

The simulation of passengers' time-space characteristics using ticket sales records with insufficient data

J.-C. Jong & E.-F. Chang

*Civil & Hydraulic Engineering Research Center,
Sinotech Engineering Consultants, Inc, Taiwan*

Abstract

It is a very common approach for any business to analyze their historical sales records to adjust operation strategy. Recently, Taiwan Railway Administration, a government-owned railway operator, faces serious competitions from other transportation systems. It becomes very urgent for the operator to modify its timetables to meet demand patterns for increasing revenue. The key issue is how to estimate passengers' time-space characteristics. For railroads with advanced automatic fare collection systems and simple service patterns, the estimation of passenger flow may not be difficult since detailed travel information is available. However, for systems with mixed traffic and insufficient ticket sales records, it requires a scientific method to deduce actual travel patterns from limited information. This study tried to establish such a model to reconstruct the time-space distribution of passenger flow. The model has been applied to Taiwan Railway Administration to estimate passenger flow. The result is very useful for decision makers to assess the utilization of train capacities and to adjust service plans, such as adding/deleting train services, changing stopping patterns, or modifying service termini. The proposed model can be applied to other railroads with mixed traffic operations and insufficient ticket sales records.

Keywords: passenger flow estimation, ticket sales records.

1 Introduction

Taiwan Railway Administration (TRA) is the oldest railway operator that provides intercity, regional and commuter train services in Taiwan. In early years when TRA was the only inter-city transportation provider, timetable preparations



were production-oriented. TRA provided transportation services mainly according to resources availability (e.g., trains and crews). Passenger demand was only a small issue in TRA's consideration list. During last few decades, several transportation systems such as freeways and high-speed rail commenced their services successively. These new transportation systems were strong competitors for TRA. Recently, TRA attempted to employ marketing-oriented approach for preparing more attractive timetables to increase revenue. To do this, TRA should accurately estimate the demand characteristics of its customers, including boarding time, origin station, and destination station.

In general, three different approaches are often used to estimate passenger demand, including marketing research, transportation planning model, and historical data analysis [3]. Among them, historical data analysis has great potential for accurately characterizing passengers' travel behaviour in a railway system with a relatively low cost. The cost for collecting data decreases as the computerization level of ticket sales systems increases. For a modern computerized ticketing system, the sales record log is automatically generated every day. Consequently, many researches prefer this approach for estimating passenger demand. However the ticketing systems in TRA are not as advanced as the Automatic Fare Collection (AFC) systems in modern urban transit systems. Currently, TRA accepts four different kinds of tickets. Only tickets with designated train numbers have detailed travel records. The others have limited information, such as lack of train number or exact time to enter/leave the railway system.

This study aims at developing a simulation model and a computer program to reconstruct passengers' time-space characteristics. The model estimates passenger walking time and assigns each passenger to an appropriate train based on limited information stored in the sales records. When the simulation is completed, the number of passengers on each train and each railway section can be calculated. Capacity insufficiency or service oversupply can also be identified. The computer program provides several 2D and 3D charts to display the time-space transitions of passenger flow. The resulting information is very useful for railway operators to adjust their train service plans and timetables to increase revenue.

2 Literature review

Analysis of historical ticketing records can be used to investigate passengers' travel behaviour in a railway system. Previous studies usually apply this method to urban railway systems where detailed system logs of ticket gates are available. For example, Myojo [4, 5] proposed a model to estimate passenger flow in a large and complicated urban railway network using origin-destination (OD) matrix data from ticket gates. The passengers' route choices (including trains and train lines) were determined by a logit model (A similar approach can also be found in Hirai and Tomii [2]). To verify the proposed method, the study compared the estimated results with the number of passengers reported by train conductors who used a visual count. The correlation coefficients between them

are about 0.95~0.83. Nagasaki et al. [6] also proposed a similar approach to estimate passenger flow but formulated passengers' route choices as a shortest path problem. Nagasaki's model considered not only journey time and transfer barrier factors but also congestion factors. In addition, his model also evaluated train schedules from the viewpoint of passengers. Note that both Myojo's and Nagasaki's studies used aggregated OD matrix data without exact arrival/departure time for each passenger. Therefore, both models cannot reproduce the detailed time-space characteristics of passenger flow.

