A Study on Train Travel Time Simulation
Focused on Detailed Dwell Time Structure and On-site Inspections

Shigeaki Adachi, Hideyuki Yoshino, Masahito Koresawa, Giancarlos Troncoso Parady, Kiyoshi Takami and Noboru Harata

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 24, 2019
A Study on Train Travel Time Simulation Focused on Detailed Dwell Time Structure and On-site Inspections

Shigeaki Adachi a,1, Hideyuki Yoshino b, Masahito Koresawa b, Giancarlos Troncoso Parady a, Kiyoshi Takami a, Noboru Harata a

a Department of Urban Engineering, The University of Tokyo
Engineering building 14, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan
1 E-mail: adachi@ut.u-tokyo.ac.jp, Phone: +81 (3) 3837-7318
b Department of Operation, Tokyo Metro Co., Ltd.
3-19-6 Higashi-ueno, Taito-ku, Tokyo 110-8614 Japan

Abstract
In order to reduce passenger congestion during morning rush hour, railway companies in the Tokyo metropolitan area have increased the number of trains. On the other hand, once a train exceeds a dwell time due to sudden events such as passengers rushing onto a train, passengers agglomerating in specific cars and doors, objects getting caught in doors etc., delays propagate to subsequent trains quickly. To evaluate daily train transport stability and countermeasures against train delays, a train travel time simulation model is needed. However, it has been difficult so far to replicate the occurrence of sudden events and the fluctuations in passenger demand. In this paper, we use detailed data based on dwell time structure and on-site inspections to construct a train travel time simulator. In addition, we evaluate several case-studies of timetable adjustments and passenger demand variations.

Keywords
Train delay, Train travel time simulation, Ticket gate ingress and egress record data, Smirnov-Grubbs test, Dwell time

1 Introduction
Railway companies in Tokyo metropolitan area of Japan have increased the number of trains to alleviate passenger congestion and improve train delays during morning rush hour. However, train headways are limited by the capacity of the signalling system. Under such circumstances, train delays propagate to subsequent trains because of short headways. Furthermore, during boarding and alighting, when small sudden events such as passengers rushing onto a train, passengers agglomerating in specific cars and doors, or objects getting caught in doors occur, dwell times are extended.

Train travel time simulation models have been constructed so far. Railway simulation using traffic record data has been studied by Carey (1999), Hürlimann (2004), Van der Meer (2010), Graffagnino (2012). Furthermore, Hansen et al. (2014) have studied various kinds of train simulators focusing on railway system functions. Janecek (2010) studied simulations focusing on changes in the infrastructure and timetable. Ushida et al. (2011) developed a chromatic diagram visualized reflecting train delays as colours. In terms of train delay measures, Yamamura (2013 & 2014) and Adachi (2016) have studied various kind of measures against train delays on one of the most congested lines in Tokyo and evaluate those of effects on operation stability.

However, these studies have mainly focused on railway system, simulator functions and
train delay measures. So far, it has been difficult to consider daily passenger demand and the effect of small sudden events which occur frequently. Thus, consideration of these detailed elements is important to improve daily train operations, adjustment of planned train timetables, and passenger demand control. We have focused on composition of dwell time, and the relationship between passenger demand and dwell time including such sudden events that has not been well studied so far.

In this paper, we construct a detailed train travel time simulation focusing on the Tokyo Metro Tozai Line, which is one of the most congested lines in Tokyo.

2 Train diagram composition

Train head way is constructed by dwell time and minimum headway and buffer time. In a dense timetable such as lines running in the city center, buffer times are set at almost minimal, therefore once dwell time extends, buffer time becomes negative. This means that train delays propagate to subsequent trains.

Dwell time is segmented into 4 parts: passenger alighting time \( (A) \), passenger boarding time \( (B) \), door closing confirmation time \( (C) \), and safety confirmation time \( (S) \). In terms of door closing confirmation time \( (C) \), station staff judge timing of door closing at the end of passenger boarding. After passenger board, the staff give a signal to close doors to the conductor, and the conductor close the doors. After door close, station staff confirm the safety along cars and give a signal for departure to the conductor. This operation time is defined as safety confirmation time \( (D) \). The most time-consuming door to alight and board affects sum of passenger alighting time \( (A) \) and passenger boarding time \( (B) \).

