

1 **Network Performance Model for Urban Rail Systems**

2

3

4 **Baichuan Mo**

5 Research Assistant

6 Department of Civil and Environmental Engineering

7 Massachusetts Institute of Technology

8 77 Massachusetts Ave, Cambridge, MA 20139

9 Tel: +1 857-999-5906; Email: baichuan@mit.edu

10

11 **Zhenliang Ma, Ph.D.** (Corresponding Author)

12 Assitant Professor

13 Department of Civil and Environmental Engineering

14 Monash University

15 Clayton, VIC 3800, Australia

16 Tel: +61 175109768; Email: mike.ma@monash.edu

17

18 **Haris N. Koutsopoulos, Ph.D.**

19 Professor

20 Department of Civil and Environmental Engineering

21 Northeastern University

22 Boston, MA 02115, United States

23 Tel: +1 617-373-6263; Email: h.koutsopoulos@northeastern.edu

24

25 **Jinhua Zhao, Ph.D.**

26 Associate Professor

27 Department of Urban Studies and Planning

28 Massachusetts Institute of Technology

29 77 Massachusetts Ave, Cambridge, MA 20139

30 Tel: +1 617-253-7594; Email: jinhua@mit.edu

31

32

33 Word count:5,766 words text + 3 tables × 250 words (each) = 6,516 words

34

35

36

37

38

39

40 Submission Date: July 31, 2019

**1 ABSTRACT**

2 This paper proposed a general data-driven Network Performance Model (NPM) for daily network  
3 performance monitoring using smart card data. The major component of NPM is a schedule-based  
4 network loading model with strict capacity constraints. The potential applications of NPM include  
5 estimating crowding patterns, diagnosing crowding sources and evaluating network resilience. A  
6 method for calibrating train capacity is introduced, which explicitly recognizes that capacity may  
7 be different at different stations depending on the congestion level. Case studies are conducted  
8 based on the Mass Transit Railway (MTR) system in Hong Kong. The NPM model was first  
9 validated using denied boarding survey and exit demand data. Then, we demonstrated its capability  
10 in performance monitoring by analyzing the spatio-temporal crowding patterns and evaluating the  
11 network resilience.

12

13 *Keywords:* Transit network loading; Performance monitoring; Effective capacity; Event-driven  
14 simulation

Under Review

## 1 INTRODUCTION

### 2 Motivation

3 Urban rail transit systems are the principal means of public transportation in many of the metropoli-  
4 tan areas. Due to its high reliability, large capacity, and low pollution, rail transit services continue  
5 to grow along with rising demand. However, in the context of large-scale network and high-volume  
6 demand, recurrent congestion in peak hours and sudden incidents in the system are becoming a  
7 major concern. Ensuring normal operations is important to keep the smooth and efficient urban  
8 mobility. Dynamic Crowd Management Systems (DCMS) are means to deal with system opera-  
9 tion of near capacity, including crowding and incidents. It can assist the operators to understand,  
10 inform and improve the transit services, especially under capacity constraints. There are many  
11 sub-problems under the DCMS framework, which covers different dimensions of management. In  
12 this study, we focus on the "performance monitoring" problem. Network performance monitoring  
13 means obtaining the level of service and operation information of the network (e.g. train load) and  
14 analyzing the service quality, which is crucial for transportation agencies to identify congestion,  
15 evaluate the system, and adjust operating strategies.

### 16 Related work

17 Performance monitoring can be conducted at two different levels. One is at data-description level  
18 and another one at in modeling level. For the data-description level, researchers only analyze  
19 the data which can be directly obtained, such as automated fare collection (AFC) data and au-  
20 tomated vehicle location (AVL) data. All performance indicators are derived from the raw data.  
21 For example, Trépanier et al. (1) calculated the performance indicators of network supply (e.g.,  
22 vehicle-kilometers, vehicle-hours, commercial speed) from the AFC data. Ma and Wang (2) de-  
23 veloped a data-driven platform for transit performance measurement using AFC and AVL data.  
24 Indicators such as speed, travel time reliability, station demand are reported. The drawback for  
25 data-description based performance models is that, only the measurements which is directly avail-  
26 able in raw data can be calculated. More detailed information, such as vehicle load, is not available.  
27 In terms of modeling level method, the performance monitoring is usually achieved by the network  
28 loading and network assignment models. In network loading framework, the passengers' travel be-  
29 havior is known and then we load passengers to the train accordingly. Network assignment models  
30 usually take into account passengers' travel behavior through the user equilibrium assumption. The  
31 models typically perform many network loading operations to update passengers' travel behaviors  
32 until convergence. Regardless of what framework to use, the network loading is a key component  
33 for modeling-level performance monitoring.

34 In the context of both transit network loading and assignment, the models can be divided  
35 into two categories, the frequency-based models and the schedule- or timetable-based models.  
36 Frequency-based approaches consider services based on different lines. The average headway or  
37 average link travel time are defined for each line, while the train schedules are not considered ex-  
38 plicitly. Therefore, this approach cannot provide detailed performance information at the train level  
39 (e.g. load of a specific train), only average line performances and flows (3–7). The schedule-based  
40 approach uses individual train arrival/departure times to model the network operations explicitly.  
41 Thus, it can capture the dynamic interactions between supply and demand, and provide perfor-  
42 mance metrics at the train level (8). Many schedule-based models have been proposed in the  
43 literature. For example, Nuzzolo et al. (8) proposed a dynamic schedule-based assignment model  
44 for transit networks, which simulates the within-day and day-to-day learning process passengers'

1 route choice. Grube et al. (9) developed an event-based network loading model. Nuzzolo et al. (3)  
2 proposed a mesoscopic transit modeling framework named "DYBUS2", which provides real-time  
3 short-term predictions of transit performance. Recently, Yao et al. (10) used an agent-based simu-  
4 lation model to conduct network assignment. The model is applied to a large-scale network of the  
5 Beijing subway system.

6 Despite the extensive literature on transit performance models, there are still research gaps.  
7 1) Most of these network loading models are run based on a fixed time step (e.g. 1 second). These  
8 discrete time models are usually computationally intensive (11). The states of a metro system  
9 are affected by significant events (e.g. train arrival and departure). So the event-driven paradigm  
10 is more suitable for modeling urban rail networks by providing higher computational efficiency  
11 (9). 2) The performance metrics are not systematically and completely reported. For example,  
12 an important indicator for the network its resilience (12), that is, how the network performs under  
13 unexpected disruptions. None of the previous models can support network resilience evaluation. 3)  
14 Most of network loading procedures are developed to mainly serve the transit assignment model.  
15 They do not focus on the performance evaluation. Thus, the model calibration (e.g. for path shares  
16 and train capacities) is usually neglected. And the empirical model validations and real-world  
17 applications are rarely reported.

