T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

Title	Fundamental diagram of urban rail transit considering train–passenger interaction
Authors	Toru Seo, Kentaro Wada, Daisuke Fukuda
Citation	Transportation
Pub. date	2022, 4
Creative Commons	Information is in the article.

Powered by T2R2 (Science Tokyo Research Repository)

Transportation
https://doi.org/10.1007/s11116-022-10281-0

Fundamental diagram of urban rail transit considering train–passenger interaction

Toru Seo[1] · **Kentaro Wada**[2] · **Daisuke Fukuda**[3]

Accepted: 18 March 2022
© The Author(s) 2022

Abstract
Urban rail transit often operates with high service frequencies to serve heavy passenger demand during rush hours. Such operations can be delayed by two types of congestion: train congestion and passenger congestion, both of which interact with each other. This delay is problematic for many transit systems, since it can be amplified due to the interaction. However, there are no tractable models describing them; and it makes difficult to analyze management strategies of congested transit systems in general and tractable ways. To fill this gap, this article proposes simple yet physical and dynamic model of urban rail transit. First, a fundamental diagram of transit system (i.e., theoretical relation among train-flow, train-density, and passenger-flow) is analytically derived considering the aforementioned physical interaction. Then, a macroscopic model of transit system for dynamic transit assignment is developed based on the fundamental diagram. Finally, accuracy of the macroscopic model is investigated by comparing to microscopic simulation. The proposed models would be useful for mathematical analysis on management strategies of urban rail transit systems, such as optimal dynamic pricing for travel demand management.

Keywords Public transport · Rush hour · Fundamental diagram · Macroscopic fundamental diagram · Dynamic transit assignment

Introduction

Urban rail transit systems such as metro is handling significant transportation needs of metropolitan areas (Vuchic 2005). Its most notable usage is the morning commute, in which heavy passenger demand is concentrated in a short time period. It is known that such transit systems often suffer from delays caused by congestion, even if no serious incidents or accidents occur (Kato et al. 2012; Tirachini et al. 2013; Kariyazaki et al. 2015). Therefore,

This paper is an enhanced version of Seo et al. (2017a, 2017b).

✉ Toru Seo
seo.t.aa@m.titech.ac.jp

[1] Tokyo Institute of Technology, 2-12-1-M1-13 O-okayama, Meguro, Tokyo 152-8552, Japan

[2] University of Tsukuba, 3F1211, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

[3] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Published online: 22 April 2022

appropriate management of transit systems is required; especially, travel demand management for mass transit systems has been gaining attention recently (Halvorsen et al. 2019; Huan et al. 2021).

One of the approaches to find management strategies of transit systems is theoretical analysis with simplifications, such as use of certain static models with constant travel time (de Cea and Fernández 1993; Tabuchi 1993; Kraus and Yoshida 2002; Tian et al. 2007; Gonzales and Daganzo 2012; Trozzi et al. 2013; de Palma et al. 2015a, b). In this approach, general policy implications can be obtained thanks to the simplicity and tractability of the analysis. However, they may not be sufficient to investigate dynamic (i.e., time-varying) operation and demand management strategies.

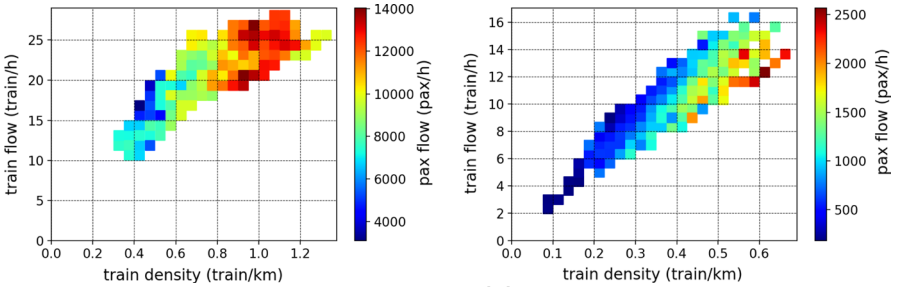
In congested transit systems, dynamical interaction among trains and passengers plays essential roles to determine the system's operational behavior, and the travel time can be dynamically and significantly changed due to this interaction. For instances, there are two types of congestion in transit systems:

- *train-congestion*: congestion involving consecutive trains using the same tracks,
- *passenger-congestion*: congestion of passengers who are boarding to a train, namely, bottleneck congestion at the doors of a train while it is stopped at a station (Lam et al. 1998; Wada et al. 2012; Kariyazaki et al. 2015),[1]

and these two types of congestion interact with each other and cause delay (Newell and Potts 1964; Kusakabe et al. 2010; Wada et al. 2012; Kato et al. 2012; Tirachini et al. 2013; Kariyazaki et al. 2015; Cuniasse et al. 2015). The "knock-on delay" is a well-known phenomenon describing "a train's delay which is caused by other trains in front of it" (Carey and Kwieciński 1994). Some of the knock-on delays can be considered as caused by the dynamic train–passenger interaction (Kariyazaki et al. 2015). Specifically, trains are delayed due to the large number of boarding and alighting passengers at stations, and such delay propagates to the subsequent trains—this can be a train equivalent of the "bus bunching" (Newell and Potts 1964; Daganzo 2009). For example, assume that passenger-congestion happened temporally due to high demand. It would extend the dwelling time of a train at a station. Then, this extended dwell time would interrupt the operation of subsequent trains, and cause train-congestion on the track. It would deteriorate the passenger throughput, and thus the passenger-congestion at stations would intensify. This kind of dynamical phenomena cannot be captured by static models.

A consequence of such dynamical passenger–train interaction can be found in macroscopic states of transit systems. Figure 1 shows observed 3-dimensional relations among states of transit systems, that is, train-flow (train/h), train-density (train/km), and passenger-flow (passenger/h). The visualization is based on the concepts of the fundamental diagram (Greenshields 1935) of urban traffic. Although Fig. 1a and 1b show data from completely different transit systems with completely different scales (e.g., the number of cars per one train was 10 for Tokyo and 1–6 for Boston), they have similar patterns. First, as the passenger-flow increases, the train-density increases; this could be a result of transit operators responding to increased passenger-demand. Second, as the passenger-flow increases, the average service speed of trains and train-flow decreases; this could be a result of the aforementioned congestion due to the interaction among trains and passengers. It would be

[1] Note that passenger-congestion differs from in-vehicle *passenger-crowding* (Kumagai et al. 2020), which results in discomfort due to standing and crowding, but is not necessarily cause any delay directly.



(a) Tokyu Den-en-toshi Line, Tokyo, Japan **(b)** Boston Subway Red Line, Massachusetts, US

Fig. 1 Observed 3-dimensional relations among train-flow, train-density, and passenger-flow. The *x*-coordinates represent the train-flow (i.e., the number of trains per a kilometer), the *y*-coordinates represent the train-flow (i.e., the number of passing trains per an hour), the slope from the origin to an arbitrary point indicates the train-speed (i.e., the average speed of trains) of that point, and the color of each point represents the passenger-flow (i.e., average demand of passengers per an hour) of that point. "pax" in the figure is an abbreviation for "passenger". Data sources: Fukuda et al. (2019), Zhang and Wada (2019), Tokyu Railways, and Massachusetts Bay Transportation Authority

preferable if we have a theoretical model of this phenomena, because it would be useful to obtain general principles on transit operations; however, to our knowledge, such a model does not exist in the literature.[2]

This study derives a theoretical relation among the state variables of transit systems similar to Fig. 1 based on the microscopic operation principles. It is modeled as a *fundamental diagram (FD)*, which is a well-known concept in the vehicular traffic flow theory. The original FD describes relation between vehicular flow and density, and it can be used to describe dynamic evolution of traffic by combining with other principles in a tractable manner (Lighthill and Whitham 1955; Richards 1956; Mahmassani et al. 1984; Geroliminis and Daganzo 2007). In fact, several recent studies have employed FDs of transit systems to describe train-congestion by modeling the relation between train-flow and train-density (Cuniasse et al. 2015; Corman et al. 2019; de Rivera and Dick 2021). The novel feature of this study is the incorporation of passengers in an analytical way. Furthermore, this study develops a dynamic transit assignment method based on the proposed FD.

This study proposes tractable models of the dynamics of urban rail transit considering the physical interaction between train-congestion and passenger-congestion. In "Microscopic Model of Rail Transit System" section, a microscopic model of a rail transit system is introduced based on a passenger boarding model and a train cruising model. In "Fundamental Diagram of Rail Transit System" section, the operation performance of the microscopic model is analyzed. Specifically, a mathematically tractable relation among train-flow, train-density, and passenger-flow is derived—namely, a fundamental diagram (FD). The model can be also viewed as a variation of 3-dimensional macroscopic fundamental

[2] Several detailed operation models have been proposed to capture the detailed mechanism of the dynamics of interaction (see Vuchic 2005; Koutsopoulos and Wang 2007; Parbo et al. 2016; Cats et al. 2016; Li et al. 2017; Alonso et al. 2017; Cunha et al. 2021, and references therein), and these have been used to develop efficient operation schemes. However, these models are often based on microscopic simulation, and thus their purposes tend to be case-specific optimization and evaluation. It would be difficult to use them to derive the relation depicted in Fig. 1 or to obtain general or strategic policy implications for management strategies, as they are essentially complex and intractable.

diagram (Mahmassani et al. 1984; Geroliminis and Daganzo 2007) with an analytical derivation. This is the key contribution of this study. In "Dynamic Model Based on Fundamental Diagram" section, a macroscopic loading model of a transit system is developed based on the proposed FD. The model describes the aggregated behavior of trains and passengers in a urban-scale spatial domain based on the FD. In Section "Validation of the Macroscopic Model", the approximation accuracy and other properties of the proposed macroscopic model are investigated through a comparison with microscopic simulation. Section "Conclusion" concludes this article. Note that empirical validation of the proposed model based on actual data is out of scope of this study. Such validation is now being conducted by some of the authors and preliminary results that support the model have been obtained (Fukuda et al. 2019; Zhang and Wada 2019).

