Research Article

Estimating Wait Time and Passenger Load in a Saturated Metro Network: A Data-Driven Approach

Hezhou Qu,^{1,2} Xiaoyue Xu,¹ and Steven Chien^{3,4}

¹School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China

²National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu, Sichuan 611756, China

³College of Transportation Engineering, Chang'an University, Xi'an, Shaanxi 710064, China

⁴John A. Reif, Jr. Department of Civil and Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA

Correspondence should be addressed to Steven Chien; chien@njit.edu

Received 6 February 2020; Revised 22 March 2020; Accepted 6 May 2020; Published 1 August 2020

Academic Editor: Yu Jiang

Copyright © 2020 Hezhou Qu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The service quality of public transit, such as comfort and convenience, is an important factor influencing ridership and fare revenue, which also reflects the passengers' perception to the transit performance. Passengers are frustrated while waiting to board a crowded train especially during the peak hours, while the fail-to-board (FtB) situation commonly exists. The service performance measures determined by deterministic passenger demand and service frequency cannot reflect the perceived service of passengers. With the automatic fare collection system data provided by Chengdu Metro, we develop a data-driven approach considering the joint probability of spatiotemporal passenger demand at stations based on posted train schedule to approximate passenger travel time (e.g., in-vehicle and out-of-vehicle times). It was found that the estimated wait time can reflect the actual situation as passengers FtB. The proposed modeling approach and analysis results would be useful and beneficial for transit providers to improve system performance and service planning.

1. Introduction

Urban rail rapid transit, also called metro, plays an important role to transport a vast number of passengers, especially in the peak period. In the past decade, metro networks have been expanding rapidly. Due to a dramatic increase of ridership, those networks are overcrowding over space and time, which has become a major problem affecting system efficiency as well as service quality.

The quality of service (QoS) for fixed-route transit consists of two categories: "Availability" and "Comfort and Convenience" [1]. Service availability reveals a minimum requirement for transit being a travel option, while comfort and convenience contribute to users' satisfaction to the service and their likelihood of using it. For a metro system, passenger load, service reliability, and travel time are such indices associated with comfort and convenience. Some

passengers even weight more on crowdedness in the train than on travel time or distance [2–4]. In addition, the service measures were determined in previous studies using deterministic passenger demand and service frequency, which cannot reflect the perceived QoS of passengers.

From transit users' perspective, the passenger load affects the comfort during their travel, which varies over space and time in a network. The passenger load on a transit vehicle affects the comfort of the on-board vehicle portion of a transit trip in terms of both being able to find a seat and in overall crowding levels within the vehicle, represented by average space per standee and load factor. From transit operators' perspective on the other hand, the variation of passenger load may affect the system performance such that service frequency or number of cars per train need to justify for easing congestion and improving comfort [1]. It is desirable to know the spatiotemporal distributions of

passenger flows entering and exiting a metro system, so that appropriate service schedule as well as frequency can be determined [5–7]. A recent study [8] indicated that many transit users rely on real-time information (i.e., estimated travel time via smartphone) to make a travel choice. The accuracy of estimated travel time, consisting of in-vehicle and out-of-vehicle times, is critical for transit users to travel effectively.

The automatic fare collection system (AFCS) has been commonly applied (i.e., Metro Card in New York City, Oyster Card in London, etc.). Passenger flows entering and exiting metro stations can be analyzed with the AFCS transaction records, while wait time can be approximated in conjunction with the train schedule. However, transfer time is difficult to estimate because the itinerary of passengers (i.e., travel path) is unknown. In addition, overcrowding discouraged passengers boarding the train. This situation may result in fail-to-board (FtB), and extra wait/transfer time shall be expected. Previous research in estimating wait/transfer time under this context is still very limited.

The objective of this study is to develop a data-driven approach applied for approximating passenger itinerary, spatiotemporal passenger load (e.g., average space per standee and load factor), and travel time with the AFCS data considering station layout, train schedule, and the probability of FtB. Passenger travel time consists of in-vehicle and out-of-vehicle times (e.g., walking, wait, and transfer times).

The remainder of this paper is organized as follows. Section 2 describes previous research and practices related to this study. Sections 3 and 4 discuss the development of the proposed model and solution approach. Then, in Section 5, the numerical analysis based on the data provided by Chengdu Metro is conducted. Finally, the research findings and future studies are concluded.

2. Literature Review

Many studies have been conducted on analyzing transit system performance and service quality. Xin et al. [9] applied the methods suggested by Transit Capacity and Quality of Service Manual (TCQSM) to evaluate the QoS indicators (e.g., service frequency and coverage, hours of service, and transit-auto travel time) for Grand River Transit in Canada. Bunker [10] developed a method to profile route-based load factor as on-board passenger comfort and investigated the time series correlation between the load factor and passenger average travel time. Furth et al. [11] analyzed wait time and passenger load of a bus transit using the automatic vehicle location (AVL) and automatic passenger counter data to improve the performance and management of transit systems. For overcrowding situations, many passengers waiting on the platform find it difficult to board the immediate coming train due to limited train capacity and must wait for another available one. The TCQSM method [1] seems to be unable to accurately estimate wait time for a saturated metro network because the impact of FtB was not considered.

The AFCS data has been widely applied in public transit systems for OD demand estimation [12], timetable design [13], passenger flow assignment [14, 15], passenger behavior

analysis [16], and transfer coordination [17–19]. Since passengers swipe the metro card only at the gates of origin and destination stations, detailed information, such as the location where a passenger makes a transfer, is unknown. Thus, wait/transfer time is very difficult to estimate, especially under overcrowding situations.

The spatiotemporal passenger load information is important for metro operation and service planning. Sun et al. [20] developed a regression model to estimate passenger demand of a transit line, using the minimum travel time between each pair of origin and destination stations. Zhang et al. [21] estimated walking time between the gate of a station and the platform and transfer time. With AFCS, AVL, and general transit feed specification data, Luo et al. [22] developed a method to generate passenger load profiles for transit vehicles, based on passenger check-in time and vehicle arrival time. However, none of the models discussed above considered FtB.

