u, v, r) \in \mathcal{F}, t \in \mathcal{T}. \quad (16)$$

Now consider the flow conservation at a transfer station. All arriving passengers at a transfer station of a path are the onboard passengers from the last leg. Therefore, we use a similar way to define the transfer

flow conservation: the cumulative number of onboard passengers from the last leg should be larger than the cumulative number of boarding passengers at the transfer station, and the number of boarding passengers at the transfer station is simply $z_{t'}^{u,v,r,i}$ with $i \in \mathcal{I}^{(u,v,r)} \setminus \{1\}$. Hence, flow conservation constraints at a transfer station are

$$\sum_{\{t': t_{\min} \leq t' + \Delta_{t'}^{u,v,r,i} \leq t\}} z_{t'}^{u,v,r,i} \leq \sum_{\{t': t_{\min} \leq t' + \delta_{t'}^{u,v,r,i-1} \leq t\}} z_{t'}^{u,v,r,i-1} \quad \forall (u, v, r) \in \mathcal{F}, i \in \mathcal{I}^{(u,v,r)} \setminus \{1\}, t \in \mathcal{T}. \quad (17)$$

Note that $z_{t'}^{u,v,r,i}$ is defined as the onboard passengers for vehicles **departing** at time t' . Therefore, $t' + \delta_{t'}^{u,v,r,i-1}$ is the alighting time for passengers at leg $i-1$ (which is also the transfer demand arrival time at leg i as we assume transfer walk time is within a time interval τ and is negligible). $t' + \Delta_{t'}^{u,v,r,i}$ is the boarding time for passengers at leg i .

The objective is to minimize the total travel time for all passengers in the system. Total travel time can be decomposed into waiting time and in-vehicle time.

4.3.4. In-Vehicle Time. Total in-vehicle time is simply the onboard flow multiplied by the travel time on each leg:

$$\text{IVT}(Q_t^{u,v,r}(\mathbf{x}), \mathbf{z}) = \sum_{(u,v,r) \in \mathcal{F}} \sum_{i \in \mathcal{I}^{u,v,r}} \sum_{t \in \mathcal{T}} z_t^{u,v,r,i} \cdot T_{u,v,r,i,t}^{\text{IVT}} \quad (18)$$

where $T_{u,v,r,i,t}^{\text{IVT}}$ is the in-vehicle time of leg (u, v, r, i) of the vehicle departing at time t .

4.3.5. Waiting Time. There are two causes of waiting time: (1) waiting time because of vehicle headways and (2) waiting time resulting from being left behind. During a specific time interval t , all left-behind passengers would have a waiting time of τ . All boarding passengers, assuming uniform arrival, have an average waiting time that is half of the time interval (i.e., $\frac{\tau}{2}$). Therefore, the total waiting time for passengers at station s and time t can be formulated as

$$\text{WT}_{s,t} = \tau(AD_{s,t} + XD_{s,t} - BD_{s,t}) + \frac{\tau}{2}(BD_{s,t+1} - BD_{s,t}), \quad (19)$$

where $AD_{s,t}$ represents the **cumulative arriving demand** at station s **up to** time t , $XD_{s,t}$ represents the **cumulative transferring demand** at station s **up to** time t , and $BD_{s,t}$ represents the **cumulative boarded demand** at station s **up to** time t . Hence, $(BD_{s,t+1} - BD_{s,t})$ represents the total number of boarding passengers at time t and station s , and $(AD_{s,t} + XD_{s,t} - BD_{s,t})$ represents the total number of left-behind passengers at station s and time t . Finally,

the total system waiting time is

$$WT(Q(x), z) = \sum_{s \in \mathcal{S}} \sum_{t=1}^T WT_{s,t}. \quad (20)$$

The cumulative arriving demand $AD_{s,t}$ is simply all arriving passengers with origin s up to time t :

$$AD_{s,t} = \sum_{\{(u,v,r):u=s\}} \sum_{t'=t^{\min}}^t (Q_t^{u,v,r}(x) + f_t^{u,v,r}) \quad \forall s \in \mathcal{S}, t \in \mathcal{T}, \quad (21)$$

where \mathcal{S} is the set of all stations.

The cumulative transferring demand is all passengers alighting at station s from their previous leg $i-1$ for their next leg i :

$$XD_{s,t} = \sum_{\{(u,v,r,i) \in \text{Trans}(s)\}} \sum_{\{t': t^{\min} \leq t' + \delta_t^{u,v,r,i-1} \leq t\}} z_{t'}^{u,v,r,i-1} \quad \forall t \in \mathcal{T}, \quad (22)$$

where $\text{Trans}(s)$ is the set of legs that transfer at station s .

The cumulative boarded demand is all passengers who successfully board a vehicle at station s at time t . Defining $\text{Board}(s)$ as the set of all legs with boarding station s , we have

$$BD_{s,t} = \sum_{\{(u,v,r,i) \in \text{Board}(s)\}} \sum_{\{t': t^{\min} \leq t' + \Delta_t^{u,v,r,i} \leq t\}} z_{t'}^{u,v,r,i} \quad \forall t \in \mathcal{T}. \quad (23)$$

Taking everything into consideration, the total travel time in the system is $WT(Q(x), z) + IVT(Q(x), z)$. Assuming everyone in the system will board the first available vehicle, the total system travel time can be obtained by minimizing $WT + IVT$:

$$(\text{LP-STT}) \quad STT(Q(x)) := \min_z WT(Q(x), z) + IVT(Q(x), z) \quad (24a)$$

$$\text{s.t. Constraints (10), (12), (15), (16), and (17),} \quad (24b)$$

$$z_t^{u,v,r,i} \geq 0 \quad \forall t \in \mathcal{T}, (u,v,r) \in \mathcal{F}, i \in \mathcal{I}^{u,v,r}. \quad (24c)$$

4.3.6. Incident Specification. Equation (24) is a general formulation of the optimal flow problem. Now we will introduce how the incident-specific information is incorporated into this problem. We assume the incident causes a service disruption on a specific line (if only several stations are interrupted, we can separate the line into multiple lines so that the assumption always holds). The service disruption on a line can be seen as the stop of vehicles for a period of time. The vehicle stopping can be captured by the parameters $\Delta_t^{u,v,r,i}$, $\delta_t^{u,v,r,i}$, and $K_{l,t}$. Specifically, a long stop due to an incident can be seen as an increase in travel time from the terminal to downstream stations (i.e., increase in $\Delta_t^{u,v,r,i}$ and $\delta_t^{u,v,r,i}$). Moreover, because there is no vehicle dispatching during the incident, we set $K_{l,t} = 0$ for the

corresponding time and line. In this way, we can model the incident without changing the formulation.

4.4. Eliminating Stochasticity with a Single-Point Approximation

With the LP formulation for $STT(\cdot)$, we can reformulate the S-IPR-P problem (9) as

$$(\text{S-IPR-P}) \quad \min_{x \in \mathcal{X}} \Psi \cdot \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r} \cdot V_p^r + \mathbb{E}_{Q(x)} \left[\min_{z \in \mathcal{Z}(Q(x))} WT(Q(x), z) + IVT(Q(x), z) \right], \quad (25)$$

where $\mathcal{Z}(Q(x))$ is the feasible region for the LP-STT problem (i.e., Constraints (24b) and (24c)). The S-IPR-P problem has an outer and inner minimization problem. However, they cannot be directly combined because of the expectation operator. In this study, we propose to remove the expectation operator using a single-point approximation. Specifically, we wish to find a path flow $q \in \mathcal{D}(x)$ such that

$$\mathbb{E}_{Q(x)} \left[\min_{z \in \mathcal{Z}(Q(x))} WT(Q(x), z) + IVT(Q(x), z) \right] \approx \min_{q \in \mathcal{D}(x), z \in \mathcal{Z}(q)} WT(q, z) + IVT(q, z), \quad (26)$$

where $\mathcal{D}(x)$ is some feasible region that is determined by the recommendation x such that the single-point approximation is reasonable. The intuition for defining $\mathcal{D}(x)$ is to make sure that under the recommendation x , the distribution of $Q(x)$ is concentrated to the mean and the final solution q^* is close to the mean. In this case, we could use a single point to represent the expectation. We define two new concepts, ϵ -feasibility and Γ -concentration, which match with the intuition.

Definition 1 (ϵ -Feasible Flows). A flow $q_t^{u,v,r}$ is ϵ -feasible if and only if

$$|q_t^{u,v,r} - \mu_t^{u,v,r}(x)| \leq \epsilon_t^{u,v,r}, \quad \forall (u,v,r) \in \mathcal{F}, t = t^{\min}, \dots, T, \quad (27)$$

where

$$\mu_t^{u,v,r}(x) := \mathbb{E}[Q_t^{u,v,r}] = \sum_{p \in \mathcal{P}^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r, \quad (28)$$

and $\epsilon_t^{u,v,r}$ is a small positive constant.

If q is ϵ -feasible, it means q is close to the expectation of the actual flow $Q(x)$ under recommendation strategy x . ϵ -feasibility results in a direct linear constraint on q .

Definition 2 (Γ -Concentrated Flows). A flow $q_t^{u,v,r}$ is Γ -concentrated if and only if it is ϵ -feasible and for any constant $a > \epsilon_t^{u,v,r}$, we have

$$\mathbb{P}[|Q_t^{u,v,r} - q_t^{u,v,r}| \geq a] \leq \left(\frac{\Gamma_t^{u,v,r}}{a - \epsilon_t^{u,v,r}} \right)^2 \quad \forall (u,v,r) \in \mathcal{F}, t = t^{\min}, \dots, T, \quad (29)$$

where $\Gamma_t^{u,v,r}$ is a small positive constant.

If q is Γ -concentrated, it means that the probability that $Q_t^{u,v,r}$ and $q_t^{u,v,r}$ are very different (i.e., with a difference greater than a) is bounded from above, suggesting that $Q_t^{u,v,r}$ is concentrated around $q_t^{u,v,r}$.

Notice that we can modify the definition of Γ -concentration using the following proposition:

Proposition 1. *The Γ -concentration inequality (29) holds if the variance of $Q_t^{u,v,r}$ is bounded from above by $(\Gamma_t^{u,v,r})^2$. Mathematically,*

$$\begin{aligned} \text{Var}[Q_t^{u,v,r}] &= \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r \cdot (1 - \pi_{p,r'}^r) \\ &\leq (\Gamma_t^{u,v,r})^2. \end{aligned} \quad (30)$$

The variance formulation is based on the fact that $(x_{p,r'})^2 = x_{p,r'}$ and $\text{Cov}[\mathbb{1}_{p,r'}, \mathbb{1}_{p,r''}] = 0$ if $r' \neq r''$. The proof of Proposition 1 is a direct result of Chebyshev's inequality and the triangle inequality (see Online Appendix C).

Then, we can define $\mathcal{D}(x)$ as the set of path flows that are both ϵ -feasible and Γ -concentrated:

$$\mathcal{D}(x) := \left\{ q \geq 0 : q_t^{u,v,r} \geq \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r - \epsilon_t^{u,v,r}, \right. \\ \left. \forall (u,v,r) \in \mathcal{F}, t \in \mathcal{T} \right. \quad (31)$$

$$\left. q_t^{u,v,r} \leq \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r + \epsilon_t^{u,v,r}, \right. \\ \left. \forall (u,v,r) \in \mathcal{F}, t \in \mathcal{T} \right. \quad (32)$$

$$\left. \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r \cdot (1 - \pi_{p,r'}^r) \leq (\Gamma_t^{u,v,r})^2, \right. \\ \left. \forall (u,v,r) \in \mathcal{F}, t \in \mathcal{T} \right\}. \quad (33)$$

Constraints (31), (32), and (33) are all linear, which keeps the tractability of the formulation.

With the approximation of (26), we can drop the expectation operator, and combine the inner and outer minimization problem. Then, the approximated individual path recommendation problem considering preference (A-IPR-P) can be reformulated as

$$\begin{aligned} \text{(A-IPR-P)} \quad & \min_{x \in \mathcal{X}, q \in \mathcal{D}(x), z \in \mathcal{Z}(q)} -\Psi \cdot \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} x_{p,r} \\ & \cdot V_p^r + \text{WT}(q, z) + \text{IVT}(q, z). \end{aligned} \quad (34)$$

The quality of the approximation and other theoretical analysis regarding ϵ -feasibility and Γ -concentration will be discussed in Section 5. For now, we just need to know that a single-point approximation is proposed to transform the original stochastic optimization with decision-dependent distributions to tractable mixed integer linear programming (MILP).

For a better understanding of the ϵ -feasibility and Γ -concentration, let us consider a simple example where there is only one path flow Q and the system travel time is $\text{STT}(Q) = \alpha \cdot Q$. In this case, $\mathbb{E}_Q[\text{STT}(Q)] = \alpha \cdot \mathbb{E}[Q] = \text{STT}(\mathbb{E}[Q])$. This shows the benefit of ϵ -feasibility: if q is ϵ -feasible and $\epsilon = 0$, we would have $\mathbb{E}_Q[\text{STT}(Q)] = \text{STT}(q)$, and the single-point approximation becomes exact. Hence, the ϵ -feasibility helps to make the approximated system travel time close to the mean.

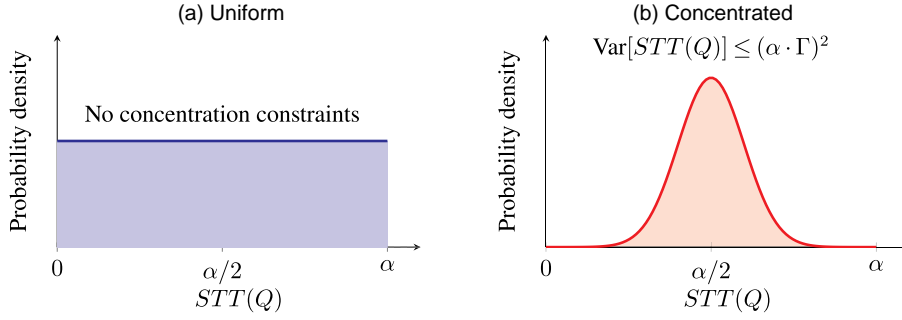
In terms of Γ -concentration, let us consider an extreme scenario where Q is uniformly distributed in $[0, 1]$ under the optimal recommendation strategy x^* . In this case, the distribution of $\text{STT}(Q)$ will be uniformly in $[0, \alpha]$ (Figure 5(a)). Then, x^* becomes meaningless because, under this recommendation, the actual system travel time has too many variations. However, if we add Γ -concentration to Q (i.e., $\text{Var}[Q] \leq \Gamma^2$), we would have $\text{Var}[\text{STT}(Q)] \leq (\alpha \cdot \Gamma)^2$, and the distribution of $\text{Var}[\text{STT}(Q)]$ will be concentrated (Figure 5(b)) toward the mean value, making the recommendation of minimizing the mean value meaningful.

4.5. Solving the Problem by BD

The structure of (34) allows us to efficiently solve it by BD (Benders 1962). The basic idea of BD is to decompose the problem into a master problem and a subproblem and solve these problems iteratively. The decision variables are divided into difficult variables, which in our case are the binary variables x and a set of easier variables, the continuous q and z . At each iteration, the master problem determines one possible leader decision x . This solution is used in the subproblem to generate optimality cuts or feasibility cuts, which are added to the master problem. In this study, the master problem decides the recommendation strategies, which are a MILP of a smaller scale and can be solved efficiently using existing solvers. The subproblem reduces to the LP-STT problem (24) with one more linear constraint (still linear programming). This format makes the BD an appropriate algorithm for the original problem. As this is a conventional method, we put the formulation details in Online Appendix H.

