Investigation and Estimation of Train Dwell Time for Timetable Planning

'Jyh-Cherng JONG, 'En-Fu CHANG
Sinotech Engineering Consultants, Inc., Taipei, Taiwan (R.O.C.)'

Abstract
The determination of train dwell time is an essential step in planning railway timetables. The magnitude of dwell time affects not only train operation efficiency but also punctuality. This paper proposes a novel survey technique to investigate and estimate dwell time. First, mobile devices, such as notebook PCs, mobile phones, and personal digital assistants (PDAs) are used in field survey to collect data. Statistical approaches that include confidence intervals and multiple linear regressions are then employed to estimate dwell time. This technique was applied to the Taiwan Railway Administration (TRA). Passenger flow time was found to be the most important factor affecting dwell time and is the focus of this paper. The relationship between passenger flow time and the number of passengers was statistically significant in the proposed linear regression models. Moreover, the adjusted R-squares of the regression models for different train types range between 0.86 and 0.95. The results demonstrated that the models have very good explanatory capabilities in estimating passenger flow time. As the proposed survey technique is relatively low-cost and easy to implement, it may help timetable planner to effectively determine adequate train dwell time without using advanced systems, such as automatic fare collection or automatic passenger counting systems.

1. Introduction
For efficient services, trains are scheduled to stop at stations with adequate dwell time for passenger boarding and alighting. Determining dwell time is crucial to the planning of timetables since insufficient dwell time would cause train delays, while excessive dwell time would lead to inefficient operation. In general, dwell time is defined as the time during which a train remains stopped at a railway station, primarily to allow passengers to board and alight. However, a train may stay at a station due to train dispatching rules, such as overtaking, meeting or insufficient headway [1]. In contrast to passenger flow time, dwell time for dispatching is easy to estimate since the required headway is determined by signaling system. Thus, the study focuses on the estimation of passenger flow time.

Previous studies has shown that train dwell time was affected by multiple factors, such as door opening/closing time, passenger flow (i.e. boarding and alighting) time, and signal headway. In most of these studies, the number of boarding and alighting passengers was considered to estimate dwell time [3, 6, 10]. Koffman’s [5] took into account the number of on-board passengers and fare collection methods (pre-paid or on-board payment). Lin and Wilson [6] also considered the number of on-board passengers (especially standees) to estimate dwell time. Other factors such as width/number of car doors, height/gap of platforms, and gender/age of passengers, etc were also discussed in previous studies. However, these factors had not been used with quantitative models. Note that not the entire dwell time can be utilized for boarding or alighting. Some systems, such as Calgary’s light rail transit and the TTC subway, have an enforced safety delay of a few seconds once the doors have closed before the train can move [10]. Consequently, a piece of dwell time will be lost.

According to the above discussions, passenger flow time is the largest component of dwell time and deserves an in-depth analysis. The development of a passenger flow time model relies on accurate field data. In past studies, counters and stopwatches were typically used to record the number of passengers and passenger flow time [8]. The total passenger flow time is usually estimated by calculating the time interval between the last passenger to board and the first passenger to alight. The average flow time for each passenger is then calculated by the total flow time divided by the number of passengers. However, it is impossible for passengers to be continuously passing through a car door while the train is stopped. Thus, counters and stopwatches in a traditional field survey are not effective in recording the flow time for each passenger.

Since field surveys are labor intensive, some studies have used transaction log data from automatic fare collection (AFC) systems to estimate the number of alighting/boarding passengers [2, 9]. However, except for on-board payment systems (e.g., LRT system), estimating passenger flow time
Challenge F: Even more trains even more on time

from AFC records is complex and less precise. This is because only two accurate time stamps of entering and leaving paid area are insufficient to guess which train a passenger may take in a complex railway system. Even if a model can estimate the exact train a passenger actually boards, it is still unknown which door the passenger chooses. The distribution of door selection is usually non-uniform. For example, the ratio of busiest door usage to average door usage is suggested to range between 1.2 and 1.5 in the “Rail Transit Capacity” report [10].

In order to accurately calculate the number of passengers for each door, automatic passenger counting (APC) systems have been applied to collect detailed passenger flow data [7, 9]. The sensor is generally placed on top of the doors in an APC system. Advanced sensors can detect the passage and direction of people, and then send these data to the Data-Logger. The APC collects not only how many people board or alight but also through which door and at what time a passenger does. Theoretically, this approach has potential to collect the most detailed data for passenger flow. However, the accuracy and precision of APC need to be further validated and evaluated [4].