The FtB situation has become an increasingly important issue as the demand increases in metro systems. Zhu et al. [23] used the maximum likelihood method and a Bayesian inference method to estimate the probability of FtB with the AFCS data. Later, they [24] developed a probabilistic model for assigning passengers to trains and estimated the number of skipped trains because of overcrowding. With a bilevel regression model, Miller et al. [25] estimated the number of FtB passengers and calculated the shortage of cumulative capacity. However, the studies discussed above did not consider transfer passengers in a large metro network.

In addition to FtB, it is also important to consider transfer efficiency which would significantly affect metro system performance and level of service [26]. However, it is difficult to estimate transfer demand at stations because the AFCS data do not have passengers' itinerary information but have the records of them entering and exiting from stations. For a complicated network where passengers need to make multiple transfers to reach their destinations [27], estimating transfer demand becomes even more challenging. Passengers would select transfer locations to the best of their interests (i.e., less travel time or distance, greater reliability and comfort, etc.). Few studies investigated spatiotemporal travel-path choice under overcrowding situations. Kusakabe et al. [28] estimated the boarded trains by railway passengers whose travel-path choice was determined by the shortest travel path with the minimum wait, egress, and transfer times. Sun and Schonfeld [29] developed a schedule-based path-choice model for a rail transit network considering the probability of FtB. Zhao et al. [30] developed a probabilistic model to estimate route choices, assuming that the number of trains waited for by passengers at origin and transfer stations and the itineraries being chosen in a specific time obey the polynomial distribution. Both studies assumed that the numbers of FtB passengers at different stations are independent, which might contradict the reality because the train rode by a passenger at a transfer station is dependent on the train the passenger rode at the origin station.

While estimating wait time, previous studies considered idealized conditions, either ignoring FtB [31] or assuming random passenger arrivals [32]. Recently, more studies

applied AFCS data for estimating wait time. With a Bayesian inference model, Zhang and Yao [33] estimated walking, wait, transfer, and in-vehicle times as those follow truncated normal distributions. Considering the correlation of travel times between pairs of stations, Lee et al. [34] analyzed those travel time components with a decomposition method. However, transfer demand was not considered. Tavassoli et al. [35] estimated passenger wait time for nontransfer and transfer passengers, while assuming the travel path is given. However, the paths with more than one transfer were not considered. Considering passenger arrival distribution, Ingvardson et al. [36] assumed that wait time follows a joint uniform and beta distribution, but the effect of FtB to wait time was not assessed. In general, the studies discussed earlier did not consider the situations in which passengers may have multiple path choices with transfers to reach their destinations.

This study proposes a data-driven approach to approximate passenger itineraries and associated travel times (i.e., the sum of in-vehicle and out-of-vehicle times) for a saturated metro network with the AFCS data. The service indicators such as passenger load (i.e., average space per standee and load factor) and travel time components (e.g., walking, wait, and transfer times) can be estimated for monitoring system operation and improving service planning.

3. Methodology

In a metro system, passenger travel time consists of in-vehicle time and out-of-vehicle time. The in-vehicle time can be determined by train schedule if the passenger itinerary is known. On the other hand, the out-of-vehicle time consists of walking time and wait time. The walking time considered here includes access and egress walking time and transfer walking time (if needed), which can be determined by walking distance divided by walking speed. Hence, wait time is out-of-vehicle time less walking time. Since the passenger's itinerary is not available in the AFCS data, passenger's in-vehicle and out-of-vehicle times are difficult to estimate, especially under overcrowding situations. A data-driven approach is proposed and discussed here for approximating passenger itinerary with the use of AFCS data and train schedule. Thus, some performance measures concerning QoS (i.e., wait time, average space per standee, and load factor) can be analyzed.

3.1. General Network. A general metro network consists of a set of routes defined as R and a set of stations defined as N . A route consists of two directional lines (e.g., outbound and inbound), and a set of lines is denoted as L . Each station is given a unique ID. For example, the station ID on line 1 begins with 1 and ends at N_1 , and then the station ID on line 2 begins with $N_1 + 1$ and ends at N_2 , and so on. Hence, the station IDs of line l are $N_{l-1} + 1$ through N_l . For line l , the trains are indexed by m (e.g., $m = 1, 2, \dots, M_l$), where M_l is the number of trains running on line l within the study time

period. Note that a set of transfer stations denoted as S also carries IDs indexed by s .

3.2. Assumptions. The assumptions considered in this study are described as follows and the definitions of variables and model parameters are summarized in Table 1:

- (1) Train dispatching and running follow the posted timetable
- (2) Walking time is determined by walking distance divided by average walking speed
- (3) Passengers will not stay at stations except for waiting for trains
- (4) The number of passengers making more than two transfers is small and negligible

3.3. Passenger Itinerary. Figure 1 shows potential itineraries for a passenger, whose access time is consumed by walking from the entry gate to the platform. The wait time is consumed at the platform for the train, considering FtB. The egress time is determined by the passenger who walks from the platform at the destination to the exit gate. For transfer passengers, transfer time consists of walking time and wait time. The walking time is dependent on the distance between the platforms of connecting trains, while the wait time is incurred at the platform for boarding the pick-up train. In this study, passengers are classified into nontransfer and transfer groups.

A nontransfer passenger, after swiping the metro card at entry station, will walk to the platform (i.e., access time is required), wait for boarding the immediate coming train m when the train m is not full, or might wait longer for later trains. After boarding the train, the passenger will spend in-vehicle time and then arrive at the destination station. Finally, the passenger will alight from the train and walk to the gate (i.e., egress time is required), swipe the card, and then exit. Under assumption 3, the in-vehicle and out-of-vehicle times of the passenger as shown in Figure 1(a) can be determined.