In contrast to the macroscopic models proposed by Myojo [4, 5] and Nagasaki et al. [6], Barry et al. [1] proposed a microscopic methodology to estimate OD tables from the AFC records of MetroCard in New York City. The results were used for other purpose such as traffic assignment in transportation planning model. Zhao and Rahbee [7] also used AFC data to analyze the behaviour of each passenger and integrated the records with Automatic Data Collection system and Automatic Vehicle Location system to estimate rail passenger OD matrix. Both studies did not estimate passenger flow on rail links and trains.

Since TRA is not equipped with advanced AFC systems, the information stored in the ticketing systems is insufficient. In addition, TRA provides train services with different classes, each of which has different stopping patterns and operation speeds. Even in the same class, the stopping patterns and service termini for different trains are not identical. Consequently, the models found in the literature cannot be directly applied to TRA. To overcome the problem, this paper introduces a microscopic simulation model to estimate passenger flow with insufficient data and mixed traffic. When the simulation is completed, detailed information about the number of onboard passengers and the flow on each section during different time intervals can be estimated. The following section will introduce the features of TRA ticketing records. The detailed model is discussed in section 4.

3 Features of TRA ticketing records

In early years, TRA used a manual approach to deal with ticket transactions. All activities including ticket sales, checking, and inspection were performed by TRA staff. At that time there were no electronic records. In the past few years, TRA installed several ticketing systems, including booking systems, ticket vending machines, ticket gate systems, etc. However, these systems are designed mainly for accounting purpose. The information stored is insufficient to analyze passenger behaviour. Since the study aims to precisely estimate passenger flow, it is important to understand the status and limitations of the ticketing records in TRA before developing the model.

There are many types of tickets in TRA. They can be divided into two major categories. The first ticket type has a designated train number. Passengers with these tickets are only allowed to board a specific train as the train number marked on the tickets. For convenience, these tickets are referred to as designated tickets. Note that designated tickets do not imply that seats are

reserved. Passengers may purchase designated tickets without reserved seats at lower prices. The second ticket type has no designated train number. Passengers with these tickets can board any trains of the same train class as shown on the tickets. These tickets are called non-designated tickets.

In general, designated tickets are only used for express trains that provide intercity services. Designated tickets with reserved seats can be purchased two weeks before the train service and seats can be booked at the same time. Non-designated tickets are used for any trains, such as commuter trains, local trains, and even express trains. Non-designated tickets have several variations, including one-way tickets and prepaid tickets. Both smart cards (RFID based) and season tickets (magnetic based) are prepaid tickets. Figure 1 shows the classifications of TRA tickets. The information stored in each ticket will be introduced in the following two sections.

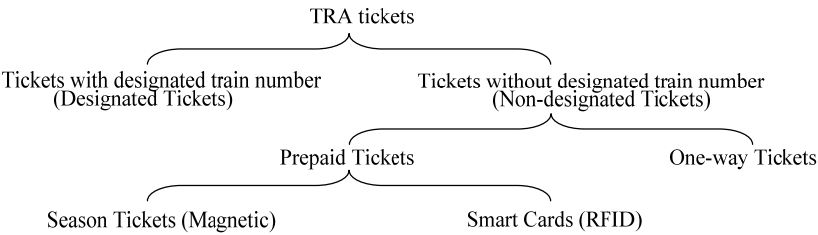


Figure 1: Classifications of TRA tickets.

3.1 Electronic records of designated tickets

Electronic records of designated tickets exist in ticket sales and gate systems. Trip information stored in ticket sales system includes the date of train service, train number, origin station, destination station, etc. Gate system records the time when the passenger passes through ticket gates. TRA provides several convenient ways for purchasing designated tickets. The process for purchasing tickets can be divided into three stages: booking, payment, and ticket pick-up. Transactions at each stage are all recorded. Figure 2 shows the flow chart for purchasing designated tickets. The simplest way to purchase a designated ticket is via ticketing offices or vending machines at stations since the three stages can be finished at the same time. If passengers book tickets through internet or phone voice, they have several options to pay for the tickets. The method of payment will decide how to pick up their tickets. The transaction records in internet/phone voice booking system, ticket vending machine system, ticketing offices and post offices were all collected for the study.

