Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–11

DOI: 10.1177/0361198120914309
journals.sagepub.com/home/trr
SAGE

Capacity-Constrained Network Performance Model for Urban Rail Systems

Baichuan Mo[1], Zhenliang Ma[2], Haris N. Koutsopoulos[3], and Jinhua Zhao[4]

Abstract

This paper proposes a general network performance model (NPM) for monitoring the performance of urban rail systems using smart card data. NPM is a schedule-based network loading model with strict capacity constraints and boarding priorities. It distributes passengers over the network given origin-destination demand, operations, route choice, and effective train capacity. A Bayesian simulation-based optimization method for calibrating the effective train capacity is introduced, which explicitly recognizes that capacity may be different at different stations depending on congestion levels. Case studies with data from the Mass Transit Railway network in Hong Kong are used to validate the model and illustrate its applicability. NPM is validated using survey data on left-behind passengers and exiting passenger flow extracted from smart card data. The use of NPM for performance monitoring is demonstrated by analyzing the spatial-temporal crowding patterns in the system and evaluating dispatching strategies.

Increases in ridership are outpacing capacity in many large urban rail transit systems, including Hong Kong, London, New York, and Beijing (*1*). Crowding at stations and on trains is a concern because of its impact on safety, service quality, and operating efficiency. Monitoring network performance (e.g., waiting time on platforms, load on trains, etc.) is essential to help agencies and operators understand the system, inform passengers, and improve operating strategies.

Compared with traditional survey methods for service performance evaluation, data from automated fare collection (AFC) and automated vehicle location (AVL) systems provide ample opportunities for analysis in areas such as travel behavior, operations planning, and monitoring, and so forth (*2*, *3*). For performance monitoring, some performance indicators can be directly derived from automated data, including vehicle-kilometers, vehicle-hours, travel time reliability, OD (origin–destination) demand, and so forth (*4*, *5*). However, the problem of determining vehicle load and passenger waiting times is not trivial. Recently, a number of methods have been proposed to monitor passenger waiting times, passengers left behind at stations, and vehicle loads using AFC and AVL data (*6*). Network loading or assignment models can also be used. The main difference between network loading and assignment models lies in their behavioral assumptions. Network loading models assume that travel choices are known, while assignment models estimate the travel choices through user equilibrium criteria (network loading is a key component of transit assignment models).

Network assignment models are mainly used for planning applications. Nuzzolo et al. (*7*) proposed a dynamic schedule-based assignment model to simulate the within-day and day-to-day learning process of passengers' route choices. Nuzzolo et al. (*8*) proposed a mesoscopic transit modeling framework named DYBUS2 to provide real-time short-term predictions of network performance. Subsequently, Yao et al. (*9*) developed an agent-based simulation model for the Beijing metro system. The applicability of these models for performance monitoring

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA
[2]Public Transport Research Group (PTRG), Department of Civil Engineering, Monash University, Victoria, Australia
[3]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA
[4]Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA

Corresponding Author:
Zhenliang Ma, mike.ma@monash.edu

is limited, however. For performance monitoring, the interest is in the performance of the system on a particular day, which represents just one realization of operating conditions. Therefore, finding an equilibrium solution is not actually applicable. Network loading models are more appropriate for this purpose.

Network loading models provide detailed performance at different levels (station, line, train, passenger) given OD flows, path choices, and operating conditions. Therefore, they are suitable for modeling the performance of the network on a particular day. For example, Grube et al. (*10*) developed an event-based network loading model to simulate metro systems in Santiago de Chile. However, since such approaches have to be applied at the network level, they require assumptions about train capacity (as assignment models also do). The problem of determining actual train capacity is not trivial. It has been found that train capacity may vary depending on the crowding levels in trains and on platforms (*11*, *12*). Ma et al. (*6*) showed that ignoring this variability of train capacity results in biased estimates of passengers left behind, highlighting the importance of network loading models to capture this varied capacity in their representation of the system.

The paper develops a capacity-constrained, schedule-based network performance model (NPM), introduces a flexible train capacity model (termed effective capacity), and proposes an optimization methodology to calibrate effective train capacities using AFC data. NPM models detailed passenger trajectories, including access and egress, queuing, transferring, boarding, alighting, and being left behind. The contribution of this paper is threefold:

- Develops a data-driven metro network performance model with explicit effective train capacity constraints and boarding priorities, where the effective train capacity varies by station depending on congestion levels
- Proposes a simulation-based optimization method to calibrate the effective train capacity using AFC and AVL data
- Validates the NPM and demonstrates its applicability using data from the heavily-congested metro system in Hong Kong.

The remainder of the paper is organized as follows. The second section introduces the network performance model. The route choice and effective train capacity models, as well as the calibration methodology, are discussed in the third section. Case studies are presented in the fourth section to validate the NPM performance and demonstrate its functionality. The final section concludes the paper and discusses future research directions.

Network Performance Model

Figure 1 provides an overview of the main structure of NPM. It consists of four main components: the input, the network loading engine, the output of various performance indicators, and the calibration engine. The calibration engine provides the capability to calibrate model parameters using available real-life data. Green squares in the input section represent the parameters to be calibrated. Gray squares in the output section indicate model outputs that are also directly observable from AFC data (and therefore, can be used for calibration and validation purposes). The calibration engine compares the output journey times and OD exit flows with the observed values (ground truth) and uses the difference to adjust model parameters (e.g., effective capacity).

Inputs

The NPM inputs include dynamic OD demand, path choice fractions, train movement information, train capacity, and access/egress/transfer walking time. Table 1 summarizes the data inputs. It is assumed that the system is closed. The AFC data contains passengers' tap-in and tap-out times and stations (the complete OD entry demand by time period). The timetable provides the planned train arrival and departure information, while the AVL data provides the actual times. The walk time is assumed to be normally distributed with the mean and variance calculated from field observations.