For transfer trips after alighting from a delivery train as shown in Figure 1(b), passengers will walk to the platform and wait for a pick-up train. Note that extra wait time is expected if the capacity of the immediate coming train is insufficient, which could be estimated if the probability of passengers FtB is known. The blue solid line represents the case that a passenger boards the pick-up train based on the estimated egress time with assumption 3. The branches with orange dash lines represent the case where passengers could board the immediate coming train m or board the train $m + 1$ (or $m + 2$) due to FtB. Therefore, the passenger with one transfer as shown in Figure 1(b) has three potential itineraries (e.g., Itinerary 1, Itinerary 2, or Itinerary 3).

Similarly, passenger itineraries with more than one transfer can be analyzed. Note that the access, egress, and transfer walking times are dependent on the layout of each metro station and walking speed. Since passengers only

TABLE 1: Notations.

Variable	Unit	Description
A_{mil}/D_{mil}	—	Arrival/departure time of train m at station i of line l
B_{mlij}	pass	Number of passengers who board train m from station i to station j of line l
C_l	pass	Capacity per train of line l
c_{ls}	pass	Number of seats per train of line l
d, i, j	—	Indices of stations
g, h, k	—	Number of trains skipped by passengers
K_i/K_s	—	Weighted average number of skipped trains at station i/s
L_{dl}	km	The length of the link connecting stations d and $d+1$ of line L
l, l', l''	—	Indices of lines
m, m', m''	—	Indices of trains
$n_i/n'_i/n'_s$	—	Maximum number of trains skipped by passengers at station $i/i'/s$
P_{mlij}	—	Probability of boarding train m of line l from station i to station j
p	—	Index of itineraries
Q_{mdl}	pass	Number of passengers in train m on link d of line l
$r_{ll's}$	sec	Weighted average transfer wait time from line l to line l' at station s
$r_{pll's}$	sec	Wait time at station s for transfer passengers from line l to line l' using itinerary p
s	—	Indices of transfer stations
t_{ai}	sec	Access time from the entrance to the platform at station i
t_{ej}	sec	Egress time from the platform at station j to the exit
t_i/t_j	—	Times when passengers swipe the card at entrance/exit
t_{vij}	sec	In-vehicle time from station i to station j
t_{wi}	sec	Weighted average wait time at station i
t_{wik}	sec	Wait time at station i skipped for k trains
W_{ik}	pass	Number of passengers who skipped k train at station i
$w_{ll's}$	sec	Transfer walking time from line l to line l' at station s
α_{ik}	—	Probability of the k trains skipped by passengers at station i
$\alpha_{ik}(p)$	—	Weight of skipping k trains at station i for boarding train $m+k$ on itinerary p
$\alpha_{pik}/\alpha_{p,s1,h}$	—	Probability of itinerary p for passengers who skipped k/g trains at station i/s_1
β_p/γ_p	—	Probability of itinerary p for transfer passengers with one transfer/two transfers
ψ_{ml}	m ²	Floor area for standees of train m on line l
φ_{mdl}	m ² /p	Average space per standee in train m on link d of line l
δ_{mdl}	—	Load factor of train m on link d of line l
δ_{ml}	—	Average load factor of train m on line l

swipe the card when entering and exiting the gates, the choice of transfer point(s) is unknown.

Assume that a passenger enters station i of line l at time t_i and would walk t_{ai} seconds to the platform. If this passenger could board the immediate coming train m , the constraint formulated as equation (1) must be sustained:

$$D_{m-1,i,l} \leq t_i + t_{ai} < D_{mil}. \quad (1)$$

Note that equation (1) is to ensure that the arrival time of the passenger is between the interdeparture times of trains $m-1$ and m . However, under overcrowding situation, the passenger will wait for more than one train and the resulting wait time will be discussed later.

Based on assumption 3, after a passenger arrives at destination station j , he/she would walk t_{ej} seconds from platform to the exit at time t_j . Thus, the passenger who boards the train m must satisfy

$$A_{mjl} \leq t_j - t_{ej} < A_{m+1,j,l}. \quad (2)$$

For a passenger who needs to transfer at station s , the relation among the delivery train arrival time, transfer walking time, and pick-up train departure time will be the concern to infer the possible itinerary. When the departure time of train m' of line l' at station s ($D_{m'sl'}$) is greater than

the arrival time of train m of line l (A_{msl}) plus transfer walking time ($w_{ll's}$), the transfer is able to be made successfully. Thus,

$$D_{m'sl'} \geq A_{msl} + w_{ll's}. \quad (3)$$

Note that all possible itineraries for a passenger can be found using equations (1)–(3).

3.4. In-Vehicle Time. The in-vehicle time, denoted as t_{vij} , can be obtained with train timetable and passenger's itinerary, using

$$t_{vij} = A_{mjl} - D_{mil}, \quad (4)$$

where A_{mjl} and D_{mil} represent the arrival and departure times of train m on line l at stations j and i , respectively.

3.5. Out-of-Vehicle Time. As discussed earlier, out-of-vehicle time consists of walking time and wait time. The average walking speed can be determined empirically. Hence, the access, egress, and transfer walking times are dependent on the layout of each station. Wait time (e.g., at origin and transfer stations) is a critical factor affecting the variation of

FIGURE 1: Space-time diagrams of passenger itineraries. (a) Passenger itineraries without transfer. (b) Passenger itineraries with one transfer.

travel time, which is dependent on the crowding conditions at stations.