5. Theoretical Analysis

The key idea under our proposed method is using a single point to approximate the expectation of the system travel time (26). A natural question is how good the approximation is. In Section 5.1, we show that the gap of the approximation is bounded from above with ϵ -feasibility and Γ -concentration constraints. In Section 5.2, we show that this approximation can be seen as a way of approximating the recourse function in a two-stage stochastic optimization. In Section 5.3, we provide further theoretical analysis to show that ϵ -feasibility

Figure 5. (Color online) Illustration for the Impact of Γ -Concentration Constraints on the System Travel Time Distribution

and Γ -concentration are related to expectation and chance constraints in a typical stochastic optimization formulation, respectively.

5.1. Approximation Quality Analysis

Let (q^*, z^*, x^*) be the optimal solution of our approximated A-IPR-P formulation (34). Given the optimal recommendation strategy x^* , let the corresponding random variable of path flows be $Q(x^*)$. Let $\mathbb{P}_{Q(x^*)}(\cdot)$ be the probability density function of $Q(x^*)$. The expectation of $STT(Q(x^*))$ can be expressed as

$$\mathbb{E}_{Q(x^*)}[STT(Q(x^*))] = \sum_{\hat{q} \in \mathcal{Q}(x^*)} \left[\min_{z \in \mathcal{Z}(\hat{q})} WT(\hat{q}, z) + IVT(\hat{q}, z) \right] \cdot \mathbb{P}_{Q(x^*)}(\hat{q}), \quad (35)$$

where $\mathcal{Q}(x^*)$ is the support of the random variable $Q(x^*)$, defined as

$$\mathcal{Q}(x^*) = \left\{ \hat{q} \geq 0 : \hat{q}_t^{u,v,r} = \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'}^* \cdot \hat{1}_{p,r'}^r, \forall \hat{1}_{p,r'}^r \in \{0, 1\}, \right. \\ \left. \sum_{r \in \mathcal{R}_p} \hat{1}_{p,r'}^r = 1, \forall (u, v, r) \in \mathcal{F}, t \in \mathcal{T} \right\}. \quad (36)$$

It represents the set of all possible values of network flows given recommendation x^* . $\hat{1}_{p,r'}^r$ is a binary deterministic variable that can be either zero or one (i.e., the possible realization of $\mathbb{1}_{p,r'}^r$).

$\mathbb{E}_{Q(x^*)}[STT(Q(x^*))]$ is the expected system travel time given recommendation strategy x^* . However, our model is optimized using (34), where the minimization is conducted over the model-evaluated system travel time under the path flow q^* :

$$STT(q^*) = \min_{z \in \mathcal{Z}(q^*)} WT(q^*, z) + IVT(q^*, z). \quad (37)$$

It is worth analyzing the relationship between the model-evaluated system travel time ($STT(q^*)$) and the expected system travel time ($\mathbb{E}_{Q(x^*)}[STT(Q(x^*))]$). This analysis tells us how well our proposed approach can approximate the real system performance indicator.

Lemma 1. $STT(q) = \min_{z \in \mathcal{Z}(q)} WT(q, z) + IVT(q, z)$ is continuous in terms of q if the set of optimal flows is bounded (i.e., there are a limited number of flow patterns that permit the optimal system travel time).

The proof is a direct implementation of Berge's Maximum Theorem (Sundaram 1996) and is shown in Online Appendix D. Lemma 1 implies that a small change in the path flows only results in small changes in system travel time. Because the system travel time is usually bounded from above given a finite-scale transit network, small flow changes should not yield infinite changes in the system travel time. Hence, $STT(q)$ should also have a bounded gradient. Combining the continuity property in Lemma 1, we conclude that $STT(q)$ is Lipschitz continuous. That is, there exists a constant L such that, for any network flows q_1 and q_2 , we have

$$|STT(q_1) - STT(q_2)| \leq L \cdot \|q_1 - q_2\|_1. \quad (38)$$

Proposition 2. Let (q^*, z^*, x^*) be the optimal solution of the A-IPR-P problem (34). $Q(x^*)$ is the random variable of path flows. The difference between the model-evaluated system travel time and the expected system travel time is bounded from above if

- the set of optimal flows is bounded (to implement Lemma 1),
- the network flows are bounded from above (i.e., there exist $q^{\text{Max}} < \infty$ such that $\hat{q} \leq q^{\text{Max}}, \forall \hat{q} \in \mathcal{Q}(x^*)$), and
- $|q^* - \mathbb{E}[Q(x^*)]| \leq \epsilon$ (ϵ -feasibility) and $\text{Var}[Q(x^*)] \leq \Gamma^2$ (Γ -concentration),

where $\mathbb{E}[Q(x^*)] := (\mathbb{E}[Q_i(x^*)])_{(u,v,r) \in \mathcal{F}, t \in \mathcal{T}}$, $\epsilon := (\epsilon_t^{u,v,r})_{(u,v,r) \in \mathcal{F}, t \in \mathcal{T}}$, $\Gamma := (\Gamma_t^{u,v,r})_{(u,v,r) \in \mathcal{F}, t \in \mathcal{T}}$. The bound of the difference is determined by both ϵ and Γ . Mathematically,

$$|\mathbb{E}_{Q(x^*)}[STT(Q(x^*))] - STT(q^*)| \\ \leq 2L \cdot \|\epsilon\|_1 + L \cdot (\|\mathbb{E}[Q(x^*)]\|_1 + \|q^{\text{Max}}\|_1 + 2\|\epsilon\|_1) \cdot \|\Gamma\|_2^2. \quad (39)$$

Remark 1. Proposition 2 shows that even if the model is optimized on a realization of the system travel time (not the expectation), as long as we impose the

ϵ -feasibility and Γ -concentration, the model-evaluated system travel time and the expected system travel time will be similar if ϵ and Γ are small. It provides a quality guarantee of the approximation.

Remark 1 raises a natural question on how we set the values for ϵ and Γ . Ideally, we wish them to be as small as possible to have a good approximation quality. Based on the definition of ϵ -feasibility (i.e., (31) and (32)), ϵ can be as small as zero without violating any constraints. Therefore, we should set $\epsilon = 0$ for better approximation quality. However, when Γ is too small, though we have a lower variance and approximation gap, the recommendation strategies will be very restricted (or even infeasible). In that case, the quality of the recommendation may not be good. In summary, ϵ should be set as zero. For Γ , though small values of Γ are preferred for the approximation qualities, there are also trade-offs for the solution quality and problem feasibility. We may need to test different settings to select the best hyperparameters.

As the selection of Γ could affect the feasibility of the problem, we propose the following method to set its value. Consider the group of passengers in $\mathcal{P}_t^{u,v}$. The largest and smallest possible variance of the total variance $\sum_{r \in \mathcal{R}^{u,v}} \text{Var}[Q_t^{u,v,r}]$ can be obtained as

$$\begin{aligned} (\Gamma_t^{u,v,\text{Max}})^2 &:= \max_{x \in \mathcal{X}} \sum_{r \in \mathcal{R}^{u,v}} \text{Var}[Q_t^{u,v,r}(x)] \\ &= \sum_{p \in \mathcal{P}_t^{u,v}} \max_{r' \in \mathcal{R}_p} \left(\sum_{r \in \mathcal{R}_p} x_{p,r'} \cdot \pi_{p,r'}^r \cdot (1 - \pi_{p,r'}^r) \right), \end{aligned} \quad (40)$$

$$\begin{aligned} (\Gamma_t^{u,v,\text{Min}})^2 &:= \min_{x \in \mathcal{X}} \sum_{r \in \mathcal{R}^{u,v}} \text{Var}[Q_t^{u,v,r}(x)] \\ &= \sum_{p \in \mathcal{P}_t^{u,v}} \min_{r' \in \mathcal{R}_p} \left(\sum_{r \in \mathcal{R}_p} x_{p,r'} \cdot \pi_{p,r'}^r \cdot (1 - \pi_{p,r'}^r) \right). \end{aligned} \quad (41)$$

Therefore, the values of Γ should satisfy

$$\Gamma_t^{u,v,\text{Min}} \leq \sum_{r \in \mathcal{R}^{u,v}} \Gamma_t^{u,v,r} \leq \Gamma_t^{u,v,\text{Max}}, \quad \forall (u,v) \in \mathcal{W}, t \in \mathcal{T}. \quad (42)$$

Note that both $\Gamma_t^{u,v,\text{Min}}$ and $\Gamma_t^{u,v,\text{Max}}$ can be precalculated given π . Then, we can modify the Γ -concentration constraints (33) to

$$\sum_{r \in \mathcal{R}^{u,v}} \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}_p} x_{p,r'} \cdot \pi_{p,r'}^r \cdot (1 - \pi_{p,r'}^r) \leq \sum_{r \in \mathcal{R}^{u,v}} (\Gamma_t^{u,v,r})^2, \quad \forall (u,v,r) \in \mathcal{F}, t \in \mathcal{T}. \quad (43)$$

With the new constraints, we can directly set the value of $\sum_{r \in \mathcal{R}^{u,v}} \Gamma_t^{u,v,r}$ in $[\Gamma_t^{u,v,\text{Min}}, \Gamma_t^{u,v,\text{Max}}]$. It is worth noting that because $\text{Var}[Q_t^{u,v,r}] \geq 0$ for all $(u,v,r) \in \mathcal{F}$ and $t \in \mathcal{T}$, imposing an upper bound on the summation

$(\sum_{r \in \mathcal{R}^{u,v}} \text{Var}[Q_t^{u,v,r}])$ also implies an upper bound of each single element (just the specific values of the upper bound are different). Hence, all the derivations above still hold.

5.2. Connections to Two-Stage Stochastic Optimization

The original S-IPR-P problem (25) needs to solve both recommendation strategy x and onboarding flow z ; a possible alternative formulation is a two-stage stochastic optimization, where the first stage is to determine the recommendation and the second stage is to determine z .

A typical way to solve the two-stage stochastic optimization is to construct an approximation $S\hat{T}T(x)$ for $\mathbb{E}_{Q(x)}[STT(Q(x))]$. Then, we solve

$$\min_{x \in \mathcal{X}} \Psi \cdot \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r} \cdot V_p^r + S\hat{T}T(x) \quad (44)$$

for the first stage instead.¹ From this perspective, we can treat ϵ -feasibility and Γ -concentration as a way of constructing $S\hat{T}T(x)$, that is,

$$S\hat{T}T(x) := \min_{q \in \mathcal{D}(x), z \in \mathcal{Z}(q)} WT(q, z) + IVT(q, z). \quad (45)$$

Therefore, combining (44) and (45) as a one-stage optimization problem yields our A-IPR-P formulation (34).

5.3. Connections to Expectation and Chance Constraints

Considering the original S-IPR-P formulation (25), the inner minimization problem (i.e., $\min_{z \in \mathcal{Z}(Q(x))} WT(Q(x), z) + IVT(Q(x), z)$) has uncertainties in both objective and constraints. A typical way to solve it is to transform the original constraints into expectation and chance constraints. For the convenience of analysis, we reformulate the inner minimization in the following general format:

$$\text{(Inner-Minimization)} \quad \min_z g(Q(x), z) \quad (46a)$$

$$\text{s.t. } h_j(Q(x), z) \leq 0 \quad \forall j \in \mathcal{J}, \quad (46b)$$

where $g(Q(x), z) := WT(Q(x), z) + IVT(Q(x), z)$ is the objective function. $h_j(\cdot)$ is the constraint function. \mathcal{J} is the set of constraint indices. $h_j(\cdot)$ is defined such that $\{z : h_j(Q(x), z) \leq 0, \forall j \in \mathcal{J}\} \Leftrightarrow \mathcal{Z}(Q(x))$. Then, we can also approximate the expected system travel time as

$$\mathbb{E}_{Q(x)}[STT(Q(x))] \approx \min_z \mathbb{E}_{Q(x)}[g(Q(x), z)] \quad (47a)$$

$$\text{s.t. } \mathbb{E}_{Q(x)}[h_i(Q(x), z)] \leq 0 \quad \forall i \in \mathcal{J} \quad (\text{Expectation constraints}) \quad (47b)$$

$$\text{or/and } \mathbb{P}_{Q(x)}[h_j(Q(x), z) \leq 0] \geq \eta \quad \forall j \in \mathcal{J} \quad (\text{Chance constraints}), \quad (47c)$$

where η is a predefined parameter for the probability

guarantee of the constraints. However, the formulations in (47) are in general hard to solve except that we have the closed-form expressions for $\mathbb{E}_{Q(x)}[\cdot]$ and $\mathbb{P}_{Q(x)}[\cdot]$ (or using some approximation techniques for the constraints). In this section, we aim to show that ϵ -feasibility and Γ -concentration are highly related to the expectation and chance constraints, respectively. Therefore, adding them to the constraints would help with the approximation of the expected system travel time.

Proposition 3. *Define*

$$STT_{SO} := \min_z \{ \mathbb{E}[g(Q(x), z)] : \mathbb{E}[h_j(Q(x), z)] \leq 0, \forall j \in \mathcal{J} \}, \quad (48)$$

$$STT_{EP}(\epsilon) := \min_{z, q} \{ g(q, z) : h_j(q, z) \leq 0, \forall j \in \mathcal{J}, |q - \mathbb{E}[Q(x)]| \leq \epsilon \}, \quad (49)$$

where STT_{SO} is the optimal solution of the stochastic optimization problem with expectation constraints. STT_{EP} is the optimal solution of the proposed approach with the ϵ -feasibility constraint. If $\epsilon = 0$, and $g(\cdot)$ and $h_j(\cdot)$ are both convex functions (corresponding to the convex optimization), we have

$$STT_{EP}(\epsilon = 0) \leq STT_{SO}. \quad (50)$$

The proof is based on Jensen's inequality and is shown in Online Appendix F. Proposition 3 shows that the ϵ -feasibility constraint is highly related to the expectation constraints (47b). Applying the ϵ -feasibility constraint will give a lower bound of the original problem with the expectation constraints. Proposition 3 is also related to the certainty-equivalent (or mean-field) variant of a stochastic optimization problem. When $\epsilon = 0$, we have $STT_{EP}(\epsilon = 0) = STT(\mathbb{E}[Q(x)])$. That is, we directly use the mean to solve the original stochastic optimization problem.

Now we will show that Γ -concentration is highly related to the chance constraint.

Proposition 4. *If $h_j(\cdot)$ is Lipschitz continuous with ℓ_2 norm, that is, there exists a positive constant C such that, for all q_1, q_2 , and z , $|h_j(q_1, z) - h_j(q_2, z)| \leq C \cdot \|q_1 - q_2\|_2$, then define the constraint set of the Γ -concentration as*

$$\mathcal{H}_\Gamma(Q(x)) = \{ z : \mathbb{E}[h_j(Q(x), z)] \leq 0, \forall j \in \mathcal{J}, \text{Var}[Q(x)] \leq \Gamma^2 \}, \quad (51)$$

and define the set of a weaker chance constraint as

$$\mathcal{H}_C(Q(x)) = \left\{ z : \mathbb{E}[h_j(Q(x), z)] \leq 0, \mathbb{P} \left[h_j(Q(x), z) \leq \frac{C \cdot \|\Gamma\|_2}{\sqrt{1-\eta}} \right] \geq \eta, \forall j \in \mathcal{J} \right\}. \quad (52)$$

Then, we have $\mathcal{H}_\Gamma(Q(x)) \subseteq \mathcal{H}_C(Q(x))$. That is, the Γ -concentration constraints can deduce a weaker version of the chance constraints.