The main feature of designated tickets is that the train number is determined once the ticket is booked. Therefore, it is easy to find out when and where the passenger gets on and gets off the train. Note that the records of cancelled tickets must be taken into account because tickets may be cancelled by passengers for any reasons at any time before the train departs. If the model ignores the logs of cancelled tickets, passenger flow will be overestimated.

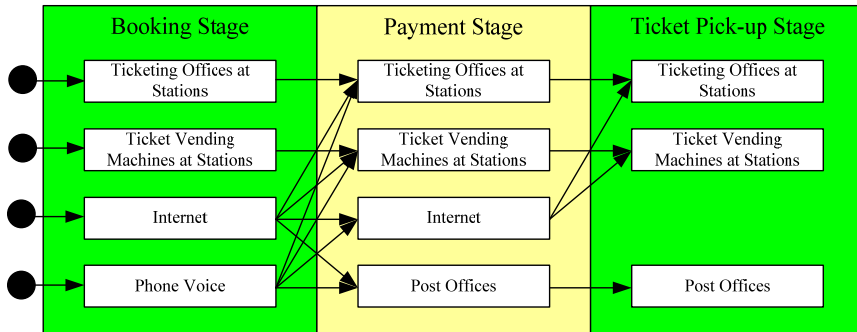


Figure 2: Flow chart for purchasing designated tickets.

3.2 Electronic records of non-designated tickets

A passenger with a non-designated ticket is allowed to board any train of the same class as shown on his/her ticket. The trip information stored in the sales records only consists of train class, origin station, and destination station. Due to lack of train number, it is difficult to judge which train the passenger actually boards. This study attempted to estimate the train number by other information. The information depends on the type of the non-designated ticket and is illustrated below:

- (1) **One-way Ticket:** One-way tickets can be purchased via ticketing offices or ticket vending machines at stations. A passenger with a one-way ticket is allowed to board any train of the designated class once on the transaction day. Two records may be useful to guess the train that the passenger should take. The first one is the ticket transaction time and the other one is the time stamp record when the passenger passes through a magnetic ticket gate. Unfortunately, not all stations in TRA are equipped with ticket gates. In such circumstances, passengers must show their tickets to station crews in order to enter the paid area. Even at stations where ticket gates are installed, passengers are not required to go through them. Therefore, transaction time is the only reliable reference to estimate the train that a one-way ticket passenger should board. Note that the purchase log and the cancel log may exit in different ticket sales systems. For example, a passenger may purchase a non-designated ticket via a vending machine and then cancel the ticket through a ticketing office. Accordingly, the serial number of the non-designated ticket is the only key to trace the two databases.
- (2) **Season Ticket:** Season ticket is a kind of prepaid ticket. A passenger with a season ticket is admitted to board any trains fifty times within two months for a specific origin and destination pair. The passenger is required to enter/leave paid areas through magnetic ticket gates. Thus, the time stamps and the station names are the keys for estimating passenger flow.
- (3) **Smart Card:** Smart Card is also a kind of prepaid ticket. It is based on RFID technology and has its own gates. The characteristics of Smart Cards are

similar to season tickets. For some reasons, the electronic records in Smart Card system have no time stamps when passengers enter stations.

3.3 Summary

Since the ticketing systems in TRA are designed mainly for accounting purpose, the electronic records are insufficient for estimating passenger flow. This study surveyed other electronic records in ticket gate system to supplement the information. Table 1 summarizes the important data used in this study for estimating passenger flow. Records for designated tickets and one-way tickets are from ticket sales system. “Quantity” is the amount of tickets in one purchase. If the value is negative, it means that those tickets have been cancelled. Records for season tickets and smart cards are collected from the ticket gate system. Thus, there is no “quantity” or “ticket cancellation” information.