3.5.1. Wait Time for Nontransfer Passengers. After a passenger arrives at the platform of station i , the immediate coming train m can be determined by equation (1), while train $m+k$ actually rode by the passenger can be determined by equation (2). Considering overcrowding situations, the wait time (t_{wik}) for train $m+k$ is expressed by

$$t_{wik} = D_{m+k,i,l} - t_{ai} - t_i. \quad (5)$$

Note that k is zero if the passenger is able to board the immediate coming train m . The number of passengers

skipped for k trains at origin station i denoted as W_{ik} can be estimated using the AFCS data, and the probability of skipping k trains denoted as α_{ik} can be determined by

$$\alpha_{ik} = \frac{W_{ik}}{\sum_k W_{ik}}, \quad k = 0, 1, 2, \dots, n_i, \quad (6)$$

where n_i is the maximum number of skipped trains by passengers at station i . The number of skipped trains at station i , denoted as K_i , is the weighted sum of skipped trains expressed by

$$K_i = \sum_{k=0}^{n_i} k \alpha_{ik}. \quad (7)$$

3.5.2. Wait Time for Transfer Passengers. According to the AFCS transaction records, train timetables, and the layout of the metro network, a set of potential itineraries of a passenger with one transfer is P . In addition to enter and exit stations, itinerary p ($p \in P$) also includes the station(s) to make transfer as well as the delivery and the pick-up train IDs. For instance, the pick-up train m' of line l' at transfer station s can be determined by the exit transaction record and equation (2). Hence, the departure time of train m' at s ($D_{m'sl'}$) can be determined, while the arrival time of delivery train $m+k$ at station i of line l ($A_{m+k,s,l}$) must satisfy the following condition:

$$A_{m+k,s,l} \leq D_{m'sl'} - w_{ll's}, \quad k = 0, 1, 2, \dots, n'_i. \quad (8)$$

Since the immediate coming train m of line l at station i can be inferred by equation (1) and the entry transaction records, the number of skipped trains (i.e., k) can be deduced using equation (8). For instance, the latest train that the transfer passenger must take to arrive at station s is train $m + n'_i$, and the maximum number of trains to skip is n'_i . Since station s can be inferred by the pick-up train schedule, the exit transaction record at destination station j , and equation (2), the key is to find the delivery train boarded at the origin station i . Hence, the probability of itinerary p is analogue to the probability of the train that passenger boarded at station i . Thus, the probability of itinerary p , denoted as β_p , that a passenger travels from station i to station j via s , who skips k trains at i , can be obtained by

$$\beta_p = \alpha_{pik} = \frac{\alpha_{ik}(p)}{\sum_p \alpha_{ik}(p)}, \quad (9)$$

where α_{pik} represents the probability that the transfer passenger can board train $m+k$ (skipped k trains) at origin station i on itinerary p . $\alpha_{ik}(p)$ is the weight of skipping k trains at station i for boarding train $m+k$ on itinerary p , which can be derived by equation (6). Considering the latest arrival time of a delivery train with equation (8), the maximum number of skipped trains (n'_i) at station i can be determined.

Under overcrowding situations, the wait time of a transfer passenger at origin station i (denoted as t_{wi}) is the weighted average of wait time considering the probability of passenger who would take itinerary p (i.e., k trains will be skipped at station i) and the associated wait time (t_{wik}) at station i . Thus,

$$t_{wi} = \sum_p t_{wik} \beta_p, \quad p \in P. \quad (10)$$

The number of skipped trains at station i (K'_i) is the weighted sum of skipped trains for all itinerary p multiplied by the associated probability β_p . Thus,

$$K'_i = \sum_p k \beta_p, \quad p \in P. \quad (11)$$

For itinerary p , the transfer wait time at station s is denoted as $r_{pll's}$ is $D_{m'sl'}$ less $A_{m+k,s,l}$ and then less the transfer walking time ($w_{ll's}$). Thus,

$$r_{pll's} = D_{m'sl'} - A_{m+k,s,l} - w_{ll's}. \quad (12)$$

Similarly, the transfer wait time at s denoted as $r_{ll's}$ is the weighted average of transfer wait times. Thus,

$$r_{ll's} = \sum_p r_{pll's} \beta_p, \quad p \in P. \quad (13)$$

Since the pick-up train is fixed but the delivery train is still uncertain, the number of skipped trains at station s is dependent on the arrival time of the delivery train. If a passenger boards train $m+k$ at station i and arrives at transfer station s and takes $w_{ll's}$ seconds to walk between platforms of connecting trains, the immediate pick-up train $m' - g$ whose departure time is $D_{m'-g,s,l'}$ can be determined by

$$D_{m'-g,s,l'} \geq A_{m+k,s,l} + w_{ll's}, \quad g = 0, 1, 2, \dots, n'_s, \quad (14)$$

where n'_s represents the maximum number of skipped trains at station s . Note that the pick-up train m' is known from equation (2) and the exit transaction record. The number of skipped trains (g) can be determined by equation (14). Similarly, the number of skipped trains at station s denoted as K'_s can be determined by equation (11).

For itinerary p with two transfers (e.g., the 1st and the 2nd transfer stations are denoted as s_1 and s_2 , respectively), a joint probability method is proposed to estimate passenger itinerary. Note that the 2nd pick-up train at s_2 can be determined by the exit transaction record and equation (2); therefore, the purpose is to estimate the delivery train at i and 1st pick-up train at s_1 . The probability of itinerary p (γ_p) is dependent on the probability that passenger boarding delivery train $m+k$ at origin station i (denoted as α_{pik}) and the 1st pick-up train $m' + h$ at s_1 (denoted as $\alpha_{p,s1,h}$) can be determined by

$$\gamma_p = \frac{\alpha_{pik} \alpha_{p,s1,h}}{\sum_p \alpha_{pik} \alpha_{p,s1,h}}, \quad (15)$$

where $\alpha_{p,s1,h}$ can be derived similarly as discussed for α_{pik} . Note that the departure/arrival time ($D_{m'+h,s1,l'}/A_{m'+h,s2,l'}$) of the 1st pick-up train $m' + h$ at s_1/s_2 must satisfy the following conditions, and the immediate coming and latest pick-up trains at s_1 can be identified:

$$\begin{aligned} D_{m'+h,s1,l'} &\geq A_{m+k,s1,l} + w_{l'l',s1}, & h = 0, 1, 2, \dots, n'_{s1}, \\ A_{m'+h,s2,l'} &\leq D_{m'',s2,l''} - w_{l'l'',s2}, & h = 0, 1, 2, \dots, n'_{s1} \end{aligned} \quad (16)$$

where n'_{s1} represents the maximum number of skipped trains at station s_1 . Similarly, passenger wait time and the number of skipped trains by passengers with two transfers can be determined using equations (10), (11), and (13).