Microscopic Model of Rail Transit System

This section introduces a microscopic model of rail transit system, from which we derive the FD in Section "Fundamental Diagram of Rail Transit System". It consists of two microscopic operation principles, namely, a *passenger boarding model* which describes the train's dwell behavior at a station for passenger boarding and a *train cruising model* which describes the cruising behavior on the railroad. This microscopic model has been proposed by Wada et al. (2012) to analyze train bunching.

Rail Transit Operation Principles

Consider a railway system on a single line track, where trains and stations are indexed by m and i, respectively. We assume that all trains stop at every station. Let $t_{m,i}$ be the arrival time of train m at station i. Then, a dynamical system that represents each train motion is given by

$$t_{m,i+1} = t_{m,i} + b_{m,i} + c_{m,i} \quad \forall m, i \tag{1}$$

where $b_{m,i}$ is the passenger boarding time of train m at station i, and $c_{m,i}$ is the trip time of train m between stations i and $i+1$, which are determined by the two operational submodels (see Figure 2).

The passenger boarding time is modeled using a queuing model. That is, the flow-rate of passenger boarding is assumed to be constant, μ_p; and there is a buffer time (e.g., time required for door opening/closing), g_b, for the dwell time. Then, the dwell time of a train at a station, $b_{m,i}$, is represented as

$$b_{m,i} = \frac{q_{p,i} h_{m,i}}{\mu_p} + g_{b,i}, \tag{2}$$

where $q_{p,i}$ is the (possibly time-dependent) passenger demand flow rate at station i, $h_{m,i} \equiv t_{m,i} - t_{m-1,i}$ is the time-headway, and thus $q_{p,i} h_{m,i}$ is the number of waiting passengers at the station.[3] This can be considered as a special case of Lam et al. (1998), which have

[3] In reality, there are passengers alighting a train, in addition to ones boarding. By carefully distinguishing the two types of passengers and replacing the terminology in the main text, the discussions in the main text are valid and the final results are not altered. For example, "the number of boarding passengers" can be replaced with "the sum of the number of boarding passengers and the number of alighting passengers".

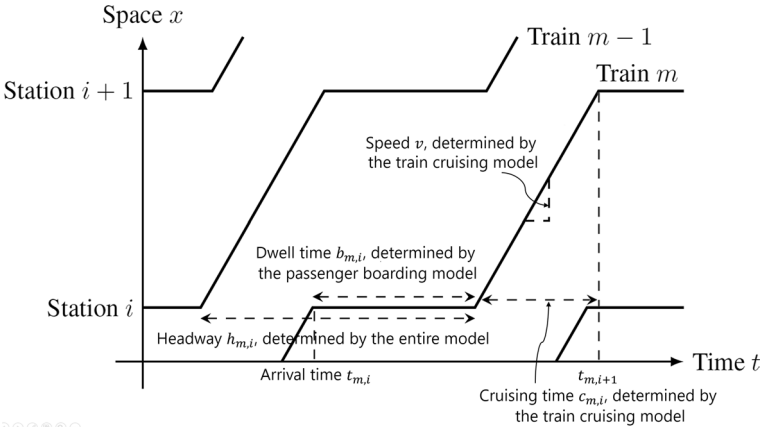


Fig. 2 Illustration of the microscopic model of rail transit system

empirically analyzed the boarding time in Hong Kong metro in crowded situations. All passengers waiting for a train at a station are assumed to board the first train arrived.

The cruising behavior of a train is modeled using the Newell's simplified car-following model (Newell 2002).[4] In this model, a train travels by maintaining the minimum safety clearance. Specifically, $x_{m,i}(t)$, position of a train m between stations i and $i+1$ at time t, is described as

$$x_{m,i}(t) = \min\left\{x_{m,i}(t-\tau) + v\tau, x_{m-1,i}(t-\tau) - \delta\right\}, \tag{3}$$

where $m-1$ indicates the preceding train of train m, τ is the physical minimum time-headway, v is the desired cruising speed that is determined by a (fixed-block or moving-block) signal control system, and δ is the minimum spacing.

We call the traffic is in *free-flowing regime* if the train travels between stations at the free-flow speed v_f (the maximum speed of the track), i.e., the train motion is represented by the first term with $v = v_f$ in the minimum operation of Eq. (3). We call the traffic is in *congested regime* if the train is required to decrease its speed to maintain both the safety headway τ and distance δ. In this case, the second term of Eq. (3) is active. The speed profile in this regime may differ in different train operators (i.e., signal control systems) and drivers. We employ one of the simplest approximations of this speed profile, that is, the train travels between stations at a constant speed while maintaining the minimum safety clearance.

Footnote 3 (continued)

However, it will complicate the discussions; therefore, we ignore passengers alighting a train.

For the boarding time function (2), more complicated functional forms such as a nonlinear function in which the boarding time is significantly extended in overcrowded situations may be considerable instead of the linear function. In fact, Seo et al. (2019) have derived another FD based on such a nonlinear boarding time function, but its functional form is extremely complicated. In order to simplify the discussion, we employ the linear function in this paper.

[4] Newell's simplified car-following model is a special case of the well-known road traffic flow model, the Lighthill–Whitham–Richards (LWR) model (Lighthill and Whitham 1955; Richards 1956; Newell 1993). Although the LWR model is known as a "macroscopic" model based on continuum fluid approximation, Newell (2002) showed that it is equivalent to a microscopic car-following model proposed by his paper.

Validity of Assumptions

In this model, the dwell time of a train is determined by the number of boarding passengers, not by a pre-determined timetable. Although this seems like inconsistency between the proposed model and actual schedule-based train operations, this can be considered as a reasonable approximation of average operation pattern of schedule-based train operations. The reasons are as follows. First, in a congested urban areas, it is common that passenger boarding time is not negligible and occasionally delay transit operation, as reviewed in Section "Introduction". Therefore, in order to maintain a scheduled operation based on a timetable, this timetable has to be determined considering the passenger demand (e.g., Niu and Zhou 2013). Consequently, the dwelling time in such timetable can be considered as similar to the proposed passenger boarding model (2) where $q_{p,i}h_{m,i}$ is interpreted as an average number of waiting passenger and g_b is interpreted as a buffer time to deal with fluctuation of the demand.

Second, the passenger boarding model with a constant capacity is consistent with the modelling of ordinary pedestrian flows for a fixed-width bottleneck (Lam et al. 1998; Hoogendoorn and Daamen 2005). Meanwhile, there is no stock capacity for passengers in the presented model; in other words, a train can transport infinite number of passengers. This is a limitation of the current model; however, unless the passenger demand is excessive level (e.g., where not all of the waiting passenger can board an arriving train), this limitation will not be problematic.

We have assumed that all passengers waiting for a train at a station board the first train arrived. However, in very crowded situations in reality, not all passengers can board the first train and they will form a queue in a platform. This phenomenon is not directly captured by the proposed model, but it can be captured by extending it. Specifically, a new model that describes queuing at the platform can be developed and attached to the proposed model, so that the outflow from the platform queuing model becomes the passenger boarding flow $q_{p,i}$ in the proposed model. This extension is left for the future work.

The train cruising model (3) can be considered as a lower order but a reasonable approximation of railway operation in a macroscopic scale. First, the model's fundamental operating principle is consistent with practical controls and existing studies (Carey and Kwieciński 1994; Higgins and Kozan 1998; Huisman et al. 2005): each train has to maintain a headway and a spacing that are greater than the given minimum ones. Second, the car-following model is almost equivalent to the "moving block signal control" (Dicembre and Ricci 2011). In fact, de Rivera and Dick (2021) and Corman et al. (2019) have employed car-following approximations of railways with moving block systems to analyze their performance. Additionally, the "fixed block signal control" can be considered as an inefficient moving block signal and approximated as a car-following model in the macroscopic scale (Li et al. 2005; Kariyazaki et al. 2015). The other assumption in the cruising model is the constant speed assumption in the congested regime. This approximates a train operation with low acceleration rates for ensuring the comfort of passengers and less energy loss. In addition, this assumption can describe the congested situation where the train stops between stations, on average, as we will show in the numerical experiment in Section "Validation of the Macroscopic Model".

Fundamental Diagram of Rail Transit System

In this section, we derive an FD of a rail transit system described by the microscopic model formulated in Section "Microscopic Model of Rail Transit System". The FD is defined as the relation among train-flow, train-density, and passenger-flow.

Steady State of Rail Transit System

We consider the *steady state* of the proposed microscopic model. The steady state is defined as a traffic state that does not change over time and/or space. In the other words, in the steady states, traffic state variables (flow, density, and average speed) in a certain time–space region are invariant with respect to temporal and/or spatial change of the position of the time–space region. There may exist a special functional that relates traffic state variables in steady state traffic; if such a function exists, it is often called as an FD of the traffic (Daganzo 1997).

To derive the traffic state in the steady state, let us consider a homogeneous rail transit system in which the stations and passenger demands are homogeneously distributed over the line, i.e., $l_i = l$, $q_{p,i} = q_p$ and other parameters (g_b, v_f, τ, and δ) are the same for all stations and trains. Then, the steady state is defined as a state that, for a given steady passenger demand q_p, the time-headway between successive trains, H, is time-independent. Note that $q_p < \mu_p$ must be satisfied; otherwise, passenger boarding will never end.