The proof uses Chebyshev's inequality and is shown in Online Appendix G. Proposition 4 shows that Γ -concentration constraints are tighter than a weaker version of the chance constraints. When Γ is sufficiently small, we have $\mathbb{P} \left[h_j(Q(x), z) \leq \frac{C \cdot \|\Gamma\|_2}{\sqrt{1-\eta}} \right] \approx \mathbb{P}[h_j(Q(x), z) \leq 0]$. In this case, we derive the chance constraints from the Γ -concentration constraints.

6. Actual Case Study

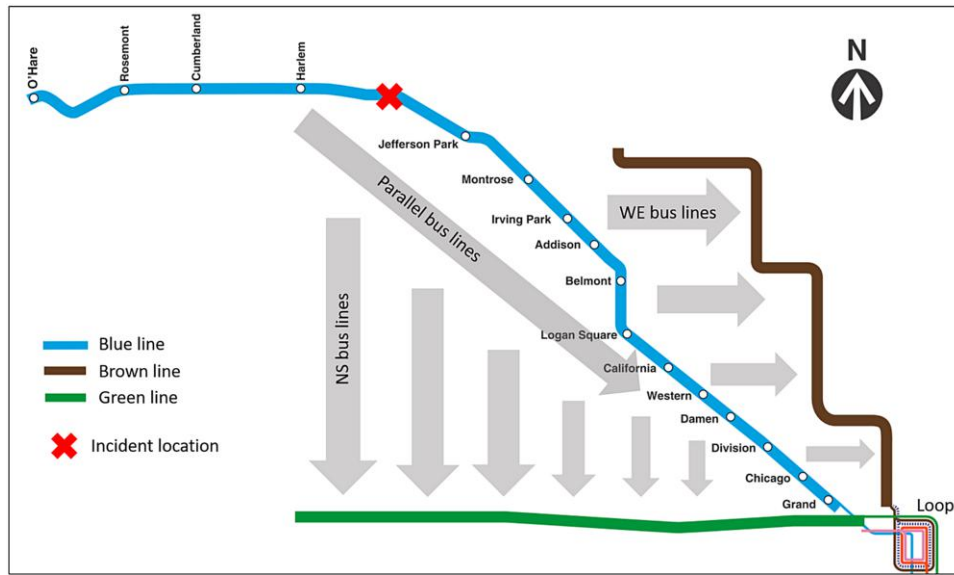
In this section, we implement the proposed algorithm in a real-world disruption case happening in the CTA urban rail system. We first introduce the background of the case study in Section 6.1, followed by model parameters, benchmark models, and experiment designs. The numerical results are summarized in Section 6.6.

6.1. Case Study Design

We consider an actual incident on the Blue Line of the CTA system (Figure 6). The incident started at 8:14 a.m. and ended at 9:13 a.m. on February 1, 2019, because of infrastructure issues between Harlem and Jefferson Park stations (the red X in the figure) that led to a whole Blue Line suspension. During the disruption (morning hours), the destination for most of the passengers is the "Loop" in the central business district (CBD) in Chicago. There are four alternative paths to the Loop: (1) using the Blue Line (i.e., waiting for the system to recover), (2) using the parallel bus lines, (3) using the North-South (NS) bus lines to transfer to the Green Line, and (4) using the West-East (WE) bus lines to transfer to the Brown Line. Based on the service structure, the route sets $\mathcal{R}^{(u,v)}$ for each OD pair (u, v) can be constructed.

In the case study, we divide the time into $\tau = 5$ -minute equal-length intervals, and focus on solving the problem at $t = 1$ (i.e., beginning of the incident). We assume that the set of passengers to receive recommendations (\mathcal{P}) consists of all passengers with their intended origins at the Blue Line and destinations in the Loop. A simulation model (Mo et al. 2020) is used to get the system state up to time $t = 1$ (i.e., the incident time 8:14 a.m.) and generate $\hat{z}_t^{u,v,r,i}$ and Ω_1 . The recommendation strategy covers passengers departing between $t = 1$ and $T^D = 23$, approximately one hour after the end of the incident (9:13 a.m.). The analysis period is set as $t^{\min} = -13$ and $T = 34$, approximately one hour before $t = 1$ and after T^D , providing enough buffer (warm-up and cool-down time) for passengers in \mathcal{P} to finish their trips. As demand and incident duration predictions are out of the scope of this paper, we simply use the actual demand and incident duration for all experiments. Our other work (Mo et al. 2023a) proposes to use robust and stochastic optimization to address demand and incident duration uncertainty, respectively.

Figure 6. (Color online) Case Study Network



The synthetic conditional probability matrix π used for the case study was generated as follows. During the incident, the CTA did not provide specific path recommendation information. For every individual, we assume that their actual path choices (referred to as the “status quo” choices) reflected their inherent preferences. Online Appendix I presents the method and results of inferring passengers’ status quo choices during the disruption using smart card data (Mo, Koutsoopoulos, and Zhao 2022). Around 49% of the passengers chose to wait whereas others either took the parallel buses or transferred to the rails. The basic idea is to track their tap-in records when entering the Blue Line and nearby bus routes, and compare them with their historical travel histories to get the transfer information.

Given the status quo choices, we assume that the “true” passenger p ’s inherent utility for path r is given by

$$V_p^r = \begin{cases} 1 + v_p^r & \text{if } r \text{ is } p\text{'s actual} \\ & \text{path choice} \\ v_p^r & \text{otherwise,} \end{cases} \quad \forall p \in \mathcal{P}, r \in \mathcal{R}_p, \quad (53)$$

where v_p^r is drawn uniformly from $\mathcal{U}[0,1]$. Equation (53) indicates every path has a random utility v_p^r normalized to $0 \sim 1$. The chosen path has an additional utility value of one. We assume that the impact of the recommendation of r' on the utility of path r is

$$I_{p,r,r'}^r = \begin{cases} \text{Drawn from } \mathcal{U}[0,5] & \text{if } r = r' \\ 0 & \text{otherwise,} \end{cases} \quad \forall p \in \mathcal{P}, r, r' \in \mathcal{R}_p. \quad (54)$$

Equation (54) means that the utility of the path recommended (i.e., $r = r'$) has an additional positive impact

drawn uniformly from $\mathcal{U}[0,5]$. The utilities of paths not being recommended ($r \neq r'$) do not change. Given (53) and (54), we can generate the conditional probability π using (5). It is worth mentioning that the above assumptions for generating synthetic passenger prior preferences are based on two reasonable principles: (1) The passenger’s actual chosen paths have a higher inherent utility. (2) Recommendations of a path can increase its probability of being chosen.

6.2. Parameter Settings

The convergence gap threshold for Benders decomposition is set as 1×10^{-8} . The ϵ -feasibility and Γ -concentration parameters are set with different values to test the performance. Specifically, we set $\epsilon_t^{u,v,r} = \epsilon \cdot q_t^{u,v,r}$ for $\epsilon \in \{0, 0.01, 0.03, 0.05, 0.1\}$, indicating different levels of deviations on path flows. For Γ -concentration, we implement the modified constraint in (43), and set $\sum_{r \in \mathcal{R}^{u,v}} (\Gamma_t^{u,v,r})^2 = \Gamma \cdot ((\Gamma_t^{u,v, \text{Max}})^2 - (\Gamma_t^{u,v, \text{Min}})^2) + (\Gamma_t^{u,v, \text{Min}})^2$ for $\Gamma \in \{0, 0.25, 0.5, 0.75, 1\}$. Note that $\Gamma = 0$ indicates that the model is forced to choose the lowest variance recommendations, whereas $\Gamma = 1$ means that there are no constraints on the variance.

6.3. Benchmark Models

There are two benchmark path choice scenarios we use for comparison purposes.

6.3.1. Status Quo Path Choices. This scenario provides the status quo situation which does not include any recommendations. It represents the worst case. In this scenario, no behavior uncertainty is considered because this is based on the actual path choices realized by passengers.

6.3.2. Capacity-Based Path Recommendations. The capacity-based path recommendations aim to recommend different paths to passengers according to the available capacity of paths. Specifically, for a path in OD pair (u, v) and time t , its capacity is the total available capacity of all vehicles passing through the first boarding station of the path during the time period. For example, for a path consisting of an NS bus route and the Green Line, the path capacity is the total available capacity of all buses at the boarding station of the NS bus route during time interval t . The available capacity can be obtained from a simulation model using historical demand as the input or using historical passenger count data. The available capacity for the Blue Line (the incident line) depends on modified operations during the incident (i.e., the service suspension is considered). When no vehicles operate on the Blue Line during time interval t , the path capacity is zero. Let the available capacity of the path (u, v, r) at time t be $C_t^{u,v,r}$. The set of all capacity-based recommendation strategies is defined as

$$\mathcal{X}^{\text{Cap}} = \left\{ x \in \mathcal{X} : \frac{\sum_{p \in \mathcal{P}_t^{u,v}} x_{p,r}}{|\mathcal{P}_t^{u,v}|} = \frac{C_t^{u,v,r}}{\sum_{r' \in \mathcal{R}^{u,v}} C_t^{u,v,r'}}, \forall (u, v, r) \in \mathcal{F}, t \in \mathcal{T} \right\}. \quad (55)$$

We randomly choose one strategy $x^{\text{Cap}} \in \mathcal{X}^{\text{Cap}}$ to implement, which is equivalent to randomly selecting passengers to receive path recommendations such that the proportion of passengers being recommended path r is the same as its available capacity proportion out of all paths in that OD pair (u, v) and time t .

6.4. System Travel Time Evaluation

Given a recommendation strategy x , as mentioned above, the actual system travel time is a random variable because of the passenger behavior uncertainty. To obtain the mean and standard deviation of the system travel time, we generate multiple passenger choice realizations based on π and x . For each generated passenger choice ($\hat{\mathbf{1}}_{p,r}^r$), the realized path flows are

$$\hat{q}_t^{u,v,r} = \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \hat{\mathbf{1}}_{p,r'}^r, \forall (u, v, r) \in \mathcal{F}, t \in \mathcal{T}. \quad (56)$$

The corresponding realized system travel time is $STT(\hat{q}) = \min_{z \in \mathcal{Z}(\hat{q})} WT(\hat{q}, z) + IVT(\hat{q}, z)$. This process is repeated with multiple realizations, providing the sample mean and standard deviation of the system travel time under recommendation strategy x .

6.5. Experiment Design

As this paper considers various components (such as individual path recommendations, passengers' path preferences, behavior uncertainty, etc.), it is useful to test different components separately to identify the impact of each one. Hence, we design the following test

cases, each one with specific parameter settings to systematically evaluate the impacts of each component.

6.5.1. Model Performance Compared to Benchmark Models. The most straightforward model validation is to evaluate the effect of reducing system travel time. In this test case, we set $\Psi = 0$, meaning that we ignore the passengers' preferences and focus only on minimizing system travel time. The results of this test case are discussed in Section 6.6.2 with different values of Γ and ϵ .

6.5.2. The Benefit of Considering Behavior Uncertainty.

In this test case, we evaluate the importance of incorporating behavior uncertainty in the model. The model without behavior uncertainty assumes that passengers take the recommended path. The recommendation strategy is obtained by solving (9) with $\pi_{p,r'}^r = 1$ if $r = r'$. Similarly, we set $\Psi = 0$. Note that, when we evaluate the recommendation strategy, the behavior uncertainty is still considered in generating the system travel time (see Section 6.4). The results of this test case are shown in Section 6.6.3.

6.5.3. Impact of Considering Passenger Preferences.

In all the above tests, $\Psi = 0$ is used, focusing on the system travel time. In this test case, we evaluate the model performance under different values of Ψ in order to assess the impact of considering passenger preferences. The results of this test case are discussed in Section 6.6.4.

6.6. Numerical Results

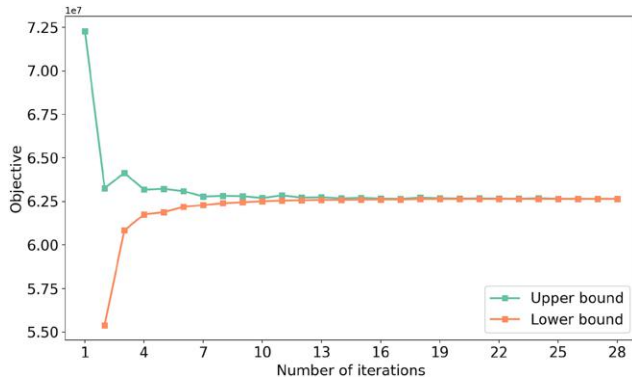
In this section, we summarize the experimental results of the proposed model compared with other benchmarks. The computational time and system travel time comparison are shown in Sections 6.6.1 and 6.6.2. Sections 6.6.3 and 6.6.4 discuss the impact of considering behavior uncertainties and passenger preferences.

6.6.1. Model Convergence and Computational Time.

Figure 7 shows the convergence of the BD algorithm. As expected, the lower bound of the model keeps increasing, whereas the upper bound, after dropping significantly in early iterations, exhibits some fluctuations. The model converges after 28 iterations with a relative gap of less than 1×10^{-8} . The number of optimality cuts was 28, and no feasibility cut was generated.

Table 1 compares the computational time of the Benders decomposition and off-the-shelf solvers. The BD algorithm was implemented using Julia 1.6 with the Gurobi 9.1 solver (Gurobi Optimization, LLC 2021) on a personal computer with the I9-9900K CPU. The total computational time is 17.8 seconds (master problem 8.2 seconds + subproblem 9.6 seconds), which is more efficient than directly using the mixed integer programming

Figure 7. (Color online) Convergence of the Benders Decomposition



(MIP) solvers, including Gurobi (Gurobi Optimization, LLC 2021), CPLEX (International Business Machines Corporation 2009), GLPK (GNU linear programming kit) (Makhorin 2008), and CBC (coin or branch-and-cut) (Forrest and Lougee-Heimer 2005).

6.6.2. Model Performance Compared with Benchmark Models. In this section, we compare the system travel times under the proposed individual path recommendations (without postadjustment) and two benchmark models. All travel times (except for the status quo that is deterministic) are calculated based on 10 replications using the randomly sampled actual path choices based on the given recommendation (see Section 6.4). We only chose 10 replications because the results are relatively stable and the standard deviations (std.) are small.

Table 2 shows that the proposed model (IPR) significantly reduces the average travel time (ATT) in the system compared with the status quo for most of the hyperparameter settings. The best model with $\epsilon = 0.0, \Gamma = 1.0$ has a 6.6% reduction in travel times of all passengers in the system, and for passengers in the incident line (i.e., passengers who received the recommendation, \mathcal{P}), the average travel time reduction is 19.2%. Our model also outperforms the capacity-based benchmark path recommendation strategy, which reduces the travel time of all passengers by 2.5% and incident line passengers by 15.9%. We also compare individual-level travel time savings. Results are elaborated in Online Appendix J.

Table 1. Computational Time Comparison

Solver	CPU time (sec)	Gap
BD	17.8	0.000%
CPLEX	65.7	0.000%
GLPK	562.6	0.000%
Gurobi	55.1	0.000%
CBC	425.4	0.000%

Comparing different parameter settings, with the decreasing values in Γ , the model has a tighter constraint on the variance. We also observe decreasing trends in the standard deviations. However, the lower value in Γ will make the recommendation strategies more restricted. Hence, the quality of the recommendations is worse with higher system travel times (for all passengers and incident line passengers). Specifically, when $\Gamma = 0$, the model is forced to choose the smallest variance recommendations, resulting in significantly bad solutions (even worse than the capacity-based recommendations). In terms of ϵ , as we analyzed in Section 5.1, the best value should be zero because it provides a better approximation gap and does not impose additional constraints on the model. The experiment results also validate this under various values of Γ .