Network Loading

Figure 2 summarizes the main structure of the network loading model. Three objects are defined: trains, queues, and passengers. Trains are characterized by routes, runs,

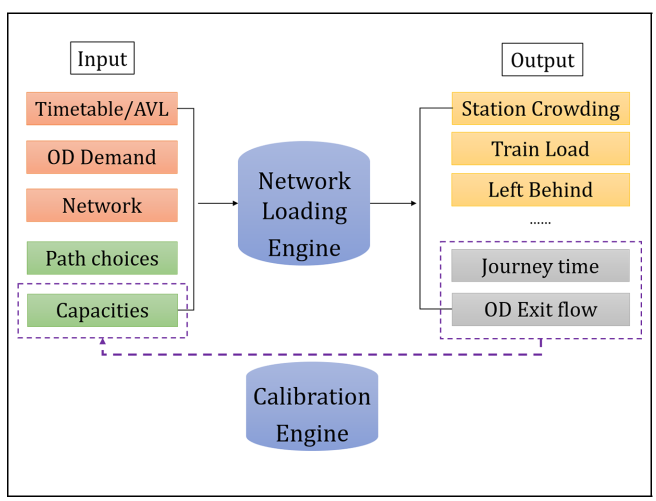

Figure 1. Structure of the network performance model.

Table 1. Input Variables and Data Sources

Input variable	Source
OD entry demand	AFC data
Path choice	Section 2.2.4
Train movement	Timetable or AVL data
Train capacity	Section 2.2.5
Access/egress/ transfer walk time	Measured by on-site observations or estimated from AFC/AVL data (*13*)

Note: OD = origin–destination; AFC = automated fare collection; AVL = automated vehicle location.

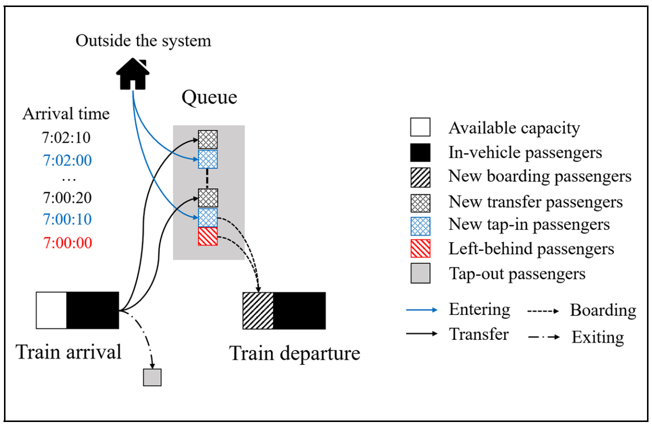

Figure 2. Structure of the network loading model.

current locations, and capacities. Passengers are queued based on their arrival times. Three different types of passengers are represented: left-behind passengers who were denied boarding from previous trains, new tap-in passengers from outside the system, and new transfer passengers from other lines. The left-behind passengers are usually at the head of the queue.

An event-based modeling framework is used to load the passengers onto the network. Two types of events are considered: train arrivals and train departures. The events are sorted by time and processed sequentially until all events are successfully completed during the analysis period. When a train arrives at a station, the offloaded passengers either transfer or exit. Transfer passengers join the boarding queue. When a train departs a station, passengers are loaded on the train up to its available capacity based on a first-come-first-served (FCFS) principle.

Preprocessing. Train event lists (arrivals and departures) are generated according to the actual train movement data from AVL (or timetable). Each event contains a train identity (ID), occurrence time, and location (platform). A passenger is randomly assigned to a route based on the corresponding path choice probability estimated from a path choice model. Random access and egress times are generated given corresponding distributions.

Train Arrivals. For an arrival event, the train offloads passengers who reach their destination or need to transfer at the station and updates its state (e.g., train load and in-vehicle passengers). For passengers who reach their destinations, their tap out times are calculated by adding their egress time. For those who transfer at the station, their arrival times at the next platform are calculated based on the transfer time distribution. The transfer passengers are added to the waiting queue in order of their arrival times.

Train Departures. For departure events, the queue on the platform is updated by the new tap-in passengers, that is, passengers who arrive at the platform after the last train departed are added into the queue based on their arrival times. Passengers board the train according to the FCFS principle until the train reaches its capacity. Passengers who cannot board are left behind and wait in the queue for the next train. The states of the train and the waiting queue are updated accordingly.

Route Choice. Route choice is usually modeled using the discrete choice framework, which assumes that decision makers maximize their utilities when making choices (*14*). The multinomial logit (MNL) model is a typical example of discrete choice models. For path choice problems, the C-logit model is often used. The C-logit is a variation of the MNL model which corrects for the fact that alternatives may not be independent because of path overlap. The C-logit incorporates an additional "cost" attribute, the commonality factor (CF), in the utility (*15*). The probability of choosing path i is given by:

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

where X_i is the attribute vector of path i, such as in-vehicle time, number of transfers, and so forth. $\mathbb{W}$ is the set of all alternative paths for the same OD pair. β_X and β_{CF} are the corresponding coefficients to be estimated. CF_i is the commonality factor of path i, defined as:

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

where $L_{i,j}$ is the number of common stations of paths i and j. L_i and L_j are the number of stations for paths i and j, respectively. γ is a positive constant which is determined based on empirical studies (*16*). In practice, the estimation of the route choice model parameters is based on survey data. Recent studies have also proposed estimation methods based on AFC data (*17–20*).

Effective Capacity. Train capacity is a vague concept. Normally trains may not reach their designed physical capacity for various reasons (e.g., passengers may decide not to board because of the crowding [*12*]). Therefore, assuming a fixed physical capacity may not be a reasonable assumption in real-world situations. This paper introduces the concept of *effective capacity*, which is the train capacity actually being utilized under crowding situations. Effective capacity is determined by three factors: (a) distribution of waiting passengers on the platform, (b) train load and distribution across the train, and (c) passengers' willingness to board a crowded train. Thus, train capacity is not constant but may vary across stations. The term effective capacity is used to differentiate it from the physical fixed train capacity. Effective capacity, as defined in this paper, is dynamic and changes depending on the crowding state of the train and the platform.