Transit systems under different steady states are illustrated as time–space diagrams in Fig. 3. In each sub-figure, train m arrives at and departs from station i, then travels to station $i+1$ at cruising speed v, and finally arrives at station $i+1$. The main differences between each sub-figure are density of trains and, consequently, traffic regime. In Fig. 3a, the density is small so that the speed v is equal to the free-flow speed v_f and h_f is greater than zero; therefore the state is classified into the free-flowing regime. In Fig. 3b, the density is medium so that the speed is equal to v_f and h_f is equal to zero; therefore, the state is classified into the critical regime. In Fig. 3c, the density is large so that the speed is less than v_f; therefore, the state is classified into the congested regime.

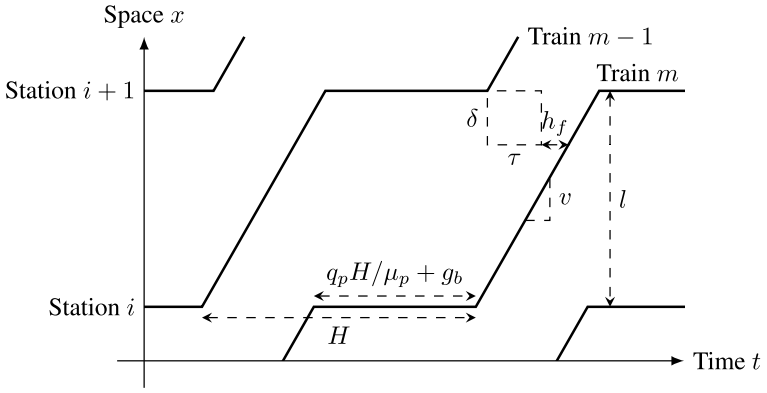
Fundamental Diagram

In general, the followings are considered as the traffic state variables of a rail transit system:

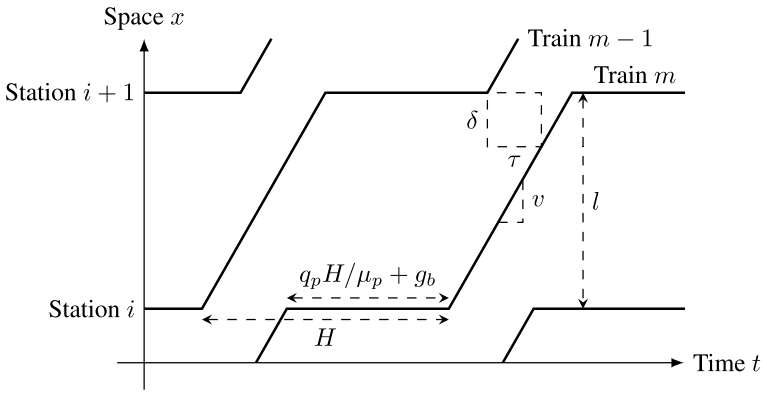
- train-flow q,
- train-density k,
- train-mean-speed $\bar{v}$,
- passenger-flow q_p,
- passenger-density k_p,
- passenger-mean-speed $\bar{v}_p$.

Among these, there are three independent variables: for example, the combination of q, k, and q_p. This is because of the identities $q = k\bar{v}$ and $q_p = k_p\bar{v}_p$, and $\bar{v} = \bar{v}_p$.[5]

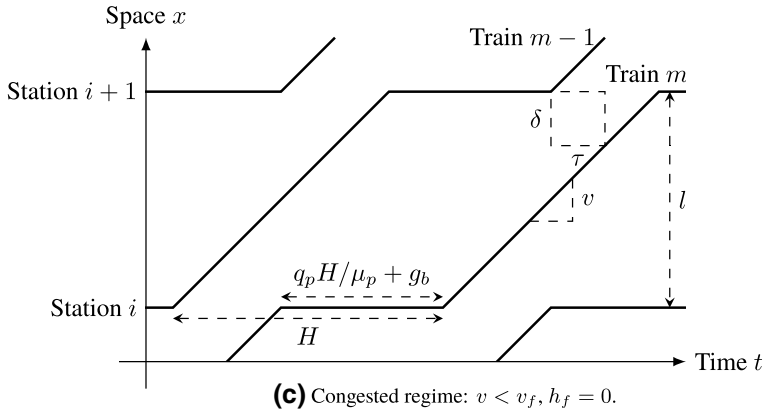
[5] Note that the mean speed $\bar{v}$ differs from the cruising speed v; the former takes the dwelling time at a station and cruising between stations into account, whereas the latter only considers the cruising time.



(a) Free-flowing regime: $v = v_f, h_f > 0$.



(b) Critical regime: $v = v_f, h_f = 0$.



(c) Congested regime: $v < v_f, h_f = 0$.

Fig. 3 Time–space diagrams of rail transit system under steady states

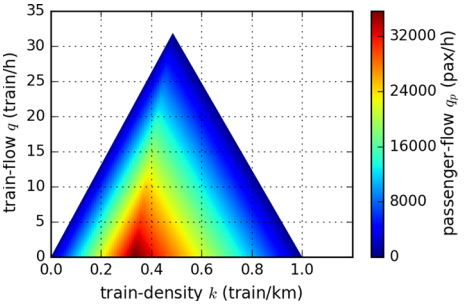


Fig. 4 Numerical example of the FD

Now suppose that the relation among the independent variables of the traffic state under every steady state is expressed using a function Q as

$$q = Q(k, q_p). \tag{4}$$

The function Q is regarded as an FD of the rail transit system. In fact, by assuming that the rail transit operation principle follows Eqs. (2) and (3), the FD function is analytically derived as

$$Q(k, q_p) = \begin{cases} \dfrac{lk - q_p/\mu_p}{g_b + l/v_f}, & \text{if } k < k^*(q_p), \\ -\dfrac{l\delta}{(l-\delta)g_b + \tau l}(k - k^*(q_p)) + q^*(q_p), & \text{if } k \geq k^*(q_p), \end{cases} \tag{5}$$

with

$$q^*(q_p) = \frac{1 - q_p/\mu_p}{g_b + \delta/v_f + \tau}, \tag{6}$$

$$k^*(q_p) = -\frac{(l-\delta)/v_f - \tau}{(g_b + \delta/v_f + \tau)\mu_p l} q_p + \frac{g_b + l/v_f}{(g_b + \delta/v_f + \tau)l}, \tag{7}$$

where $q^*(q_p)$ and $k^*(q_p)$ represent train-flow and train-density, respectively, at a critical state with passenger-flow q_p. For the derivation, see Appendix A. Although the FD equations (5)–(7) look complicated, they represent a simple relation: a piecewise linear (i.e., triangular) relation between q and k under fixed q_p. See Fig. 4 for a numerical example of the FD which we will explain later.

Discussions

The FD is interpreted as a function that describes transit operation performance (train-flow q, headway $H = 1/q$, and mean-speed $\bar{v} = q/k$) under a given train supply (train-density k) and passenger demand (passenger-flow q_p) for the given technical parameters of the transit system $(\mu_p, g_b, v_f, \tau, \delta, l)$. Therefore, it can be considered as a similar concept to the macroscopic fundamental diagram (MFD) (Geroliminis and Daganzo 2007; Daganzo 2007), which describes a road network throughput under a given number of vehicles and technical parameters of the road network.

Table 1 Parameters of the numerical example

Parameter	Value
u	70 km/h
τ	1/70 h
δ	1 km
μ_p	36000 pax/h
g_b	10/3600 h
l	3 km

Numerical Example

First of all, for ease of understanding, we show a numerical example of the FD in Fig. 4. The parameter values are presented in Table 1. In the figure, the horizontal axis represents train-density k, the vertical axis represents train-flow q, and the plot color represents passenger-flow q_p. The slope of the straight line from a traffic state to the origin represents the mean speed $\bar{v}$ of the state.

For example, the figure can be read as follows. Suppose that the passenger demand per station is $q_p = 16000$ (pax/h). If the number of trains in the transit system is given by the train-density $k = 0.3$ (train/km), then the resulting train traffic has a train-flow of $q \simeq 15$ (train/h) and a mean speed of $\bar{v} \simeq 50$ (km/h). This is the traffic state in the free-flowing regime. There is a congested state corresponding to a free-flowing state: for the aforementioned state with $(q, k, \bar{v}) \simeq$ (15 veh/h, 0.3 veh/km, 50 km/h), the corresponding congested state is (15 veh/h, 0.55 veh/km, 27 km/h). The critical state under $q_p = 16000$ (pax/h) is (22 veh/h, 0.42 veh/km, 52 km/h). Notice that this state has the fastest mean speed under the given passenger demand. The triangular q–k relation mentioned before is clearly shown in the figure; the "left edge" of the triangle corresponds to the free-flowing regime, the "top vertex" corresponds to the critical regime, and the "right edge" corresponds to the congested regime.

By comparing the theoretical FD (Fig. 4) with the actual data (Fig. 1), some similarities can be found. The two features found in the actual data (as the passenger-flow increases, the train-density increases; and as the passenger-flow increases, the train-speed and train-flow decrease) can be interpreted that the actual data are from a part of free-flow regime of the theoretical FD. Furthermore, in the high train-density regime in the Tokyo data (Fig. 1a), we observed a slight drop of passenger-flow; this might be a congested regime of the theoretical FD. From these results, we can say that the theoretical model explains the actual data to some extent.

Detailed Features of Fundamental Diagram

The FD has the following theoretical features which are analytically derived from Eq. (5). They can easily be found in the numerical example in Fig. 4.