6.6.3. Benefits of Considering Behavior Uncertainty. In this section, we aim to compare the model with and without considering the behavior uncertainty. The model without behavior uncertainty assumes that all passengers follow the recommended path when designing the recommendation (but they may not in reality).

Table 3 shows the comparison of average travel times for the two models. The IPR model with BU uses $\epsilon = 0, \Gamma = 1$ as they are the best hyperparameters. As expected, considering behavior uncertainty in the path recommendation design achieves a smaller travel time for all passengers and incident line passengers. Note that, though the 0.94% reduction (around 15 seconds saved per passenger) is relatively small, considering the large number of passengers in the system, the total travel time savings are still significant.

6.6.4. Impact of Respecting Passengers' Prior Preferences. In this section, we evaluate the impact of different values of Ψ in terms of respecting passengers' prior preferences. Besides the system travel time, we also evaluate the total utility, defined as the sum of the prior utilities of the recommended path:

$$TU(x) = \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} x_{p,r} \cdot V_{p,r}. \quad (57)$$

Note that the maximum value of $TU(x)$ is achieved when every passenger is recommended their preferred path (i.e., the path with the highest prior utility, $V_{p,r}$). Denote this maximum value as TU^{\max} . The relative ratio of total utility, $\frac{TU(x)}{TU^{\max}}$, represents the fraction of the total (prior) utility that the recommendation has achieved.

Another indicator is the number of passengers recommended their preferred path (denoted as $NP(x)$). Similarly, we also define the proportion of passengers

Table 2. ATT Comparison for Different Models

Models	ATT (all passengers)		ATT (incident line passengers)	
	Mean (min)	Std. (min)	Mean (min)	Std. (min)
Status quo	28.318	N.A.	40.255	N.A.
Capacity-based	27.609 (−2.5%)	0.033	33.848 (−15.9%)	0.165
IPR model ($\epsilon = 0.0, \Gamma = 1.0$)	26.454 (−6.6%)	0.019	32.526 (−19.2%)	0.204
IPR model ($\epsilon = 0.0, \Gamma = 0.75$)	26.461 (−6.5%)	0.019	32.563 (−19.1%)	0.183
IPR model ($\epsilon = 0.0, \Gamma = 0.5$)	26.492 (−6.4%)	0.019	32.776 (−18.6%)	0.167
IPR model ($\epsilon = 0.0, \Gamma = 0.25$)	26.649 (−5.9%)	0.017	33.768 (−16.1%)	0.145
IPR model ($\epsilon = 0.0, \Gamma = 0.0$)	28.315 (−0.0%)	0.015	39.950 (−0.76%)	0.126
IPR model ($\epsilon = 0.01, \Gamma = 1.0$)	26.455 (−6.6%)	0.019	32.529 (−19.2%)	0.199
IPR model ($\epsilon = 0.03, \Gamma = 1.0$)	26.459 (−6.6%)	0.018	32.539 (−19.2%)	0.191
IPR model ($\epsilon = 0.05, \Gamma = 1.0$)	26.460 (−6.6%)	0.017	32.555 (−19.1%)	0.187
IPR model ($\epsilon = 0.10, \Gamma = 1.0$)	26.467 (−6.5%)	0.017	32.574 (−19.1%)	0.178
IPR model ($\epsilon = 0.01, \Gamma = 0.5$)	26.493 (−6.4%)	0.019	32.779 (−18.6%)	0.165
IPR model ($\epsilon = 0.03, \Gamma = 0.5$)	26.495 (−6.4%)	0.018	32.786 (−18.6%)	0.163
IPR model ($\epsilon = 0.05, \Gamma = 0.5$)	26.498 (−6.4%)	0.017	32.791 (−18.5%)	0.147
IPR model ($\epsilon = 0.10, \Gamma = 0.5$)	26.504 (−6.4%)	0.017	32.805 (−18.5%)	0.128
IPR model ($\epsilon = 0.01, \Gamma = 0.25$)	26.653 (−5.9%)	0.017	33.771 (−16.1%)	0.147
IPR model ($\epsilon = 0.03, \Gamma = 0.25$)	26.655 (−5.9%)	0.017	33.801 (−16.0%)	0.144
IPR model ($\epsilon = 0.05, \Gamma = 0.25$)	26.677 (−5.8%)	0.017	33.812 (−16.0%)	0.137
IPR model ($\epsilon = 0.10, \Gamma = 0.25$)	26.691 (−5.7%)	0.017	33.830 (−16.0%)	0.141

Notes. Numbers in parentheses represent percentage travel time reduction compared with the status quo. The best model results are bolded.

recommended their preferred path (i.e., $\frac{NP(x)}{|\mathcal{P}|}$, where $|\mathcal{P}| = 5,827$ in the case study).

Figure 8 shows the results for different values of Ψ . All models use $\epsilon = 0, \Gamma = 1$. The x -axis is plotted in a log scale. In Figure 8(a), the average travel time for all passengers and incident line passengers increases with the increase of Ψ , which is as expected because the larger value of Ψ means that the recommendation generation focuses more on satisfying the passenger's inherent preferences than on minimizing the system travel time. Similarly, in Figure 8(b), as expected, both $TU(x)$ and $NP(x)$ increase with the increase in Ψ . When $\Psi = 10^5$, the average travel time of the incident line passengers increases by 21.3%, which is close to the status quo scenario. This is because we generate passengers' prior utilities based on the status quo choices. Figure 8(b) shows that nearly all passengers in \mathcal{P} are recommended their preferred path when $\Psi = 10^5$.

Figure 8 illustrates the trade-off between respecting passengers' preferences and reducing system congestion. When the value of Ψ is relatively small (e.g., less than 10^3), increasing Ψ can effectively increase the total utility and the number of passengers recommended

their preferred path. Meanwhile, the system travel time only slightly increases. However, when Ψ is large (e.g., greater than 10^4), increasing Ψ significantly increases the system travel time, but the impact on increasing the passenger's utility is limited. The reason may be that, in the system, there are some passengers whose preferred paths are not at the capacity bottlenecks. Hence, when Ψ is small, the optimal solution recommends those passengers use their preferred paths without significantly impacting the system travel time. When Ψ is large, passengers are recommended to use their preferred paths even if these paths are highly congested, causing a significant increase in the system travel time. The results imply that a reasonable value of Ψ should be relatively small. With small Ψ , most of the passengers (e.g., more than 70%) are recommended to use their preferred paths without significantly reducing the system efficiency.

6.7. Impact of Different Behavior Uncertainty Scenarios

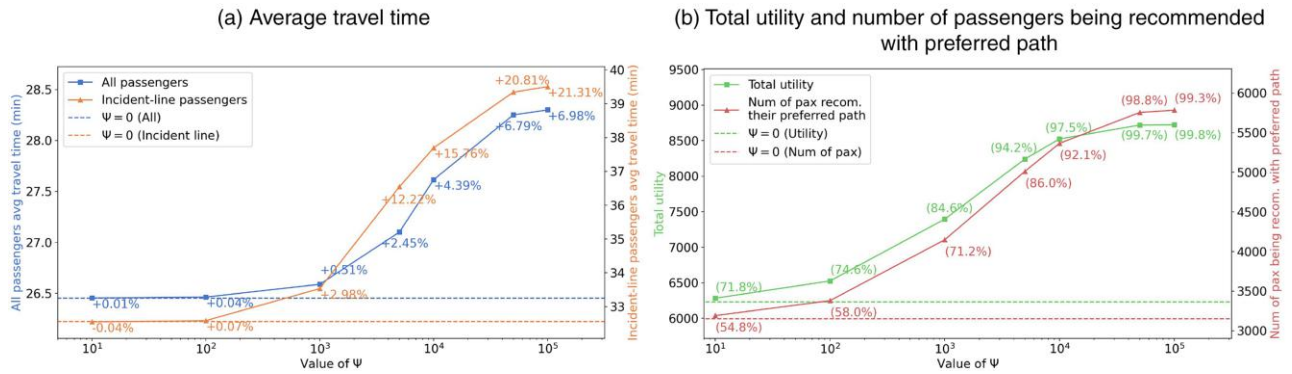
The model's performance may vary according to the level of passengers' uncertainties. In this section, we test different scenarios on the passenger's conditional

Table 3. ATT Comparison with and Without BU

Models	ATT (all passengers)		ATT (incident line passengers)	
	Mean (min)	Std. (min)	Mean (min)	Std. (min)
IPR model (w.o. BU)	26.706	0.026	32.852	0.122
IPR model (w. BU)	26.454 (−0.94%)	0.019	32.526 (−1.0%)	0.204

Note. Numbers in parentheses represent percentage travel time reduction compared with the IPR model without (w.o.) BU.

Figure 8. (Color online) Impact of Different Values of Ψ on Results



Notes. The percentage change in panel (a) is compared with the scenario of $\Psi = 0$. The percentage in parentheses in panel (b) represents the relative ratio of total utility and proportion of passengers recommended their preferred path, respectively. pax., passengers; recom., recommended.

path choice probabilities. We modify the impact of the recommendation of path r' on the utility of path r as

$$I_{p,r,r'} = \begin{cases} \text{Drawn from } \mathcal{U}[I^{\text{Min}}, I^{\text{Max}}] & \text{if } r = r' \\ 0 & \text{otherwise,} \end{cases} \quad \forall p \in \mathcal{P}, r, r' \in \mathcal{R}_p. \quad (58)$$

Besides our previous testing of $I^{\text{Min}} = 0, I^{\text{Max}} = 5$, we added the following additional scenarios:

- Smaller recommendation impact: passengers are less affected by the recommendation of the system. We consider $I^{\text{Min}} = 0, I^{\text{Max}} = 3$ and $I^{\text{Min}} = 0, I^{\text{Max}} = 1$.
- Random recommendation impact: passengers may be positively or negatively affected by the recommendation. We consider $I^{\text{Min}} = -5, I^{\text{Max}} = 5, I^{\text{Min}} = -3, I^{\text{Max}} = 3$, and $I^{\text{Min}} = -1, I^{\text{Max}} = 1$.

Results are shown in Table 4. The IPR model with BU uses $\epsilon = 0, \Gamma = 1$. Each number is the average of 10 replications for travel time evaluations. The results highlight that incorporating behavior uncertainty in the IPR model leads to significant improvements in travel time performance, especially under scenarios with larger behavior uncertainties (e.g., $I^{\text{Min}} = -5, I^{\text{Max}} = 5$). The average travel time for incident line passengers decreases from 36.67 minutes to 33.58 minutes (−8.4%). The model is robust in handling both positive and negative recommendation influences.

6.8. Impact of Incident Durations

In this section, we test the model’s performance under different incident durations. Note that the actual incident duration is 59 minutes. We assume the demand patterns are the same for all incident scenarios. Results

Table 4. Impact of Different Behavior Uncertainty Scenarios

Scenario [$I^{\text{Min}}, I^{\text{Max}}$]	Model	ATT (all passengers)		ATT (incident line passengers)	
		Mean (min)	Std. (min)	Mean (min)	Std. (min)
[0, 5]	Capacity-based	27.609	0.033	33.848	0.165
	IPR (w.o. BU)	26.706	0.026	32.852	0.122
	IPR (w. BU)	26.454	0.019	32.526	0.204
[0, 3]	Capacity-based	27.777	0.034	34.927	0.181
	IPR (w.o. BU)	27.177	0.031	34.475	0.131
	IPR (w. BU)	26.771	0.020	33.843	0.211
[0, 1]	Capacity-based	28.202	0.037	37.037	0.201
	IPR (w.o. BU)	28.049	0.034	36.978	0.136
	IPR (w. BU)	27.535	0.022	36.202	0.215
[−5, 5]	Capacity-based	28.252	0.044	36.667	0.221
	IPR (w.o. BU)	28.001	0.038	36.176	0.142
	IPR (w. BU)	26.714	0.026	33.578	0.210
[−3, 3]	Capacity-based	28.324	0.045	37.236	0.217
	IPR (w.o. BU)	28.123	0.041	36.807	0.139
	IPR (w. BU)	27.018	0.028	34.931	0.214
[−1, 1]	Capacity-based	28.418	0.047	37.864	0.214
	IPR (w.o. BU)	28.394	0.042	37.761	0.137
	IPR (w. BU)	27.864	0.030	36.857	0.222

Note. The best model results are bolded.

are shown in Table 5. The IPR model with BU uses $\epsilon = 0, \Gamma = 1$.

We find that, across all durations analyzed (30, 59, 90, and 120 minutes), the IPR model considering behavior uncertainty consistently yields the lowest ATT, highlighting its effectiveness in mitigating the adverse effects of incidents compared with the other approaches. Moreover, the effect is more prominent in cases with longer incidents.

7. Synthetic Case Study

In the synthetic case study, we focus on testing the model's generalizability to different network scales and incident situations.

7.1. Case Study Design

Consider a general three-line synthetic network (Figure 9). Each line has N stations, numbered from 1 to N . For any stations $n \in [N]$ on rail line 1 (where $[N] = \{1, 2, \dots, N\}$), people can transfer to rail line 2 or 3 by walking, providing them alternatives during disruptions. For simplicity, we assume station 1 on rail line 1 is the only destination people are willing to go to (e.g., the downtown area for the morning commute), and only rail line 1 has demand. For all OD pairs $(n, 1)$, $\forall n \in [N] \setminus \{1\}$, the demand $d_i^{n,1}$ is assumed to be the same (i.e., uniform demand). We assume there is an incident happening between stations $\lceil N/2 \rceil + 1$ and $\lceil N/2 \rceil$ on rail line 1, lasting for one hour (the location of the incident will vary in Section 7.4), and there is a shuttle bus serving between stations n and 1 during the disruptions. All individuals have four options: (1) waiting at rail line 1 until it recovers, (2) transferring to rail line 2, (3) transferring to rail line 3, and (4) transferring to the shuttle bus.

The uniform demand is set as 25 passengers per OD pair per hour. The vehicle capacities for rail lines 1, 2, and 3 and shuttle buses are 500, 300, 300, and 40,

respectively. The headway under normal operation times for rail lines 1, 2, and 3 is 10 minutes, 12 minutes, and 13 minutes, respectively. The headway for the shuttle bus during the disruption is 8 minutes. The transfer walking time between rail lines is 10 minutes, and from the rail line to the shuttle bus is three minutes. The vehicle travel time between any two stations n and $n - 1$ for rail lines 1, 2, and 3 is five minutes, seven minutes, and eight minutes, respectively, and for the shuttle bus is 10 minutes. These settings generally tell us that rail line 1 has the most carrying capacity, followed by rail lines 2 and 3, and shuttle buses. The path recommendation hyperparameters are set as $\epsilon = 0, \Gamma = 1, \Psi = 0$. Other assumptions and setups are the same as in the actual case study. In the following testing, we vary $N = 2$ (i.e., two-station network) to $N = 20$ to test the model performance extensively.

7.2. Computational Time

The computational results are summarized in Table 6. The relative performances of BD and the Gurobi solver vary across network sizes. For small-scale scenarios (e.g., $N = 2$ to $N = 10$), Gurobi consistently outperforms BD with lower CPU times. However, as the network size increases beyond $N = 12$, the efficiency of BD becomes more apparent. Whereas Gurobi's computational time grows rapidly with increasing problem size—reaching over 450 seconds at $N = 20$ —BD demonstrates more moderate growth, maintaining a substantial advantage in large-scale instances. This indicates that BD scales more efficiently and is better suited for solving large-scale IPR problems.