Based on previous studies, two factors are included in the effective capacity (C^e) model: the current train load when a train arrives at the platform (denoted as L) (*11*); and the number of queuing passengers on the platform (denoted as Q) (*21*). The base capacity of train i is $C_i = \theta_0 n_i$, where n_i is the number of cars of train i. C_i can be seen as the train load that represents acceptable service standards. At congested stations, passengers may still board a train even if it is already crowded (*11*), which makes the actual train load exceed C_i. Therefore, the effective capacity of train i at platform j ($C^e_{i,j}$) can be formulated as:

$$C^e_{i,j} = \begin{cases} \theta_0 n_i + \theta_1 L_{i,j} + \theta_2 Q_j & \text{if platform } j \text{ is in the list of congested stations} \\ \theta_0 n_i & \text{otherwise} \end{cases} \quad \forall i,j \quad (3)$$

The congested stations and time periods can be identified using AFC data (*6*). The term "platform" means a combination of station + line + direction. θ_0, θ_1, and θ_2 are the parameters to be estimated. θ_0 is a measurement of the service standard (passengers/car). C^e at congested stations and for congested trains is expected to be higher than that of stations/trains with less crowding, therefore θ_1 and θ_2 should be positive. Although a linear model is used here, the proposed approach is quite general and can accommodate more complex relationships between C^e, L, and Q.

Model Calibration

The calibration approach is illustrated using the effective capacity model, assuming two available types of ground-truth information: observed OD exit flows and observed journey time distribution (JTD). The calibration problem is formulated as an optimization problem. The objective function has two parts: the square error between model-derived OD exit flows and the observations, and the difference between model-derived and observed JTD. The optimization problem is formulated as:

$$\min_{\theta_1, \theta_2} w_1 \sum_{i,j,t} (q^{i,j_t} - \tilde{q}^{i,j_t})^2 + w_2 \sum_{i,j,t} D_{\mathrm{KL}}(f_{i,j_t}(x) \| \tilde{f}_{i,j_t}(x)) \quad (4a)$$

$$\text{s.t. } q^{i,j_t}, f_{i,j_t}(x) = \text{Network Loading}(\theta_0, \theta_1, \theta_2) \quad \forall i,j,t \quad (4b)$$

where q^{i,j_t} represents the number of passengers arriving from station i and exiting at station j during time interval t (i.e., OD exit flows). $\tilde{q}^{i,j_t}$ is the observed OD exit flow extracted from AFC data. w_1 and w_2 are the weights to balance the scale and the importance of the two parts. $f_{i,j_t}(x)$ is the probability density function of the estimated JTD of passengers who come from station i and exit at station j during time interval t. $\tilde{f}_{i,j_t}(x)$ is the observed JTD obtained from AFC data. Equation 4b indicates that q^{i,j_t} and $f_{i,j_t}(x)$ are obtained from the network loading model with θ_0, θ_1, and θ_2 as inputs. The difference of the two distributions is expressed using Kullback-Leibler divergence (D_{KL}):

$$D_{\mathrm{KL}}(f_{i,j_t}(x) \| \tilde{f}_{i,j_t}(x)) = \int_x f_{i,j_t}(x) \cdot \log \frac{f_{i,j_t}(x)}{\tilde{f}_{i,j_t}(x)} \mathrm{d}x. \quad (5)$$

Solving the problem in Equation 4 is a black-box optimization problem because of the non-analytical nature of the network loading process. In this study, a Bayesian simulation-based optimization (BSO) method (*22*) is applied. The BSO works by constructing a posterior distribution (surrogate function) that best approximates the objective function. As the number of observations grows, the posterior distribution improves, and the algorithm becomes more certain of which regions in the parameter space are worth exploring. Given the general optimization approach, more sophisticated effective capacity models could also be explored using the proposed BSO method.

Applications

NPM can be used to monitor performance in four dimensions: measuring crowding, diagnosing crowding causes,

Table 2. Crowding Indicators

Train	Train load
Platform	Waiting time Number of times left behind Left-behind rate (% of passengers left behind) Queue length
Link	Link flow

evaluating dispatching strategies, and evaluating network resilience.

Crowding Patterns

Crowding is one of the most important metrics for evaluating the level of service, safety, and so forth. The crowding indicators, directly obtained from NPM outputs, are summarized in Table 2. All indicators are time-dependent with flexible aggregated intervals. The left-behind rate is the probability of not boarding on the first train, which can be calculated as the number of passengers who have been left behind at least once divided by the total number of boarding passengers at the platform during a specific time period. The number of "times left behind" is the number of trains missed after the first train because of crowding on the trains.

Other service quality indicators can also be output by NPM, such as the availability of seats, the number of standing passengers, and journey time reliability.

Crowding Sources

Since the NPM models passengers' travel behavior at the individual level, the complete trajectories of all passengers are recorded. To diagnose the formation of crowding, the NPM can trace the sources of passengers on each link and passengers exiting at each station. The information about where passengers come from and how they contribute to loads at critical links of the system can inform operators to develop specific demand management strategies, such as promotions (*2*), peak pricing, and so forth.

Dispatching Strategies

Train dispatching strategies (e.g., headway adjustment, express trains) are basic instruments transit operators use on the supply side to deal with crowding or improve service reliability. Evaluating different train dispatching strategies can provide useful insights for improving service performance. NPM can be used to analyze network performance under different train dispatching strategies, such as operating express trains at different times during the peak. Such trains may skip stops to provide more capacity at crowded stations.