As mentioned, the traffic state of a transit system is categorized into three regimes (free-flowing, critical, and congested), as in the standard traffic flow theory. Therefore, there is a critical train-density $k^*(q_p)$ for any given q_p. Train traffic is in the free-flowing regime if $k < k^*(q_p)$, in the critical regime if $k = k^*(q_p)$, or in the congested regime otherwise. The

congested regime can be considered as inefficient compared with the free-flowing regime, because the congested regime takes more time to transport the same volume of passengers. The critical regime is the most efficient in the sense that its travel time (i.e., $1/\bar{v}$, $1/\bar{v}_p$) and in-vehicle crowding (i.e., number of passengers per train, q_p/q) are minimum under a given passenger demand. However, the critical regime requires more trains (i.e., higher train-density) than the free-flowing regime; therefore, it may not be the most efficient if the operation cost is taken into account.

Even in the critical regime, the mean speed $\bar{v}$ is inversely proportional to passenger demand q_p. This means that travel time increases as passenger demand increases. In addition, the size of the feasible area of (q, k) narrows as q_p increases. Thus, the operational flexibility of the transit system declines as the passenger demand increases.

Flow and density of trains in the critical regime satisfy the following relations:

$$q^*(q_p) = \frac{l}{(l-\delta)/v_f - \tau} k^*(q_p) - \frac{1}{(l-\delta)/v_f - \tau}. \tag{8}$$

(Here, we have assumed $(l-\delta)/v_f - \tau \neq 0$.) Therefore, the critical regime is represented as a straight line whose slope $(l/[(l-\delta)/v_f - \tau])$ is either positive or negative in the q–k plane. This implies a qualitative difference between transit systems. Specifically, if the slope is positive, a transit operation with constant train-density would transition from the free-flowing regime to the congested regime as passenger demand increases (Fig. 4). On the contrary, if the slope is negative, such an operation would transition from free-flowing to congested as passenger demand *decreases*. This seems paradoxical, but it is actually reasonable because the operational efficiency can be degraded if the number of trains is excessive compared to passenger demand.

The FD describes an transit system's performance under a steady state operation as mentioned. Under the presence of well-designed adaptive control strategies, such as schedule-based and headway-based control (Daganzo 2009; Wada et al. 2012), the steady state is likely to be realized. This is because the aim of such adaptive control is usually to eliminate bunching—in other words, such control makes the operation steady. Therefore, it can be expected that the FD could be useful to describe average performance of actual transit system, which is usually not steady due to heterogeneity among passenger demand and train supply. This issue is numerically validated in Section "Results".

Last but not least, it is worth mentioning that all parameters in the proposed model have an explicit physical meaning. Therefore, the parameter calibration required to approximate an actual transit system is relatively easy.

Relation to the Macroscopic Fundamental Diagram

The proposed FD resembles the MFD (Geroliminis and Daganzo 2007; Daganzo 2007) and its extensions (e.g., Geroliminis et al. 2014; Chiabaut 2015) as mentioned. They are similar in the following sense. First, they both consider dynamic traffic. Second, they both describe the relations among macroscopic traffic state variables in which the traffic is not necessarily steady or homogeneous at the local scale (i.e., they use area-wide aggregations based on Edie's definition; see Appendix B). Third, they both have unimodal relations, meaning that there are free-flowing and congested regimes, where the former has higher performance than the latter; in addition, there is a critical regime where the throughput is maximized. Therefore, it is expected that existing approaches for MFD applications, such as modeling, control and the optimization of transport

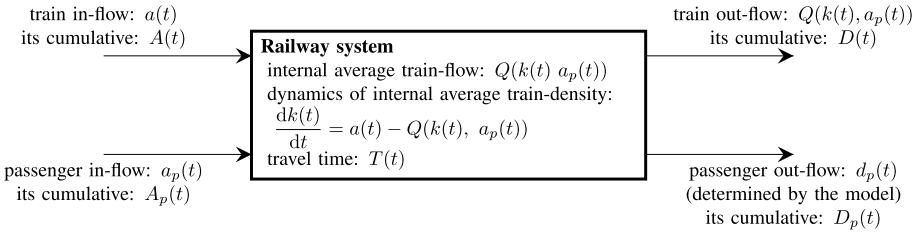


Fig. 5 Railway system as an input–output system

systems (e.g., Daganzo 2007; Geroliminis and Levinson 2009; Geroliminis et al. 2013; Fosgerau 2015), are also suitable for the proposed transit FD.

However, there are substantial differences between the proposed FD and the existing MFD-like concepts. In comparison with the original MFD (Geroliminis and Daganzo 2007; Daganzo 2007) and its railway variant (Cuniasse et al. 2015), the proposed FD has an additional dimension, that is, passenger-flow. In comparison with the three-dimensional MFD of Geroliminis et al. (2014), which describes the relations among total traffic flow, car density, and bus density in a multi-modal traffic network, the proposed FD explicitly models the physical interaction among the three variables. In comparison with the passenger MFD of Chiabaut (2015), which describes the relation between passenger flow and passenger density when passengers can choose to travel by car or bus, in the proposed FD, passenger demand can degrade the performance (i.e., speed) of the vehicles because of the inclusion of the boarding time.

Dynamic Model Based on Fundamental Diagram

Recall that the proposed FD describes the relationship among traffic variables under the steady state. It means that the behavior of a dynamical system in which demand and supply change over time is not described solely by the FD. This feature is the same as in the road traffic FD and MFDs. In this section, we formulate a model of urban rail transit operation where the demand (i.e., passenger-flow) and supply (i.e., train-density) change dynamically. In this proposed model, individual train and passenger trajectories are not explicitly described; therefore, the model is called macroscopic.

The proposed model is based on an *exit-flow model* (Merchant and Nemhauser 1978; Carey and McCartney 2004) in which the proposed FD is employed as the exit-flow function. Specifically, the transit system is considered as an input–output system, as illustrated in Fig. 5. The exit-flow modeling approach is often employed for area-wide traffic approximations and analysis using MFDs, such as optimal control to avoid congestion (Daganzo 2007) and analyses of user equilibrium and social optimum in morning commute problems (Geroliminis and Levinson 2009). The advantage of this approach is that it would be possible to conduct mathematically tractable analysis of dynamic, large-scale, and complex transportation systems, where the detailed traffic dynamics are difficult to model in a tractable manner—this is the case for transit operations.

Formulation

Let $a(t)$ be the inflow of trains to the transit system, $a_p(t)$ be the inflow of passengers, $d(t)$ be the outflow of trains from the transit system, and $d_p(t)$ be the outflow of passengers, on time t respectively. We set the initial time to be 0. Let $A(t)$, $A_p(t)$, $D(t)$, and $D_p(t)$ be the cumulative values of $a(t)$, $a_p(t)$, $d(t)$, and $d_p(t)$, respectively (e.g., $A(t) = \int_0^t a(s)\mathrm{d}s$). Let $T(t)$ be the travel time of a train that entered the system at time t, and let its initial value $T(0)$ be given by the free-flow travel time under $q = a(0)$ and $q_p = a_p(0)$. To simplify the formulation, the trip length of the passengers is assumed to be equal to that of the trains.[6] This means that $T(\cdot)$ is the travel time of both the trains and the passengers. These functions are interpreted as follows:

- $a(t)$: trains' departure rate from their origin station at time t.
- $a_p(t)$: passengers' arrival rate at the platform of their origin station at time t.
- $d(t)$: trains' arrival rate at their final destination station at time t.
- $d_p(t)$: passengers' arrival rate at their destination station at time t.
- $T(t)$: travel time of a train and passengers from origin (departs at time t) to destination. Note that the arrival time at the destination is $t + T(t)$.

Therefore, in reality, $a(\cdot)$ and $a_p(\cdot)$ will be determined by the transit operation plan and passenger departure time choice, respectively. Then, $d(\cdot)$, $d_p(\cdot)$, and $T(\cdot)$ are endogenously determined through the operational dynamics.

In accordance with exit-flow modeling, the train traffic is modeled as follows. First, the exit flow $d(t)$ is assumed to be

$$d(t) = Q(k(t), a_p(t)) \tag{9}$$

where the FD function $Q(\cdot)$ is considered to be an exit-flow function.[7] This means that the dynamics of the transit system are modeled by taking the conservation of trains into account as follows:

$$L\frac{\mathrm{d}k(t)}{\mathrm{d}t} = a(t) - Q(k(t), a_p(t)), \tag{10}$$

where L represents the length of the transit route. This exit-flow model has been employed in several studies to represent the macroscopic behavior of a transportation system (e.g., Merchant and Nemhauser 1978; Carey and McCartney 2004; Daganzo 2007). Note that the average train-density $k(t)$ is defined as

$$k(t) = \frac{A(t) - D(t)}{L}, \tag{11}$$

[6] This assumption is reasonable if the average trip length is shared by trains and passengers. If they are different, a modification such as $T_p(t) = T(t)/\lambda$, where λ is the ratio of average trip length of the passengers to that of the trains, would be possible.

[7] If n_p is considered as the sum of the number of passengers who are boarding and alighting (as mentioned in note 3), we can simply define $d(t)$ to be equal to $Q(k(t), a_p(t) + d_p(t))$. Such a model is also computable using a similar procedure.

which is consistent with Eq. (10). Based on above functions and equations, $d(t)$ and $D(t)$ are sequentially computed—in other words, the train traffic is computed using the initial and boundary conditions and the exit-flow model based on the FD.