7.3. Performance

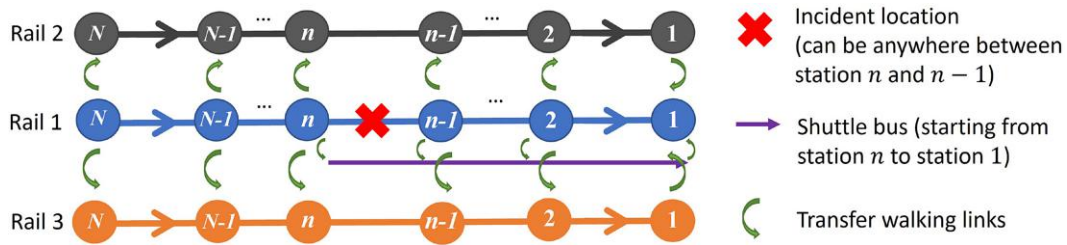
The results in Table 7 demonstrate that the proposed IPR model consistently outperforms the capacity-based method in terms of ATT across all tested network sizes. Specifically, the IPR model achieves noticeable reductions in ATT, ranging from 13.3% at $N = 4$ to as much as

Table 5. Impact of Different Incident Durations

Incident duration	Model	ATT (all passengers)		ATT (incident line passengers)	
		Mean (min)	Std. (min)	Mean (min)	Std. (min)
30 min	Capacity-based	26.038	0.031	30.538	0.155
	IPR (w.o. BU)	24.976	0.024	29.753	0.115
	IPR (w. BU)	24.802	0.017	29.539	0.178
59 min (actual)	Capacity-based	27.609	0.033	33.848	0.165
	IPR (w.o. BU)	26.706	0.026	32.852	0.122
	IPR (w. BU)	26.454	0.019	32.526	0.204
90 min	Capacity-based	29.643	0.037	38.770	0.170
	IPR (w.o. BU)	28.721	0.028	36.953	0.124
	IPR (w. BU)	28.431	0.021	36.204	0.211
120 min	Capacity-based	32.714	0.040	44.476	0.182
	IPR (w.o. BU)	30.746	0.031	41.473	0.131
	IPR (w. BU)	30.325	0.025	40.588	0.228

Note. The best model results are bolded.

Figure 9. (Color online) Synthetic Case Study Design



15.0% at $N = 2$. However, as the network scale increases (e.g., from $N = 12$ to $N = 20$), the relative improvement gradually diminishes, with the reduction dropping to just 1.8% at $N = 20$. This trend can be attributed to increasing demand and tighter remaining capacity in larger-scale scenarios. Under such conditions, the flexibility of the IPR model becomes more limited, as fewer excess resources are available to reassign passengers optimally. Hence, although the IPR model maintains better performance, its advantage over the capacity-based approach narrows as the system becomes more saturated.

7.4. Impact of Incident Location

To evaluate the impact of the incident location, we choose scenarios with $N \in \{8, 12, 16, 20\}$. For each scenario, we consider three different incident locations: Upstream (incident between stations $\lceil 3N/4 \rceil + 1$ and $\lceil 3N/4 \rceil$), Middle (incident between stations $\lceil N/2 \rceil + 1$ and $\lceil N/2 \rceil$, i.e., same as the above), and Downstream (incident between stations $\lceil N/4 \rceil + 1$ and $\lceil N/4 \rceil$).

The results in Table 8 indicate that incident location has no significant impact on the average travel time under either the capacity-based method or the proposed IPR model. Across all scenarios ($N = 8, 12, 16, 20$) and for all three incident locations (Upstream, Middle, Downstream), the average travel time values remain relatively consistent, with only minor variations. This outcome is expected, as the model assumes that a service disruption completely blocks the entire line regardless of where the incident occurs (an assumption that reflects common real-world operational protocols). The only difference introduced by varying incident locations lies in how the shuttle bus service is designed. For Upstream incidents, the shuttle bus coverage tends to be longer, which can slightly reduce the total system travel time. However, because of the limited capacity of shuttle buses, their influence on the overall passenger travel time remains marginal.

Table 6. Computational Time Comparison on Synthetic Networks

Scenario	Solver	CPU time (sec)
$N = 2$	BD	3.18
	Gurobi	0.51
$N = 4$	BD	3.66
	Gurobi	0.70
$N = 6$	BD	4.12
	Gurobi	0.95
$N = 8$	BD	5.97
	Gurobi	1.14
$N = 10$	BD	8.18
	Gurobi	2.01
$N = 12$	BD	12.61
	Gurobi	11.55
$N = 14$	BD	16.60
	Gurobi	35.62
$N = 16$	BD	23.45
	Gurobi	80.79
$N = 18$	BD	76.52
	Gurobi	231.83
$N = 20$	BD	134.42
	Gurobi	451.32

Table 7. ATT Comparison for Different Network Scales

Scenario	Models	ATT (all passengers)	
		Mean (min)	Std. (min)
$N = 2$	Capacity-based	12.67	0.101
	IPR	10.77 (-15.0%)	0.064
$N = 4$	Capacity-based	20.91	0.142
	IPR	18.14 (-13.3%)	0.077
$N = 6$	Capacity-based	29.64	0.174
	IPR	26.95 (-9.1%)	0.083
$N = 8$	Capacity-based	39.79	0.205
	IPR	35.59 (-10.6%)	0.104
$N = 10$	Capacity-based	51.40	0.238
	IPR	46.39 (-9.7%)	0.121
$N = 12$	Capacity-based	65.81	0.266
	IPR	60.02 (-8.8%)	0.145
$N = 14$	Capacity-based	76.76	0.301
	IPR	72.26 (-5.9%)	0.169
$N = 16$	Capacity-based	86.87	0.331
	IPR	83.48 (-3.9%)	0.203
$N = 18$	Capacity-based	93.28	0.379
	IPR	89.72 (-3.8%)	0.222
$N = 20$	Capacity-based	96.93	0.403
	IPR	95.15 (-1.8%)	0.276

Note. Numbers in parentheses represent percentage travel time reduction compared with the capacity-based method.

Table 8. ATT Comparison Under Different Incident Locations

Scenario	Location	Models	ATT (all passengers)	
			Mean (min)	Std. (min)
N = 8	Upstream	Capacity-based	39.55	0.207
		IPR	35.43 (−10.4%)	0.110
	Middle	Capacity-based	39.79	0.205
		IPR	35.59 (−10.6%)	0.104
	Downstream	Capacity-based	39.91	0.203
		IPR	35.73 (−10.5%)	0.106
N = 12	Upstream	Capacity-based	65.67	0.263
		IPR	59.84 (−8.9%)	0.144
	Middle	Capacity-based	65.81	0.266
		IPR	60.02 (−8.8%)	0.145
	Downstream	Capacity-based	65.93	0.266
		IPR	60.52 (−8.2%)	0.144
N = 16	Upstream	Capacity-based	86.75	0.328
		IPR	83.29 (−4.0%)	0.201
	Middle	Capacity-based	86.87	0.331
		IPR	83.48 (−3.9%)	0.203
	Downstream	Capacity-based	86.91	0.332
		IPR	83.62 (−3.8%)	0.204
N = 20	Upstream	Capacity-based	96.74	0.398
		IPR	95.11 (−1.7%)	0.275
	Middle	Capacity-based	96.93	0.403
		IPR	95.15 (−1.8%)	0.276
	Downstream	Capacity-based	97.08	0.406
		IPR	95.22 (−1.9%)	0.273

Note. Numbers in parentheses represent percentage travel time reduction compared with the capacity-based method.

8. Conclusion and Discussion

This study proposes an individual path recommendation model during PT service disruptions with the objective of minimizing total system travel time and respecting passengers' path choice preferences. Passengers' behavior uncertainty in path choices given recommendations is also considered in the formulation. The original IPR formulation yields a stochastic optimization with decision-dependent distributions. We propose a single-point approximation method to eliminate the expectation operator by introducing two new concepts: ϵ -feasibility and Γ -concentration. The approximation yields a tractable single-stage mixed integer linear formulation, which can be solved efficiently with Benders decomposition. The approximation gap is proved to be bounded from above. Additional theoretical analysis shows that ϵ -feasibility and Γ -concentration are strongly connected to expectation and chance constraints in a typical stochastic optimization formulation, respectively.

The proposed approach is demonstrated in a real-world case study using data from an urban rail disruption in the CTA system and a synthetic case study with varied network sizes and incident locations. In the real-world case study, results show that the proposed IPR model reduces the average travel times in the system by 6.6% compared with the status quo and by 4.2% compared with a capacity-based benchmark model. In

the synthetic case study, the proposed model shows 15.0%–1.8% lower system travel time compared with the capacity-based method, depending on the network sizes and demand situations.

Future studies can be pursued in the following directions. First, it is possible to extend the current framework with more complex recommendation compositions. The challenges in implementing the more general framework stem from the quantification of the posterior path choice probabilities. Future studies may conduct corresponding surveys to calibrate passengers' responses to the recommendations. Second, future studies may consider different sources of uncertainty (including incident duration, in-vehicle time, demand, etc.) for a more realistic modeling framework. Third, future studies could extend beyond individual path recommendations to a codesign framework that integrates personalized routing with service adjustments, such as bus rerouting and shuttle deployment. This approach will optimize both passenger travel experience and operational efficiency by balancing travel time, accessibility, and cost considerations.

Acknowledgments

The authors thank the Chicago Transit Authority for their support and data availability for this research.

Endnote

¹ The approximation is then updated based on the second-stage solutions.

References

- Abdelgawad H, Abdulhai B (2012) Large-scale evacuation using subway and bus transit: Approach and application in city of Toronto. *J. Transportation Engrg.* 138(10):1215–1232.
- Bardou O, Boutheimy S, Pagès G (2009) Optimal quantization for the pricing of swing options. *Appl. Math. Finance* 16(2):183–217.
- Bartl D, Mendelson S (2022) On Monte-Carlo methods in convex stochastic optimization. *Ann. Appl. Probab.* 32(4):3146–3198.
- Ben-Akiva ME, Lerman SR (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press Series in Transportation Studies, vol. 9 (MIT Press, Cambridge, MA).
- Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numer. Math.* 4(1):238–252.
- Bertsimas D, Mundru N (2023) Optimization-based scenario reduction for data-driven two-stage stochastic optimization. *Oper. Res.* 71(4):1343–1361.
- Bertsimas D, Sian Ng Y, Yan J (2020) Joint frequency-setting and pricing optimization on multimodal transit networks at scale. *Transportation Sci.* 54(3):839–853.
- Bhumika, Das D (2022) MARRS: A framework for multi-objective risk-aware route recommendation using multitask-transformer. Golbeck J, Harper FM, Murdock V, Ekstrand M, Shapira B, Basilico J, Lundgaard K, Oldridge E, eds. *RecSys '22 Proc. 16th ACM Conf. Recommender Systems* (Association for Computing Machinery, New York), 360–368.
- Böhmová K, Mihalák M, Pröger T, Srámek R, Widmayer P (2013) Robust routing in urban public transportation: How to find reliable journeys based on past observations. Frigioni D, Stiller S, eds. *ATMOS 13th Workshop Algorithmic Approaches Transportation Model. Optim. Systems* (Schloss Dagstuhl—Leibniz-Zentrum fuer Informatik, Wadern, Germany), 27–41.

- Bruglieri M, Bruschi F, Colorni A, Luè A, Nocerino R, Rana V (2015) A real-time information system for public transport in case of delays and service disruptions. *Transportation Res. Procedia* 10:493–502.
- Chen D, Ong CS, Xie L (2016) Learning points and routes to recommend trajectories. *Proc. 25th ACM Internat. Conf. Inform. Knowledge Management* (Association for Computing Machinery, New York), 2227–2232.
- Chen C, Li L, Li M, Li R, Wang Z, Wu F, Xiang C (2022) cuRL: A generic framework for bi-criteria optimum path-finding based on deep reinforcement learning. *IEEE Trans. Intelligent Transportation Systems* 24(2):1949–1961.
- Cheng Y, Zhou Q, Lai Y (2019) TaxiC: A taxi route recommendation method based on urban traffic charge heat map. Liu X, Mrissa M, Zhang L, Benslimane D, Ghose A, Wang Z, Bucchiarone A, Zhang W, Zou Y, Yu Q, eds. *Service Oriented Comput. ICSOC 2018 Workshops* (Springer International Publishing, Cham, Switzerland), 322–334.
- Cox A, Prager F, Rose A (2011) Transportation security and the role of resilience: A foundation for operational metrics. *Transport Policy* 18(2):307–317.
- Dai R, Liu J, Xu X (2024) Guidance information release strategy under disruptions in urban rail transit. Meng L, Qian Y, Bai Y, Lv B, Tang Y, eds. *Proc. 11th Internat. Conf. Traffic Transportation Stud.* (Springer, Singapore), 399–407.
- Dai J, Yang B, Guo C, Ding Z (2015) Personalized route recommendation using big trajectory data. *2015 IEEE 31st Internat. Conf. Data Engrg.* (IEEE, New York), 543–554.
- Danaf M, Becker F, Song X, Atasoy B, Ben-Akiva M (2019) Online discrete choice models: Applications in personalized recommendations. *Decision Support Systems* 119:35–45.
- Delling D, Goldberg AV, Pajor T, Werneck RF (2011) Customizable route planning. Pardalos PM, Rebennack S, eds. *Experiment. Algorithms SEA 2011* (Springer, Berlin), 376–387.
- Drusvyatskiy D, Xiao L (2023) Stochastic optimization with decision-dependent distributions. *Math. Oper. Res.* 48(2):954–998.
- Fang B, Yang Z, Wang S, Di X (2024) TravelLLM: Could you plan my new public transit route in face of a network disruption? Preprint, submitted July 20, <https://arxiv.org/abs/2407.14926>.
- Forrester J, Lougee-Heimer R (2005) CBC user guide. Smith JC, ed. *Emerging Theory, Methods, and Applications, Tutorials in Operations Research* (INFORMS, Catonsville, MD), 257–277.
- Geisberger R, Sanders P, Schultes D, Delling D (2008) Contraction hierarchies: Faster and simpler hierarchical routing in road networks. McGeoch CC, ed. *Experiment. Algorithms WEA 2008* (Springer, Berlin), 319–333.
- Goel V, Grossmann IE (2006) A class of stochastic programs with decision dependent uncertainty. *Math. Programming* 108(2): 355–394.
- Guo X, Caros NS, Zhao J (2021) Robust matching-integrated vehicle rebalancing in ride-hailing system with uncertain demand. *Transportation Res. Part B Methodological* 150:161–189.
- Guo X, Mo B, Koutsopoulos HN, Wang S, Zhao J (2024) Robust transit frequency setting problem with demand uncertainty. *IEEE Trans. Intelligent Transportation Systems* 25(10):13305–13322.
- Gurobi Optimization, LLC (2021) Gurobi optimizer reference manual. <https://www.gurobi.com>.
- Han X, Sun Q, Xing Y, Gao ZY, Zhang HM (2024) Reducing strategic uncertainty in transportation networks by personalized routing advice: A route-choice laboratory experiment. *Travel Behav. Soc.* 34:100701.
- Hellemo L, Barton PI, Tomasgard A (2018) Decision-dependent probabilities in stochastic programs with recourse. *Comput. Management Sci.* 15(3):369–395.
- Homem-de Mello T, Bayraksan G (2014) Monte Carlo sampling-based methods for stochastic optimization. *Surveys Oper. Res. Management Sci.* 19(1):56–85.
- Høyland K, Kaut M, Wallace SW (2003) A heuristic for moment-matching scenario generation. *Comput. Optim. Appl.* 24:169–185.
- Hu J, Li H (2019) A new clustering approach for scenario reduction in multi-stochastic variable programming. *IEEE Trans. Power Systems* 34(5):3813–3825.
- Huang F, Xu J, Weng J (2020) Multi-task travel route planning with a flexible deep learning framework. *IEEE Trans. Intelligent Transportation Systems* 22(7):3907–3918.
- International Business Machines Corporation (2009) *IBM ILOG CPLEX V12.1: User's Manual for CPLEX* (International Business Machines Corporation, Armonk, NY).
- Ji S, Wang Z, Li T, Zheng Y (2020) Spatio-temporal feature fusion for dynamic taxi route recommendation via deep reinforcement learning. *Knowledge Based Systems* 205:106302.
- Jung JY, Blau G, Pekny JF, Reklaitis GV, Eversdyk D (2004) A simulation based optimization approach to supply chain management under demand uncertainty. *Comput. Chemical Engrg.* 28(10):2087–2106.
- Karaman S, Frazzoli E (2011) Sampling-based algorithms for optimal motion planning. *Internat. J. Robotics Res.* 30(7):846–894.
- Keutchanyan J, Ortman J, Rei W (2023) Problem-driven scenario clustering in stochastic optimization. *Comput. Management Sci.* 20(1):13.
- Lai Y, Lv Z, Li KC, Liao M (2018) Urban traffic Coulomb's law: A new approach for taxi route recommendation. *IEEE Trans. Intelligent Transportation Systems* 20(8):3024–3037.
- Lim N, Hooi B, Ng SK, Wang X, Goh YL, Weng R, Tan R (2021) Origin-aware next destination recommendation with personalized preference attention. *Proc. 14th ACM Internat. Conf. Web Search Data Mining* (Association for Computing Machinery, New York), 382–390.
- Liu S, Jiang H, Chen S, Ye J, He R, Sun Z (2020) Integrating Dijkstra's algorithm into deep inverse reinforcement learning for food delivery route planning. *Transportation Res. Part E Logist. Transportation Rev.* 142:102070.
- Liu H, Tong Y, Zhang P, Lu X, Duan J, Xiong H (2019) Hydra: A personalized and context-aware multi-modal transportation recommendation system. *Proc. 25th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery, New York), 2314–2324.
- Mahmassani HS (1984) Uncertainty in transportation systems evaluation: Issues and approaches. *Transportation Planning Tech.* 9(1):1–12.
- Makhori A (2008) GLPK (GNU linear programming kit). Accessed March 25, 2024, <http://www.gnu.org/s/glpk/glpk.html>.
- Mehrotra S, Papp D (2013) Generating moment matching scenarios using optimization techniques. *SIAM J. Optim.* 23(2):963–999.
- Mendler-Dünner C, Perdomo J, Zmic T, Hardt M (2020) Stochastic optimization for performative prediction. Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds. *Adv. Neural Inform. Processing Systems 33 (NeurIPS 2020)* (Curran Associates, Inc., Red Hook, NY), 4929–4939.
- Mirchevska V (2013) Behavior modeling by combining machine learning and domain knowledge. PhD thesis, IPS Jožef Stefan, Ljubljana, Slovenia.
- Mo B (2022) Toward a resilient public transportation system: Effective monitoring and control under service disruptions. PhD thesis, Massachusetts Institute of Technology, Cambridge.
- Mo B, Koutsopoulos H, Zhao J (2022) Inferring passenger responses to urban rail disruptions using smart card data: A probabilistic framework. *Transportation Res. Part E Logist. Transportation Rev.* 159:102628.
- Mo B, Koutsopoulos HN, Shen ZJM, Zhao J (2023a) Robust path recommendations during public transit disruptions under demand uncertainty. *Transportation Res. Part B Methodological* 169:82–107.
- Mo B, Ma Z, Koutsopoulos HN, Zhao J (2020) Capacity-constrained network performance model for urban rail systems. *Transportation Res. Rec.* 2674(5):59–69.