Although AFC and APC systems could collect passenger data relatively easily, both AFC and APC require considerable capital investments and not all railway agencies are equipped with such advanced systems, especially for traditional regional railways. To present, neither approach is widely used by railroad operators. For this reason, this study proposes a novel survey technique to collect detailed passenger flow data without advanced AFC and APC systems. Surveyors use notebook PCs or personal digit assistants (PDAs) with customized software instead of counters and stopwatches. The software integrates conventional counters and timers to assist surveyors in collecting detailed data. A post statistical program is then performed to analyze passenger flow data. The proposed survey technique and the statistical methods can be used by railroad operators to plan for adequate train dwell time with quantitative evaluation methods.

2. Proposed Survey Technique

The proposed survey technique consists of two procedures. First is the collection of the detailed time stamps of important events, such as train stop/start, door open/close, and passenger boarding/alighting. Second is the analysis of the log files collected in the field survey. The two procedures will be introduced in the following two sections.

2.1 Investigation Method

To trace detailed states of trains and passengers, the study developed customized software to assist surveyors in collecting data. Figure 1 illustrates the snapshots of the software. The left one runs on PC/Windows compatible system, and the right one runs on Windows Mobile device, such as a notebook PC, mobile phone or PDA. The user interfaces are similar in both versions.

![Figure 1: Snapshots of investigation devices.](image)

In the field survey, each surveyor collects information for one train door. Before a train stops at the platform, surveyors must type in the train ID (the information can be obtained from the timetable or the
passenger information display system), car number, and the door designation. After confirming above information, surveyors must click the “Start Investigation” button to begin the survey process.

Events that occur during train dwell time include (1) train stop, (2) door open, (3) passenger alighting, (4) passenger boarding, (5) door close, and (6) train start, as shown in Figure 2. Each event corresponds to a button in the software. When the train stops, surveyors must immediately press the “Train Stop” button to record the time stamp of the event. The “Door Opened” button is then clicked when train door is opened. Surveyors then click corresponding buttons to record the time stamps of passenger boarding and alighting. Finally, surveyors click “Door Closed” and “Train Start” at the moments when the corresponding events happen. Once the field survey is finished, the software has recorded not only the number of passengers, but also the time stamps of the above-mentioned six events. Finally, the software will sort all events by time sequence and export the output in text mode for further analysis. Figure 3 is an example of an output file that records train ID, car number, door number, and the time stamps for each event.

2.2 Estimation Method

Dwell time is the time interval between train arrival and departure. To make the following analysis clear, we divide dwell time into three parts (see Figure 2):
- Type I Lost Time: The time interval between train stop and door opened.
- Passenger Flow Time: The time interval during which passengers alight from and board the train.
- Type II Lost Time: The time interval between door closed and train start.

During the first and third intervals of train dwell time, passengers are unable to board or alight. Type I lost time can be obtained by calculating the time difference between train stop and door opened. Similarly, Type II lost time can be obtained by computing the difference between train start and door closed. The amount of lost time may vary among different trains and doors. This study employed descriptive statistics and confidence intervals to estimate lost time.

The second interval (i.e., passenger flow time) is the time available for passengers to alight from or board a train. Since passengers queue up inside the train for alighting and at the platform for boarding,
each door can be treated as a queuing system. Under queuing conditions, the time interval between two consecutive passengers can be considered as the flow time (door service time) for the preceding passenger. Flow time for each passenger can be calculated from the output log file. For example, the time interval between box C and box D in Figure 4 is regarded as the flow time for alighting, and that between box E and box F is considered as the flow time for boarding. If the time interval between two consecutive events of alighting and/or boarding is more than five seconds (a predetermined parameter), this pair of samples is ignored since the following passenger should be under free flow condition without staying in the queue in this situation. After flow times for each passenger have been collected, descriptive statistics and regression models are applied for estimation.

Figure 4: Pre-processing of investigation log

3. Application

This study applied the proposed technique and analysis methods to assist Taiwan Railway Administration (TRA), a government-owned railway operator, in analyzing train dwell time and passenger flow time. Currently, TRA provides intercity, regional and commuter train services. The doors of commuter trains are wider than those of intercity trains. In contrast to commuter trains, intercity trains have more seats and less space for standees than commuter trains. Photographic comparisons of intercity trains and commuter trains are presented in Table 1.