Network Resilience

Network resilience is the ability of the system to provide and maintain an acceptable level of service in the face of disruptions and other challenges to normal operations, such as incidents, large-scale natural disasters, special events, and so forth. Metro systems in large cities are facing more and more service disruptions because of increasing demand and aging infrastructure. These problems cause serious safety concerns and service performance deterioration. The responsible agencies are using various strategies, from demand management to infrastructure improvements, to prevent disruptions and mitigate their impacts. Approaches based on passenger information are still emerging as strategies transit agencies use to deal with disruptions. NPM can be used to analyze network resilience by comparing the performance indicators (e.g., waiting time) given different actions that operators may take when disruptions occur, such as providing information and dispatching shuttle buses.

Case Study

The NPM was demonstrated and validated using data from the Mass Transit Railway (MTR) system in Hong Kong (Figure 3). The system serves the urbanized areas of Hong Kong Island, Kowloon, and the New Territories, and consists of 11 lines with 218.2 km (135.6 mi) of rail, 159 stations, including 91 heavy rail stations and 68 light rail stops. The network serves over 5 million trips on an average weekday. For the urban heavy rail lines, trip transactions are recorded when entering and exiting the system. The Admiralty (ADM) station is one of the most crowded stations, with high volumes of passengers boarding and transferring there. In this case study, the airport express and light rail transit services were not considered since they are separated from the urban heavy rail lines. Passengers who enter the urban heavy rail lines from the airport express and light rail services need to tap-in again.

System Settings

AFC data from a weekday in March 2017 were used to generate the OD entry demand and conduct effective capacity calibration. Since AVL data is not available for all lines, the timetable was used to provide train movement information (the actual train movements may differ from the timetable). Considering the high on-time performance of the MTR system (99.9% on-time rate) (*23*), this is a reasonable approximation. Since the evening

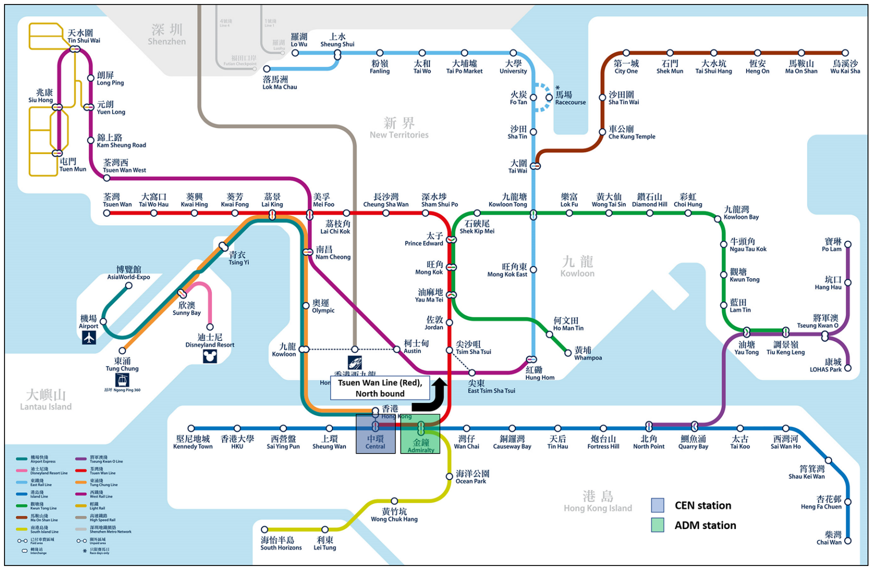

Figure 3. Hong Kong Mass Transit Railway (MTR) metro system map.

Table 3. Route Choice Model Estimation Results

	Estimate	Standard error	*t*-value	
In-vehicle time	−0.147	0.011	−13.64	***
Relative walk time	−1.271	0.278	−4.56	***
Number of transfers	−0.573	0.084	−6.18	***
CF	−3.679	1.273	−2.89	**
$\rho^2 = 0.54$				

Note: CF = commonality factor.
*** = *p*<.01; ** = *p*<.05.

peak is the most congested period, only the period from 17:00 to 20:00 was considered for the model application. The warm-up and cool-down times are both set as one hour. The running time is about 15 min on a personal computer with a 3.6 GHz CPU and 32 GB of RAM.

Model Calibration

The route choice model used to calculate path choice fractions for various OD pairs was estimated using data from a survey of MTR users (*16*). A total of 31,640 passengers participated in the survey, with 26,996 valid responses. The model estimation results are shown in Table 3. The main explanatory variables are the total in-vehicle time, relative walk time, and the number of transfers. The relative walk time is defined as the total walk time (access + transfers + egress) divided by the map distance of the path. All variables are statistically significant with expected signs. Routes with high in-vehicle, walk, and transfer times are less likely to be chosen by passengers. Based on the estimated parameters, the path shares for all paths for the OD pairs in the MTR system were calculated.

The optimization problem (Equation 4) is used for effective capacity calibration. The weights in the objective function were set to $w_1 = 1$ and $w_2 = 1000$ for the error in OD exit flows and JTD, respectively. The optimal coefficients are $\theta_0^* = 231.6$, $\theta_1^* = 0.0732$, and $\theta_2^* = 0.0607$. The value of θ_0^* is close to the MTR standard (230 passengers/car). The signs of θ_1^* and θ_2^* are consistent with the discussion in the previous section.

For comparison purposes, a fixed train capacity model is used as the benchmark to compare with the effective capacity model. The fixed capacity for train i at platform j ($C_{i,j}^f$) is defined as:

$$C_{i,j}^f = \theta_f n_i \qquad \forall i,j \tag{5}$$

Three different values of θ_f were tested for comparison, that is, $\theta_f = 230$, $\theta_f = 245$, and $\theta_f = 260$ passengers/car.