## 18 **Paper objectives and organization**

19 The paper aims to develop a general network performance model (NPM) used for performance  
20 monitoring at the platform and train levels. The main characteristics of the model are: 1) it cap-  
21 tures the fine-grained passenger travel behaviors, such as access and egress walking, queuing,  
22 transferring, boarding, alighting and left behind. 2) It is based on an event-driven framework (13),  
23 which is more efficient for large scale networks. 3) It can output complete performance metrics  
24 (e.g. train load, left behind rate) in both spatial and temporal dimensions. The contributions of this  
25 paper are threefold.

- 26 • First, a network resilience analysis model is proposed and incorporated into the NPM, which  
27 allows us to capture the impact of station closure in the network.
- 28 • Second, a simulation-based optimization method for calibrating train capacity is introduced,  
29 which explicitly recognizes that capacity may be different at different stations depending on  
30 the congestion level.
- 31 • Third, the model is calibrated, validated and applied using the real-world data in Hong Kong  
32 Mass Transit Railway (MTR) system, which demonstrate the capabilities of performance mon-  
33 itoring.

34 The remainder of the paper is organized as follows. The schedule-based NPM for transit  
35 network is introduced in Section 2, which includes the data preparation and network loading.  
36 Section 3 presents the model calibration methods for route choice and train capacity. The functions  
37 of NPM for performance monitoring are demonstrated in Section 4. The model implementation  
38 in Hong Kong MTR networks are presented in Section 5. Section 6 presents the conclusion and  
39 discussion for this work.

## 40 **NETWORK PERFORMANCE MODEL**

### 41 **Input**

42 NPM requires multiple inputs including dynamic OD demand, path choice fractions, train move-  
43 ment information, train capacity and access/egress/transfer walking time. We summarize all inputs

1 data sources in Table 1. AFC data contains people’s tap-in and tap-out time and stations, which can  
 2 provide the complete OD entry demand. As for train movements, the time table can provide the  
 3 planned train arrival and departure information, while the AVL data can provide the actual ones.  
 4 Therefore, if AVL data is available, it is more suitable for performance monitoring. The walking  
 5 time is assumed to be normally distributed where the mean and variance are calculated from the  
 6 observations.

**TABLE 1:** Input Variables and Data Sources

Input variables	Sources
OD entry demand	AFC data
Path choice	See Section 3.1
Train movement	Time table or AVL data
Train capacity	See Section 3.2
Access/egress/transfer walking time	Measured by on-site observations

## 7 Network loading

8 An event-driven framework is used to conduct the network loading. Two types of events are con-  
 9 sidered in this model, train arrival and train departure. The events are sorted by time and processed  
 10 sequentially until all events are successfully processed during the simulation time period. New and  
 11 transferring passengers join the waiting queue on the platform and board a train based on the First  
 12 Come First Serve (FCFS) criteria. The number of boarding passengers depends on the available  
 13 train capacity. The model details are shown in following.

### 14 *Preprocessing*

15 We first set up the simulation period with a warm-up and cool-down time. The warm-up and  
 16 cool-down time are set based on the scale of the network (e.g.30 minutes or 1 hour). We then  
 17 generate the train event lists (arrivals and departures) within the simulation period according to the  
 18 timetable or the actual train movement data from AVL. Each event contains the occurrence time  
 19 and place (platform), as well as the train ID. Then all passengers are assigned a route based on  
 20 the corresponding path choice probability. To simplify the access walking process, we assign the  
 21 access walking time for each passenger based on their walking time distribution of tap-in stations.  
 22 Then the platform arrival time of all passengers can be calculated in advance.

### 23 *Train arrival*

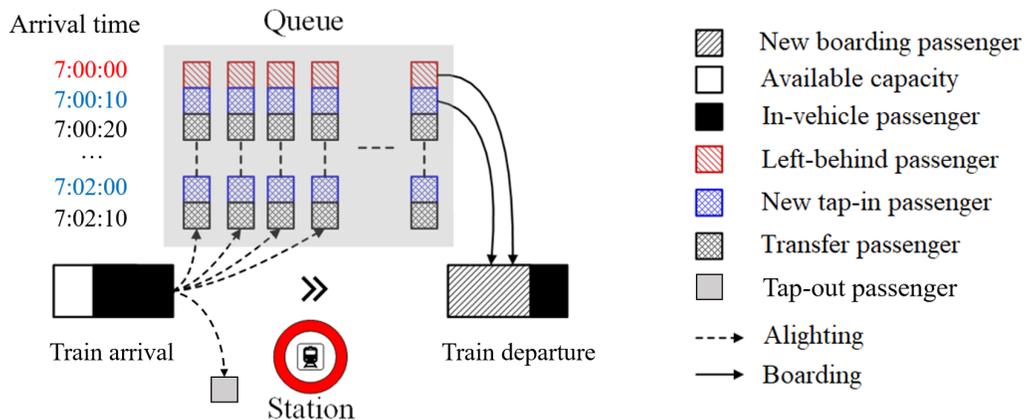
24 To process the arrival event, the train offloads passengers who reach their destination or need to  
 25 transfer at this station, and updates its state (e.g. train load and in-vehicle passengers). For pas-  
 26 sengers who reached their destination, their tap out time is calculated based on the egress walking  
 27 time distribution. For those who transfer at this station, their arrival time for the next platform is  
 28 calculated based on the transfer time distribution. These passengers are then added to the waiting  
 29 queue of that platform based on their arrival time.

### 1 *Train departure*

2 If the event is departure, the waiting queue of this platform is updated by accumulating the new tap-  
 3 in passengers, that is, passengers who enter and arrive at the platform after the last train departed  
 4 are added into the queue based on their arrival time. Next, passengers will board the train based  
 5 on an FCFS boarding priority rule until the train reach the capacity. Passengers who cannot board  
 6 will be left behind and wait in the queue for the next train. The state of the train and the waiting  
 7 queue at the platform are then updated accordingly.

### 8 *Model structure*

9 Figure 2 shows the diagram of the model structure. Three objects are defined: trains, waiting  
 10 queues and passengers. Passengers are queued based on their arrival time. Three different types  
 11 of passengers are shown in the queue, the left-behind passengers who were denied boarding by  
 12 previous trains, new tap-in passengers, and transfer passengers from other stations. The left-behind  
 13 passengers are usually at the head of the queue. When the train arrives at the station, the offloaded  
 14 passengers either transfer or tap-out. Transfer passengers will join the queue. When the train  
 15 departs from the station, passengers from the queues are loaded up to the train capacity based on  
 FCFS rule.