Furthermore, it takes 2 seconds for doors to open after arriving at a station. According to these definitions, dwell time at station \( i \) of train \( j \) is defined as (1). All times are given in seconds.

\[
D_{i,j} = 2(\text{sec}) + \max_{k,l}(A_{i,j,k,l} + B_{i,j,k,l}) + C_{i,j} + S_{i,j}
\]

(1)

\( 2(\text{sec}) \): Door opening time of 2 seconds
\( A_{i,j,k,l} \): Alighting time at station \( i \) of train \( j \), car No. \( k \), door No. \( l \)
\( B_{i,j,k,l} \): Boarding time at station \( i \) of train \( j \), car No. \( k \), door No. \( l \)
\( C_{i,j} \): Closing confirmation time at station \( i \) of train \( j \)
\( D_{i,j} \): Dwell time at station \( i \) of train \( j \)
\( S_{i,j} \): Safety confirmation time at station \( i \) of train \( j \)

3 Factors influencing each time to construct dwell time

To build a detailed train travel time simulation, it is necessary to know what kind of factors influence each time to construct dwell time. Factor affecting composed time are illustrated in Figure 1. Alighting and boarding times are influenced by the number of passengers and by the number of passengers in a car. In Tozai line, some trains have wider car than usual cars. This width also affects alighting and boarding times.

In terms of door closing operations, when staff judge the timing in some station, multiple station staff members cooperate due to curved nature of some platform and depending on the congestion situation on the platform. Door closing confirmation time fluctuates depending on these characteristics.

After door close, staff confirm safety along cars in the same way as during door closing
confirmation operations. Safety confirmation time also fluctuates depending on these characteristics.

We estimated each time model in dwell time considering the abovementioned causes.

4 Train Travel Time Simulation Outline

The outline of the simulation is illustrated in Figure 2. In the initial condition, the simulation starts with 78 trains running on the Tozai line in direction of the city center between 6:30 to 10:00 distributed along stations.

First, departure times and number of passengers for each car at each starting station are input. Departure time data is acquired from train traffic record data which is obtained from electric circuit on a track at each station. Passenger number data is acquired from a five-day on-site inspection conducted on November 2015. Then, each time that makes up dwell time is estimated for each train.

In terms of the number of passengers alighting and boarding, ticket gate ingress and egress count record data aggregated in 30-minute intervals is utilized. Using these data, the number of alighting passengers is allocated to each train and car based on the number of passengers on arrival. The number of boarding passengers is allocated based on train headways. Furthermore, the calculated number of passengers is allocated to each door based on rate of door utilization observed during the on-site inspections. We model alighting and boarding times using linear regression analysis.

Door closing confirmation time and safety confirmation time are estimated based on on-site inspection results. Especially during door closing confirmation time, there are some small sudden events such as passengers rushing onto a train, passengers agglomerating in specific cars and doors, objects getting caught in doors etc. These events must be considered to build a more detailed simulation. In this study, these events are applied by Smirnov-Grubbs test.

Running time is calculated depending on whether the buffer time is negative or positive. Minimum headways are determined by the signalling system, so excess of planned running times is influenced by the negative buffer time at each station.
4.1 Estimation of alighting and boarding times

To estimate alighting and boarding times, the number of passengers should be calculated. Ticket gate egress and ingress data is utilized to estimate them. The cumulative distribution is approximated by Gompertz curve (Figure 3) which has property of left-right asymmetry around the inflection point and is defined as \( y = a \cdot b^e^{-ce} \). This property can precisely express each time zone passenger demand. The number of egress and ingress passengers in second-scale are derived by the curve. In a precise sense, time differences between ticket gate and train door should be considered. In this simulation, the time difference between ticket gate and the most time-consuming door to alight and board is considered.