For a complicated metro network, passengers may face multipath choice decision with transfers. It is necessary to identify potential train(s) for passengers to ride at origin as well as transfer stations considering spatiotemporal constraints, such as train arrival times at different stations of a line and different lines of the network as well as the possibility of FtB. According to the number of transfers associated with each path (also called itinerary), the probability

of each itinerary can be calculated by equations (9) and (15). Then, the average wait time and number of skipped trains can be estimated.

3.6. Quality of Service (QoS). To analyze the performance measures concerning QoS, we investigate *passenger wait time*, *average space per standee*, and *load factor* in different levels of details (e.g., link and train). As discussed earlier, wait time can be estimated by equations (5), (10), and (13). Other equations applied to assess average space per standee and load factor are discussed next.

3.6.1. Average Space per Standee. Average space per standee (ASPS) in square feet (or meters) per passenger can be used to describe the level of crowding on board of the vehicle and that for train m on link d (link connects station d to $d + 1$) of line l , denoted as ϕ_{mdl} , can be calculated by

$$\phi_{mdl} = \frac{\psi_{ml}}{Q_{mdl} - c_{ls}}, \quad (17)$$

where ψ_{ml} represents the floor area for standees. c_{ls} represents the number of seats per train of line l , and Q_{mdl} represents the number of passengers which can be derived by

$$Q_{mdl} = \sum_{i=N_{l-1}+1}^d \sum_{j=d+1}^{N_l} B_{mlij} P_{mlij}, \quad (18)$$

where B_{mlij} and P_{mlij} represent the number of passengers from station i to station j and the associated probability of them boarding train m of line l , respectively. Table 2 describes levels of crowding for transit vehicles [1], along with potential implications for passenger perspective.

3.6.2. Load Factor. Load factor at a point is defined as the ratio of passengers transported to spaces offered at maximum schedule load [37], which is a normalized measure used in this study. Therefore, one of the important indicators of system productivity is the load factor for link (δ_{mdl}) and train (δ_{ml}), which can be defined as the total passenger-miles traveled divided by the total capacity-miles provided. Thus,

$$\delta_{mdl} = \frac{Q_{mdl}}{C_l}, \quad (19)$$

$$\delta_{ml} = \frac{\sum_{d=1}^{N_l-1} Q_{mdl} L_{dl}}{\sum_{d=1}^{N_l-1} C_l L_{dl}},$$

where L_{dl} represents the length of the link connecting stations d and $d + 1$ of line l . C_l represents the capacity per train of line l .

4. Solution Approach

Considering the variation of travel time, the in-vehicle and out-of-vehicle times should be determined based on time varying demand and supply attributes. Since train

running times between stations and dwell times at stations are based on posted timetables, the in-vehicle time is relatively stable compared to out-of-vehicle time. A data-driven approach is developed to estimate the probability of passenger itinerary and the associated wait time as well as passenger load using the AFCS data and train schedule. The step procedure is discussed as follows and illustrated in Figure 2:

Step 1. Data inputs

- (i) Extract entry and exit transaction records from the AFCS data
- (ii) Input the train timetables, average access, and egress time at each station, as well as walking times at transfer stations

Step 2. Data preprocessing

- (i) Eliminate the abnormal trip data missing either entry or exit record
- (ii) Classify passengers into nontransfer and transfer trips according to the AFCS transaction records and the layout of metro network

Step 3. Extract the data of nontransfer trips

- (i) Identify the boarding train using exit transaction records and estimate wait time at the origin station
- (ii) Determine the number of skipped trains based on the schedule of immediate coming train
- (iii) Calculate the probability of the trains skipped by passengers and the associated wait time

Step 4. Extract the data of trips with one transfer

- (i) Identify the pick-up train by exit transaction records and infer the latest train that a passenger may board at the origin station by transfer connection constraint
- (ii) Determine the earliest coming train at the origin station by entry transaction records
- (iii) Generate the set of potential delivery trains and all potential itineraries
- (iv) Determine the maximum number of skipped trains at the origin and transfer stations and extract the probability of skipped trains at the origin station
- (v) Calculate the probability of each potential itinerary and the associated wait time and the number of skipped trains

Step 5. Extract the data of trips with two transfers

- (i) Identify the sets of potential delivery and the 1st pick-up train similarly to Step 4 and generate all potential itineraries
- (ii) Extract the probability of skipped trains at the origin and the 1st transfer station and calculate the probability of each potential itinerary
- (iii) Calculate wait time and the number of skipped trains at the origin and transfer stations

Step 6. Results output

TABLE 2: Levels of crowding.

Level	Space per standee		Passenger perspective
I	$>10.8 \text{ ft}^2/\text{p}$	$>1.00 \text{ m}^2/\text{p}$	Passengers are able to spread out
II	$5.4\text{--}10.8 \text{ ft}^2/\text{p}$	$0.5\text{--}1.0 \text{ m}^2/\text{p}$	Comfortable standing load that retains space between passengers
III	$4.3\text{--}5.3 \text{ ft}^2/\text{p}$	$0.40\text{--}0.49 \text{ m}^2/\text{p}$	Standing load without body contact
IV	$3.2\text{--}4.2 \text{ ft}^2/\text{p}$	$0.30\text{--}0.39 \text{ m}^2/\text{p}$	Occasional body contact
V	$2.2\text{--}3.1 \text{ ft}^2/\text{p}$	$0.20\text{--}0.29 \text{ m}^2/\text{p}$	Approaching uncomfortable conditions
VI	$<2.2 \text{ ft}^2/\text{p}$	$<0.20 \text{ m}^2/\text{p}$	Crush loading conditions

Note. See exhibits 5–17 [1].

FIGURE 2: Framework of data-driven approach.

- (i) Calculate the value of performance measures, including wait time, number of skipped trains, average space per standee, and load factor
- (ii) Output the results

5. Case Study

The studied Chengdu Metro network consists of 6 routes intersecting at 14 transfer stations as shown in Figure 3. Route 7 is a ring that intersects routes 1, 2, 3, 4, and 10. Hence, the maximum number of transfers per trip is less than or equal to two. The ridership on a typical weekday is more than three million. The size of train units (i.e., cars per train) varies with different routes.