The passenger traffic is derived as follows. By the definition of the travel time of trains,

$$A(t) = D(t + T(t)) \tag{12}$$

holds. As $A(t)$ and $D(t)$ have already been obtained, the travel time $T(t)$ such that Eq. (12) holds is computed. Then, $D_p(t)$ and $d_p(t)$ are computed from the definition of the travel time of passengers, which is also $T(t)$:

$$A_p(t) = D_p(t + T(t)). \tag{13}$$

Discussion

The proposed macroscopic model computes train out-flow $d(t)$ and passenger out-flow $d_p(t)$ based on the FD function $Q(\cdot)$, the initial and boundary conditions $a(t)$, $a_p(t)$, and $T(0)$. The advantages of the proposed model are twofold. First, it is a dynamic model, meaning that it can describe dynamic changing (i.e., time-varying) passenger demand, travel time, fee, and so on. This is a significant advantage compared with static models, in which these variables are static and constant. Second, it is analytically tractable, so that optimal control strategy may be derived as an analytical solution. This is a significant advantage compared with microscopic simulation models, which are usually not analytically tractable and computationally costly. Thus, the proposed model might be suitable for some specific purposes, such as to derive the optimal solution for dynamic pricing under time-varying passenger demand.

It is reasonable to expect that the proposed model can accurately approximate the macroscopic behavior of a transit operation with high-frequency operation (i.e., small time-headway) under moderate changes in demand and/or supply, although the FD is derived based on the steady state assumption. This is because exit-flow models are reasonably approximate a dynamical system's behavior when the changes in inflow are moderate compared with the relaxation time of the system. In the next section, the quantitative accuracy of the model is validated through numerical experiments.

Validation of the Macroscopic Model

In this section, we validate the quantitative accuracy of the macroscopic model by comparing its results with that of the microscopic model (i.e., Eqs. (2) and (3)).

Simulation Setting

The parameter values of the transit operation are listed in Table 1 for both the microscopic and macroscopic models. The railroad is considered to be a one-way corridor. The stations are equally spaced at intervals of l, and there are a total of 10 stations. Trains enter the railroad with flow $a(t)$; in the microscopic model, a discrete train enters the railroad from the upstream boundary station if $\lfloor A(t) \rfloor$ (i.e., integer part of $A(t)$) is incremented. In the microscopic model, trains leave the railroad from the downstream boundary station without any

restrictions, other than the passenger boarding and minimum headway clearance. Passengers arrive at each station with flow $a_p(t)$.

The functions $a(t)$ and $a_p(t)$ are exogenously determined to mimic morning rush hours, with each having a peak at $t = 2$. The flow before the peak time increases monotonically, whereas the flow after the peak time decreases monotonically—in other words, the so-called S-shaped $A(t)$ and $A_p(t)$ (c.f., Fig. 7) are considered. The parameters of these functions are the minimum train supply $a^{\min}$, the maximum train supply $a^{\max}$, the minimum passenger demand $a_p^{\min}$, and the maximum passenger demand $a_p^{\max}$. The functional forms are described in Appendix C. The simulation duration is set to 4 h for the baseline scenario in Section "Baseline scenario" and to 8 h for the sensitivity analysis in Section "Sensitivity analysis of the demand/supply conditions" (the reason will be explained later).

The microscopic model without any control is asymptotically unstable, as proven by Wada et al. (2012); this means that time-varying demand and supply always cause train bunching, making the experiment unrealistic and useless. Therefore, the headway-based control scheme proposed by Wada et al. (2012) is implemented in the microscopic model to prevent bunching and stabilize the operation. This scheme has two control measures: holding (i.e., extending the dwell time) and an increase of free-flow speed, similar to Daganzo (2009). The former is activated by a train if its following train is delayed, and is represented as an increase in g_b in the microscopic model. The latter is activated by a train if it is delayed, and is represented as an increase in v_f up to a maximum allowable speed $v_{\max}$. In this experiment, $v_{\max}$ is set to 80 km/h and v_f is 70 km/h. This control scheme can be considered realistic and reasonable, as similar operations are executed in practice. See Appendix D for further details of the control scheme. Note that the boundary conditions are the trajectory of the first train $x_0(t)\,\forall t$, the initial position of all the trains $x_m(0)\,\forall m$ (this is converted to the departure time of all the trains from the most upstream station $x_m(t_m^0)\,\forall m$ where t_m^0 denotes the departure time of train m), and the passenger demand to each station q_p. Also note that the microscopic model can be considered as an operation under planned frequency ($1/a(t)$) with a buffer time-based stabilization or regularization scheme (Appendix D).

Results

First, to examine how well the proposed model reproduces the behavior of the transit system under time-varying conditions, the results for the baseline scenario are presented in Section "Baseline scenario". Then, a sensitivity analysis of the demand/supply conditions is conduced and applicable ranges of the proposed model are investigated in Section "Sensitivity analysis of the demand/supply conditions".

Baseline scenario

The baseline scenario with parameter values $a^{\min} = 10$ (train/h), $a^{\max} = 15$ (train/h), $a_p^{\min} = 0.1\mu_p$ (pax/h), and $a_p^{\max} = 0.5\mu_p$ (pax/h) is investigated first. This situation can be considered as ordinary crowded. A solution of the microscopic model is shown in Fig. 6 as a time–space diagram. The colored curves represent the trajectories of each train traveling in the upward direction while stopping at every station. Around the peak time period ($t = 2$), train congestion occurs; namely, some of the trains stop occasionally between stations in order to maintain the safety interval. The congestion is caused by heavy passenger demand; therefore, the situation during rush hour is reproduced.

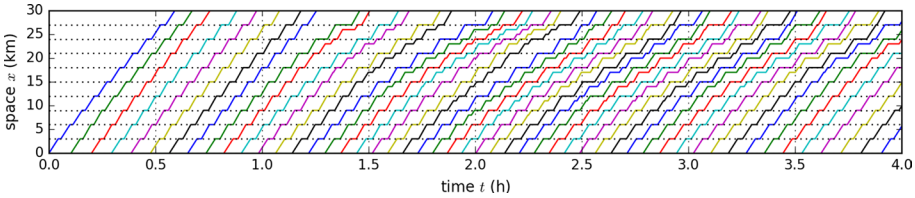


Fig. 6 Result of the microscopic model in the baseline scenario

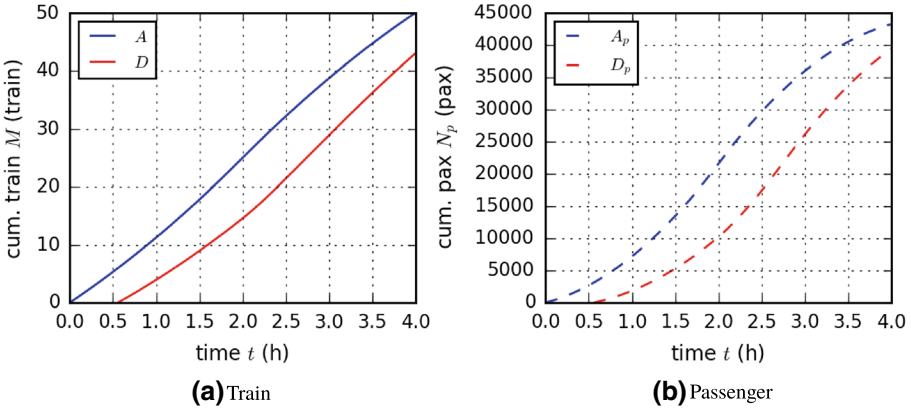
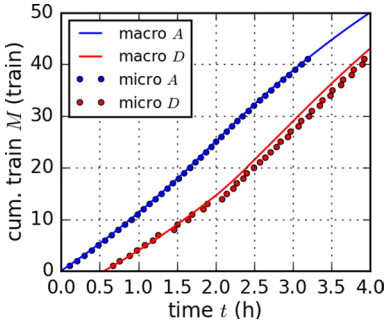


Fig. 7 Result of the macroscopic model in the baseline scenario

Fig. 8 Comparison between the macroscopic and microscopic models in the baseline scenario



The result given by the macroscopic model is shown in Fig. 7 as cumulative plots. Fig. 7a shows the cumulative curves for the trains, where the blue curve represents the inflow A and the red curve represents the outflow D. Fig. 7b shows those of passengers in the same manner. Congestion and delay are observed around the peak period (it is more remarkable in the passenger traffic). For example, during the peak time period, $d_p(t)$ is less than $a_p(t)$ and $a_p(t')$, where t' is time such that $t = t' + T(t')$. This means that the throughput of the transit system is reduced by the heavy passenger demand. Consequently, $T(t)$ is greater during peak hours than in off-peak periods such as $T(0)$, meaning that delays occur due to the congestion.

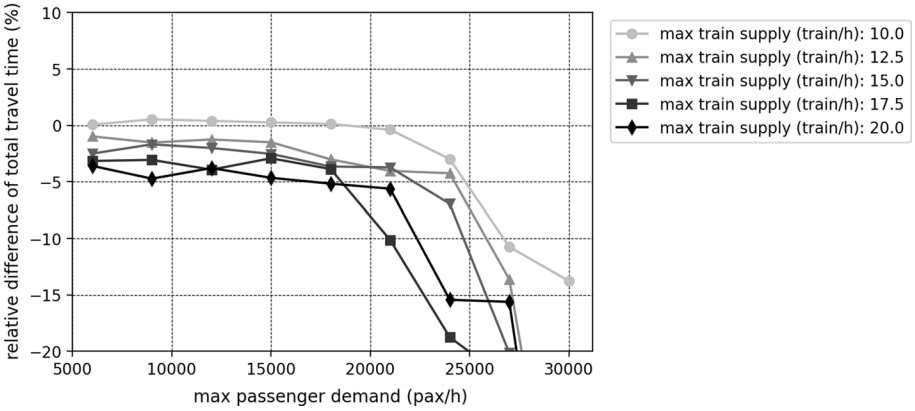


Fig. 9 Comparison between the microscopic and macroscopic models under different demand/supply conditions

The macroscopic and microscopic models are compared in terms of the cumulative number of trains in Fig. 8. In the figure, the solid curves denote the macroscopic model and the dots denote the microscopic model. The macroscopic model reproduced the microscopic model's result accurately; the relative difference of total travel time was roughly 4%. For example, the congestion and delay during the peak time period are captured very well. However, there is a slight bias: the macroscopic model gives a slightly shorter travel time. This is mainly due to the large-scale unsteady state (i.e., train bunching) generated in the microscopic model; the delay caused by such large-scale bunching cannot be recovered by the microscopic model under the implemented headway-based control scheme (for details, see Appendix D). It means that if the control is schedule-based, the bias could be reduced.