In terms of the ticket gate egress, the data has OD record for each 30-minute time interval, and boarding direction of egress passenger is observed. In this model, it is no need to grasp detailed from-where-to-where information, it is sufficient to grasp the number of alighting and boarding passengers. To distribute egress passengers to each train, the total number of egress passengers are calculated as following (2) to (4). Furthermore, train direction to the city center is defined as \( A \) and train direction to the suburbs is defined as \( B \).
In general, the number of alighting passengers for each train is influenced by those of number of passengers on arrival defined as $\text{ArrCon}_d,i,j$ which are set as on-site inspection results at starting station on November 2015. Given that $\text{NA}_{l,t,A}$ is the number of alighting passengers at station $i$ on time zone $t$ for direction $A$, the equations are expressed as (5) and (6).

$$\text{NA}_{d,i,j} = \text{NA}_{d,i} \cdot (\text{ArrCon}_{d,i,j} / \text{ArrCon}_{d,i,\text{SJ}})$$ \hspace{1cm} (5)

$$\text{ArrCon}_{d,i,\text{SJ}} = \sum_{j \in \text{SJ}} \text{ArrCon}_{d,i,j}$$ \hspace{1cm} (6)

$\text{NA}_{d,i,j}$: The number of alighting passengers on day $d$ at station $i$ on train $j$  
$\text{ArrCon}_{d,i,j}$: Number of passengers on arrival on day $d$ at station $i$ on train $j$  
$\text{SJ}$: Set of train $j$
On the other hand, ticket gate ingress data doesn’t have OD record. Thus, the number of boarding passengers and rate of direction $A$ are calculated using on-site inspection data and $Eg_{d,i,t,A}$. The number of passengers at the time of departure for direction $A$ is calculated that the number of passengers on arrival plus alighting passengers minus boarding passengers.

The equations are expressed as (7) and (8). To simulate on the day, which is not inspection days, $DRb_{d,i,A}$ is adopted as average rate.

$$Ing_{d,i,t,A} = (\sum_{j\in S_j I_t} DepCon_{d,i,j} - \sum_{j\in S_j I_t} ArrCon_{d,i,j}) + Eg_{d,i,t,A}$$

(7)

$$DRb_{d,i,A} = \frac{\sum_t Ing_{d,i,t,A}}{\sum_{j\in S_j I_t} Ing_{d,i,t,A\&B}}$$

(8)

$Ing_{d,i,t}$: The number of ingress passengers on day $d$ at station $i$ on time zone $t$

$DepCon_{d,i,j}$: The number of passengers at the time of departure on day $d$ at station $i$ of train $j$ for direction $A$

$DRb_{d,i,A}$: Average rate of direction $A$ on day $d$ at station $i$ on time zone $t$

$S_j I_t$: Set of train $j$ at station $i$ on time zone $t$

Using $DRb_{d,i,A}$ and ticket gate ingress data, the number of boarding passengers each train is calculated as following (9) to (11). Since it is difficult to grasp how long it takes for passengers to get on the train during dwell time, then the number of boarding passengers each train is defined as the cumulative numbers between subsequent train’s arrival time and following train’s arrival time.

$$Ing_{d,i,t,A} = Ing_{d,i,t,A\&B} \cdot DRb_{d,i,A}$$

(9)

$$Y_{d,i}(s) = Y’_{d,i}(s - Mov_{i,A})$$

(10)

$$NB_{d,i,j} = Y_{d,i,j}(s_{d,i,j}) - Y_{d,i,j}(s_{d,i,j-1})$$

(11)

$DRb_{d,i,A}$: Rate of direction $A$ on day $d$ at station $i$ on time zone $t$

$Y’_{d,i}(s)$: Function of cumulative distribution approximated $Ing_{d,i,t,A}$

$Y_{d,i}(s)$: Function of cumulative distribution adjusted the time difference on $Y’_{d,i}(s)$

$NB_{d,i,j}$: The number of boarding passengers on day $d$ at station $i$ on train $j$ for direction $A$

To distribute alighting and boarding passengers to each car and door, utilization rate of cars and doors must be estimated. Utilization rate of car each station is estimated from car weight data acquired between October 2015 and December 2015. And utilization rate of each door is grasped from the on-site inspection results. Both rates are implemented as fixed average value on the simulator. Estimated dwell times almost depend on specific door, but considering given normal random value of estimation error in the models, the target doors are subject to variation.
4.2 Alighting time model

In terms of alighting time, two significant parameters are adopted, one is the number of alighting passengers and second is wider doors described earlier. To create the model, we utilize the video recording data which records passenger alighting and boarding on the platform at each station. Alighting number of passengers counting data, which is each 391 samples, is given by video data. The number of passengers and wider door are determined from on-site inspections. In fact, passenger flow on platform affects dwell time. However, it is assumed that the model expresses the effects due to on-site inspection results including the flow.