The input data are mainly from the AFCS and train timetables. At each station, metro staffs are placed at platforms to assist boarding and alighting passengers and to ensure the timeliness of train departures. Other input data, such as walking time, were estimated based on actual walking distance and speed via the data collected at each station.

5.1. Analysis of Passenger Flow

5.1.1. Ridership Distribution over Time. The ridership distribution over time on a typical weekday showed obviously bimodal peaks as shown in Figure 4, which are the AM peak (i.e., 8:00 am~9:00 am) and off-peak (i.e., 12:00 pm~1:00 pm). It is worth noting that the number of passengers accessing the metro network at the AM peak is nearly 15% of daily ridership, which leads to a very crowding situation.

5.1.2. Passenger Flows Entering and Exiting Stations. The passengers entering and exiting the metro system during AM peak are shown in Figure 5. The largest entering volume was found at XP station (Route 2, 10,300 pass/hr), and the largest exiting volume was at 3TFS station (Route 1, 17,000 pass/hr). Most stations with larger entering flows concentrate near the residential zones, while the major exiting flows fall in the stations near the employment areas.

5.2. Analysis of Travel Time

5.2.1. Travel Time Components. The assessment of travel time components for passengers is shown in Figure 6. In general, the average travel time of transfer passengers is obviously longer than that of nontransfer passengers because of longer travel distance and expected wait/transfer time. In-vehicle time accounts for 69.7% of the travel time for nontransfer passenger (Figure 6(a)). Wait time at origin station is the second longest (19.9% of travel time), while the entry and exit walking times are minors and around 5.0% of travel time, respectively. For passengers with one transfer, in-vehicle time accounts for 70.0% of the travel time, and the percentage of out-of-vehicle time is 30.0%, in which the transfer time ratio is over 15.0% (see Figure 6(b)). For passengers with two transfers, the average travel time is obviously the longest; however, the percentage (55.2%) of in-

vehicle time is lower because of increased transfer time as shown in Figure 6(c).

5.2.2. Distribution of Wait Time. Figure 7 illustrates the probability distribution of wait time at different time period. The peaks represent the greatest probability of wait time, and the wait time distributions are similar between the AM peak and PM peak. About 60% of passengers spent less than 3 minutes to wait at station in AM and PM peaks. In general, more passengers experience less wait time at peak periods than that at off-peak due to the smaller headways.

The probability (P_{t_w}) density obeys the Gaussian distribution, which is $P_{t_w} = ae^{-((t_w-b)/c)^2}$, and t_w represents wait time. The fitting results such as the value of parameters (a , b , and c), the Sum Square Error (SSE), R -squared (R^2), Adjusted R -squared (Adjusted R^2), and Root Mean Square Error (RMSE) are shown in Table 3. The peak value $t_{w\max}$ after calculating the derivatives of the fitting functions represents the value of wait time with highest probability. At AM peak, $t_{w\max}$ is 1.98 minutes; at PM peak, $t_{w\max}$ is 2.05 minutes; at off-peak, $t_{w\max}$ is 2.88 minutes; at all day, $t_{w\max}$ is 2.80 minutes.

5.3. Average Wait Time and Number of Skipped Trains on Line 1. The average wait times at stations as an example on line 1 at AM peak and off-peak are shown in Figure 8. Compared with Figure 9, it is found that the average wait time increases as the number of trains that passengers needed to wait for increases. In general, the average wait time at off-peak is longer than that at AM peak because the headway in off-peak is longer than that in the AM peak, except for a few stations (i.e., TFS, SCG, NJQ, and SRS). The average wait time at many stations (i.e., TFS, SCG, SRS, etc.) is longer than the wait time estimated by previous methods [1] in the peak because of overcrowding, while that is significantly reduced in the off-peak. Note that, as indicated in Figures 8 and 9, there are no data at WGS and SC stations, which are the end terminals for the feeder and main line services, respectively, as shown in Figure 3.

Figure 9(a) illustrates the probability of the number of skipped trains by passengers at stations of line 1 (WJN-SC) at AM peak. The number of FtB passengers varies with time and location, which reveals the crowding situation of the trains and the platform. During AM peak, passenger demand increases and sometimes exceeds the train capacity. Therefore, some FtB passengers must wait for later train(s). TFS, SCG, and SRS stations are deemed as bottleneck stations on line 1, which are all transfer stations. It is found that some passengers skip more than 5 trains and wait time is more than 13 minutes as the headway was 130 seconds. For the off-peak, Figure 9(b) illustrates the similar information to Figure 9(a). It reveals that the status of FtB was eased because of lighter demand.

5.4. Average Wait Time and Number of Skipped Trains at TFS Station on Line 1. TFS station located in the center of Chengdu city is a transfer station intersected by Routes 1 and

FIGURE 3: The Chengdu Metro network.

FIGURE 4: Ridership distribution over time on a typical weekday.

2. The daily volumes consist of nearly 41,000 nontransfer passengers and 180,000 transfer passengers. Taking line 1 as an example, the load factor of the link TFS-JJH affects the number of skipped trains at TFS as shown in Figure 10(a). It

was found that as the load factor increases, the number of skipped trains increases.

As shown in Figure 10(b), the red and blue lines represent the average wait times estimated with the developed method

FIGURE 5: Passenger flows at stations during AM peak. (a) Entry pass. (b) Exit pass.

FIGURE 6: Travel time components for passengers. (a) Nontransfer passengers. (b) Passengers with one transfer. (c) Passengers with two transfers.

FIGURE 7: Probability distributions of wait time. (a) AM peak. (b) PM peak. (c) Off-peak. (d) All day.