In general, train number provides precise information about the train that a designated ticket passenger intends to board. Thus, designated tickets have sufficient information to estimate passenger flow under the assumption that all passengers follow TRA’s regulations. Secondly, we can deduce what train a season ticket holder may take based on actual train schedules and the time stamps of passing through ticket gates. The accuracy rate of determining the boarding train for a smart card record is lower than that for a season ticket record since the entering time for the smart card holder is not recorded. Finally, the estimation of the boarding train for a one-way ticket record is the least precise since only transaction time information can be used to judge which train the ticket holder may take.

Table 1: Useful information for different ticket types.

	Designated Ticket	One-way Ticket	Season Ticket	Smart Card
Item	Origin station	Origin station	Origin station	Origin station
	Destination station	Destination station	Destination station	Destination station
	Train number	Transaction time	Entering time ²	Leaving time
	Train departure time	Train class	Leaving time ³	
	Quantity	Quantity		
		Serial number ¹		

- 1. “Serial number” is used to trace the records of cancelled tickets.
- 2. “Entering time” is the time when a passenger enters the paid area through a ticket gate at his/her departure station.
- 3. “Leaving time” is the time when a passenger leaves the paid area through a ticket gate at his/her arrival station.

4 Simulation model

This study developed a simulation model to estimate passenger flow characteristics based on the limited information summarized in Table 1. The assumption and the framework of the proposed model are explained in this section. Some important issues for the simulation process are also addressed.

4.1 Assumption

The assumptions for the proposed simulation are summarized below:

- (1) The rail network in TRA is not complicated. The great majority of trips have only one route between origin and destination stations. The model assumes that passengers always choose a train that stops at both their origin and destination stations without any transfers.
- (2) A passenger who boards a train without any ticket is prohibited. Such illegal behaviors should not happen frequently. These trips can not be counted since there are no electronic records for illegal passengers.
- (3) It is also an illegal behaviour that a passenger purchases a train ticket with lower class and then boards a higher class train. The model assumes that passengers always board trains according to their tickets.
- (4) Passengers with one-way tickets will arrive at platform and prepare to board a train within a certain time window after purchasing the tickets.
- (5) Passengers with season tickets will arrive at platform and prepare to board a train within a certain time frame after passing through magnetic ticket gates.
- (6) Passengers with smart cards will take trains that arrives their destination stations within a certain time window before they leave the stations.

4.2 Framework

Figure 3 shows the framework of the simulation model. The proposed model consists of two major components: Train Traffic Simulator (TTS) and Passenger Flow Simulator (PFS). The first one simulates the movement of trains. The second one simulates the flow of passengers.

To make the model widely applicable, the TTS accepts two different types of inputs: the planned timetable and the actual train schedule. The planned timetable is easier to collect than the actual train schedule, but the assignment of passengers to trains for the former is less precise than that for the latter. If train punctuality is close to 100%, the simulation results will be similar. The PFS extracts useful information from various databases and combines them into four tables. The trips of cancelled tickets must be deducted from normal trips to avoid overestimation.

Passengers must board a train to move from their origin stations to destination stations. The process to assign each passenger to an appropriate train is the core of the simulation. Passengers with designated tickets are assigned to the designated train. The assignment of passengers with non-designated tickets is more complicated. The study defined “walking time” to represent the time for a passenger to walk to platform after purchasing a ticket or passing through the entry gate. The model employs a uniform distribution to generate walking time for each passenger. Actual value will be determined by parameters a and b in Equation (1).