**FIGURE 1:** Diagram of the Model Structure

16

## 17 **MODEL CALIBRATION**

### 18 **Route choice**

19 The route choice belongs to the general class of discrete choice problems. This framework assumes  
 20 the goal of an individual is to maximize the utility of his/her choice. The choice probability as  
 21 a closed-form expression is also known as the Multinomial Logit (MNL) model. However, in  
 22 the urban rail system, the alternative routes might overlap in certain degree, which violates the  
 23 independence of irrelevant alternative (IIA). assumption of MNL model. A variation of the MNL,  
 24 known as the C-Logit, can be used to correct for this violation. The C-Logit model shares the  
 25 computational and estimation efficiency of the standard Logit model. The basic idea is to deal with  
 26 similarities among overlapping paths through an additional "cost" attribute, called the commonality

1 factor (CF), in the utility function (14). The probability of choosing path  $i$  can be formulated below.

$$P_i = \frac{\exp(\beta_X \cdot X_i + \beta_{CF} \cdot CF_i)}{\sum_{j \in \mathbb{W}} \exp(\beta_X \cdot X_j + \beta_{CF} \cdot CF_j)}, \quad (1)$$

2 where  $X_i$  are the attributes for path  $i$ , such as in-vehicle time, number of transfers, etc..  $\mathbb{W}$  is the  
3 path set of the OD pair that path  $i$  belongs to.  $\beta_X$  and  $\beta_{CF}$  are the corresponding coefficients to be  
4 estimated.  $CF_i$  is the commonality factor of path  $i$ . The  $CF_i$  can be expressed as following.

$$CF_i = \ln \sum_{j \in \mathbb{W}} \left( \frac{L_{i,j}}{L_i L_j} \right)^\gamma, \quad (2)$$

5 where  $L_{i,j}$  is the number of common stations of path  $i$  and  $j$ .  $L_i$  and  $L_j$  are the number of stations  
6 for path  $i$  and  $j$ , respectively.  $\gamma$  is a positive constant which is determined based on empirical  
7 studies (15).

8 To target in a general NPM framework, we should not rely on the survey data as the input  
9 because obtaining it is time-consuming and cost-inefficient. Thus, we are developing a AFC data-  
10 based path choice calibration method (16) which does not rely on external data sources. In the  
11 future we will combine the path choice calibration module into the proposed NPM. But for this  
12 paper, as the main purpose is to illustrate the overall framework, the survey-based path shares are  
13 used.

#### 14 **Effective capacity**

15 The capacity of a train is a vague concept. Normally the train cannot reach its designed capacity  
16 because passengers may deliberately deny to board and due to the lack of space or inability to get  
17 a seat (17). Therefore, a fixed physical capacity may be hard to capture the real-world situation. In  
18 this paper, we define a concept termed *effective capacity*, which is the number of passengers in the  
19 train after waiting passenger boarding the train while some of them are left behind. It corresponds  
20 to the concept of willingness to board proposed by Liu et al. (18). This concept comes from the fact  
21 that although some queuing passengers cannot board a crowded train in this station, when the same  
22 train arrives at the next station, some other passengers could still get on it (18). So, the "capacity"  
23 of a train is not a constant and may vary across stations. We use the term *effective capacity* in this  
24 study to differentiate with the physical fixed capacity. Note that the effective capacity in this paper  
25 is station-specific, rather than train-specific.

26 Based on previous research, we choose two factors to model the effective capacity ( $EC$ ).  
27 One is the current train load when the train enter the station (denoted as  $L$ ) (18). Another is the  
28 number of queuing passengers in the platform (denoted as  $Q$ ) (19). Denote the standard train  
29 capacity as  $C$ . For congested stations, the actual train load always exceeds  $C$  according to the  
30 observations. Since the effective capacity phenomenon usually appears in crowding stations (18),  
31 the train capacity for uncongested stations is set as  $C$ . In this paper, the list of congested stations  
32 is provided by the operator based on where left-behind often occurs. We formulate the effective  
33 capacity for platform  $i$  ( $EC_i$ ) as a simple linear equations (Eq. 3). The term *platform* in this study  
34 means a combination of station+line+direction. The relationship between  $EC$  and  $L, Q$  may be  
35 more complex than the assumption. But exploring the expression of  $EC$  is beyond the scope of this  
36 paper.

$$EC_i = \begin{cases} C + \beta_1 L_i + \beta_2 Q_i & \text{if platform } i \text{ is in list of congested stations} \\ C & \text{otherwise} \end{cases} \quad (3)$$

1  $\beta_1$  and  $\beta_2$  are the parameters to estimate. We expect both  $\beta_1$  and  $\beta_2$  to be positive. According to the  
 2 psychological effect described in (18), for the train leaving from platform  $i$  with some passengers  
 3 being left behind, if the passengers waiting at station  $i + 1$  cannot board the train, it would give  
 4 them the impression that they would never get on any following trains. So, this passenger will  
 5 have much larger motivation to board the train than passengers in platform  $i$ . Therefore, what  
 6 we observe is, even if the train is already very crowded that passengers waiting at platform  $i$  do  
 7 not board, the passenger waiting at station  $i + 1$  will still get on it. Thus,  $\beta_1$  should be positive  
 8 because  $Q_{i+1} > Q_i$  and  $EC_{i+1} > EC_i$ . The same philosophy can be used to explain the sign of  $\beta_2$ .  
 9 If there were many passengers waiting at the platform, people in the tail of the queue would give  
 10 some pressure to people in the head of the queue and push them to board. Thus,  $EC$  in congested  
 11 stations should be higher than the uncongested stations, which means  $\beta_2$  should be positive as well.

12 Calibrating the value of  $\beta$  is a black-box optimization problem because the network loading  
 13 is a non-analytical process. In this study, a Bayesian Simulation-based Optimization (BSO) method  
 14 (20) is applied. The BSO works by constructing a posterior distribution (surrogate function) that  
 15 best describes the objective function. As the number of observations grows, the posterior distribu-  
 16 tion improves, and the algorithm becomes more certain of which regions in parameter space are  
 17 worth exploring. In this study, we use two variables with real-world observations to construct the  
 18 objective function. One is the OD exit flow, another is the journey time distribution (JTD). We di-  
 19 vide the simulation period into several time intervals of width  $\tau$  (e.g. 15 min). Denote  $t = 1, 2, \dots, T$   
 20 as the index of each time interval. We define the OD exit flow for origin  $i$ , destination  $j$  and time  
 21 interval  $t$  as  $q^{i,j,t}$ , which is the number of passengers come from station  $i$  and exit at station  $j$  during  
 22 time interval  $t$  ( $t$  is the index of exit time).  $q^{i,j,t}$  is the output of the network loading model, which  
 23 can also be obtained from the AFC data since we know people's tap-out time. Denote the observed  
 24 OD exit flow as  $\tilde{q}^{i,j,t}$ . Within the time interval  $t$ , we can also obtain the ground truth JTD for each  
 25 OD pair because all passengers' travel time is available in the AFC data. Denote the estimated  
 26 JTD for origin  $i$ , destination  $j$  and time interval  $t$  as  $p_{i,j,t}(x)$ , and the corresponding ground-truth  
 27 JTD derived from AFC data as  $\tilde{p}_{i,j,t}(x)$ . We can formulate the difference of two distributions as  
 28 Kullback-Leibler (KL) divergence ( $D_{KL}$ ), that is:

$$D_{KL}(p_{i,j,t}(x) || \tilde{p}_{i,j,t}(x)) = \int_x p_{i,j,t}(x) \cdot \log \frac{p_{i,j,t}(x)}{\tilde{p}_{i,j,t}(x)} dx. \quad (4)$$