In the alighting time regression model, explanatory variables are the number of alighting passengers and the presence or absence of wider door. The equation is expressed as (12). Figure 4 and Table 1 show the results.

\[ A_{i,j,k,l} = \alpha + \beta_1 \cdot NA_{i,j,k,l} + \beta_2 \cdot \text{Wide} + \epsilon \] (12)

\( A_{i,j,k,l} \): Alighting time at station \( i \) of train \( j \), car No. \( k \), door No. \( l \)  
\( NA_{i,j,k,l} \): The number of alighting passengers at station \( i \) of train \( j \), car No. \( k \), door No. \( l \)  
\( \text{Wide} \): Wider door dummy  
\( \alpha, \beta_1, \beta_2 \): Parameter  
\( \epsilon \): Error term

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.89</td>
<td>19.97</td>
<td>1.56E-61</td>
</tr>
<tr>
<td>Number of alighting passengers</td>
<td>0.52</td>
<td>43.58</td>
<td>1.5E-151</td>
</tr>
<tr>
<td>Wider door dummy</td>
<td>-1.52</td>
<td>-6.19</td>
<td>1.5E-09</td>
</tr>
</tbody>
</table>

R²: 0.83  
Sample: 391 trains

The result obtains good fit by R²0.83, however there is variability between measured value and estimated value due to uncertain passenger flow. Therefore, the estimated value of alighting time is given by adding the normal random value of estimation error.
4.3 Boarding time model

In terms of boarding time, two parameters are estimated, the number of boarding passengers and the number of passengers in the car as of departure. To create the model, we utilize the video recording data as is the case with alighting model. Boarding number of passengers counting data has also 391 samples.

In the boarding time model, since boarding time tends to extend due to congestion, and this distribution increase towards one side the dependent variable is log-transformed. Explanatory variables are the number of boarding passengers and the number of passengers in the car as of departure. The wider door dummy is insignificant in the boarding time model. The equation is expressed as (13).

\[
\log B_{i,j,k,l} = \alpha + \beta_1 \cdot NB_{i,j,k,l} + \beta_2 \cdot DepCon_{i,j,k} + \epsilon
\]  

\(B_{i,j,k,l}\): Boarding time at station i of train j, car No. k, door No. l
\(NB_{i,j,k,l}\): The number of boarding passengers at station i of train j, car No. k, door No. l
\(DepCon_{i,j,k}\): Number of passengers at departure time at station i of train j, car No. k
\(\alpha, \beta_1, \beta_2\): Parameter
\(\epsilon\): Error term

Figure 5 shows the relationship between measured value and estimated value in boarding time. The result obtains good fit from \(R^2 0.67\), however there is variability between measured value and estimated value due to uncertainty passenger flow. Therefore, the estimated value of boarding time is given by adding the normal random value of estimation error.

Table 2: Result of boarding time model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.63</td>
<td>17.56</td>
<td>3.37E-51</td>
</tr>
<tr>
<td>Number of boarding passengers</td>
<td>0.030</td>
<td>26.96</td>
<td>6.1E-91</td>
</tr>
<tr>
<td>Number of passengers as of departure</td>
<td>0.00051</td>
<td>2.36</td>
<td>0.019</td>
</tr>
</tbody>
</table>

\(R^2: 0.67\) Sample: 391 trains
4.4 Estimation of door closing confirmation time
Door closing confirmation time depends on station staff operations. To estimate door closing confirmation time, normal random numbers were simulated based on the distribution observed during the on-site inspections at each station. Moreover, detailed analysis of the time should consider small sudden events that happen frequently. The events are considered that a dwell time excess degree is discerned by Smirnov-Grubbs test based on long term dwell time records.