TABLE 3: Fitting results.

	a	b	c	SSE	R^2	Adjusted R^2	RMSE
AM peak	0.2201	1.978	2.022	0.01712	0.8475	0.8366	0.02473
PM peak	0.2196	2.054	2.039	0.01456	0.8672	0.8577	0.02281
Off-peak	0.1798	2.882	2.801	0.00821	0.9101	0.9037	0.01712
All day	0.1585	2.797	3.136	0.01031	0.8686	0.8592	0.01919

and the method suggested by TCQSM, respectively, while the green line is headway. It was found that the wait time estimated by the proposed data-driven approach is much higher and closer to the wait time that passengers experienced.

5.5. Analysis of Passenger Load

5.5.1. Average Space per Standee. In the QoS analysis, TCQSM provides general guidance for passenger load standards that can be expressed as an average during a peak 15-, 30, or 60-min period [1]. As shown in Figure 11, the bidirectional (inbound and outbound) ASPS during AM peak is produced (8:00–9:00 AM). The segments colored in green through red represent the ASPS from high to low, which also reflects the average passenger load in metro network.

The distributions of passenger flows in and out of stations were heterogeneous over space and time. For Route 1, the crowding situations existed on the links of downtown or

employment concentration areas in both directions. Most routes have similar crowding situations in the downtown area as Route 1, except loop Route 7.

The ASPS of any train can be obtained if the number of boarding passengers, number of seats, and the standing area per train are known. Therefore, the crowding situation of any coming trains is able to be displayed. Some passengers might be willing to skip the immediate coming train and wait a little longer for the next one (if less crowded).

The ASPS and associated standard deviation (SD) on each link of line 1 in the AM peak are illustrated in Figure 12. It indicates that the trains were not congested and there were no standees on the links of WJN-SXL, HSP-SC (main line), and SH-WGS (feeder line) because of high ASPS and low SDs. However, the trains running between stations NRS and 3TFS are constantly crowded with low ASPS and SD, where the ASPS on the links between stations LMS and JCP represents uncomfortable conditions (e.g., Levels V and VI according to Table 2).

FIGURE 8: Average wait time at stations on line 1 (WJN-SC).

FIGURE 9: Probability of number of skipped trains by passengers on line 1 (WJN-SC). (a) AM peak. (b) Off-peak.

FIGURE 10: Number of skipped trains and average wait time at TFS station on line 1. (a) Number of skipped trains. (b) Average wait time.

FIGURE 11: Average space per standee (8:00–9:00 AM).

5.5.2. Load Factor. The spatiotemporal load factor of each train on line 1 (WJN-SC) is shown in Figure 13. The empty areas in Figure 13 represent no service provided since line 1 offers the feeder line and main line train services. The color

of a train profile represents the load factor on each link. The congestion level of each train is illustrated in a space-time diagram, in which the red segments with high load factors appeared during AM and PM peaks.

FIGURE 12: Average space per standee and standard deviation on links of line 1 (AM).

FIGURE 13: Spatiotemporal load factor for each train on line 1 (WJN-SC).

The spatiotemporal distribution of passenger flow is obviously uneven on line 1, which indicates that the load factor at AM peak is more serious than that at PM peak. For AM peak, more trains have high load factor with longer distance on line 1, while in PM peak, less trains have high load factor with shorter distance. The arrival times related to the work starting times in the AM are focused on, but the off-work time in the PM varies within a longer period. Some passengers work late, and others leave early.

The information about the congestion level of any coming train and station is useful for operator to draft the operation strategy such as the short-turn service, enhancing the train capacity, and proposing reasonable passenger flow limited measures at some crowding stations. In addition, passengers could obtain more accuracy travel time and make more appropriate travel plan.

6. Conclusions

With the AFCS data, the spatiotemporal passenger entry and exit distributions at metro stations can be determined. This study presents a data-driven approach to estimate the probabilities of passenger itineraries considering FtB for a congested metro network. This information is critical for estimating in-vehicle and out-of-vehicle travel times (e.g., walking, wait, and transfer times) as well as spatiotemporal passenger loads in different details (i.e., train, link, etc.). This information would be extremely useful in justifying and/or optimizing service planning (i.e., frequency, train size, service patterns, etc.) to elevate metro's QoS. A case study was conducted using the AFCS data provided by Chengdu Metro in China. The results of the analysis suggest that the probability of wait time obeys the Gaussian distribution. The average wait time, particularly at the AM peak, is much

higher than a half of headway applied in previous studies. The spatiotemporal wait time and the number of skipped trains at each station were able to be estimated effectively.

Several performance indicators related to passenger load were assessed, which include passenger flow, average space per standee, and load factor of any train running on the network. These indicators are critical for various extensions in advanced travel planning, travel time prediction, QoS assessment, developing strategic operating plans, demand forecasting, and dynamic passenger assignment in the context of Mobility as a Service (MaaS). As an immediate extension of this study, the path-choice survey will be conducted and the collected data with train AVL information may be applied to verify the results derived in this study.

Data Availability

The automatic fare collection system (AFCS) data and existing train timetables used to support the findings of this study are unavailable for the public due to confidentiality agreements with research collaborators. Supporting data can only be available to bona fide researchers subject to a nondisclosure agreement.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was financially supported by the Key Research and Development Plan of the Ministry of Science and Technology, China (Grant No. 2018YFB1601402) and the National Engineering Laboratory of Integrated Transportation Big Data Application Technology, China (Grant No. CTBDAT201910). The authors appreciate the data support from Chengdu Metro.