Sensitivity analysis of the demand/supply conditions

The accuracy of the macroscopic model regarding the dynamic patterns of demand/supply is now examined. This is worth investigating it quantitatively, because it is qualitatively clear that the exit-flow model is valid if the speed of demand/supply changes is "sufficiently" small as discussed in Section "Discussion". Specifically, the sensitivity of the peak passenger demand $a_p^{\max}$ and train supply $a^{\max}$ is evaluated by assigning various values to these parameters. The simulation duration is set to 8 h to take the residual delay after $t = 4$ (h) in some scenarios into account. The other parameters are the same as in the baseline scenario.

The results are summarized in Fig. 9 and Table 2. It shows the relative difference in total travel time (TTT) of trains between the microscopic and macroscopic models for various peak train supply $a^{\max}$ and peak passenger demand $a_p^{\max}$. The minimum train supply and passenger demand are set as $a^{\min} = 10$ (train/h) and $a_p^{\min} = 0.1\mu_p = 6000$ (pax/h). The relative difference can be considered as an error index of the macroscopic model. The negative values indicate that TTT of the macroscopic model is smaller.

According to the results in Fig. 9 and Table 2, the accuracy of the macroscopic model is high when the maximum passenger demand is not extremely large. This is an expected result, as the speed of demand change is slow in these cases. TTT given by the macroscopic model is almost always less than that of the microscopic model; this might be due to

Table 2 Comparison of total travel time (TTT) of the microscopic and macroscopic models under different demand/supply conditions

(a) Max train supply 10.0 train/h									
Max pax demand (pax/h)	6000	9000	12000	15000	18000	21000	24000	27000	30000
TTT of macro model (train h)	40.9	41.9	42.9	43.8	44.8	45.8	46.7	47.7	48.7
TTT of micro model (train h)	40.9	41.7	42.7	43.7	44.8	46.0	48.2	53.4	56.4
(b) Max train supply 12.5train/h									
Max pax demand (pax/h)	6000	9000	12000	15000	18000	21000	24000	27000	30000
TTT of macro model (train h)	43.3	44.2	45.2	46.1	47.1	48.1	49.0	50.0	51.0
TTT of micro model (train h)	43.7	44.9	45.7	46.8	48.6	50.1	51.2	57.9	91.5
(c) Max train supply 15.0 train/h									
Max pax demand (pax/h)	6000	9000	12000	15000	18000	21000	24000	27000	30000
TTT of macro model (train h)	45.5	46.5	47.5	48.4	49.4	50.4	51.3	52.3	53.2
TTT of micro model (train h)	46.7	47.3	48.4	49.7	51.3	52.3	55.1	65.5	78.1
(d) Max train supply 17.5 train/h									
Max pax demand (pax/h)	6000	9000	12000	15000	18000	21000	24000	27000	30000
TTT of macro model (train h)	47.8	48.8	49.8	50.7	51.7	52.6	53.6	54.6	55.5
TTT of micro model (train h)	49.4	50.3	51.8	52.3	53.8	58.6	66.0	70.6	88.3
(e) Max train supply 20.0 train/h									
Max pax demand (pax/h)	6000	9000	12000	15000	18000	21000	24000	27000	30000
TTT of macro model (train h)	50.1	51.1	52.0	53.0	54.0	54.9	55.9	56.9	58.3
TTT of micro model (train h)	52.0	53.6	54.1	55.6	56.9	58.2	66.1	67.4	119.9

the aforementioned inconsistency between the steady state assumption of the macroscopic model and headway-based control of the microscopic model.

The relative error increases suddenly when the demand exceeds a certain value, around 20000–22000. This sudden change is a result of extraordinary large-scale train bunching in the microscopic model. This bunching often occurs in cases with excessive passenger demand, such as $a_p^{\max} > \mu_p/2$. These excessive passenger demand can be considered as extraordinary and unrealistically high. For example, a train need to stop at a station for more than 10 min to process the passenger demand with 20000 pax/h. Additionally, the dwell time becomes longer than the cruising time between adjacent stations in this situation; in the real world, this does not occur even in rush hours .

As for the sensitivity of the train supply $a(\cdot)$, there is a weak tendency for faster variations in supply to cause larger errors. This is also an expected result. In any case, the error is small.

From these results, we conclude that the proposed model accurately reproduced the microscopic model's results accurately under ordinary passenger demand. On the other hand, the proposed model could not reproduce extraordinary and unrealistic situations with excessive train bunching. This might be acceptable for representing transit systems during normal rush hours.

Conclusion

The main contribution of this paper is that it analytically derived a closed-form expression of FD of rail transit systems based on microscopic operation principles. The FD determines operation performance of rail transit systems (i.e., flow, headway, and mean speed) based on supply of trains and passenger demand. Furthermore, this paper proposed an efficient, macroscopic dynamic assignment method based on the FD. The results of numerical experiments showed that the proposed method accurately reproduced results of a microscopic simulation model in macroscopic scales under realistic situations.

Specifically, the following three models of an urban rail transit system have been analyzed in this paper:

- *Microscopic model:* A model describing the trajectories of individual trains and passengers based on Newell's car-following model and passenger boarding model. This is represented in Eqs. (2) and (3), and is solved using simulations.
- *Fundamental diagram:* An exact relationship among train-flow, train-density, and passenger-flow in the microscopic model under a steady state. This is represented in Eqs. (4)–(7). It is a closed-form equation.
- *Macroscopic model:* A model describing train and passenger traffic using an exit-flow model whose exit-flow function is the FD. This is represented in Eqs. (9), (11), (12), and (13), and is solved using simple simulations.

The FD and macroscopic model are the original contributions of this study, whereas the microscopic model was proposed by Wada et al. (2012).

The FD itself implies several insights on transit system, such as relation between mean speed of the system and passenger demand. In addition, according to the results of the numerical experiment, the macroscopic model can reproduce the behavior of the microscopic model accurately, except for cases with unrealistically excessive demands. The proposed model is able to analyze dynamical changes in operation and demand (which is not possible for existing static models) while keeping the analytical tractability high (which is usually not possible for existing detailed simulation). Thus, we can expect that the model can contribute for obtaining general policy implications on management strategies of rail transit systems, such as pricing and control for morning commute problems.

Following future works are considerable. First, rigorous empirical validation on the existence of the FD is required. In fact, several preliminarily results on it have been reported (Fukuda et al. 2019; Zhang and Wada 2019) as shown in Fig. 1. Second, as an application of the FD, analysis of operation and demand management for transit systems is important. For example, the morning commute problem (Zhang et al. 2021) has been analyzed, and its departure time choice equilibrium and an optimal dynamic timetable have been derived. Third, queuing phenomena on platforms are important in highly crowded situations as explained in Section "Validity of Assumptions". This could be done by adding a platform queueing model to the outer layer of the exit flow model. Forth, extensions to more complicated and realistic railway operation, such as skip-stop operations and variable stopping patterns are considerable. Such heterogeneous operation schemes could be modeled by the following approach. When derive the FD equation, we need to consider train operation in a specific spatial-temporal region which is the minimum unit component of steady operation; in the current paper, we consider two trains and two stations as in Fig. 2. If the operation is heterogeneous, we can consider a larger spatial-temporal region to obtain

the minimum unit component of the operation. After defining such spatial-temporal region, we may able to derive an FD equation by employing a similar approach to this paper. Control theoretical approaches (Van Breusegem et al. 1991; Farhi 2019; Schanzenbacher et al. 2021) may be also useful to extend the proposed model to consider such heterogeneity as well as bottlenecks and railway networks.

Appendix A: Derivation of FD

This appendix describes derivation of the FD expressed in Eqs. (5)–(7). Consider a looped rail transit system[8] under steady state operation. Let L be the length of the railroad, S be the number of the stations, M be the number of trains, H be the time-headway of the operation, t_b be the dwelling time of a train at a station, t_c be the cruising time of a train between adjacent stations, and q_p be the passenger demand flow rate per station. Note that the distance between adjacent stations l is L/S and the number of passengers boarding a train at each station is q_pH.