$$W_i = U(a, b) \quad (1)$$

where W_i = the walking time of passenger i

a = minimum walking time to platform

b = maximum walking time to platform



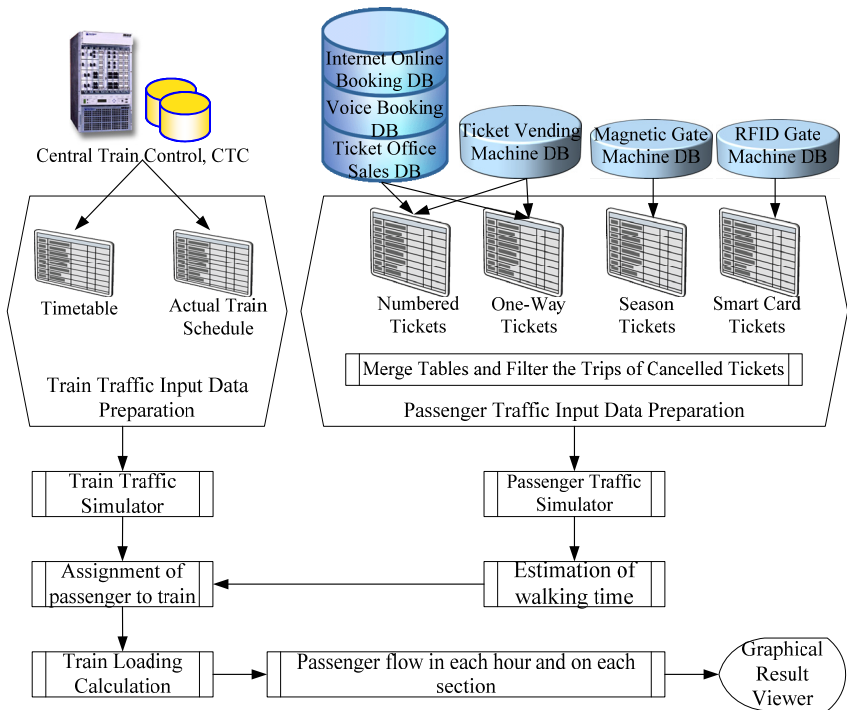


Figure 3: Framework of the simulation model.

For one-way tickets and season tickets, the example of assigning a passenger to an appropriate train is shown in Figure 4. Assume that a passenger purchases a ticket from station D to station B at T_1 . The model then generates walking time and adds it to T_1 . Let T_2 be the time when the passenger arrives at platform. The model will search forward for the first train of the recorded train class that stops at both stations D and B, and then assigns the passenger to the train. For example, the departure time of train 1 is earlier than the arrival time of the passenger at the platform of station D. It is impossible for the passenger to board train 1. Train 2, train 3 and train 4 do not stop at both stations B and D. Thus, train 5 is the most likely train that the passenger may board in this example.

The assignment of passengers to trains for smart card tickets is similar to that for one-way tickets and season tickets. However, the entering time of smart cards is not recorded. Thus, the model will search backward for an appropriate train from the time that a smart card holder leaves his/her destination station. The concept is shown in Figure 5. In this example, train 1 is the most likely train that the passenger may take.

When all passengers are assigned to trains, the number of passengers on each train and each rail section can be calculated. In addition, every train has detailed time and space information. Passenger flow at any time and any space can be estimated.



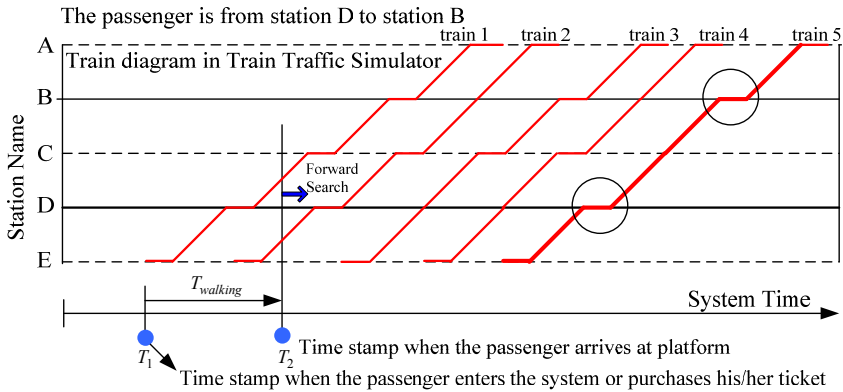


Figure 4: Train assignment for one-way tickets.

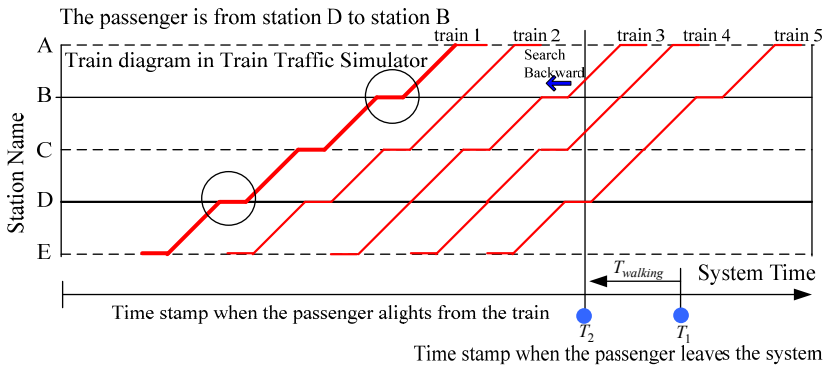


Figure 5: Train assignment for smart card.