29 Therefore, the BSO can be formulated as

$$\begin{aligned} \min_{\beta_1, \beta_2} \quad & w_1 \sum_{i,j,t} (q^{i,j,t} - \tilde{q}^{i,j,t})^2 + w_2 \sum_{i,j,t} D_{KL}(p_{i,j,t}(x) || \tilde{p}_{i,j,t}(x)) \\ \text{s.t.} \quad & q^{i,j,t}, p_{i,j,t}(x) = \text{NPM}(\beta_1, \beta_2) \quad \forall i, j, t \end{aligned} \quad (5)$$

30 where  $w_1$  and  $w_2$  are the weights to balance the scale of two terms. The effective capacity is  
 31 an interesting phenomenon which can be observed in many real-world metro systems, including  
 32 MTR. However, we admit this article only deal with it in a rough way. Based on our numerical  
 33 tests, considering the effect of  $L$  and  $Q$  can make the model better fit the real-world observation

1 compared to a fixed physical capacity. In the future, we will propose more sophisticated method  
2 to better model this problem.

### 3 **PERFORMANCE MONITORING**

4 The proposed NPM can be used to monitor the performance at three dimensions: estimating crowd-  
5 ing patterns, diagnosing crowding causes and evaluating network resilience. These three dimen-  
6 sions reflect the interests of both operators and customers.

#### 7 **Estimating crowding patterns**

8 Crowding patterns are one of the major components of the performance metrics. The crowding  
9 indicators are summarized in Table 2. All indicators are time-dependent with flexible aggregated  
10 intervals. To better visualize the indicators for real-world practice, a web-based visualization tool  
11 was developed, which takes the output of NPM as input. The example of visualization tool is  
12 shown in the case study.

**TABLE 2:** Performance Indicators

Train	Train load
	Waiting time
	Denied boarding times
Platform	Left behind rate
	Number of boarding passengers
	Number of arrival passengers
	Queue length
Station	OD entry flow
	OD exit flow
Link	Link flow

#### 13 **Diagnose crowding sources**

14 To diagnose the reason of crowding, the visualization tool also helps to trace the sources of passen-  
15 gers in each link. For each link, we can trace where and when these passengers come from, which  
16 are grouped by OD pairs and time intervals. Knowing the crowding causes can assist operators  
17 to release station-specific flow guidance information to relieve the congestion.

#### 18 **Evaluate network resilience**

19 In network science, resilience is defined as the ability to provide and maintain an acceptable level  
20 of service in the face of faults and challenges to normal operation (21). NPM can be used to analyze  
21 the network resilience by comparing the performance indicators (e.g. waiting time) between the  
22 system with unplanned disruption and the normal system.

23 In the NPM, there is a module to add link disruption to the system for a specific time.  
24 When a link disruption happens, we assume all passenger can receive this message. The line  
25 where the disrupted link located will stop working. Passengers who need to use the link will  
26 alight from the train and wait at the alighting station. Passengers who enter the system after the  
27 link disruption will wait at the origin station if they are going to use the link until the blockage

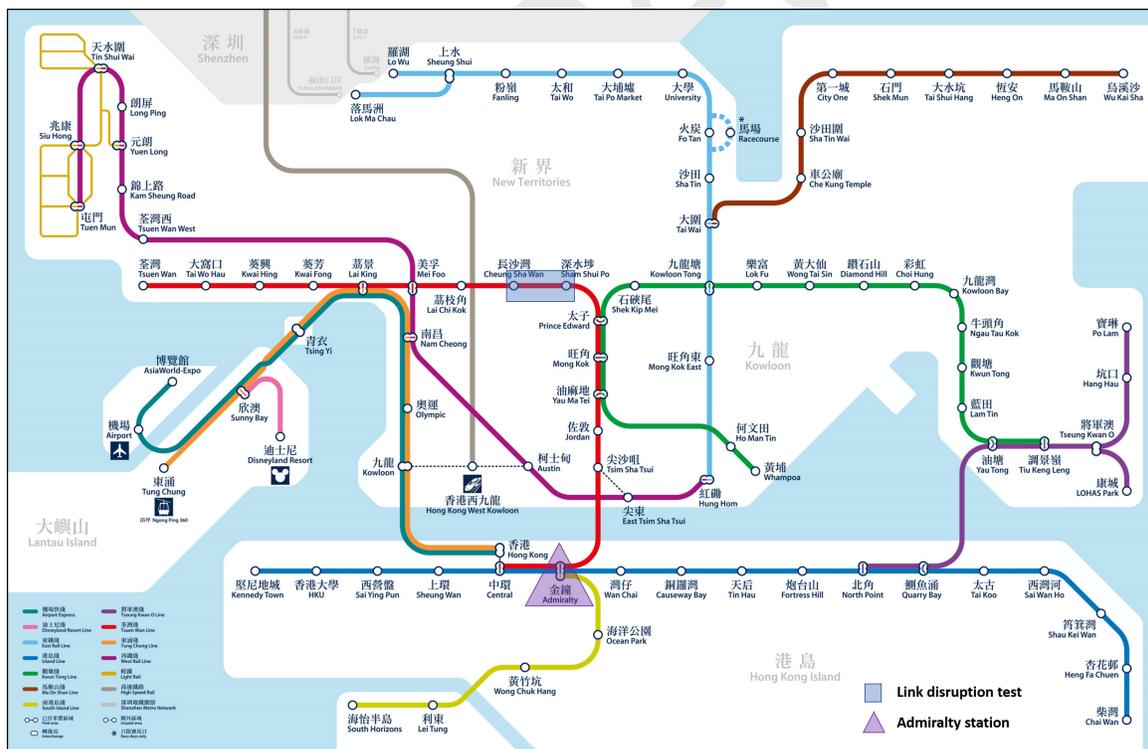
1 is cleared. Passengers who do not use the disrupted link will not be affected. This assumption  
 2 may be a simplification passengers' behavior because the dynamic route change is not considered.  
 3 However, given that most of the incidents in metro systems only last for several minutes, this  
 4 assumption is still reasonable because passengers can accept for short time waiting. More complex  
 5 situations with passengers' route change and operators' guidance can be incorporated in the future.  
 6 The case study for network resilience analysis is shown in Section 5.4.