- Mo B, Ma Z, Koutsopoulos HN, Zhao J (2023b) Ex post path choice estimation for urban rail systems using smart card data: An aggregated time-space hypernetwork approach. *Transportation Sci.* 57(2):313–335.
- Mo B, Von Franque MY, Koutsopoulos HN, Attanucci JP, Zhao J (2022) Impact of unplanned long-term service disruptions on urban public transit systems. *IEEE Open J. Intelligent Transportation Systems* 3:551–569.
- Mo B, Wang QY, Moody J, Shen Y, Zhao J (2021) Impacts of subjective evaluations and inertia from existing travel modes on adoption of autonomous mobility-on-demand. *Transportation Res. Part C Emerging Tech.* 130:103281.
- Mo B, Zheng Y, Guo X, Ma R, Zhao J (2025) Robust binary and multinomial logit models for classification with data uncertainties. *Eur. J. Oper. Res.* 327(2):577–591.
- Niederreiter H, Winterhof A (1992) Quasi-Monte Carlo methods. *Applied Number Theory* (Springer, Berlin), 185–306.
- Perdomo J, Zrníc T, Mandler-Dünner C, Hardt M (2020) Performance prediction. Daumé H III, Singh A, eds. *Proc. 37th Internat. Conf. Machine Learn.*, Proceedings of the Machine Learning Research, vol. 119 (PMLR, New York), 7599–7609.
- Qu B, Yang W, Cui G, Wang X (2019) Profitable taxi travel route recommendation based on big taxi trajectory data. *IEEE Trans. Intelligent Transportation Systems* 21(2):653–668.
- Qu M, Zhu H, Liu J, Liu G, Xiong H (2014) A cost-effective recommender system for taxi drivers. *Proc. 20th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery, New York), 45–54.
- Quijano-Sánchez L, Cantador I, Cortés-Cediel ME, Gil O (2020) Recommender systems for smart cities. *Inform. Systems* 92:101545.
- Reza RM, Ali ME, Cheema MA (2017) The optimal route and stops for a group of users in a road network. Hoel E, Newsam S, Ravada S, Tamassia R, Trajcevski G, eds. *Proc. 25th ACM SIGSPATIAL Internat. Conf. Adv. Geographic Inform. Systems* (Association for Computing Machinery, New York), 1–10.
- Roelofsen D, Cats O, van Oort N, Hoogendoorn S (2018) Assessing disruption management strategies in rail-bound urban public transport systems from a passenger perspective. *Proc. 14th Conf. Adv. Systems Public Transport (CASPT)* (Brisbane, Australia).
- Song X, Danaf M, Atasoy B, Ben-Akiva M (2018) Personalized menu optimization with preference updater: A Boston case study. *Transportation Res. Rec.* 2672(8):599–607.
- Stentz A (1994) Optimal and efficient path planning for partially-known environments. *Proc. 1994 IEEE Internat. Conf. Robotics Automation* (IEEE, New York), 3310–3317.
- Subramanyam A, Mufalli F, Láinez-Aguirre JM, Pinto JM, Gounaris CE (2021) Robust multiperiod vehicle routing under customer order uncertainty. *Oper. Res.* 69(1):30–60.
- Sundaram RK (1996) *A First Course in Optimization Theory* (Cambridge University Press, Cambridge, UK).
- Tan Z, Xu M, Meng Q, Li ZC (2020) Evacuating metro passengers via the urban bus system under uncertain disruption recovery time and heterogeneous risk-taking behaviour. *Transportation Res. Part C Emerging Tech.* 119:102761.
- Thiengburanathum P, Cang S, Yu H (2016) Overview of personalized travel recommendation systems. 2016 22nd Internat. Conf. Automation Comput. (ICAC) (IEEE, New York), 415–422.
- Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).
- Wang Q, Wallace SW (2022) Non-compliance in transit-based evacuation pick-up point assignments. *Socio-Econom. Planning Sci.* 82:101259.
- Wang S, Mo B, Zhao J (2020) Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. *Transportation Res. Part C Emerging Tech.* 112: 234–251.
- Wang J, Wu N, Zhao WX (2021) Personalized route recommendation with neural network enhanced search algorithm. *IEEE Trans. Knowledge Data Engng.* 34(12):5910–5924.
- Wang J, Yuan Z, Yin Y (2019) Optimization of bus bridging service under unexpected metro disruptions with dynamic passenger flows. *J. Adv. Transportation* 2019(1):6965728.
- Wang C, Li C, Huang H, Qiu J, Qu J, Yin L (2023) ASNN-FRR: A traffic-aware neural network for fastest route recommendation. *GeoInformatica* 27(1):39–60.
- Wang H, Li G, Hu H, Chen S, Shen B, Wu H, Li WS, Tan KL (2014) R3: A real-time route recommendation system. *Proc. VLDB Endowment* 7(13):1549–1552.
- Wong SD, Broader JC, Walker JL, Shaheen SA (2023) Understanding California wildfire evacuee behavior and joint choice making. *Transportation* 50(4):1165–1211.
- Wu F, Lyu C, Liu Y (2022) A personalized recommendation system for multi-modal transportation systems. *Multimodal Transportation* 1(2):100016.
- Xia W, Li Y, Guo W, Li S (2022) Efficient navigation for constrained shortest path with adaptive expansion control. 2022 IEEE Internat. Conf. Data Mining (ICDM) (IEEE, New York), 588–597.
- Yoon G, Chow JY (2020) Contextual bandit-based sequential transit route design under demand uncertainty. *Transportation Res. Rec.* 2674(5):613–625.
- Zhang S, Luo Z, Yang L, Teng F, Li T (2024) A survey of route recommendations: Methods, applications, and opportunities. Preprint, submitted March 1, <https://arxiv.org/abs/2403.00284>.
- Zhu X, Wang F, Chen C, Reed DD (2020) Personalized incentives for promoting sustainable travel behaviors. *Transportation Res. Part C Emerging Tech.* 113:314–331.

Individual Path Recommendation Under Public Transit Service Disruptions Considering Behavior Uncertainty - Online Appendix

Baichuan Mo^{1,2,*}, Haris N. Koutsopoulos³, Zuo-Jun Max Shen⁴, Jinhua Zhao⁵

* Corresponding author

¹Department of Civil Engineering, Tsinghua University, Beijing, China, 100084

²Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

³Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115

⁴Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, CA 94720

⁵Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 20139

Appendix A: Model extensions

In this section, we discuss several extensions of the model to accommodate more realistic/general scenarios.

A.1. Generalization of recommendations

In this study, we assume the information given to passengers is a recommended path. In reality, the recommendation system may provide a bundle of recommended paths with information like estimated in-vehicle time, waiting time, travel cost, etc. The proposed framework can be extended to handle different recommendation typologies. Figure 10 shows an example where the recommendation system will provide a composition of path and travel time information, where each composition can include different paths, different estimated waiting/in-vehicle times, etc. Then, we can change $x_{p,r}$ to $x_{p,c}$, where $x_{p,c}$ indicates whether we will present composition c to passenger p . Similarly, each c is associated with a conditional probability $\pi_{p,c}^r$ as shown in Figure 10 (the probability for passenger p to choose path r given that he/she is recommended composition c).

In this way, we only need to calibrate $\pi_{p,c}^r$ and predetermine the composition set \mathcal{C}_p for each passenger p . The overall framework proposed above can be easily adapted to the new recommendation typology by replacing $x_{p,r}$ and $\pi_{p,r'}^r$ with $x_{p,c}$ and $\pi_{p,c}^r$, respectively.

A.2. Feedback and rolling-horizon

As mentioned in Section (3.2), the whole path recommendation problem could be solved in a rolling-horizon manner. At each time interval $t \geq 1$, we update the demand, supply, and system state information, and solve the proposed framework above to get a recommendation strategy \mathbf{x} . But we only implement the $x_{p,r}$ for $p \in \mathcal{P}_t^{u,v}$, $\forall (u,v)$ (i.e., passengers departing at current time t). Specifically, the inputs can be updated as follows. The new supply information can be updated by querying the operators' database. The onboard

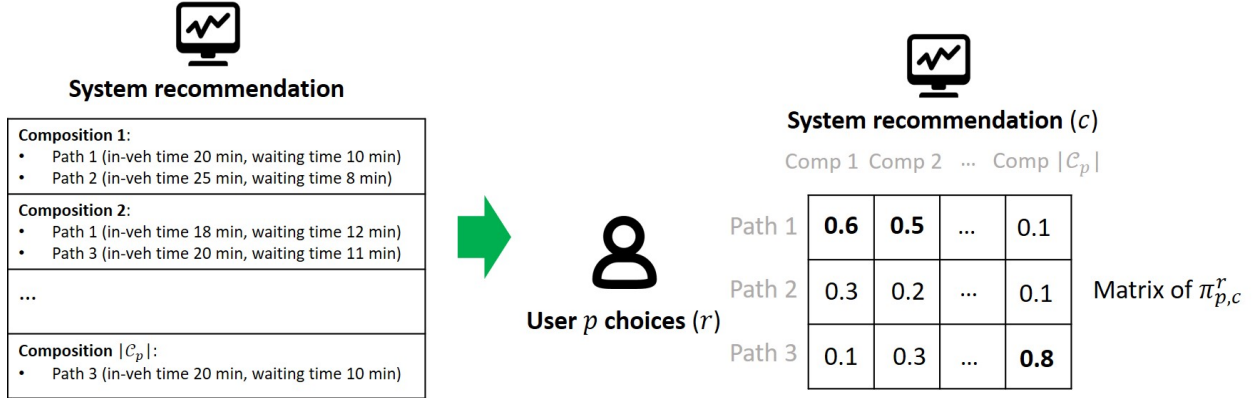


Figure 10 Illustration of the generalized recommendation typology. \mathcal{C}_p is the predetermined recommendation composition sets for passenger p

passengers Ω_1 can be updated based on new tap-in transactions and the simulation from the last time step. The predicted demand without recommendation $f_t^{u,v,r}$ can be updated by running the prediction model again with new smart card data transactions. We also need to construct the sets of passengers need recommendations in future time steps: $\mathcal{P}_{t+1}^{u,v}, \dots, \mathcal{P}_{T^D}^{u,v}$. Note that some passengers may already make recommendation requests even if they have not arrived at the system (i.e., part of the passengers in $\mathcal{P}_{t+1}^{u,v}, \dots, \mathcal{P}_{T^D}^{u,v}$ are already known). Then, we can add additional “predicted individuals” to match the predicted future demand of each OD pair (u, v) and time $t + h$ in the future with the associated preference matrix. This can be done by sampling passengers from the existing transit application database (usually these applications have passengers’ historical travel patterns). We assume all recommendations for passengers who arrived before t are fixed (even if they are not the optimal ones given the newly available information). This can be done by simply adding a new constraint $x_{p,r} = \hat{x}_{p,r}, \forall p \in \cup_{t'=1, \dots, t-1} \mathcal{P}_{t'}^{u,v}$, where $\hat{x}_{p,r}$ is the previous decision. But we can still modify the decisions for passengers in $\mathcal{P}_t^{u,v}, \mathcal{P}_{t+1}^{u,v}, \dots, \mathcal{P}_{T^D}^{u,v}$ as they have not arrived at the system by time t . The rolling horizon may also update the associated preference matrix information through passenger feedback when they have not arrived at the transit system. For example, after providing a recommendation, we can ask the passenger to respond whether he/she will actually use it or not. This feedback can be used to update the associated preference matrix.