Results

Model Validation. To validate the performance of the NPM, field observations at ADM station (Tsuen Wan Line, north direction) on the same day as the AFC data were used for comparison. The data were collected by MTR employees who counted passengers on the platform during the period 18:00–19:00 hours. Left-

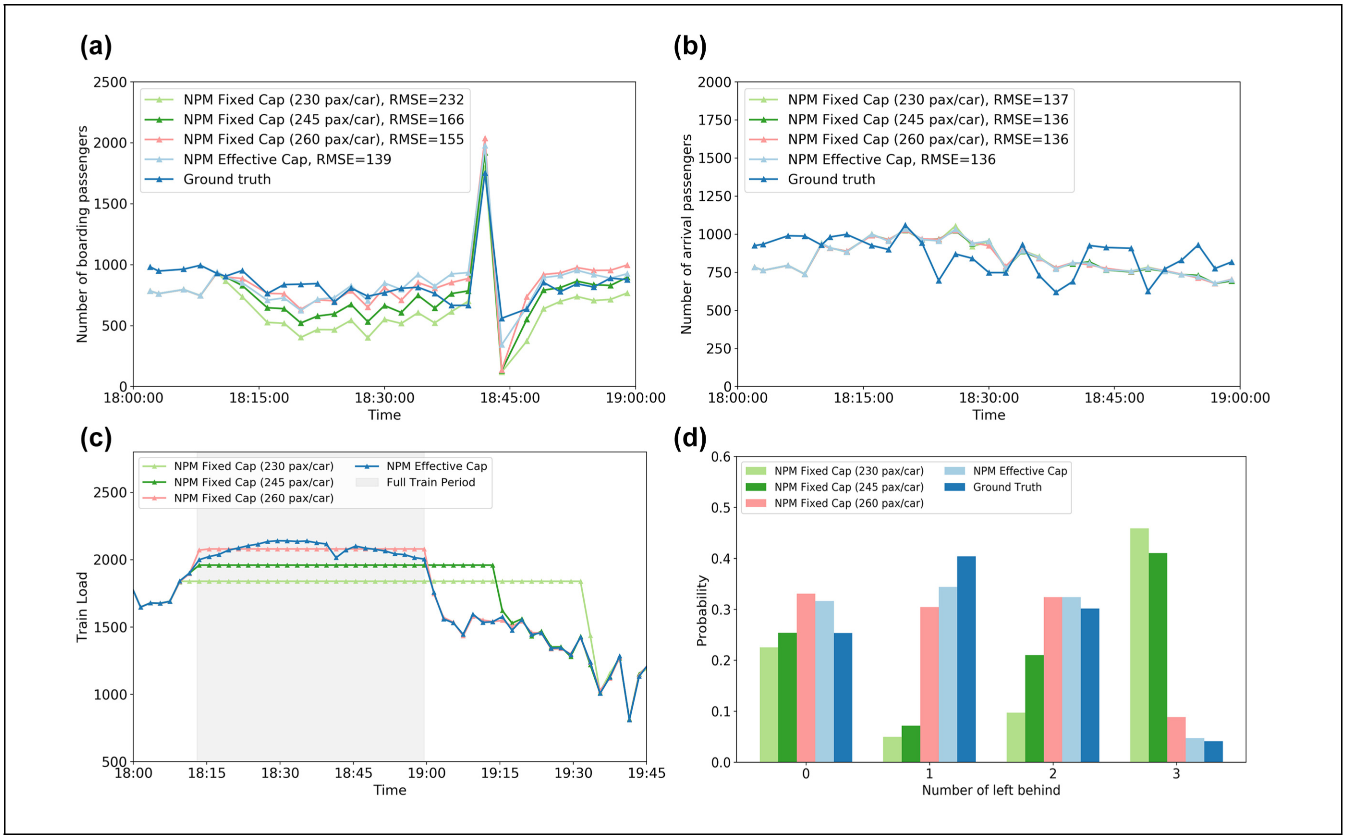

Figure 4. Model validation at the Admiralty station, Tsuen Wan Line, northbound (18:00–19:00): (*a*) boarding passengers; (*b*) arriving passengers; (*c*) train load; and (*d*) left behind.

behind passengers, the number of arriving passengers (sum of the new tap-in and transfer passengers), and the number of passengers boarding each train were recorded.

Figure 4 compares the fixed capacity model, effective capacity model, and ground truth for different indicators. The number of boarding and arriving passengers from the effective capacity model matches the ground truth observations well, as shown in Figure 4, *a* and *b*. The peak in Figure 4*a* is because of an empty train dispatched from the upstream terminal station, so that more capacity was available to serve the passengers at the crowded ADM station. The root mean square error (RMSE) of the number of boarding and arriving passengers for each train is reported. The arrival passenger curves (Figure 4*b*) for fixed capacity and effective capacity models are nearly the same. This is expected because the number of arriving passengers mainly depends on the OD demand and path shares and these two inputs are the same for the fixed and effective capacity models. However, the estimates of boarding passengers from the effective capacity model are closer to the observed values compared with the fixed capacity model (Figure 4*b*). The results support the importance of using the effective capacity since boarding passengers are directly related to train capacity.

A comparison of the train load between the fixed capacity and effective capacity models is shown in Figure 4*c* (ground truth train load data were not available). The trains at ADM station are always full from 18:15 to 19:00. Figure 4*c* shows that the effective capacity can capture the variability of the train load because of the change in crowding levels over time. Other studies, based on actual observations of train loads, also support this finding (*11*).

Figure 4*d* compares the percentage of passengers who are left behind at different times according to the models and ground-truth observations. The effective capacity model provides a more accurate estimation of left-behind passengers than the fixed capacity model, which is consistent with findings in Ma et al. (*6*).

Figure 5 compares the exit flows from the NPM against the actual observations extracted from the AFC data. The top 30 stations in relation to exit flows are displayed. The RMSE of the exit flows for each model is also reported. The results from the effective capacity model match the ground truth well and outperform the fixed capacity models.