7 **CASE STUDY**

8 The NPM was demonstrated and validated using data from Hong Kong MTR System

9 **Hong Kong MTR Network**

10 The map for Hong Kong MTR system is shown in Fig 2. In this study, the airport express and light  
 11 rail transit services were not considered since they are separated from the urban railway lines and  
 12 passengers who enter the urban railway lines from these services need to tap-in again. The system  
 13 consists of 10 lines and 114 stations, including 16 transfer stations. In this network, most transfer  
 14 stations connect only two lines. A special case is Admiralty (purple triangle in the figure) in the  
 15 Hong Kong island, where three lines pass through the same transfer station. The Admiralty station  
 16 is in the CBD area of Hong Kong, and there are serious crowding problem in peak hours due to the  
 high volume of passengers boarding and transferring at this station.



**FIGURE 2: Hong Kong MTR Metro System Map**

17

18 **System settings**

19 The AFC data in March 16th (Thursday), 2017 is used to generate the OD entry demand and  
 20 conduct the effective capacity calibration. Since AVL data is not available for all lines, the time

1 table was used to provide train movement information, which may introduce some errors. Ideally,  
 2 AVL data is more suitable for performance monitoring because it contains the actual train arrival  
 3 and departure information. Since the evening peak is the most congested period. We only consider  
 4 the period from 17:00 to 20:00 for model application. The warm-up time and cool-down time are  
 5 both set as 1 hour. The running time for is about 15 minutes in a personal computer with a 3.6GHz  
 6 CPU.

7 To test the network resilience, we add the link disruption between Cheuang Sha Wan station  
 8 and Sham Shui Po station in the Tsuen Wan Line (see the blue square in Figure 2). Since Tsuen  
 9 Wan Line is very busy during evening peak in the north bound, we hope to analyze how does link  
 10 disruption in Tsuen Wan Line affect the system. Two scenarios are considered, one is the disruption  
 11 during 17:15-17:30, when the system is not yet very congested. Another is the disruption during  
 12 18:00-18:15, when the system is already very congested.

### 13 Model Calibration

14 To estimate the route choice behavior, we launched a survey to the MTR system users. A total  
 15 number of 31,640 passengers completed the questionnaire. After filtering duplicate responses,  
 16 26,996 responses were available. The model results are shown in Table 3. The main explanatory  
 17 variables are the total in-vehicle time, relative transfer walking time and number of transfers. The  
 18 relative transfer walking time is defined as total transfer walking time divided by the map distance  
 19 of the path. All variables are statistically significant with the expected signs. Routes with high  
 20 in-vehicle time, walking time and number of transfers are less likely to be chosen by passengers.  
 21 Based on the estimated parameters, we are able to calculate path shares for all OD pairs in the  
 22 MTR system.

**TABLE 3:** Route Choice Model Estimation Results

	Estimate	Std. Error	t-value	
In-vehicle time	-0.184	0.001	-18.82	***
Relative transfer walking time	-3.204	0.209	-15.29	***
Number of transfers	-0.484	0.065	-7.49	***
CF	-2.014	0.648	-3.11	**
$\rho^2 = 0.3948$				

\*\*\*:  $p < 0.01$ ; \*\*:  $p < 0.05$ .

23 As for the *EC* calibration, the weights are set to  $w_1 = 1$  and  $w_2 = 1000$  to match the  
 24 scale of two terms in the objective function.  $C = (230/\text{car} \times \text{Number of cars per train})$  is used  
 25 according to MTR's standard. The optimal coefficients we got is  $\beta_1 = 0.0904$  and  $\beta_2 = 0.0718$ ,  
 26 which correspond to our expectations.

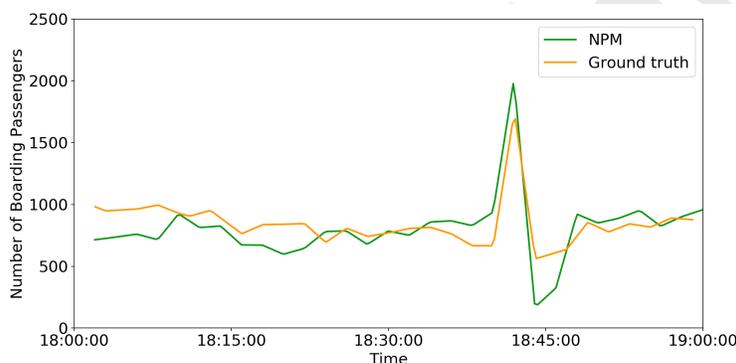
### 27 Numerical results

#### 28 Model validation

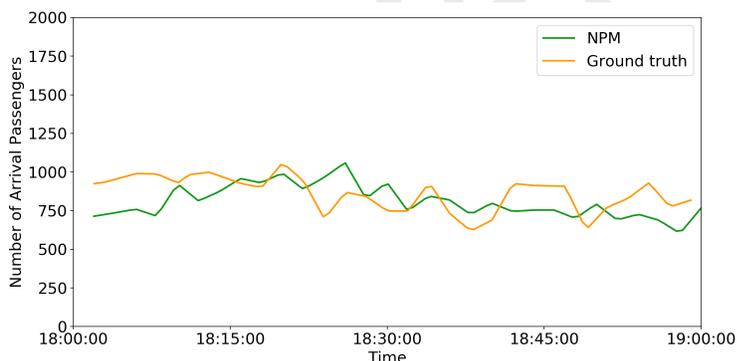
29 To validate whether the proposed NPM can replicates the actual conditions, field observation data  
 30 at Admiralty station on the same day were used for comparison. The data were collected by  
 31 MTR employees who counted passengers on the platform during 18:00-19:00. The left behind

1 rate, number of arrival passengers (sum of tap-in and transfer passengers) and number of boarding  
 2 passengers are recorded. The comparison results are shown in Figure 3. The patterns of number  
 3 of boarding and arrival passengers match the ground truth well. The peak in Figure 3a is due to  
 4 an empty train dispatched from an upstream station, so that more capacity is available to serve the  
 5 passengers at Admiralty. The empty train dispatching information is embedded into the model. In  
 6 terms of the left behind, we can see the percentage of passengers being left behind different times  
 7 can also be estimated well. Figure 4 shows the comparison of OD exit flows between the NPM  
 8 estimated ones and the true ones extracted from AFC data. The top 50 stations with higher exit  
 9 flows are displayed.

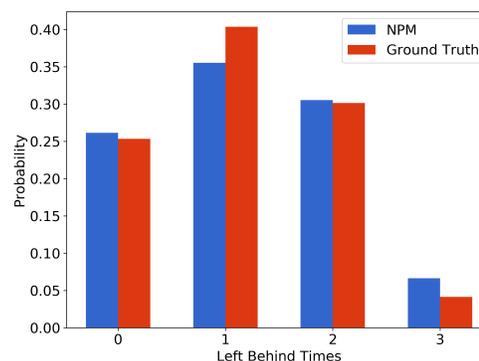
10 The model errors may come from the followings: 1) Use of timetable rather than AVL data.  
 11 2) Error in assumed path shares and effective capacity. 3) Measurement error in access, egress and  
 12 transfer times. Overall, the model can well capture the real-world situations and is effective for  
 13 performance monitoring.