Regarding train \( j \), the test statistics is defined as \( T_j \), the logarithmic value of dwell time is defined as \( X_j \), the average of logarithmic value of dwell time is defined as \( \bar{X}_j \), the standard deviation is defined as \( s_j \), the equation is expressed as (14). This judgement is focused on excess dwell time, so one sided-testing is adopted. To detect the time of sudden events precisely, we set the data that have nearly as much passenger demand at each station.

\[
T_j = \frac{X_j - \bar{X}_j}{s_j} 
\]

(14)

4.5 Estimation of safety confirmation time
Safety confirmation time also depends on station staff operations. As such, similar to door closing confirmation time, normal random numbers were simulated based on the distribution observed during the on-site inspections at each station.

4.6 Estimation of running time
To estimate running time, buffer time is considered. If the buffer is positive, the train would run following the planned running time. However, if the buffer time is negative, subsequent trains slow down or stop between stations because they are too close to the preceding train. The buffer time is determined by the signalling system design at each station. The phenomenon is illustrated in Figure 6, 7 and following (15) and (16). In figure 6, train headway (H) is segmented into 3 parts: dwell time (D), minimum headway, which is determined by signaling system each station (MH), buffer time (Bu), running time (R). The red lines are expressed actual train behavior, and red letters with dash are actual time.

Figure 6: Mechanism of train delay propagation

\[
\begin{align*}
\text{Distance} & \quad \text{Station}_{i+1} \\
\text{Station}_i & \quad \text{Train}_i \quad H_{i,j} \quad \text{MH}_i \\
\text{Station}_{i-1} & \quad \text{Train}_{i-1} \quad H'_{i-1,j+1} \quad \text{MH}_i \\
\text{Time} & \quad R_{i-1,j+1} \\
R'_{i-1,j} & \quad R_{i-1,j+1} \\
D_{i,j} & \quad D'_{i,j} \\
\end{align*}
\]
When buffer time is positive, the train driver can adjust to recover lost time, but train driver operation is different with each driver. Hence, error term $\varepsilon$ is considered recovery time or variation of train driver’s operation because the analyzed line is not operated by like an automatic train operation system. The error is given normal random value.

Figure 1 is expressed by detailed dwell time, then Figure 6 is expressed focusing on effect of train behaviors.

$$Bu'_{i,j+1}=H'_{i-1,j+1}-(R'_{i-1,j}R_{i-1,j})D'_{i,j}MH_{i,j+1} + \varepsilon < 0 \quad (15)$$

$$Bu'_{i,j+1}=H'_{i-1,j+1}-(R'_{i-1,j}R_{i-1,j})D'_{i,j}MH_{i,j+1} + \varepsilon \geq 0 \quad (16)$$

In terms of the relationship between buffer time and running time, with increasing negative buffer time, running time increases linearly (See Figure 7). Utilizing this linearity property, running time between stations is calculated.

4.7 Adjustment of train headway

In daily operations, if there is change in train headways, the control center operator adjusts the headways to prevent agglomerate of passenger congestion. If the train interval is longer than 1 minute 30 seconds and less than 2 minutes compared to the planned headway at the time of the departure, the preceding train is adjusted by a planned dwell time + 1 minute after the departure time. In the same way, the train interval is longer than 2 minutes and less than 2 minutes and 30 seconds, the adjustment time of preceding train is planned dwell time + 1 minute and 30 seconds.

In usual situations, the number of boarding passengers is calculated between arrival times. However, in the case of headway adjustment, the number of boarding passengers is calculated between arrival time of subsequent train and the time which subtract departure time of following train considered adjustment from the door closing confirmation time and the safety confirmation time.

![Figure 7: Relationship between Buffer time and Running time](image-url)
5 Assessment of simulation reproducibility

To confirm that the simulation reproducibility and its accuracy is maintained, we put into the departure time and congestion data at starting station which is the 5 days data based on the construction of the simulation, then simulate 100 times for each day. Residual error RMS (Root mean square) is adopted as the performance index.

Further, we simulated 100 times for 10 days at random excluding the 5 days. Those 10 days data are adopted from weekdays in 2015 of no small transportation troubles and no vacation periods. However, there’s no way to get some data on random days, we estimate them as follows.