Appendix B: Notations

Table 9: Notation summary

Notation	Description
<i>Model Parameters</i>	
(u, v, r, i)	The i -th leg of path r for OD pair (u, v)
t	Integer time index, $t = 1$ represents the start of the incident. Non-positive time indices indicate the time before the incident.
T^D	Time index at which the recommendation system stops working
t^{\min}	Start time index of the whole analysis period (negative by definition)
t^{end}	End time index of the incident
T	End time index of the whole analysis period (greater than T^D)

\mathcal{T}	The set of time indices of analysis and $\mathcal{T} = \{t^{\min}, t^{\min} + 1, \dots, T\}$
τ	The time duration that each time index represents
\mathcal{P}	The set of passengers that will receive the path recommendation
\mathcal{R}_p	The set of feasible paths for passenger $p \in \mathcal{P}$
$\mathcal{R}^{u,v}$	The set of feasible paths for OD pair (u, v)
V_p^r	Passenger p 's inherent preference (utility) of using path r
$I_{p,r'}^r$	The impact of the recommendation of path r' for passenger p on his/her utility of path r
$\pi_{p,r'}^r$	Conditional probability for passenger p to choose path r given that he/she is recommended with path r'
\mathcal{X}	The set of all feasible path recommendations \mathbf{x}
Ψ	The hyper-parameter to adjust the scale and balance the trade-off between system efficiency and passenger preference
$f_t^{u,v,r}$	Number of passengers with OD pair (u, v) and departure time t and using path r who are not provided path recommendations
$d_t^{u,v}$	Total number of passengers with OD pair (u, v) and departure time t
$\Delta_t^{u,v,r,i}$	Travel time between the terminal station and the boarding station of leg $((u, v, r, i))$ for vehicle departing at time t
$\delta_t^{u,v,r,i}$	Travel time between the terminal station and the alighting station of leg $((u, v, r, i))$ for vehicle departing at time t
Ω_1	The set of onboard flow indices at time $t = 1$
$\hat{z}_t^{u,v,r,i}$	Number of onboard passengers in the vehicle departing at time t in leg (u, v, r, i) when the incident happens
$\mathcal{P}_t^{u,v}$	The set of passengers with OD pair (u, v) and departure time t that will receive the path recommendation (a subset of \mathcal{P})
\mathcal{F}	The set of all (u, v, r) indices
\mathcal{S}	The set of all stations (stops)
\mathcal{W}	The set of all OD pairs
\mathcal{L}	The set of all transit lines (routes) in the system
$\mathcal{I}^{u,v,r}$	The set of legs for path r of OD pair (u, v)
Vehicle (l, t)	The vehicle departing at time t on line l
$T_{l,t}$	The time that vehicle (l, t) arrives the last station of line l
$O_{l,t,t'}$	Total number of onboard passengers at time t' for vehicle (l, t)
$K_{l,t}$	The capacity of vehicles (l, t)
$AD_{s,t}$	Cumulative arriving demand at station s up to time t
$XD_{s,t}$	Cumulative transferring demand at station s up to time t
$BD_{s,t}$	Cumulative boarded demand at station s up to time t
$\mu_t^{u,v,r}$	Expectation of $Q_t^{u,v,r}$
$\epsilon_t^{u,v,r}$	Threshold parameter for ϵ -feasibility
$\Gamma_t^{u,v,r}$	Threshold parameter for Γ -concentration
$\mathcal{D}(\mathbf{x})$	A feasible region of path flow \mathbf{q} that is determined by the recommendation \mathbf{x} such that the single-point approximation is reasonable
$\mathcal{Q}(\mathbf{x}^*)$	Support of the random path flow patterns $\mathcal{Q}(\mathbf{x}^*)$
$\Gamma_t^{u,v,\text{Min}}$	Lower bound of the sum of all $\Gamma_t^{u,v,r}$ over paths in $\mathcal{R}^{u,v}$
$\Gamma_t^{u,v,\text{Max}}$	Upper bound of the sum of all $\Gamma_t^{u,v,r}$ over paths in $\mathcal{R}^{u,v}$
$STT(\cdot)$	A pseudo function that returns the system travel time of the public transit network given a path flow pattern

Random Variables

$\mathbb{1}_{p,r}^{r'}$	Binary variable indicating whether passenger p has chosen route r or not
$Q_t^{u,v,r}$	Actual flow for path (u, v, r) at time t

$STT(\mathbf{Q})$	System travel time given path flow \mathbf{Q}
<i>Decision Variables for Optimization Models</i>	
$x_{p,r}$	Binary variable indicating whether recommending passenger p to use route r or not
$q_t^{u,v,r}$	Number of passengers in \mathcal{P} with OD pair (u, v) and departure time t and using path r
$z_t^{u,v,r,i}$	Number of passengers boarding a vehicle that had started at time t on leg (u, v, r, i)

Appendix C: Proof of Proposition 1

From the triangular inequality, we have:

$$\underbrace{|Q_t^{u,v,r} - q_t^{u,v,r}|}_{\text{LHS}} \leq |Q_t^{u,v,r} - \mu_t^{u,v,r}(\mathbf{x})| + |\mu_t^{u,v,r}(\mathbf{x}) - q_t^{u,v,r}|$$

$$\leq \underbrace{|Q_t^{u,v,r} - \mu_t^{u,v,r}(\mathbf{x})| + \epsilon_t^{u,v,r}}_{\text{RHS}}. \quad (59)$$

As $\text{LHS} \leq \text{RHS}$, the probability measure satisfies (for all $a > \epsilon_t^{u,v,r}$):

$$\mathbb{P}[\text{LHS} \geq a] \leq \mathbb{P}[\text{RHS} \geq a]. \quad (60)$$

Notice that

$$\mathbb{P}[\text{RHS} \geq a] = \mathbb{P}[|Q_t^{u,v,r} - \mu_t^{u,v,r}(\mathbf{x})| \geq a - \epsilon_t^{u,v,r}] \leq \frac{(\sigma_t^{u,v,r}(\mathbf{x}))^2}{(a - \epsilon_t^{u,v,r})^2}. \quad (61)$$

Equation (61) is based on Chebyshev's inequality. Therefore,

$$\mathbb{P}[\text{LHS} \geq a] = \mathbb{P}[|Q_t^{u,v,r} - q_t^{u,v,r}| \geq a] \leq \frac{(\sigma_t^{u,v,r}(\mathbf{x}))^2}{(a - \epsilon_t^{u,v,r})^2}. \quad (62)$$

Comparing (62) and (29), we know that to satisfy (29), we only need $\sigma_t^{u,v,r}(\mathbf{x}) \leq \Gamma_t^{u,v,r}$, which completes the proof.

Appendix D: Proof of Lemma 1

$STT(\mathbf{q})$ is obtained by solving linear programming. \mathbf{q} is a parameter in the constraints. For every $\mathbf{q} \geq 0$, the problem is feasible because of the physical meaning of the problem (i.e., assigning flows to the network) as long as the system has enough capacity (i.e., dispatching enough vehicles). Hence, the lemma directly follows Theorem 1 in Martin (1975), which implements Berge's Maximum Theorem in parametric linear programming.

Appendix E: Proof of Proposition 2

According to Lemma 1:

$$\begin{aligned} |\mathbb{E}_{\mathbf{Q}(\mathbf{x}^*)}[STT(\mathbf{Q}(\mathbf{x}^*))] - STT(\mathbf{q}^*)| &= \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)} |STT(\hat{\mathbf{q}}) - STT(\mathbf{q}^*)| \cdot \mathbb{P}_{\mathbf{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ &\leq \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)} L \cdot \|\hat{\mathbf{q}} - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{\mathbf{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}). \end{aligned} \quad (63)$$

Let us divide the support of the random variable $\mathcal{Q}(\mathbf{x}^*)$ as three mutually exclusive subsets:

$$\mathcal{Q}(\mathbf{x}^*)^{\text{Leq}} = \mathcal{Q}(\mathbf{x}^*) \cap \{\hat{\mathbf{q}} : 0 \leq \hat{\mathbf{q}} < \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \epsilon\}, \quad (64)$$

$$\mathcal{Q}(\mathbf{x}^*)^{\text{Mid}} = \mathcal{Q}(\mathbf{x}^*) \cap \{\hat{\mathbf{q}} : \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \epsilon \leq \hat{\mathbf{q}} \leq \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] + \epsilon\}, \quad (65)$$

$$\mathcal{Q}(\mathbf{x}^*)^{\text{Geq}} = \mathcal{Q}(\mathbf{x}^*) \cap \{\hat{\mathbf{q}} : \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] + \epsilon < \hat{\mathbf{q}} \leq \mathbf{q}^{\text{Max}}\}, \quad (66)$$

where $\mathcal{Q}(\mathbf{x}^*) = \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}} \cup \mathcal{Q}(\mathbf{x}^*)^{\text{Mid}} \cup \mathcal{Q}(\mathbf{x}^*)^{\text{Geq}}$.

We can calculate the summation over these three subsets separately:

(1) Bounds on the summation over $\mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}$:

$$\sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot \|\hat{\mathbf{q}} - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \leq \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot (\|\hat{\mathbf{q}} - \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 + \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \mathbf{q}^*\|_1) \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}), \quad (67)$$

which is followed by the triangle inequality. Notice that

$$\begin{aligned} \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot \|\hat{\mathbf{q}} - \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) &\leq L \cdot \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ &= L \cdot \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 \cdot \mathbb{P}[\mathcal{Q}(\mathbf{x}^*) \leq \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \epsilon] \leq L \cdot \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 \cdot \|\mathbf{\Gamma}\|_2^2, \end{aligned} \quad (68)$$

where the last inequality is the result of the following:

$$\begin{aligned} \mathbb{P}[\mathcal{Q}(\mathbf{x}^*) \leq \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \epsilon] &\leq \mathbb{P}[|\mathcal{Q}(\mathbf{x}^*) - \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]| \geq \epsilon] \\ &\leq \sum_{i \in \mathcal{F} \times \mathcal{T}} \mathbb{P}[|Q_i|_{\mathbf{x}^*} - \mathbb{E}[Q_i|_{\mathbf{x}^*}]| \geq \epsilon_i] \leq \sum_{i \in \mathcal{F} \times \mathcal{T}} (\Gamma_i)^2 = \|\mathbf{\Gamma}\|_2^2, \end{aligned} \quad (69)$$

where the inequality is followed by the union bound and the Γ -concentration property. Similarly, we have

$$\begin{aligned} \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) &\leq \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot \|\epsilon\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ &= L \cdot \|\epsilon\|_1 \cdot \mathbb{P}[\mathcal{Q}(\mathbf{x}^*) \leq \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \epsilon] \leq L \cdot \|\epsilon\|_1 \cdot \|\mathbf{\Gamma}\|_2^2, \end{aligned} \quad (70)$$

where the first inequality is due to the ϵ -feasibility.

Therefore, combining (68) and (70) leads to

$$\sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot \|\hat{\mathbf{q}} - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \leq L \cdot (\|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 + \|\epsilon\|_1) \cdot \|\mathbf{\Gamma}\|_2^2. \quad (71)$$

(2) Bounds on the summation over $\mathcal{Q}(\mathbf{x}^*)^{\text{Mid}}$:

$$\begin{aligned} \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Mid}}} L \cdot \|\hat{\mathbf{q}} - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) &\leq \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Leq}}} L \cdot (\|\hat{\mathbf{q}} - \mathbb{E}[\mathcal{Q}(\mathbf{x}^*)]\|_1 + \|\mathbb{E}[\mathcal{Q}(\mathbf{x}^*)] - \mathbf{q}^*\|_1) \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ &\leq \sum_{\hat{\mathbf{q}} \in \mathcal{Q}(\mathbf{x}^*)^{\text{Mid}}} L \cdot (\|\epsilon\|_1 + \|\epsilon\|_1) \cdot \mathbb{P}_{\mathcal{Q}(\mathbf{x}^*)}(\hat{\mathbf{q}}) \leq 2L \cdot \|\epsilon\|_1. \end{aligned} \quad (72)$$

(3) Bounds on the summation over $\mathcal{Q}(\mathbf{x}^*)^{\text{Geq}}$:

Similar to the proof of $Q(\mathbf{x}^*)^{\text{Leq}}$, notice that

$$\begin{aligned} & \sum_{\hat{\mathbf{q}} \in Q(\mathbf{x}^*)^{\text{Geq}}} L \cdot \|\hat{\mathbf{q}} - \mathbb{E}[Q(\mathbf{x}^*)]\|_1 \cdot \mathbb{P}_{Q(\mathbf{x}^*)}(\hat{\mathbf{q}}) \leq L \cdot \sum_{\hat{\mathbf{q}} \in Q(\mathbf{x}^*)^{\text{Geq}}} \|\mathbf{q}^{\text{Max}}\|_1 \cdot \mathbb{P}_{Q(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ & = L \cdot \|\mathbf{q}^{\text{Max}}\|_1 \cdot \mathbb{P}[Q(\mathbf{x}^*) \leq \mathbb{E}[Q(\mathbf{x}^*)] - \epsilon] \leq L \cdot \|\mathbf{q}^{\text{Max}}\|_1 \cdot \|\Gamma\|_2^2. \end{aligned} \quad (73)$$

Combining (73) and (70), we have

$$\begin{aligned} \sum_{\hat{\mathbf{q}} \in Q(\mathbf{x}^*)^{\text{Geq}}} L \cdot \|\hat{\mathbf{q}} - \mathbf{q}^*\|_1 \cdot \mathbb{P}_{Q(\mathbf{x}^*)}(\hat{\mathbf{q}}) & \leq \sum_{\hat{\mathbf{q}} \in Q(\mathbf{x}^*)^{\text{Geq}}} L \cdot (\|\hat{\mathbf{q}} - \mathbb{E}[Q(\mathbf{x}^*)]\|_1 + \|\mathbb{E}[Q(\mathbf{x}^*)] - \mathbf{q}^*\|_1) \cdot \mathbb{P}_{Q(\mathbf{x}^*)}(\hat{\mathbf{q}}) \\ & \leq L \cdot (\|\mathbf{q}^{\text{Max}}\|_1 + \|\epsilon\|_1) \cdot \|\Gamma\|_2^2. \end{aligned} \quad (74)$$

In summary, combining the summation over three mutually exclusive sets, we have:

$$|\mathbb{E}_{Q(\mathbf{x}^*)}[SST(Q(\mathbf{x}^*))] - SST(\mathbf{q}^*)| \leq 2L \cdot \|\epsilon\|_1 + L \cdot (\|\mathbb{E}[Q(\mathbf{x}^*)]\|_1 + \|\mathbf{q}^{\text{Max}}\|_1 + 2\|\epsilon\|_1) \cdot \|\Gamma\|_2^2. \quad (75)$$

Appendix F: Proof of Proposition 3

When $\epsilon = 0$, we have $\mathbf{q} = \mathbb{E}[Q(\mathbf{x})]$. Then:

$$G_{\text{EP}}(\epsilon = 0) = \min_z \{g(\mathbb{E}[Q(\mathbf{x})], z) : h_j(\mathbb{E}[Q(\mathbf{x})], z) \leq 0\}. \quad (76)$$

According to Jensen's inequality, we have:

$$g(\mathbb{E}[Q(\mathbf{x})], z) \leq \mathbb{E}[g(Q(\mathbf{x}), z)], \quad h_j(\mathbb{E}[Q(\mathbf{x})], z) \leq \mathbb{E}[h_j(Q(\mathbf{x}), z)], \quad \forall j \in \mathcal{J}. \quad (77)$$

Therefore, the proposed approach has a smaller objective function and a larger feasible space compared to the stochastic optimization formulation (48), which makes it a lower bound of (48).

Appendix G: Proof of Proposition 4

Step 1: We first show that if $\text{Var}[Q(\mathbf{x})]$ is bounded, then $\text{Var}[h_j(Q(\mathbf{x}), z)]$ is also bounded.