3.1 Background of Field Survey

The survey was conducted at Taipei Main Station (the busiest TRA station) in the morning peak hours (AM 7:00 – AM 9:00) for five weekdays. Since a commuter train consisted of a maximum of 8 cars, each of which has three doors on each side, the maximum number of doors being surveyed was 24. On the other hand, an intercity train could haul up to 12 cars, each of which was equipped with two doors on each side. Thus, the maximum number of doors being investigated was also 24. For this reason, 24 surveyors were allocated at each platform. Sometimes the incoming train might be shorter. In such a case, only a portion of surveyors was on duty. To ensure that filed survey was smooth and successful, a plan for surveyor rotation was created in advance based on the timetable. When a train arrived, each on-duty surveyor was responsible for one door. Since the trains did not always stop at the same location, surveyors had to move to an appropriate location for easy investigation. In the field survey, there were totally 106 trains, 1927 doors, and over 60,000 passengers being investigated. A detailed description of the samples is shown in Table 2.
Challenge F: Even more trains even more on time

Table 1: Comparison between intercity train and commuter train

<table>
<thead>
<tr>
<th></th>
<th>Intercity Train</th>
<th>Commuter Train</th>
</tr>
</thead>
<tbody>
<tr>
<td>Door Style</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layout of Car</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary of samples

<table>
<thead>
<tr>
<th>Category</th>
<th>Train Model</th>
<th>Sample Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Trains</td>
<td>Number of Boarding Passengers</td>
</tr>
<tr>
<td>Intercity Train</td>
<td>Push-Pull</td>
<td>7,523</td>
</tr>
<tr>
<td></td>
<td>DMU2800</td>
<td>1,639</td>
</tr>
<tr>
<td></td>
<td>EMU1200</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>FP/SP(Original)</td>
<td>2,941</td>
</tr>
<tr>
<td></td>
<td>FP/SP(Improved)</td>
<td>1,445</td>
</tr>
<tr>
<td>Commuter Train</td>
<td>EMU500</td>
<td>24,740</td>
</tr>
<tr>
<td></td>
<td>EMU400</td>
<td>8,028</td>
</tr>
<tr>
<td>Total</td>
<td>106</td>
<td>46,346</td>
</tr>
</tbody>
</table>

3.2 Estimation of Lost Time

The study applied interval estimation to analyze lost time. The statistics of Type I lost time are listed in Table 3. To obtain sufficient dwell time, the upper-bound of the interval at the 95% confidence level was used to estimate type I lost time. Following the same statistical approach, we found that Type II lost time was significantly larger than Type I lost time. One reason for this might be that some exogenous factors deter trains from departure. For example, the signaling system has not yet displayed a clear aspect, or the dispatcher has not yet allowed the train to depart. Since doors open and close via the same mechanism, Type II lost time in this study is assumed to equal Type I lost time. Thus, the total lost time in this study is estimated to be twice Type I lost time.

3.3 Estimation of Flow Time

The descriptive statistics show that the average flow time is about 0.97 seconds per alighting passenger and 1.47 seconds per boarding passenger. In terms of different train types, the boarding times for intercity trains and commuter trains are 1.69 seconds and 1.13 seconds, respectively, while alighting times are 1.37 seconds and 0.8 seconds, respectively. In general, boarding time is longer than alighting time, and flow time for intercity trains is also longer than that for commuter trains. The reason boarding time is longer than the alighting time is that train floors are higher than platforms at
Challenge F: Even more trains even more on time

TRA stations. Therefore, it is easier for passengers to get off a train than to get on to it. This finding is consistent with the result of the Rail Transit Capacity [10]. On the other hand, flow time for intercity trains is longer than that for commuter trains because commuter train doors are wider than those of intercity trains. As a result, it is easier for passengers to pass through commuter train doors than those of intercity trains.

If the class interval is 0.1 seconds, the frequency distributions of flow time per passenger are as illustrated in Figure 4. From Figure 4, the value that occurs most frequently (the mode) and the spread of the data are readily obtained. Like the statistical mean, the mode is a way of capturing important information about a random variable. For instances, the mode of alighting time per passenger for intercity trains is about 1.25 seconds and the distribution is almost symmetrical, while that for commuter trains is around 0.55 seconds and the distribution is skewed. Figure 4 also indicates the variation of alighting time is less than the variation of boarding time.

Table 3: Statistics of Type I lost time

<table>
<thead>
<tr>
<th>Statistics</th>
<th>EMU500</th>
<th>EMU400</th>
<th>Push-Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.110</td>
<td>3.110</td>
<td>3.038</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.058</td>
<td>0.384</td>
<td>0.105</td>
</tr>
<tr>
<td>Median</td>
<td>1.720</td>
<td>1.620</td>
<td>2.829</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.665</td>
<td>5.310</td>
<td>1.920</td>
</tr>
<tr>
<td>Variance</td>
<td>2.773</td>
<td>28.193</td>
<td>3.686</td>
</tr>
<tr>
<td>Sample Size</td>
<td>829</td>
<td>191</td>
<td>337</td>
</tr>
<tr>
<td>Interval at 95% CI</td>
<td>0.228</td>
<td>1.516</td>
<td>0.412</td>
</tr>
<tr>
<td>lower-bound of CI</td>
<td>1.996</td>
<td>2.352</td>
<td>2.832</td>
</tr>
<tr>
<td>upper-bound of CI</td>
<td>2.224</td>
<td>3.868</td>
<td>3.244</td>
</tr>
</tbody>
</table>

Figure 5: Distribution of flow time per passenger.