The number of passengers of those random days at starting station is figured out based on proportion of the 5 days average to those of degree on random days.

The number of alighting passengers each station in random days is figure out based on equation (2) to (4). Ticket gate ingress and egress data replace the 5 days data with random days data, and rate of direction is adopted average rate of direction A on the 5 days. The number of boarding passengers each station in random days is figure out based on equation (9) to (11). Ticket gate ingress and egress data replace the 5 days data with random days data, and rate of direction is adopted average rate of direction A on the 5 days. In addition, wider door is set at random.

Figure 8 shows the results of the reproducibility test. The actual average of travel time is 17 minutes and 13 seconds and standard deviation is 1 minute and 22 seconds, and
simulated that time is 17 minutes and 16 seconds and standard deviation is 1 minute and 31 seconds. High accuracy is maintained compared to references. Also, in the case of the data selected at random, those of simulated travel time is confirmed high accuracy that error between travel time and standard deviation are few seconds.

5.1 Case study for improvement of train delay
Railway companies have taken measures to improve train delay and train congestion. There are two types of measures, one is improvement of train timetable, second is distribution of passenger congestion. The former measure aims at avoiding delay propagation to subsequent trains. Important point to avoid propagation is to expand buffer times. This is also conducted by daily operation at control center.

The latter measure aims at distributing congestion agglomeration of specific cars and doors. Station staff encourage passengers to use more empty cars or use earlier trains. In 2017 summer, Tokyo metropolitan government implemented “Jisa Biz” staggered commuting campaign and many companies addressed changes in work start time during the campaign term. In 2020, the Tokyo Olympic and Paralympic Games will be held. Especially, congestion of peak-hour adding spectators would over the limit of train transportation capacity in Tokyo. The government would like to build staggered commuting as routine by 2020. Furthermore, for legacy, staggered commuting would be conductive to smooth transports and flexible lifestyles.

Utilizing the proposed simulation, we estimate the effect of staggered commuting on Tozai line focusing on one day. Passengers demand on starting station during 7:30 to 8:29 reduce 10%, and the 10% passengers are allocated to each train running on time zone 6:30 to 7:29 based on each train congestion degree. And boarding passengers during 8:00 to 8:29 and 8:30 to 8:59 reduce 10%, and the 10% passengers are allocated to each train running on time zone 7:00 to 7:29 and 7:30 to 7:59 based on each train passenger congestion degree.
The number of alighting passengers is calculated as same way of boarding case. Furthermore, in the case of 20% reduce is calculated as same way (Figure 9).

Figure 10 shows the results. The actual average travel time is 16 minutes and 20 seconds, and passenger 10% moving case is 16 minutes and 14 seconds and that of 20% moving case is 16 minutes and 11 seconds. The average travel time is alleviated due to demand moving. Particularly, before peak hour, travel time increases by 17 seconds in the case of 10% moving case, and 28 seconds in the case of 20% moving case. On the other hand, on peak hour, the maximum improvement time is 24 seconds in the case of 10% moving case, and 41 seconds in the case of 20% moving case. The effects have decent improvement, but further demand moving deal is necessary for legacy.

From this result, it is confirmed that travel time before peak hour increase temporarily but travel time at peak hour improve well and average travel time is also shortened.

6 Conclusion

We have introduced an innovative method of train travel time simulation model utilizing daily ticket gate ingress and egress data and detailed on-site inspection results. Especially, focussing on each time model in dwell time is new characteristic of the simulation. Also, utilizing past traffic record data to model sudden small events during closing confirmation
time is reproduced detailed situation. We obtained high reproducibility and confirm the usefulness of the proposed method. Basing on this method, we can obtain the effect of train transportation on lines operated density.

In the case of the staggered commuting campaign, we confirmed the effect of travel time change due to moving passenger demand. In this case, we confirmed certain level of peak hour improvement. However, for flexible commuting, staggered activities should be promoted more.

In order to contribute to the improvement of passenger congestion and train delays, further work should consider the characteristics of different lines and different situation of passenger alighting and boarding situations and simulate more cases reflecting other demand change deal. station situation and actual operations more.

References