Overall, the proposed effective capacity NPM can capture real-world situations well and has the potential to be an effective tool for performance monitoring.

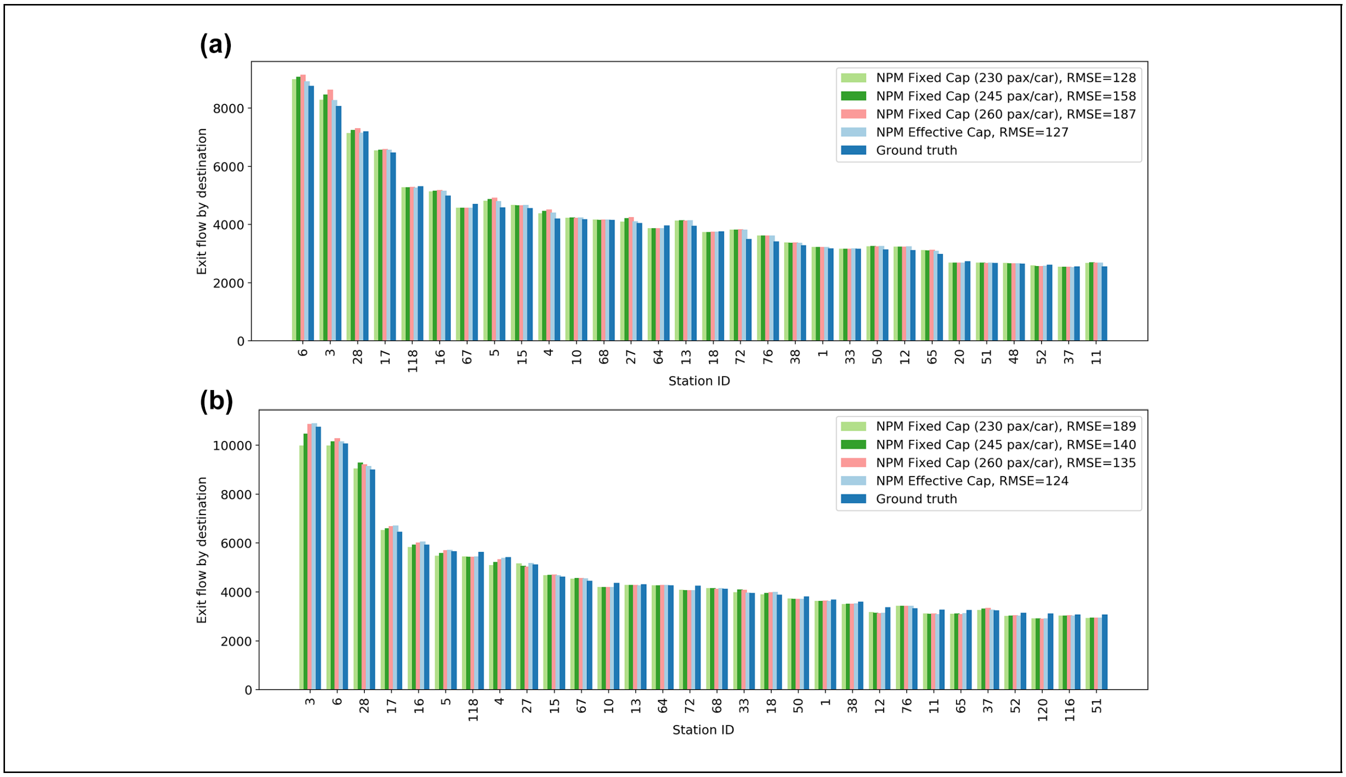

Figure 5. Exit flow comparison (18:00–19:00): (*a*) 18:00-18:30; and (*b*) 18:30-19:00.