(a) Boarding Passenger Comparison



(b) Arrival Passenger Comparison

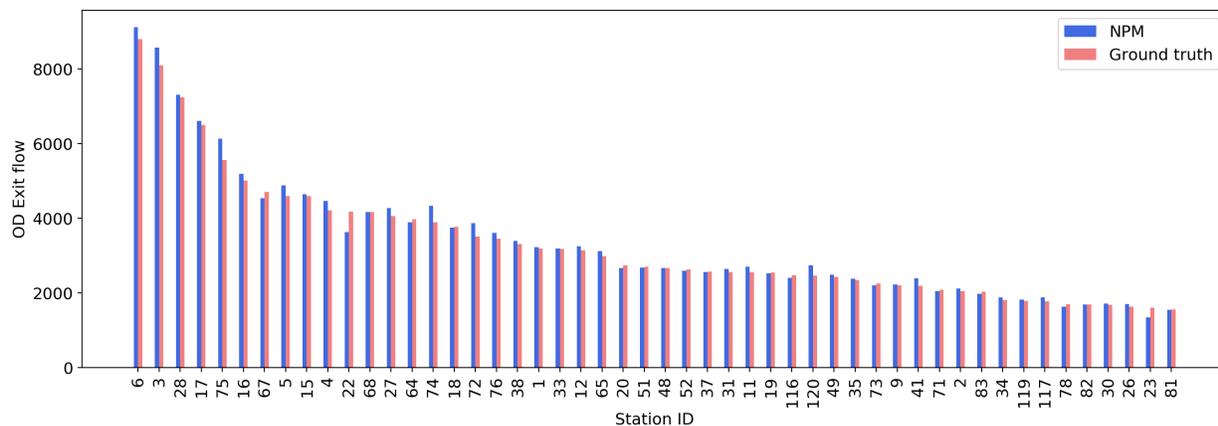


(c) Left Behind Comparison

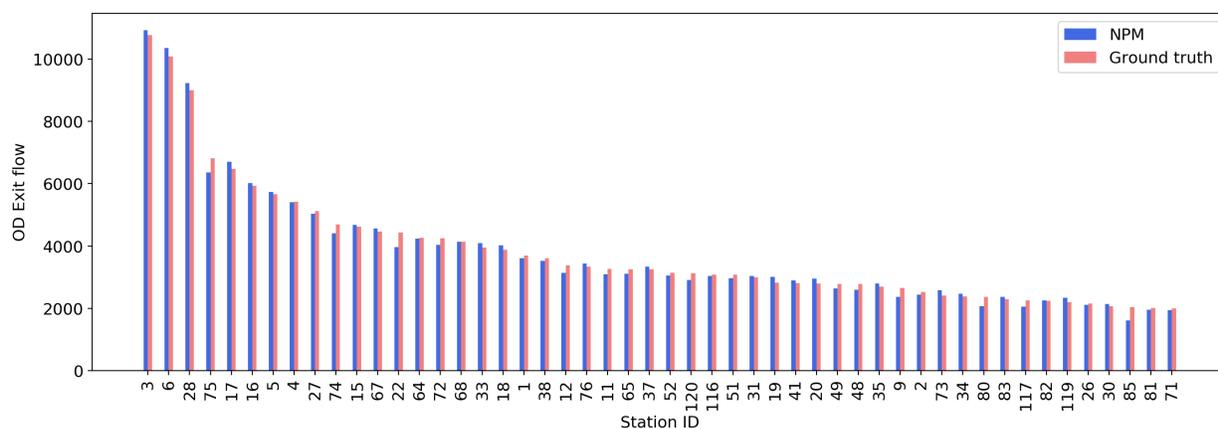
**FIGURE 3: Model Validation at the Admiralty Station (18:00-19:00)**

#### 14 *Crowding analysis*

15 The spatial distribution of OD entry flow and link flow are shown in Figure 5a. The critical stations  
 16 and links at different times can be identified. For example, the most busy station in the evening  
 17 peak is Central, located in the Central area of Hong Kong Island. During the peak hour (18:00-  
 18 19:00), more than 120,000 passengers entering this station. Similarly, the crowded lines can also