The time-headway of the operation is derived as follows. The round trip time of a train in the looped railroad is $S(t_b + t_c)$, and M trains pass the station during that time. Then, the identities $NH = S(t_b + t_c)$ and

$$H = \frac{g_b + t_c}{M/S - q_p/\mu_p} \tag{A.1}$$

hold. Moreover, by the definition of headway and Newell's car-following rule, the time-headway H must satisfy

$$H = t_b + \frac{\delta + v\tau}{v} + h_f. \tag{A.2}$$

This reduces to

$$H = \frac{g_b + \delta/v + \tau}{1 - q_p/\mu_p} + h_f. \tag{A.3}$$

The q–k relation in a free-flowing regime is derived as follows. As the train-flow is $1/H$ and train-density is M/L by definition, Eq. (A.1) is transformed to

$$q(k) = \frac{kl - q_p/\mu_p}{g_b + l/v_f}. \tag{A.4}$$

The train-flow and train-density under a critical state, (q^*, k^*), are derived as follows. By substituting $v = v_f$ and $h_f = 0$ into Eq. (A.3) and using the identity $q = k\bar{v}$, we obtain

$$q^* = \frac{1 - q_p/\mu_p}{g_b + \delta/v_f + \tau}, \tag{A.5}$$

[8] Many other assumptions are acceptable on the shape of the track as long as we can define the steady state operation shown in Fig. 3.

$$k^* = k_0 + \frac{(1 - q_p/\mu_p)(g_b + l/v_f)}{(g_b + \delta/v_f + \tau)l}, \tag{A.6}$$

where k_0 is the minimum train-density where the train-flow is zero, namely, $k_0 = q_p/(\mu_p l)$.

The q–k relation in a congested regime is derived as follows. First, the k–v relation in a congested regime is easily derived from the q–v relation (A.3) with $h_f = 0$ and the identity $q = k\bar{v}$:

$$k(v) = k_0 + \frac{(1 - q_p/\mu_p)(g_b + l/v)}{(g_b + \delta/v + \tau)l}. \tag{A.7}$$

Now, consider $\mathrm{d}q/\mathrm{d}k$, which is identical to $(\mathrm{d}q/\mathrm{d}v) \cdot (\mathrm{d}v/\mathrm{d}k)$. This is derived as

$$\frac{\mathrm{d}q}{\mathrm{d}k} = \frac{l\delta}{(\delta - l)g_b - \tau l}, \tag{A.8}$$

which is constant and negative; therefore, the q–k relation is linear in a congested regime. Then, recalling that the linear q–k curve passes the point (q^*, k^*) with a slope of $\mathrm{d}q/\mathrm{d}k$, the q–k relation in a congested regime is derived as

$$q(k) = \frac{l\delta}{(\delta - l)g_b - \tau l}\, k + q_0 \tag{A.9}$$

with

$$q_0 = q^* - \frac{\mathrm{d}q}{\mathrm{d}k} \cdot k^*. \tag{A.10}$$

Eqs. (5)–(7) are constructed based on Eqs. (A.4), (A.5), (A.6), (A.9), and (A.10).

Appendix B: Consistency of the FD and Edie's generalized definition of traffic state

It is noteworthy that Eqs. (4) and (5) are consistent with Edie's generalized definition (Edie 1963) of traffic states; because from this consistency we can confirm that the FD is consistent with the fundamental definition of traffic. For steady-state transit operation, Edie's traffic state is derived as

$$q = \frac{1}{H}, \tag{B.1}$$

$$k = \frac{q_p H/\mu_p + g_b + l/v}{lH}, \tag{B.2}$$

$$\bar{v} = \frac{l}{q_p H/\mu_p + g_b + l/v}. \tag{B.3}$$

These relations are derived by applying Edie's definition to the "minimum component of the time–space diagram" of the steady state, which is a parallelogram-shaped area in Fig. 3 whose vertexes are time–space points of (i) train m departs from station i, (ii) train m

arrives at station $i+1$, (iii) train $m-1$ arrives at station $i+1$, and (iv) train $m-1$ departs from station i. One can easily confirm that Eqs. (B.1)–(B.3) satisfy the FD equation. In fact, the FD equation is also derived from Eqs. (B.1)–(B.3) and the constraint (A.3) induced by Newell's car-following model.

Appendix C: S-shaped supply and demand functions

The train supply and passenger demand in the experiments are given by the following functions:

$$a(t) = \begin{cases} a^{\min} + (a^{\max} - a^{\min})\frac{t}{2}, & \text{if } t < 2, \\ a^{\min} + (a^{\max} - a^{\min})\frac{4-t}{2}, & \text{if } 2 \le t < 4, \\ a^{\min}, & \text{if } 4 \le t, \end{cases} \tag{C.1}$$

$$a_p(t) = \begin{cases} a_p^{\min} + (a_p^{\max} - a_p^{\min})\frac{t}{2}, & \text{if } t < 2, \\ a_p^{\min} + (a_p^{\max} - a_p^{\min})\frac{4-t}{2}, & \text{if } 2 \le t < 4, \\ a_p^{\min}, & \text{if } 4 \le t. \end{cases} \tag{C.2}$$

Both functions have a minimum value at $t = 0$ and $t \ge 4$ and a minimum value at $t = 2$, and change linearly in between.

Appendix D: Adaptive control scheme in the microscopic model

This appendix briefly explains the adaptive control scheme for preventing train bunching, proposed by Wada et al. (2012). This scheme consists of two control measures: holding at a station and increasing the maximum speed during cruising.

First, the scheme modifies the buffer time for dwelling (originally defined as g_b in Eq. (2)) of train m at station i to

$$g_b := \max\{0,\ g_b - E_m(i)\} \tag{D.1}$$

with

$$E_m(i) = (1-\alpha)\varepsilon_m(i) + \alpha\mu_p\big(\varepsilon_m(i) - \varepsilon_{m-1}(i)\big), \tag{D.2}$$

where $\varepsilon_m(i) \equiv t_m(i) - T_{m,i}$ represents the delay, $t_m(i)$ represents the time at which train m arrives at station i, $T_{m,i}$ represents the scheduled time (i.e., without delay) at which train m should arrive at station i, and $\alpha \in [0,1]$ is a weighting parameter. This scheme represents a typical holding control strategy, similar to the bunching prevention method of Daganzo (2009), which extends the dwelling time of a vehicle if the headway to the preceding vehicle is too small and vice versa.

Second, the scheme modifies the free-flow cruising speed v_f such that the interstation travel time is reduced by

$$\min\left\{l/v_f - l/v_{\max},\ \max\{0,\ E_m(i) - g_b\}\right\}. \tag{D.3}$$

This means that, in the event of a delay, the train tries to catch up by increasing its cruising speed up to the maximum allowable speed $v_{\max}$ (which implies that the free-flow speed v_f is a "buffered" maximum speed).

Meanwhile, the proposed train operation model in this study does not have a schedule—it is a frequency-based operation. Therefore, in this study, the scheduled headway in the scheme $(T_{m,i} - T_{m-1,i})$ is approximated by the planned frequency $(1/a(t_m(i)))$. Thus, we set $\alpha = 1$ and substitute $E_m(i)$ with

$$\mu_p\left(t_m(i) - t_{m-1}(i) - 1/a(t_m(i))\right). \tag{D.4}$$

The stationary state of the operational dynamics under the original scheme is basically identical to the steady state defined in Section 3.1. There may be small difference in the congested regime because of the operation scheme; however, this will not be problematic since heavily congested regime will not occur. In the case of $\alpha < 1$, the scheme makes the train operation asymptotically stable, meaning that the operation schedule is robust to small disturbances. In the case of $\alpha = 1$, the scheme prevents the propagation and amplification of delay, but does not recover the original schedule (the small 'shift' found in Fig. 8 is due to $\alpha = 1$). Note that these control measures do not interrupt passenger boarding or violate the safety clearance between trains, meaning that most of the fundamental assumptions of the proposed FD are satisfied.

Author Contributions TS: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Funding acquisition. KW: Conceptualization, Methodology, Formal analysis, Investigation, Writing - Review & Editing, Funding acquisition. DF: Conceptualization, Methodology, Investigation, Writing - Review & Editing, Funding acquisition.

Funding This work was supported by Japan Society for the Promotion of Science under KAKENHI Grant-in-Aid for Scientific Research (B) #17H03320 and #20H02268.

Availability of data and material The data used to produce Fig. 1 are partially available at Fukuda et al. (2019), Zhang and Wada (2019), and Massachusetts Bay Transportation Authority.

Code availability The code of the simulation in Section "Validation of the Macroscopic Model" is available upon request.

Declarations

Conflicts of interest The authors declare no conflicts of interest associated with this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visithttp://creativecommons.org/licenses/by/4.0/.

References

Alonso, B., Munoz, J.C., Ibeas, A., Moura, J.L.: A congested and dwell time dependent transit corridor assignment model. J. Adv. Transport. (2017)

Carey, M., Kwieciński, A.: Stochastic approximation to the effects of headways on knock-on delays of trains. Transport. Res. Part B **28**(4), 251–267 (1994)

Carey, M., McCartney, M.: An exit-flow model used in dynamic traffic assignment. Comput. Oper. Res. **31**(10), 1583–1602 (2004)

Cats, O., West, J., Eliasson, J.: A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transport. Res. Part B **89**, 43–57 (2016)

Chiabaut, N.: Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram. Transport. Res. Part B **81**, 410–420 (2015)

Corman, F., Henken, J., Keyvan-Ekbatani, M.: Macroscopic fundamental diagrams for train operations—are we there yet? In: 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems, pp. 1–8 (2019)

Cunha, J., Reis, V., Teixeira, P.: Development of an agent-based model for railway infrastructure project appraisal. Transportation (2021)

Cuniasse, P.-A., Buisson, C., Rodriguez, J., Teboul, E., de Almeida, D.: Analyzing railroad congestion in a dense urban network through the use of a road traffic network fundamental diagram concept. Public Transport. **7**(3), 355–367 (2015)

Daganzo, C.F.: Fundamentals of Transportation and Traffic Operations. Pergamon, Oxford (1997)

Daganzo, C.F.: Urban gridlock: macroscopic modeling and mitigation approaches. Transp. Res. Part B **41**(1), 49–62 (2007)