Notice that for any random variable X , we have $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$. Hence, if we take $X = h_j(Q(\mathbf{x}), z) - h_j(\mathbb{E}[Q(\mathbf{x})], z)$, we get

$$\text{Var}[h_j(Q(\mathbf{x}), z) - h_j(\mathbb{E}[Q(\mathbf{x})], z)] = \text{Var}[h_j(Q(\mathbf{x}), z)] \leq \mathbb{E}[(h_j(Q(\mathbf{x}), z) - h_j(\mathbb{E}[Q(\mathbf{x})], z))^2]. \quad (78)$$

From the Lipschitz continuity of $h_j(\cdot)$, we have

$$|h_j(Q(\mathbf{x}), z) - h_j(\mathbb{E}[Q(\mathbf{x})], z)| \leq C \|Q(\mathbf{x}) - \mathbb{E}[Q(\mathbf{x})]\|_2, \quad (79)$$

which further yields:

$$\begin{aligned} \mathbb{E}[(h_j(Q(\mathbf{x}), z) - h_j(\mathbb{E}[Q(\mathbf{x})], z))^2] & \leq C^2 \mathbb{E}[\|Q(\mathbf{x}) - \mathbb{E}[Q(\mathbf{x})]\|_2^2] \\ & = C^2 \cdot \sum_{i=1}^n \mathbb{E}[(Y_i - \mathbb{E}[Y_i])^2] = C^2 \sum_{i=1}^n \text{Var}[Y_i]. \end{aligned} \quad (80)$$

Combining with (78), we have

$$\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})] \leq C^2 \cdot \sum_{i=1}^n \text{Var}[Y_i] \leq C^2 \cdot \|\mathbf{\Gamma}\|_2^2. \quad (81)$$

Step 2: We then show that $\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})] \leq C^2 \cdot \|\mathbf{\Gamma}\|_2^2$ can deduce a weaker version of the chance constraints.

Consider Chebyshev's inequality, for a given positive number $a > 0$:

$$\mathbb{P}[|h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z}) - \mathbb{E}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]| > a] \leq \frac{\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]}{a^2}. \quad (82)$$

Equation (82) implies

$$\begin{aligned} \mathbb{P}[|h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z}) - \mathbb{E}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]| \leq a] &\geq 1 - \frac{\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]}{a^2} \\ \Rightarrow \mathbb{P}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z}) \leq a + \mathbb{E}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]] &\geq 1 - \frac{\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]}{a^2}. \end{aligned} \quad (83)$$

Since we know that $\mathbb{E}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})] \leq 0$, (83) yields:

$$\mathbb{P}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z}) \leq a] \geq 1 - \frac{\text{Var}[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z})]}{a^2} \geq 1 - \left(\frac{C}{a}\right)^2 \cdot \|\mathbf{\Gamma}\|_2^2. \quad (84)$$

Let us pick $a = \frac{C\|\mathbf{\Gamma}\|_2}{\sqrt{1-\eta}}$, we have

$$\mathbb{P}\left[h_j(\mathbf{Q}(\mathbf{x}), \mathbf{z}) \leq \frac{C\|\mathbf{\Gamma}\|_2}{\sqrt{1-\eta}}\right] \geq \eta. \quad (85)$$

Combining the two steps finishes the proof.

Appendix H: Formulation of the Benders Decomposition

H.1. Subproblem

The subproblem is derived by fixing the decision variables \mathbf{x} , and only considering the components including \mathbf{q} and \mathbf{z} .

$$[SP(\mathbf{x})] \quad \min_{\mathbf{q}, \mathbf{z}} \quad WT(\mathbf{q}, \mathbf{z}) + IVT(\mathbf{z}) \quad (86a)$$

$$\text{s.t.} \quad \text{Constraints (24b) – (24c)}, \quad (86b)$$

$$\text{Constraints (31) – (32)}. \quad (86c)$$

The objective of the dual problem of (86) is

$$\begin{aligned} D(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\iota}, \boldsymbol{\kappa}, \boldsymbol{\rho}; \mathbf{x}) &= \sum_{l \in \mathcal{L}} \sum_{t \in \mathcal{T}} \sum_{t'=t}^{T_{l,t}} K_{l,t} \alpha_{l,t,t'} + \sum_{(u,v,r) \in \mathcal{F}} \sum_{t \in \mathcal{T}} \sum_{t'=\min}^t f_{t'}^{u,v,r} \beta_t^{u,v,r} \\ &+ \sum_{(u,v,r,i,t) \in \Omega_1} \tilde{z}_t^{u,v,r,i} \gamma_t^{u,v,r,i} + \sum_{(u,v) \in \mathcal{W}} \sum_{t \in \mathcal{T}} d_t^{u,v} \iota_t^{u,v} \\ &+ \sum_{(u,v,r) \in \mathcal{F}} \sum_{t \in \mathcal{T}} \kappa_t^{u,v,r} \cdot (1 - \epsilon) \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r \\ &+ \sum_{(u,v,r) \in \mathcal{F}} \sum_{t \in \mathcal{T}} \rho_t^{u,v,r} \cdot (1 + \epsilon) \sum_{p \in \mathcal{P}_t^{u,v}} \sum_{r' \in \mathcal{R}^{u,v}} x_{p,r'} \cdot \pi_{p,r'}^r, \end{aligned} \quad (87)$$

where $\alpha, \beta, \gamma, \iota$ are the dual variables associated with constraints (15), (16), (17), (10), respectively. κ, ρ are the dual variables associated with constraints (31) and (32). Let $\mathbf{x} := (\alpha, \beta, \gamma, \iota, \kappa, \rho)$. If the dual problem of (86) is feasible and bounded with a solution \mathbf{x}^* , the following optimality cut is added to the master problem:

$$Z \geq D(\mathbf{x}^*; \mathbf{x}), \quad (88)$$

where Z is a decision variable in the master problem. If the dual problem of (86) is unbounded, and \mathbf{x}^* is an optimal extreme ray of the dual, the following feasibility cut is added to the master problem:

$$D(\mathbf{x}^*; \mathbf{x}) \leq 0. \quad (89)$$

H.2. Master problem

Let \mathcal{A}^O be the set of solutions \mathbf{x}^* of optimality cuts and \mathcal{A}^F be the set of solutions \mathbf{x}^* of feasibility cuts. At each iteration of the BD, a cut based on the solution of the subproblem is added to the respective set, and the corresponding master problem is defined as follows:

$$[MP(\mathcal{A}^O, \mathcal{A}^F)] \quad \min_{\mathbf{x} \in \mathcal{X}, Z} \quad \Psi \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r} \cdot V_{p,r} + Z \quad (90a)$$

$$\text{s.t.} \quad Z \geq D(\mathbf{x}^*; \mathbf{x}) \quad \forall \mathbf{x}^* \in \mathcal{A}^O, \quad (90b)$$

$$D(\mathbf{x}^*; \mathbf{x}) \leq 0 \quad \forall \mathbf{x}^* \in \mathcal{A}^F, \quad (90c)$$

$$\text{Constraints (33)}. \quad (90d)$$

Note that the master problem has a smaller scale compared to the original problem (because there are no \mathbf{z} and \mathbf{q}), which can be solved efficiently.

H.3. Convergence

Let $(\mathbf{x}^{(k)}, Z^{(k)})$ and $(\mathbf{q}^{(k)}, \mathbf{z}^{(k)})$ be the solutions of the master problem and subproblem, respectively, in the k -th iteration. Then, the upper ($UB^{(k)}$) and lower ($LB^{(k)}$) bounds at the k -th iteration are given by:

$$UB^{(k)} = \Psi \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r}^{(k)} \cdot V_{p,r} + WT(\mathbf{q}^{(k)}, \mathbf{z}^{(k)}) + IVT(\mathbf{z}^{(k)}), \quad (91)$$

$$LB^{(k)} = \Psi \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}_p} -x_{p,r}^{(k)} \cdot V_{p,r} + Z^{(k)}. \quad (92)$$

$LB^{(k)}$ will keep increasing as k increases because more cuts are added to the master problem. $UB^{(k)}$ does not necessarily decrease at every iteration. The convergence criterion is

$$\text{Gap}^{(k)} = \frac{UB^{(k)} - LB^{(k)}}{LB^{(k)}} \leq \text{Predetermined threshold}. \quad (93)$$

Appendix I: Inference of status quo choices

The status quo path choice inference method is based on our previous study (Mo et al. 2022), which is also similar to the trip-train method used for destination inference in open public transit systems (i.e., no tap-out).

[In the system when the incident happens]: Consider a passenger $p \in \mathcal{P}$ with an incident line tap-in record before the end of the incident, meaning that he/she was in the transit system when the incident

happens. We then track his/her next tap-in record. If his/her next tap-in is a transfer at a nearby bus or rail station, we can identify his/her chosen path based on the transfer station. We can also identify the waiting passenger if he/she continues to use the incident line to his/her intended destination inferred by his/her next tap-in records.

[Out of the system when the incident happens]: For a passenger $p \in \mathcal{P}$ with only a tap-in record in nearby bus or rail stations. He/she may be affected by the incident to change the tap-in station, or just use the service as a normal commute. To identify whether he/she was affected, we extracted his/her travel histories on previous days without incidents to get the normal commute trajectories. If his/her tap-in time and location on the incident day have never appeared in the historical records before, we treat him/her as a passenger affected by the incident and identify his/her chosen path based on the tap-in station.

For passengers in \mathcal{P} without next tap-in records or travel histories, we randomly assign him/her a status quo path based on the proportion of inferred passengers. The final resulted path shares are shown in Figure 11. Around 49% of the passengers chose to wait. 22% of them chose the parallel bus lines. Others either took NS or WE bus routes and transferred to rails.

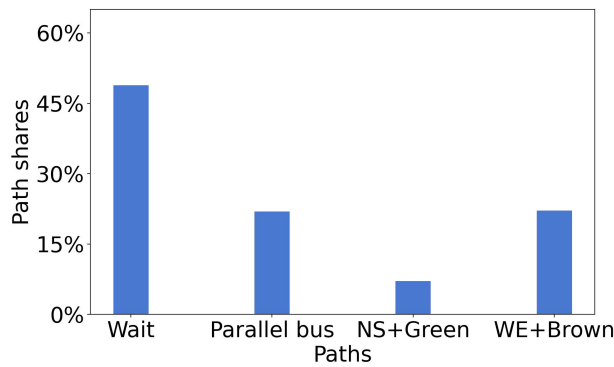


Figure 11 Path shares of the inferred status quo choices

Appendix J: Individual travel time comparison

It is worth noting that (24) does not explicitly output the travel time of passengers using different paths. The travel time of passengers using path (u, v, r) for trips departing at time t (denoted as $TT_t^{u,v,r}$) has to be obtained from the network flow patterns **after** solving (24). Specifically, consider the group of passengers using path (u, v, r) and departing at time t . Their arrival time at the destination given realized passenger flow \hat{q} (denoted as $AT_t^{u,v,r}(\hat{q})$) can be calculated as

$$AT_t^{u,v,r}(\hat{q}) = \min \left\{ \tilde{t} \in \mathcal{T}_t^{u,v,r} : \sum_{t'=t^{\min}}^{\tilde{t}} (f_{t'}^{u,v,r} + \hat{q}_{t'}^{u,v,r}) \leq \sum_{t^{\min} \leq t' + \delta_{t'}^{u,v,r, |I^{u,v,r}|} \leq \tilde{t}} z_{t'}^{u,v,r, |I^{u,v,r}|} \right\} \quad \forall t \in \mathcal{T}, (u, v, r) \in \mathcal{F}, \quad (94)$$

where $\mathcal{T}_t^{u,v,r}$ is the set of possible arrival time indices, defined as $\mathcal{T}_t^{u,v,r} = \{t' : t \leq t' \leq T\}$. Equation (94) represents the travel time calculation with cumulative demand curves at origins and destinations. $\sum_{t'=t^{\min}}^{\tilde{t}} (f_{t'}^{u,v,r} + \hat{q}_{t'}^{u,v,r})$ is the cumulative demand up to time \tilde{t} at the origin. $\sum_{t^{\min} \leq t' + \delta_{t'}^{u,v,r, |I^{u,v,r}|} \leq \tilde{t}} z_{t'}^{u,v,r, |I^{u,v,r}|}$ is the cumulative passengers arriving at the destination up to time \tilde{t} . When the cumulative arrivals at the destination are greater or equal to the cumulative demand at the origin (up to time \tilde{t}), all passengers finish the trip. So taking the minimum over \tilde{t} gives the arrival time for passengers departing at t . The travel time is then simply:

$$TT_t^{u,v,r} = AT_t^{u,v,r} - t \quad \forall t \in \mathcal{T}, (u, v, r) \in \mathcal{F}. \quad (95)$$

Figure 12 illustrates the travel time calculation.

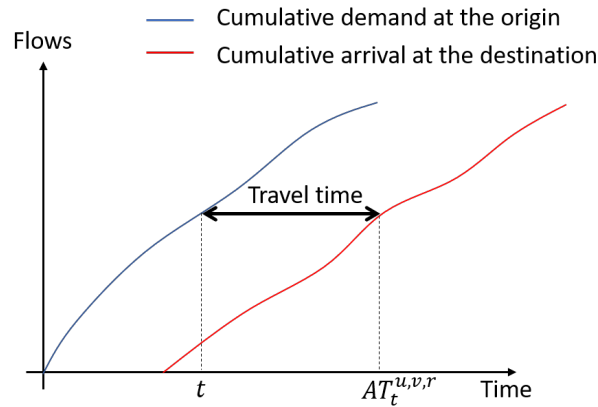


Figure 12 Illustration of travel time calculation

Based on the above formulas, we calculate the travel time saving for each individual in the system. The distribution is shown in Figure 13, where the negative values imply that the proposed model (IPR) has a lower travel time. For the comparison between IPR and status quo, we saw a sizable population has lower travel time under the recommendation, the maximum saving can be 50 minutes. For the comparison between the IPR and the capacity-based method, the most time saving is within 10 minutes. Note that there are some

passengers experiencing higher travel time. This may be due to two reasons. First, there may be multiple ways to recommend passengers that achieve the same system travel time. As the model has no sense of its original travel time, some passengers may be worse off. Second, in order to reach the system optimal, some passengers may need to switch to a worse route in order to make the system better. In order to address this issue, future studies may impose some equity-related constraints to ensure no passenger is provided with a worse path.

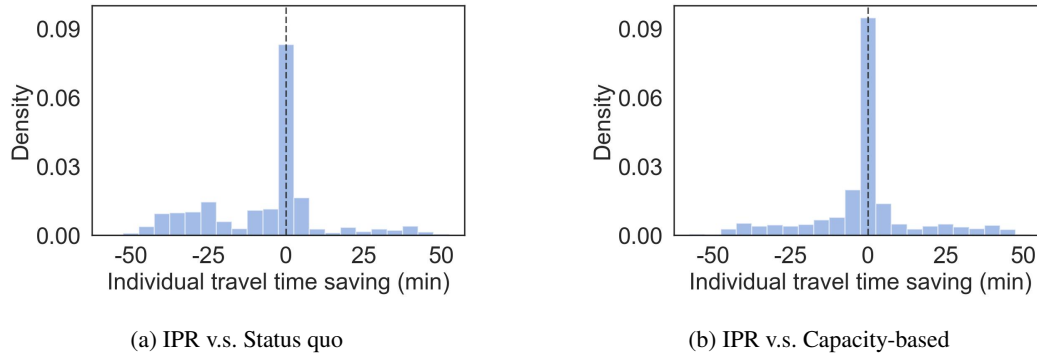


Figure 13 Distribution of travel time saving (new travel time minus the old travel time, negative values imply lower travel time).

References

- Martin D (1975) On the continuity of the maximum in parametric linear programming. *Journal of Optimization Theory and Applications* 17(3):205–210.
- Mo B, Koutsopoulos H, Zhao J (2022) Inferring passenger responses to urban rail disruptions using smart card data: A probabilistic framework. *Transportation Research Part E: Logistics and Transportation Review* 159:102628.