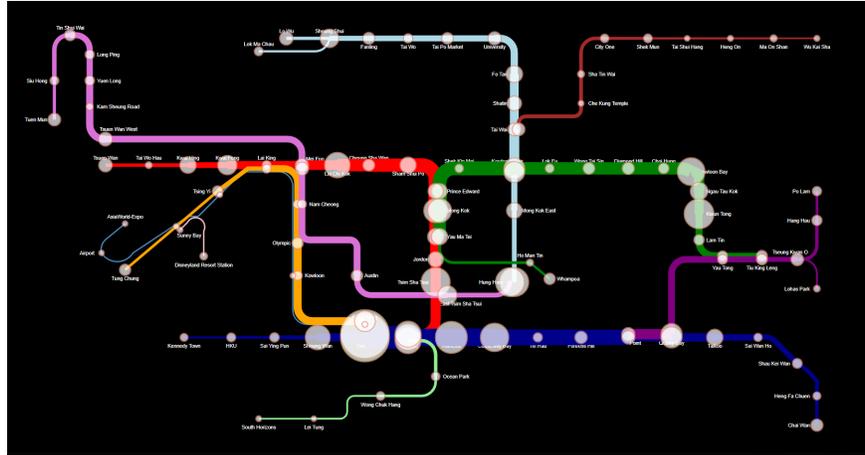
(a) 18:00-18:30



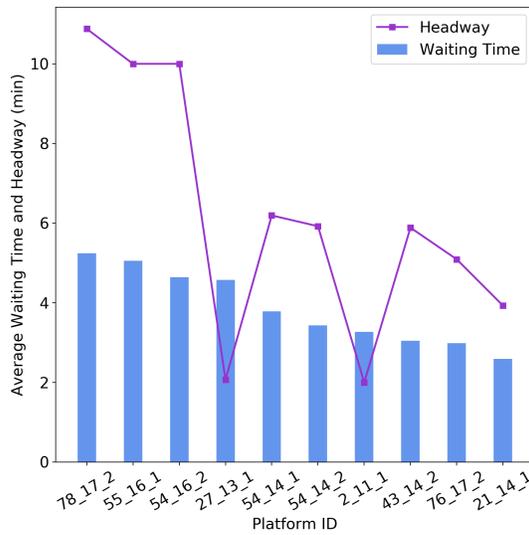
(b) 18:30-19:00

**FIGURE 4: OD Exit Flow Comparison (18:00-19:00)**

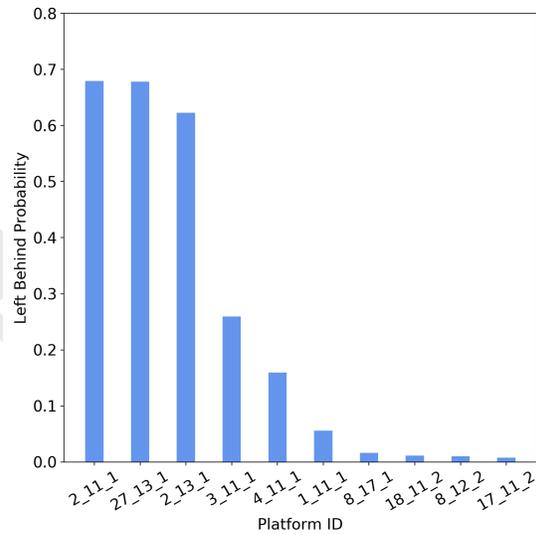
1 be identified at different times, such as the Island Line (blue), Tsuen Wan Line (red) and Kwun  
 2 Tong Line (green).  
 3 The spatial distribution of waiting time and left behind rate are shown in Figure 5b and 5c,  
 4 from which we can recognize the critical congested platforms. The left behind rate is defined as  
 5 the percentage of passengers being left behind. The platform ID in this figured is named by station  
 6 ID + line ID + direction. For example, 2\_11\_1 means the platform in Admiralty station serving the  
 7 Tsuen Wan Line in the north direction. The long waiting time may be caused either by headway  
 8 or by denied boarding. Figure 5b shows that most of the long-waiting-time platforms are due to  
 9 long headway, except for 27\_13\_1 and 2\_11\_1. Passengers have to wait for more than 1 headway  
 10 in these two platforms because of the congestion. Figure 5c shows the top 10 crowded platforms.  
 11 The most congested platform during peak hour is in Admiralty station in northbound Tsuen Wan  
 12 Line with north direction . The Wan Chai station in eastbound Island Line with (27\_13\_1) is also  
 13 very crowded with around 68% left behind rate.



(a) Spatial Distribution of OD Entry Flow and Link Flow



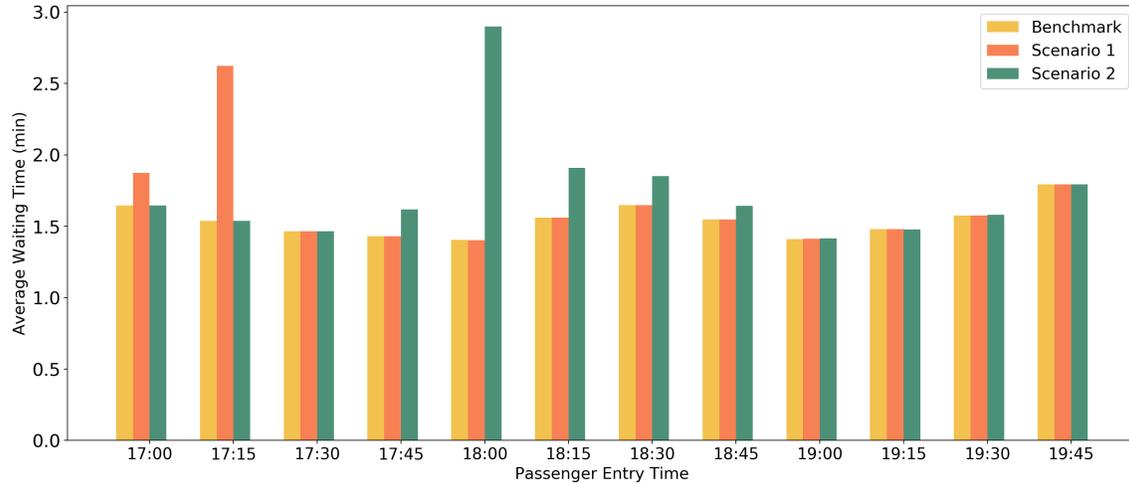
(b) Average Waiting Time



(c) Left Behind Rate

**FIGURE 5:** Network Crowding patterns (18:00-19:00)

1 *Network resilience evaluation*  
 2 Two scenarios with link disruption in Tsuen Wan Line in the 17:15-17:30 (Scenario 1) and 18:00-  
 3 18:15 (Scenario 2) periods are analyzed. The benchmark scenario is the one without link disruption.  
 4 The waiting time comparison among those three scenarios are shown in Figure 6. We find  
 5 that the disruption in 17:15-17:30 will affect the system for half an hour (passengers tap-in during  
 6 17:00-17:30), where the passengers' waiting time will recover to the normal state after 17:30.  
 7 However, if the link disruption occurs in 18:00-18:15, which is the busiest time for the network, the  
 8 whole system will take more than 1 hour to return to its normal state. Passengers who tap in during  
 9 18:00-18:15 suffer more than twice waiting time than the normal situation. The system recovery  
 10 time is an important indicator of network resilience. Thus, the proposed NPM can be used to test  
 11 the resilience of different network structures, operation/response strategies and types of incidents,  
 12 which shows more functionality than other models in the literature.



**FIGURE 6:** Waiting Time Comparison under the Link Disruption Scenarios

## 1 CONCLUSION AND DISCUSSION

2 This paper proposed a general framework for a Network Performance Model (NPM) with the  
 3 purpose of network performance monitoring. The major component of NPM is an event-driven  
 4 network loading module, which is capable of simulating passengers' walking, queuing, boarding,  
 5 and alighting processes. The applications of NPM include estimating crowding patterns, diagnos-  
 6 ing crowding sources and evaluating network resilience. A method for calibrating train capacity is  
 7 introduced, which explicitly recognizes that capacity may be different at different stations depend-  
 8 ing on the congestion level. The model is applied to the Hong Kong MTR network for illustration.  
 9 It is first validated with real-world observations, and then applied to demonstrate the capabilities  
 10 of performance monitoring. The spatial and temporal analysis of crowding patterns are conducted  
 11 from different perspectives.

12 The developments in this paper have been focused on a general framework, while the mod-  
 13 els and examples we presented in this paper still have some limitations. First, the accuracy of  
 14 NPM relies on the calibration of path choice and train capacity. As we target in a general frame-  
 15 work, future research could incorporate the AFC data-based path choice estimation into the model,  
 16 rather than calibration with survey data. Second, the effective train capacity model can be more  
 17 sophisticated. Behavioral models of passengers' willingness to board can be incorporated with the  
 18 optimization techniques. Third, the resilience evaluation module is a simplification of real-world  
 19 situation. Passengers may choose to use other modes when the incidents happen. And the operators  
 20 may provide information to passengers with recommendations. Future research should incorporate  
 21 more complex behaviors from both passenger's and operator's perspectives.

22 Many promising implications for both practice and research can be conducted in the future.  
 23 First, the model can be easily extended from performance monitoring to operations planning. By  
 24 adjusting the input, the proposed NPM is capable of evaluating different operation strategies, such  
 25 as adjustment in time table, different empty train arrangement, change in OD demand and network  
 26 structures. Second, the NPM can also be incorporated into a network assignment model. In the  
 27 case of long-term planning, the NPM can be used with a day-to-day updating framework to address  
 28 the increasing demand and change of choice behaviors with the user equilibrium assumption (22).