In addition to descriptive statistics and frequency distributions, this study also employed linear regression models to estimate passenger flow time by using the number of passengers. The basic linear regression function used in this study is given in equation (1), where $FT$ is the total passenger flow time; $P_a$ and $P_b$ are the numbers of alighting and boarding passengers, respectively.

$$FT = a + b \times P_a + c \times P_b$$  \hspace{1cm} (1)

After trying different regression functions, the best regression model found is shown in equation (2). The coefficient of intercept was removed from equation (1) because the sign of the coefficient is not reasonable. The R-square of equation (2) is about 0.904 and the adjusted R-Square is 0.82.

$$FT = 0.87678 \times P_a + 1.968336 \times P_b$$

(95.5) \hspace{1cm} (61.70)  \hspace{1cm} (2)

Since the design of intercity trains is different from commuter trains, the study also classified samples into two groups to re-run the regression model with a dummy variable. The dummy variable was treated as the intercept of the model, as shown in equation (3).
$FT = 0.855251 \times P_a + 1.457526 \times P_b + 12.29818 \times D$, \(D\) = \begin{cases} 1 & \text{: Intercity Train} \\ 0 & \text{: Commuter Train} \end{cases}$ (3)

Although equation (3) only increases the R-square value from 0.904 to 0.94 (or adjusted R-square from 0.82 to 0.88), the coefficient of partial determination of the dummy variable is about 0.303, which means that the introduction of the dummy variable reduces the variation of flow time by 30.3% given that $P_a$ and $P_b$ are already in the model. In addition, the t-value of the dummy variable was about 28.89, indicating that the dummy variable is statistically significant.

Equation (4) allocates dummy variables in the slopes of independent variables. The result is similar to that for equation (3). Both equation (3) and equation (4) indicate that the dummy variable of train type is a significant factor for estimating passenger flow time. The adjusted R-square of equation (4) is 0.95.

$FT = 1.48 + 0.76 \times P_a + 1.01 \times P_b + 0.60 \times D \times P_a + 0.58 \times D \times P_b$, \(D\) = \begin{cases} 1 & \text{: Express Train} \\ 0 & \text{: Commuter Train} \end{cases}$ (4)

Since the dummy variables suggest that passenger flow time is significant different for intercity and commuter trains, another approach to perform regression analysis is directly separating samples for intercity and commuter trains. The results are shown in equations (5) and (6) with adjusted R-square equal to 0.91 and 0.95, respectively. Therefore, if one regression equation is preferred, then equation (4) is suggested. If two regression models are favored, then equations (5) and (6) are recommended.

$FT = 2.168903 + 0.739998 \times P_a + 0.971878 \times P_b$, for samples from commuter trains \begin{cases} (9.23) \ (101.25) \ (26.00) \end{cases}$ (5)

$FT = 1.377865 \times P_a + 1.642144 \times P_b$, for samples from intercity trains \begin{cases} (110.03) \ (62.33) \end{cases}$ (6)

The same analysis was also applied using other factors, such as the elevation gap between platform and train floor and the door width. Regression analysis shows that wider train doors and smaller elevation gaps can significantly reduce passenger flow time.

### 4. Conclusions and Recommendations

This paper develops an improved survey technique and integrated software that can be used on portable devices such as notebook PCs, mobile phones or PDAs to investigate and analyze passenger flow time. The technique was applied to intercity and commuter trains operated by the TRA. After field surveys and post analysis were performed, it was found that boarding time is longer than alighting time, and that passenger flow time for intercity trains is also greater than that for commuter trains. In the hypothesis test, the number of alighting passengers and boarding passengers, the elevation gap between the platform and train floor, and the width of train doors, are all statistically significant. The regression analysis shows that the easiest and most efficient way to estimate passenger flow time is through the number of alighting and boarding passengers. If one regression model is preferred, train type can be treated as a dummy variable in the regression model. Another approach is separating the samples by train type to build regression models. Both have very good explanatory capabilities.

The proposed survey technique and analysis method significantly reduce the effort spent in collecting detailed information on train dwell time, especially for railway agencies without AFC and APC systems. Based on the data, railway agencies would be able to estimate and determine adequate dwell time when planning timetables. They could avoid serious delay caused by insufficient dwell time or inefficient operation resulted from excessive dwell time. Under limited resources, the samples collected in the field survey presented in this study may be insufficient. If the scope of the field survey were expanded, better analysis results would be obtained.
5 Acknowledgements

This paper was produced under MOTC-IOT-94-PBB005 project by the Institute of Transportation (IOT), Ministry of Transportation and Communications. We would like to extend our gratitude to IOT for its financial support.

6 References