Daganzo, C.F.: A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transp. Res. Part B **43**(10), 913–921 (2009)

de Cea, J., Fernández, E.: Transit assignment for congested public transport systems: an equilibrium model. Transp. Sci. **27**(2), 133–147 (1993)

de Palma, A., Kilani, M., Proost, S.: Discomfort in mass transit and its implication for scheduling and pricing. Transp. Res. Part B **71**, 1–18 (2015)

de Palma, A., Lindsey, R., Monchambert, G.: The economics of crowding in public transport. Working Paper (hal-01203310) (2015b)

de Rivera, A.D., Dick, C.T.: Illustrating the implications of moving blocks on railway traffic flow behavior with fundamental diagrams. Transp. Res. Part C **123**, 102982 (2021)

Dicembre, A., Ricci, S.: Railway traffic on high density urban corridors: capacity, signalling and timetable. J. Rail Transp. Plan. Manag. **1**(2), 59–68 (2011)

Edie, L.C.: Discussion of traffic stream measurements and definitions. In: Almond, J. (ed.) Proceedings of the 2nd International Symposium on the Theory of Traffic Flow, pp. 139–154 (1963)

Farhi, N.: Physical models and control of the train dynamics in a metro line without junction. IEEE Trans. Control Syst. Technol. **27**(4), 1829–1837 (2019)

Fosgerau, M.: Congestion in the bathtub. Econ. Transp. **4**, 241–255 (2015)

Fukuda, D., Imaoka, M., Seo, T.: Empirical investigation of fundamental diagram for urban rail transit using Tokyo's commuter rail data. In: TRANSITDATA2019: 5th International Workshop and Symposium (2019)

Geroliminis, N., Daganzo, C.F.: Macroscopic modeling of traffic in cities. In: Transportation Research Board 86th Annual Meeting (2007)

Geroliminis, N., Haddad, J., Ramezani, M.: Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. IEEE Trans. Intell. Transp. Syst. **14**(1), 348–359 (2013)

Geroliminis, N., Levinson, D.M.: Cordon pricing consistent with the physics of overcrowding. In: Lam, W.H.K., Wong, S.C., Lo, H.K. (eds.) Transportation and Traffic Theory 2009. Springer, pp. 219–240 (2009)

Geroliminis, N., Zheng, N., Ampountolas, K.: A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transport. Res. Part C **42**, 168–181 (2014)

Gonzales, E.J., Daganzo, C.F.: Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. Transp. Res. Part B **46**(10), 1519–1534 (2012)

Greenshields, B.D.: A study of traffic capacity. Highway Res Board Proc. **14**, 448–477 (1935)

Halvorsen, A., Koutsopoulos, H.N., Ma, Z., Zhao, J.: Demand management of congested public transport systems: a conceptual framework and application using smart card data. Transportation **47**(5), 2337–2365 (2019)

Higgins, A., Kozan, E.: Modeling train delays in urban networks. Transp. Sci. **32**(4), 346–357 (1998)

Hoogendoorn, S.P., Daamen, W.: Pedestrian behavior at bottlenecks. Transp. Sci. **39**(2), 147–159 (2005)
Huan, N., Hess, S., Yao, E.: Understanding the effects of travel demand management on metro commuters' behavioural loyalty: a hybrid choice modelling approach. Transportation (2021)
Huisman, D., Kroon, L.G., Lentink, R.M., Vromans, M.J.C.M.: Operations research in passenger railway transportation. Stat. Neerl. **59**(4), 467–497 (2005)
Kariyazaki, K., Hibino, N., Morichi, S.: Simulation analysis of train operation to recover knock-on delay under high-frequency intervals. Case Stud. Transp. Policy **3**(1), 92–98 (2015)
Kato, H., Kaneko, Y., Soyama, Y.: Departure-time choices of urban rail passengers facing unreliable service: Evidence from Tokyo. In: Proceedings of the International Conference on Advanced Systems for Public Transport 2012 (2012)
Koutsopoulos, H., Wang, Z.: Simulation of urban rail operations: application framework. Transp. Res. Rec. **2006**, 84–91 (2007)
Kraus, M., Yoshida, Y.: The commuter's time-of-use decision and optimal pricing and service in urban mass transit. J. Urban Econ. **51**(1), 170–195 (2002)
Kumagai, J., Wakamatsu, M., Managi, S.: Do commuters adapt to in-vehicle crowding on trains? Transportation (2020)
Kusakabe, T., Iryo, T., Asakura, Y.: Estimation method for railway passengers' train choice behavior with smart card transaction data. Transportation **37**(5), 731–749 (2010)
Lam, W.H.K., Cheung, C.Y., Poon, Y.F.: A study of train dwelling time at the Hong Kong mass transit railway system. J. Adv. Transp. **32**(3), 285–295 (1998)
Li, K.P., Gao, Z.Y., Ning, B.: Cellular automaton model for railway traffic. J. Comput. Phys. **209**(1), 179–192 (2005)
Li, S., Dessouky, M.M., Yang, L., Gao, Z.: Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines. Transp. Res. Part B **99**, 113–137 (2017)
Lighthill, M.J., Whitham, G.B.: On kinematic waves. II. a theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. A **229**(1178), 317–345 (1955)
Mahmassani, H.S., Williams, J.C., Herman, R.: Investigation of network-level traffic flow relationships: some simulation results. Transp. Res. Rec. **971**, 121–130 (1984)
Merchant, D.K., Nemhauser, G.L.: A model and an algorithm for the dynamic traffic assignment problems. Transp. Sci. **12**(3), 183–199 (1978)
Newell, G.F.: A simplified theory of kinematic waves in highway traffic. Transp. Res. Part B **27**(4), 281–313 (Part I, II, and III) (1993)
Newell, G.F.: A simplified car-following theory: a lower order model. Transp. Res. Part B **36**(3), 195–205 (2002)
Newell, G.F., Potts, R.B.: Maintaining a bus schedule. In: Proceedings of the 2nd Australian Road Research Board. Vol. 2 (1964)
Niu, H., Zhou, X.: Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transp. Res. Part C **36**, 212–230 (2013)
Parbo, J., Nielsen, O.A., Prato, C.G.: Passenger perspectives in railway timetabling: a literature review. Transp. Rev. **36**(4), 500–526 (2016)
Richards, P.I.: Shock waves on the highway. Oper. Res. **4**(1), 42–51 (1956)
Schanzenbacher, F., Farhi, N., Leurent, F., Gabriel, G.: Feedback control for metro lines with a junction. IEEE Trans. Intell. Transp. Syst. **22**(5), 2741–2750 (2021)
Seo, T., Wada, K., Fukuda, D.: Fundamental diagram of urban rail transit considering train–passenger interaction. arXiv preprint arXiv: 1708.02147(updated in 2021) (2017a)
Seo, T., Wada, K., Fukuda, D.: A macroscopic and dynamic model of urban rail transit with delay and congestion. In: Transportation Research Board 96th Annual Meeting (2017b)
Seo, T., Wada, K., Fukuda, D.: Transit fundamental diagram and its application to analysis of high frequency transitsystems. Proceedings of the Infrastructure Planning 60, (in Japanese) (2019)
Tabuchi, T.: Bottleneck congestion and modal split. J. Urban Econ. **34**(3), 414–431 (1993)
Tian, Q., Huang, H.-J., Yang, H.: Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system. Transp. Res. Part B **41**(6), 616–631 (2007)
Tirachini, A., Hensher, D.A., Rose, J.M.: Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. Part A **53**, 36–52 (2013)
Trozzi, V., Gentile, G., Bell, M.G.H., Kaparias, I.: Dynamic user equilibrium in public transport networks with passenger congestion and hyperpaths. Transp. Res. Part B **57**, 266–285 (2013)
Van Breusegem, V., Campion, G., Bastin, G.: Traffic modeling and state feedback control for metro lines. IEEE Trans. Autom. Control **36**(7), 770–784 (1991)
Vuchic, V.R.: Urban Transit: Operations, Planning, and Economics. Wiley, New York (2005)

Wada, K., Kil, S., Akamatsu, T., Osawa, M.: A control strategy to prevent delay propagation in high-frequency railway systems. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management) 68(5), I_1025–I_1034 (2012), (in Japanese; extended abstract in English was presented at the 1st European Symposium on Quantitative Methods in Transportation Systems and available at https://www.researchgate.net/publication/281823577)

Zhang, J., Wada, K.: Fundamental diagram of urban rail transit: An empirical investigation by Boston's subway data. In: hEART 2019: 8th Symposium of the European Association for Research in Transportation (2019)

Zhang, J., Wada, K., Oguchi, T.: Morning commute in congested urban rail transit system: A macroscopic model for equilibrium distribution of passenger arrivals. arXiv preprint arXiv:2102.13454 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Toru Seo is an Associate Professor at Tokyo Institute of Technology. He had earned his doctoral degree in 2015 at Tokyo Institute of Technology. His primary research interests are transportation science, traffic flow theory, transportation data analysis, and transportation system modeling and management.

Kentaro Wada is an Associate Professor at the Faculty of Engineering, Information and Systems, University of Tsukuba. He is also a Research Fellow at the Institute of Industrial Science, the University of Tokyo. His expertise includes dynamic traffic assignment, traffic flow theory, and transportation system design/optimization. He had his Ph.D. in Information Sciences from Tohoku University in 2013.

Daisuke Fukuda is a professor in transportation studies at the University of Tokyo in the department of civil engineering. He has his Ph.D. from the University of Tokyo in 2004 and was an associate professor from 2005 to 2020. He conducts research in travel behavior analysis and transportation economics.