References

- [1] Kittelson and Associates, Inc., Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, and Arup, *TCRP Report 165: Transit Capacity and Quality of Service Manual*, Transportation Research Board of the National Academies, Washington, DC, USA, 3rd edition, 2013.
- [2] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: understanding tube station congestion patterns from trip data," in *Proceedings of the ACM SIGKDD international Workshop on Urban computing*, pp. 134–141, New York, NY, USA, 2012.
- [3] S. Liu, Y. Liu, L. Ni, M. Li, and J. Fan, "Detecting crowdedness spot in city transportation," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1527–1539, 2012.
- [4] J. M. Bunker, "Assessment of transit quality of service with occupancy load factor and passenger travel time measures," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2535, no. 1, pp. 45–54, 2015.
- [5] Y. Y. Ulusoy, S. I.-J. Chien, and C.-H. Wei, "Optimal all-stop, short-turn, and express transit services under heterogeneous demand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2197, no. 1, pp. 8–18, 2010.
- [6] Y. Y. Ulusoy and S. I.-J. Chien, "Optimal bus service patterns and frequencies considering transfer demand elasticity with genetic algorithm," *Transportation Planning and Technology*, vol. 38, no. 4, pp. 409–424, 2015.
- [7] H.-z. Qu, S. I.-J. Chien, X.-b. Liu, P.-t. Zhang, and A. Bladikas, "Optimizing bus services with variable directional and temporal demand using genetic algorithm," *Journal of Central South University*, vol. 23, no. 7, pp. 1786–1798, 2016.
- [8] Z. Guo, "Mind the map! The impact of transit maps on path choice in public transit," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 7, pp. 625–639, 2011.
- [9] Y. Xin, L. Fu, and F. F. Saccomanno, "Assessing transit level of service along travel corridors," *Transportation Research Record*, vol. 1927, pp. 259–267, 2005.
- [10] J. M. Bunker, "How transit route passenger load and distance can together influence quality of service," in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2014.
- [11] P. G. Furth, B. Hemily, T. H. Muller, and J. G. Strathman, *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management*, Transportation Research Board of the National Academies, Washington, DC, USA, 2006.
- [12] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and destination estimation in New York city with automated fare system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1817, no. 1, pp. 183–187, 2002.
- [13] L. Sun, J. G. Jin, D.-H. Lee, K. W. Axhausen, and A. Erath, "Demand-driven timetable design for metro services," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 284–299, 2014.
- [14] W. Zhu, F. Zhou, J. Huang, and R. Xu, "Validating rail transit assignment models with cluster analysis and automatic fare collection data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2526, no. 1, pp. 10–18, 2015.
- [15] L. Hong, W. Li, and W. Zhu, "Assigning passenger flows on a metro network based on automatic fare collection data and timetable," *Discrete Dynamics in Nature and Society*, vol. 2017, Article ID 4373871, 10 pages, 2017.
- [16] A. Tirachini, L. Sun, A. Erath, and A. Chakirov, "Valuation of sitting and standing in metro trains using revealed preferences," *Transport Policy*, vol. 47, pp. 94–104, 2016.
- [17] T. Liu, A. Ceder, J. Ma, and W. Guan, "Synchronizing public transport transfers by using intervehicle communication scheme: case study," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2417, no. 1, pp. 78–91, 2014.
- [18] W. Li, R. Xu, Q. Luo, and S. Jones, "Coordination of last train transfers using automated fare collection (AFC) system data," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2209–2225, 2016.
- [19] X. Liu, M. Huang, H. Qu, and S. Chien, "Minimizing metro transfer waiting time with AFCS data using simulated annealing with parallel computing," *Journal of Advanced Transportation*, vol. 2018, Article ID 4218625, 17 pages, 2018.
- [20] L. Sun, D.-H. Lee, A. Erath, and X. Huang, "Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system," in *Proceedings of the ACM SIGKDD international Workshop on Urban computing*, pp. 142–148, New York, NY, USA, 2012.

- [21] F. Zhang, J. Zhao, C. Tian, C. Xu, X. Liu, and L. Rao, "Spatio-temporal segmentation of metro trips using smart card data," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1137–1149, 2015.
- [22] D. Luo, L. Bonnetain, O. Cats, and H. van Lint, "Constructing spatiotemporal load profiles of transit vehicles with multiple data sources," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 175–186, 2018.
- [23] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transportation Research Procedia*, vol. 23, pp. 362–379, 2017.
- [24] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "A probabilistic passenger-to-train assignment model based on automated data," *Transportation Research Part B: Methodological*, vol. 104, pp. 522–542, 2017.
- [25] E. Miller, G. E. Sánchez-Martínez, and N. Nassir, "Estimation of passengers left behind by trains in high-frequency transit service operating near capacity," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 497–504, 2018.
- [26] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach," *Transportation Research Part B: Methodological*, vol. 67, pp. 235–263, 2014.
- [27] H. Qu, X. Xu, and S. Chien, "Analysis of passenger load and wait time for a metro system with automatic fare collection data," in *Proceedings of the Transportation Research Board 98th Annual Meeting*, Washington, DC, USA, 2019.
- [28] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.
- [29] Y. Sun and P. M. Schonfeld, "Schedule-based rail transit path-choice estimation using automatic fare collection data," *Journal of Transportation Engineering*, vol. 142, no. 1, Article ID 04015037, 2016.
- [30] J. Zhao, F. Zhang, L. Tu et al., "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 790–801, 2017.
- [31] Y. Jiang and W. Y. Szeto, "Reliability-based stochastic transit assignment: formulations and capacity paradox," *Transportation Research Part B: Methodological*, vol. 93, pp. 181–206, 2016.
- [32] T. Liu and A. Ceder, "Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: a bi-objective bi-level model using deficit function approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 935–955, 2018.
- [33] Y. S. Zhang and E. J. Yao, "Splitting travel time based on afc data: estimating walking, waiting, transfer, and in-vehicle travel times in metro system," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 539756, 11 pages, 2015.
- [34] H. Lee, D. Zhang, T. He, and S. H. Son, "Metrottime: travel time decomposition under stochastic time table for metro networks," in *Proceedings of the 2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, Hong Kong, China, pp. 1–8, May 2017.
- [35] A. Tavassoli, M. Mesbah, and A. Shobeirinejad, "Modelling passenger waiting time using large-scale automatic fare collection data: an Australian case study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 500–510, 2018.
- [36] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: a smart card data analysis," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 292–306, 2018.
- [37] V. R. Vuchic, *Urban Transit: Operations, Planning, and Economics*, John Wiley & Sons, Hoboken, NJ, USA, 2017.