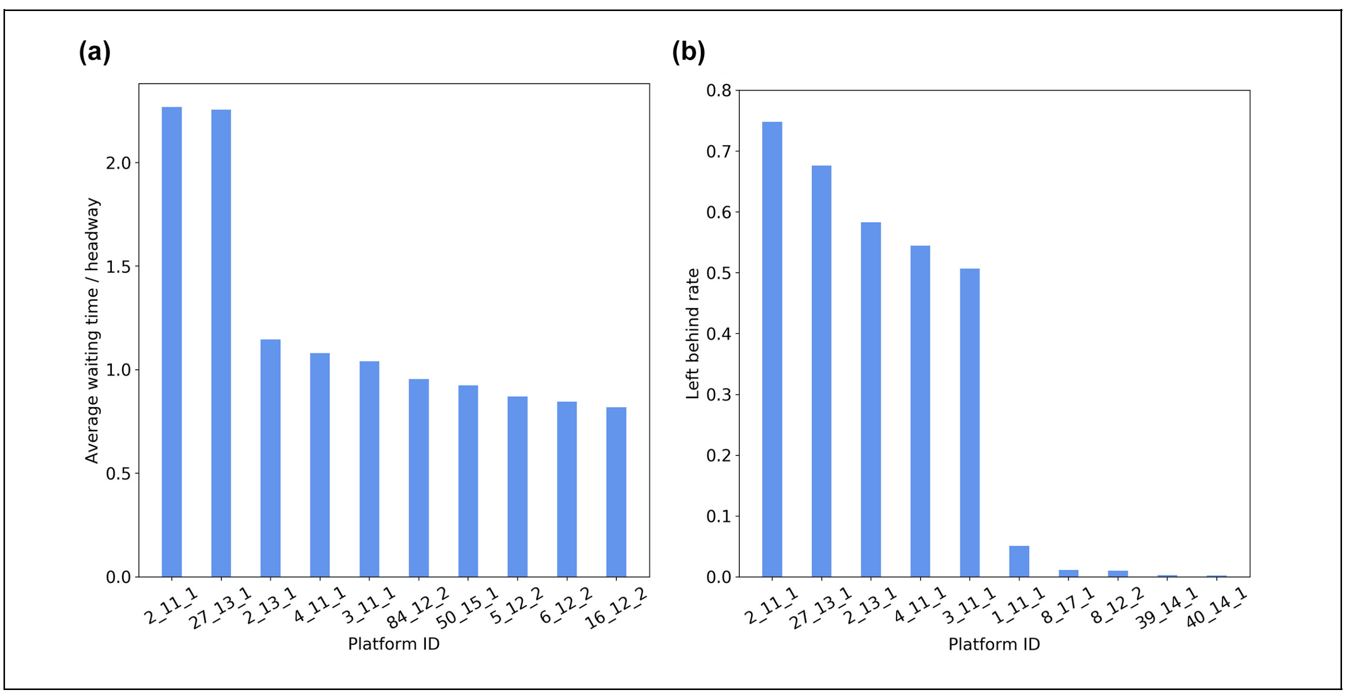

Figure 6. Network crowding patterns (18:00–19:00): (*a*) average wait time-headway ratio; and (*b*) left behind rate.

Crowding Analysis. In a congested rail system, waiting times increase because of passengers left behind because of full trains. Figure 6 shows the wait time:headway ratio and left-behind rates for the 10 most crowded platforms in the network. The wait time:headway ratio is defined as the passenger's average wait time divided by the

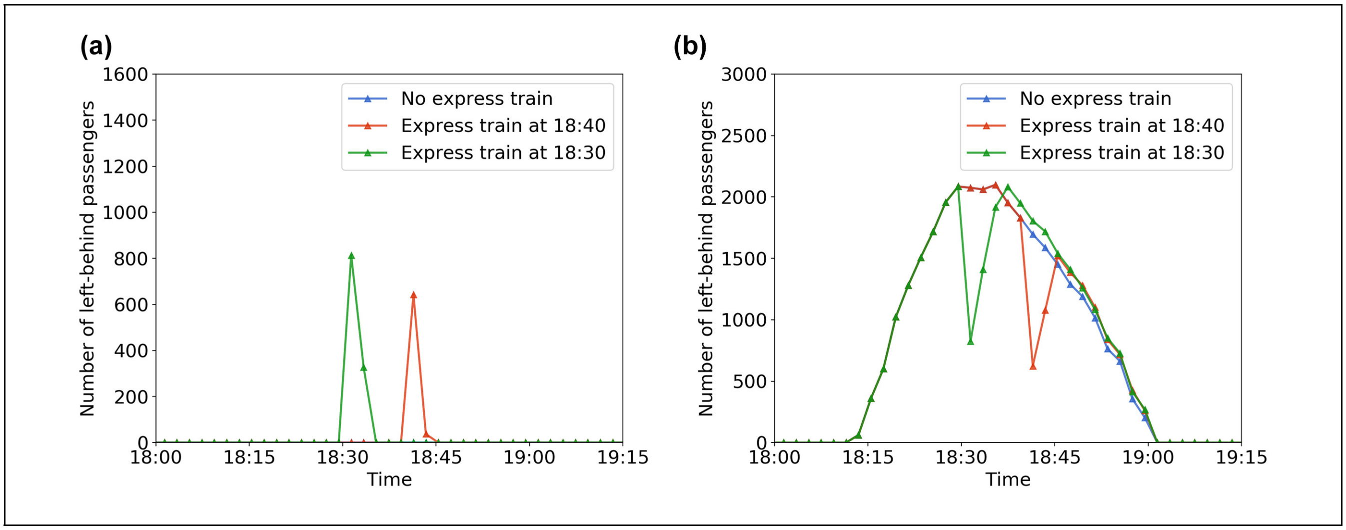

Figure 7. Comparison of left-behind passengers for different dispatching strategies: (*a*) CENTRAL (CEN) Station; and (*b*) ADMIRALTY (ADM) station.

headway. Under normal conditions for operations with small headway variabilities and no capacity constraints (assuming random passenger arrivals), the ratio has a value close to 0.5. The platform ID in Figure 6 reflects the station ID + line ID + direction. For example, 2_11_1 is the platform at ADM station serving the Tsuen Wan Line in the north direction. Figure 6*a* shows that, at platforms 27_13_1 and 2_11_1, passengers have to wait for an average of two headways. That means a passenger is expected to wait for more than two trains before being able to board. Figure 6*b* shows the top 10 platforms by their left-behind rates (probability of being left behind at least once). The most congested platform during the evening peak is at ADM station (Tsuen Wan Line northbound). The platform has a left-behind rate of about 0.75, consistent with the high wait time:headway ratio shown in Figure 6*a*.

Evaluation of Dispatching Strategy. A key application of NPM is the evaluation of different dispatching strategies. As shown in Figure 4*a*, an empty train is dispatched as an express from CEN to ADM at 18:40 to serve a large number of passengers typically waiting at ADM. NPM can be used to test how effective such strategies are in relieving congestion. For comparison purposes, two additional scenarios were also tested: (a) no express train is dispatched; and (b) the express train is dispatched at 18:30. Figure 7 compares the number of left-behind passengers at CEN and ADM, which are the first two stations on the Tsuen Wan Line, northbound. Dispatching an express train transfers the congestion from ADM to CEN. The strategy temporally decreases the number of left-behind passengers at ADM. The dispatching time does not significantly influence the crowding patterns at ADM. However, dispatching the express train at 18:30 seems to reduce the number of left-behind passengers more than at 18:40 as it targets better the peak of the crowding conditions.