5 Case study

This study developed a computer program and used real data from TRA to demonstrate the proposed model for estimating passengers' time-space characteristics. The input data and simulation result are explained as follows:

5.1 Input data

Since the operation control center in TRA did not have output module at the time the study was carried out, the actual train schedules were not available. Instead, the study employed the planned timetables as the input data. The sales records and timetables on January 4 and 5, 2009 were used for the case study. There were about 220 stations, 800 trains, and 400,000 records per day in TRA.

5.2 Simulation result

The program provides three different kinds of 2D and 3D charts to illustrate the simulation results. Basically, there are two types of passenger flow estimated by the model, i.e., node flow and link flow. The analysis of node flow focuses on the number of passengers that enter and leave stations, while the analysis of link flow focuses on passenger flow through each rail section.

Figure 6 displays the estimated hourly passenger inflow and outflow at Taipei station (the busiest station in TRA). This information is quite useful for planners to figure out the distribution of passenger flow and to identify the peak hour for Taipei station.

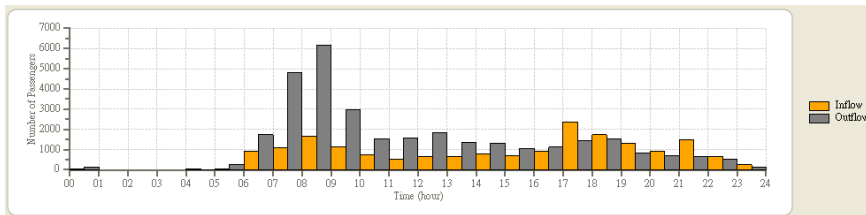


Figure 6: Hourly passenger inflow and outflow at Taipei station.

The number of onboard passengers is a key to estimate section flow. An example of passenger volume on a train along its journey is depicted in Figure 7. With this figure, the transition of passenger volume along its journey can be recognized. Figure 7 shows not only the total passenger volume, but also the components of passengers. According to the definitions in TRA, a trip shorter than 50 km is classified as a short-distance trip. If the length of a trip is between 50 km and 200 km, it is classified as a middle-distance trip. A trip whose length is more than 200 km is defined as a long-distance trip. The figure can be used to locate the Maximum Load Section (MLS) and to find out the trip combinations of different lengths on the MLS. The information may help railway operator plan and modify their train services to satisfy passenger demand. For example, providing shorter distance trains if the passenger volume near the two ends of the rail line is too low.

Figure 7 only displays the passenger volume on a specific train. To explore the variations in section flow, the study aggregated onboard passenger flows for all trains, section by section and hour by hour. In other words, onboard passenger volumes of the same section in the same hour were added together. The results are shown in Figure 8. The *x*-axis, *y*-axis, and *z*-axis represent space, time, and hourly passenger volume, respectively. This 3D chart clearly illustrates passenger flow on each section in each hour. Figure 8 also shows significant difference of passenger demands between weekday and weekend. This 3D chart may assist planner to realize passengers' time-space characteristics.

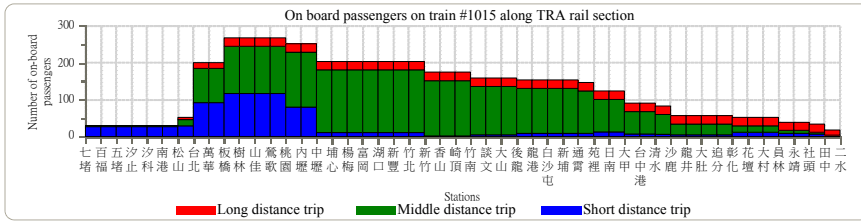


Figure 7: Onboard passenger volume of a specific train along its journey.