1 Third, by applying this model, we can further reveal other service satisfactory indicators, such as  
2 the availability of seats, the standing, walking times and the reliability of total travel time; hence,  
3 the results of this paper can applied on various choice modeling problems, serving future decision  
4 making processes.

#### 5 **AUTHOR CONTRIBUTION STATEMENT**

6 The authors confirm contribution to the paper as follows: study conception and design: B. Mo,  
7 Z. Ma, H.N. Koutsopoulos, J. Zhao; data collection: B. Mo, Z. Ma; analysis and interpretation of  
8 results: B. Mo, Z. Ma; draft manuscript preparation: B. Mo. All authors reviewed the results and  
9 approved the final version of the manuscript.

#### 10 **ACKNOWLEDGEMENTS**

11 The authors would like to thank Hong Kong Mass Transit Railway (MTR) for their support and  
12 data availability for this research.

#### 13 **REFERENCES**

- 14 1. Trépanier, M., C. Morency, and B. Agard, Calculation of transit performance measures  
15 using smartcard data. *Journal of Public Transportation*, Vol. 12, No. 1, 2009, p. 5.
- 16 2. Ma, X. and Y. Wang, Development of a data-driven platform for transit performance mea-  
17 sures using smart card and GPS data. *Journal of Transportation Engineering*, Vol. 140,  
18 No. 12, 2014, p. 04014063.
- 19 3. Nuzzolo, A., U. Crisalli, L. Rosati, and A. Comi, DYBUS2: a real-time mesoscopic transit  
20 modeling framework. In *2015 IEEE 18th International Conference on Intelligent Trans-  
21 portation Systems*, IEEE, 2015, pp. 303–308.
- 22 4. Nguyen, S. and S. Pallottino, Equilibrium traffic assignment for large scale transit net-  
23 works. *European journal of operational research*, Vol. 37, No. 2, 1988, pp. 176–186.
- 24 5. Lam, W. H.-K., Z. Gao, K. Chan, and H. Yang, A stochastic user equilibrium assignment  
25 model for congested transit networks. *Transportation Research Part B: Methodological*,  
26 Vol. 33, No. 5, 1999, pp. 351–368.
- 27 6. Cepeda, M., R. Cominetti, and M. Florian, A frequency-based assignment model for con-  
28 gested transit networks with strict capacity constraints: characterization and computation  
29 of equilibria. *Transportation research part B: Methodological*, Vol. 40, No. 6, 2006, pp.  
30 437–459.
- 31 7. Schmöcker, J.-D., A. Fonzone, H. Shimamoto, F. Kurauchi, and M. G. Bell, Frequency-  
32 based transit assignment considering seat capacities. *Transportation Research Part B:  
33 Methodological*, Vol. 45, No. 2, 2011, pp. 392–408.
- 34 8. Nuzzolo, A., F. Russo, and U. Crisalli, A doubly dynamic schedule-based assignment  
35 model for transit networks. *Transportation Science*, Vol. 35, No. 3, 2001, pp. 268–285.
- 36 9. Grube, P., F. Núñez, and A. Cipriano, An event-driven simulator for multi-line metro sys-  
37 tems and its application to Santiago de Chile metropolitan rail network. *Simulation Mod-  
38 elling Practice and Theory*, Vol. 19, No. 1, 2011, pp. 393–405.
- 39 10. Yao, X., B. Han, D. Yu, and H. Ren, Simulation-based dynamic passenger flow assignment  
40 modelling for a schedule-based transit network. *Discrete Dynamics in Nature and Society*,  
41 Vol. 2017, 2017.

- 1 11. Nunez, F., F. Reyes, P. Grube, and A. Cipriano, Simulating railway and metropolitan rail  
2 networks: From planning to on-line control. *IEEE Intelligent Transportation Systems Mag-*  
3 *azine*, Vol. 2, No. 4, 2010, pp. 18–30.
- 4 12. Silva, R., S. M. Kang, and E. M. Airoldi, Predicting traffic volumes and estimating the  
5 effects of shocks in massive transportation systems. *Proceedings of the National Academy*  
6 *of Sciences*, Vol. 112, No. 18, 2015, pp. 5643–5648.
- 7 13. Giambiasi, N. and J. C. Carmona, Generalized discrete event abstraction of continuous  
8 systems: GDEVS formalism. *Simulation Modelling Practice and Theory*, Vol. 14, No. 1,  
9 2006, pp. 47–70.
- 10 14. Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta, A modified logit route choice model  
11 overcoming path overlapping problems. Specification and some calibration results for in-  
12 terurban networks. In *Transportation and Traffic Theory. Proceedings of The 13th Interna-*  
13 *tional Symposium On Transportation And Traffic Theory, Lyon, France, 24-26 July 1996*,  
14 1996.
- 15 15. Li, W., Route and Transfer Station Choice Modeling in the MTR System, 2014, working  
16 paper.
- 17 16. Mo, B., Z. Ma, H. Koutsopoulos, and J. Zhao, Path Choice Calibration for Metro System  
18 Using Smart Card Data, 2019, working paper.
- 19 17. Preston, J., J. Pritchard, and B. Waterson, Train overcrowding: investigation of the pro-  
20 vision of better information to mitigate the issues. *Transportation research record*, Vol.  
21 2649, No. 1, 2017, pp. 1–8.
- 22 18. Liu, Z., S. Wang, W. Chen, and Y. Zheng, Willingness to board: a novel concept for mod-  
23 eling queuing up passengers. *Transportation Research Part B: Methodological*, Vol. 90,  
24 2016, pp. 70–82.
- 25 19. Massachusetts Bay Transportation Authority, *At What Level Does Crowd-*  
26 *ing Become Unacceptable*. [https://www.mbtackontrack.com/blog/](https://www.mbtackontrack.com/blog/48-at-what-level-does-crowding-become-unacceptable)  
27 [48-at-what-level-does-crowding-become-unacceptable](https://www.mbtackontrack.com/blog/48-at-what-level-does-crowding-become-unacceptable), 2016, accessed:  
28 2019-06-27.
- 29 20. Snoek, J., H. Larochelle, and R. P. Adams, Practical bayesian optimization of machine  
30 learning algorithms. In *Advances in neural information processing systems*, 2012, pp.  
31 2951–2959.
- 32 21. Sterbenz, J. P., D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and  
33 P. Smith, Resilience and survivability in communication networks: Strategies, principles,  
34 and survey of disciplines. *Computer Networks*, Vol. 54, No. 8, 2010, pp. 1245–1265.
- 35 22. Nuzzolo, A., U. Crisalli, and L. Rosati, A schedule-based assignment model with ex-  
36 plicit capacity constraints for congested transit networks. *Transportation Research Part C:*  
37 *Emerging Technologies*, Vol. 20, No. 1, 2012, pp. 16–33.