Conclusion

The paper proposed a general network performance model (NPM) for monitoring network performance. The major component of NPM is an event-driven network loading module, which is capable of simulating passengers' walking, queuing, boarding, and alighting processes. NPM can be used to infer crowding patterns and evaluate dispatching strategies. An important contribution of the paper is a method for calibrating train capacity, which explicitly recognizes that capacity may be different at different stations, depending on the crowding levels on the platform and the train. The model is applied using a case study with data from Hong Kong's MTR network. The results show that NPM is able to replicate actual conditions (based on AFC data and direct observations of crowding levels at one station). NPM was also used to evaluate the effectiveness of various dispatching strategies in reducing onboard crowding. The results highlight the importance of calibrating the train capacity and support the value of the model for performance monitoring and evaluation of operating strategies. Future research should focus on jointly calibrating the parameters of the path choice and train capacity models. The effective train capacity model can also be improved to better reflect passengers' willingness to board. Further sensitivity tests on the impacts of accuracy of inputs could also be interesting, particularly for train operations.

Acknowledgments

The authors thank the Hong Kong Mass Transit Railway (MTR) for the support from staff and supply of data for this research.

Author Contributions

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

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Koutsopoulos, H. N., Z. Ma, P. Noursalehi, and Y. Zhu. Chapter 10 - Transit Data Analytics for Planning, Monitoring, Control, and Information. In *Mobility Patterns, Big Data and Transport Analytics* (C. Antoniou, L. Dimitriou, and F. Pereira, eds.), Elsevier, Amsterdam, 2019, pp. 229–261.
2. Ma, Z., and H. N. Koutsopoulos. Optimal Design of Promotion Based Demand Management Strategies in Urban Rail Systems. *Transportation Research Part C: Emerging Technologies*, Vol. 109, 2019, pp. 155–173.
3. Nassir, N., M. Hickman, and Z.-L. Ma. A Strategy-Based Recursive Path Choice Model for Public Transit Smart Card Data. *Transportation Research Part B: Methodological*, Vol. 126, 2019, pp. 528–548.
4. Trépanier, M., C. Morency, and B. Agard. Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation*, Vol. 12, No. 1, 2009, p. 5.
5. Ma, X., and Y. Wang. Development of a Data-Driven Platform for Transit Performance Measures Using Smart Card and GPS Data. *Journal of Transportation Engineering*, Vol. 140, No. 12, 2014, p. 04014063.
6. Ma, Z., H. N. Koutsopoulos, Y. Chen, and N. H. Wilson. Estimation of Denied Boarding in Urban Rail Systems: Alternative Formulations and Comparative Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 771–778.
7. Nuzzolo, A., F. Russo, and U. Crisalli. A Doubly Dynamic Schedule-Based Assignment Model for Transit Networks. *Transportation Science*, Vol. 35, No. 3, 2001, pp. 268–285.
8. Nuzzolo, A., U. Crisalli, L. Rosati, and A. Comi. DYBUS2: A Real-Time Mesoscopic Transit Modeling Framework. *Proc., 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE, Las Palmas, Spain, 2015, pp. 303–308.
9. Yao, X., B. Han, D. Yu, and H. Ren. Simulation-based Dynamic Passenger Flow Assignment Modelling for a Schedule-Based Transit Network. *Discrete Dynamics in Nature and Society*, Vol. 2017, 2017, p. 15.
10. Grube, P., F. Núñez, and A. Cipriano. An Event-Driven Simulator for Multi-Line Metro Systems and Its Application to Santiago de Chile Metropolitan Rail Network. *Simulation Modelling Practice and Theory*, Vol. 19, No. 1, 2011, pp. 393–405.
11. Liu, Z., S. Wang, W. Chen, and Y. Zheng. Willingness to Board: A Novel Concept for Modeling Queuing up Passengers. *Transportation Research Part B: Methodological*, Vol. 90, 2016, pp. 70–82.
12. Preston, J., J. Pritchard, and B. Waterson. Train Overcrowding: Investigation of the Provision of Better Information to Mitigate the Issues. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. 2649: 1–8.
13. Zhu, Y., H. N. Koutsopoulos, and N. H. Wilson. A Probabilistic Passenger-to-Train Assignment Model Based on Automated Data. *Transportation Research Part B: Methodological*, Vol. 104, 2017, pp. 522–542.
14. Ben-Akiva, M. E., S. R. Lerman, and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*, Vol. 9. MIT Press, Cambridge, Mass., 1985.
15. Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta. A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks. *Proc., 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 1996.
16. Li, W. *Route and Transfer Station Choice Modeling in the MTR System*. Working Paper, MIT Transit Lab, 2014.
17. Sun, L., Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen. An Integrated Bayesian Approach for Passenger Flow Assignment in Metro Networks. *Transportation Research Part C: Emerging Technologies*, Vol. 52, 2015, pp. 116–131.
18. Zhu, Y., H. N. Koutsopoulos, and N. H. Wilson. Passenger Itinerary Inference Model for Congested Urban Rail Networks. *Transportation Research C: Emerging Technologies*, 2020, under review.
19. Mo, B., Z. Ma, H. Koutsopoulos, and J. Zhao. Assignment-Based Path Choice Estimation for Metro System Using Smart Card Data. *Proc., 24th International Symposium on Transportation & Traffic Theory (ISTTT)*, 2020.
20. Mo, B., Z. Ma, H. Koutsopoulos, and J. Zhao. *Calibrating Route Choice for Urban Rail System: A Comparative Analysis Using Simulation-Based Optimization Methods*. Presented at 99th Annual Meeting of the Transportation Research Board, Washington, D.C., 2020.

21. Massachusetts Bay Transportation Authority. *At What Level Does Crowding Become Unacceptable?* 2016. https://www.mbtabackontrack.com/blog/48-at-what-level-does-crowding-become-unacceptable. Accessed June 27, 2019.
22. Snoek, J., H. Larochelle, and R. P. Adams. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems*, 2012, pp. 2951–2959.
23. Mass Transit Railway, Hong Kong. *MTR Maintains 99.9 Percent On-Time Performance in 2017*. 2018. https://www.mtr.com.hk/archive/corporate/en/press_release/PR-18-009-E.pdf. Accessed February 18, 2020.