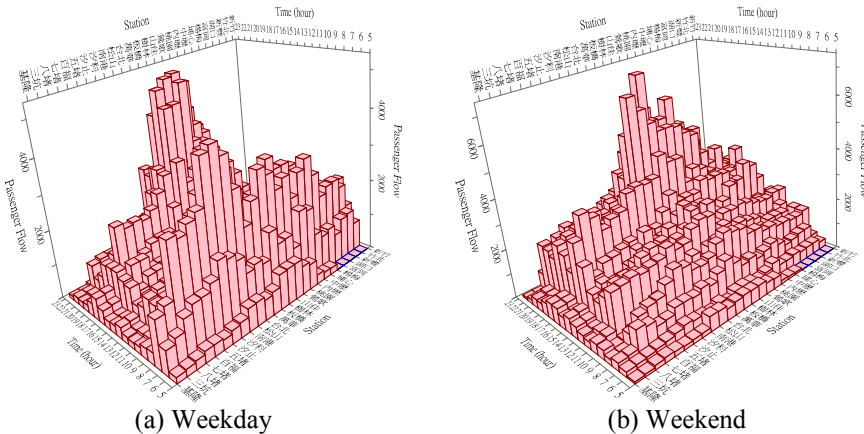


Figure 8: 3D time-space distribution of passenger flow.

6 Conclusion and suggestion

In early years, TRA assembled all station masters to hold a meeting for discussing how to modify timetables once a year. The decisions were based on their experiences without quantitative evaluation methods. Moreover, station masters only understood passenger flow at their stations. Passenger flow between adjacent stations was not supervised by station masters. The primary objective of the study is developing a simulation tool to estimate passenger flow in a rail network with insufficient ticket sales data. Through the proposed model, railway operators can calculate the hourly passenger flow at each station, the onboard passenger volume for each train, the passenger flow on each section, and the 3D time-space distribution of passenger flow. The simulation results are very useful for decision makers to adjust their train services to increase revenue. The proposed methodology can be applied to other similar railway systems with mixed traffic and different types of tickets.

Analysis of historical sales records has a restriction that it can not consider potential demands because the analysis result may be influenced by historical timetables. For example, if the passenger inflow at a station is low, the reason may be no demands or no suitable train services. Therefore, the analysis of

historical sales records is just a reference only. To investigate potential demands, it is suggested that railway operator collect and analyze the log of booking failure. The reason of booking failure may be no available seats or even no train services at the preferable time for the OD pair of a potential customer. The information will be a useful reference to decide whether to increase train capacities or to add train services for increasing revenue.

References

- [1] Barry, J. J., Newhouser, R., Rahbee, A., and Sayeda, S., "Origin and Destination Estimation in New York City Using Automated Fare System Data", *Proceedings of the 2001 TRB Planning Applications Conference*, Corpus Christi, Texas, 2001.
- [2] Hirai, C. and Tomii, N., "An Estimation Method of the Number of On-board Passengers Applicable to Evaluation of Traffic Rescheduling Plans", *Quarterly Report of RTRI*, Vol. 42, No. 4 pp.195-200, 2001.
- [3] Jong, J. C. and Suen, C. S., "A Train Service Planning Model with Dynamic Demand for Intercity Railway Systems", *Journal of the Eastern Asia Society of Transportation Studies*, Vol. 7, pp. 1598-1613, 2007.
- [4] Myojo, S., "Daily Estimation of Passenger Flow in Large and Complicated Urban Railway Network" *Proceedings of the 7th World Congress on Railway Research*, 2006a.
- [5] Myojo, S., "Method to Estimate Passenger Flow Using Stored Ticket Gate Data", *Quarterly Report of RTRI*, Vol. 47, No. 4 pp.178-181, 2006b.
- [6] Nagasaki, Y., Asuka, M., Koyama, K., "A Fast Method for Estimating Railway Passenger Flow", *Computers in Railways X*, pp. 179-187, 2006.
- [7] Zhao, J. and Rahbee, A., "Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems", *Computer-Aided Civil and Infrastructure Engineering*, p376-387, 2007.

