

Study on Improvement of Estimation Accuracy of a Shinkansen Passenger Simulator



Masumi Nakamura*



Kazutaka Ito*



Sei Sakairi*

Identifying the flow of passengers precisely is very important in transport improvement measures. The Shinkansen automatic ticket gate system records various ticket data. That data is composed of many types: origin and destination, when and where each passenger gets on and off, types of tickets, and so on. The authors studied an algorithm to estimate passenger volume and which train and seat type they use by combining ticket data with actual train operating time, and we developed a prototype simulator implementing this algorithm. In addition, we checked accuracy of estimation by comparing with numbers of passengers measured through a survey and worked to further improve the algorithm.

•Keywords: Automatic ticket gate data, Passenger volume, OD data

1 Introduction

Identifying with precision the flow of passengers in train operation is very important for planning and proposing measures to improve transport. The Shinkansen automatic ticket gate system records and retains in ticket gate usage log data (ticket gate data) information on the type of ticket (ticket type data) inserted by individual passengers. A wide variety of items is included in that data: origin and destination station, time and ticket, limited express ticket information, ticket type information, and more.

The authors studied an algorithm to estimate Shinkansen passenger volume by train and seat type by combining ticket type data with train operating time. We also developed a prototype Shinkansen passenger volume simulator implementing this algorithm. Moreover, we verified accuracy of the estimation by comparing estimations with actual passenger volume measured through an onboard survey, and we are working to gradually improve estimation accuracy.

Reports of estimated numbers of passengers riding trains (passenger volume reports) are transport volume indexes of a “cross section” defining the maximum number of people in a predefined survey section. With the method of estimation using this algorithm, passenger origin and destination (OD) data can be obtained as well, and more detailed estimation becomes possible by combining passenger volume and OD data. This paper covers improvement in estimation accuracy and verification of that accuracy in studies implemented from fiscal 2014 and 2015 based on previous studies implemented from fiscal 2002 to 2004.

2 Overview of Studies

2.1 Improvement of Algorithm (Fiscal 2014)

2.1.1 Issues and Response to Those

Timetable data was used as operation time to be combined with ticket type data in previous studies, so correction was needed according to actual operation results such as train cancellations.

However, by using actual time data delivered externally for mobile Suica (fare cards) where service started later, cancellations, delays, and the like were reflected and corrections we no longer needed. Additionally, most group passengers do not pass through automatic ticket gates, causing large downturns in passenger volume estimations. Improvements in estimation accuracy can be expected, however, by obtaining group ticket data from the Shinkansen Multi Access seat Reservation System (MARS) and inputting that. Table 1 shows input data required for simulation operations.

Table 1 Input Data Required for Operations

No.	Data type
1	Ticket gate data
2	Actual time data
3	Group ticket data

2.1.2 Processing Flow and Algorithm

Fig. 1 shows the processing flow for estimating passenger volume and OD data and describes the basic algorithm.

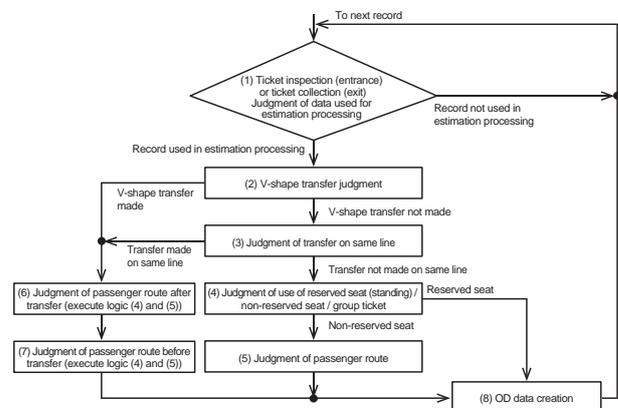


Fig. 1 Estimation Processing Flow

(1) Judgment of data used for estimation processing
Judge which data—entrance-side (ticket inspection) or exit-side (ticket collection)—to use for judging what train was ridden. Usually, exit-side data is used as there is little deviation in time from detraining to exiting the station, but exit-side data cannot

be obtained in the following situations, so judgment is made complementing with entrance-side data.

- a) If there is no automatic ticket gate at the destination station.
- b) If transferring directly to the Tokaido/Sanyo Shinkansen via the transfer gate at Tokyo Station.

(2) V-shape transfer judgment

If the entrance station and exit station are not on the same commercial line, as shown in Fig. 2, a V-shape transfer is judged to have been made.

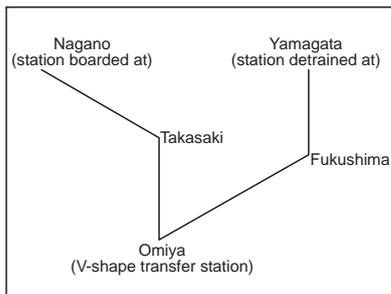


Fig. 2 Example of V-shape Transfer

(3) Judgment of transfer on same line

Judge transfer by combination of reserved seat ticket and non-reserved seat ticket.

(4) Judgment of reserved seat / non-reserved seat / group ticket
Judge type of ticket used. If, when a reserved seat ticket is used, the passenger enters much later (earlier) than the arrival (departure) time of the specified train, handle as having used a non-reserved seat.

(5) Judgment of passenger route

For ticket gate data where it is judged that a non-reserved seat was used in processing of (4), search for the closest train and assign the passenger route.

(6) Judgment of transfer train

(7) Same as above

If V-shape transfer or transfer on same line, judge passenger route before and after transfer.

(8) OD data creation

Perform processing to set extraction conditions. Make so condition settings can be selected as conventional passenger volume report format (passenger volume by section between stations) and OD format (passenger volume by section between stations by OD chart).

2.1.3 Verification of Accuracy

In order to verify accuracy, relative error was defined as in Formula (1) and calculated with the correct value being passenger volume measured visually by survey personnel riding a total of 36 trains in commercial operation.

$$\text{Relative error} = \frac{\sum_{\text{by type of seat}} \sum | \text{estimated value} - \text{measured value} |}{\sum_{\text{by type of seat}} \sum \text{measured value}} \dots (1)$$

Additionally, the percentage of all sections between stations with a relative error of 10% or greater was defined as “deviation” and calculated. Table 2 shows the average values of those calculation results.

Table 2 Results of Verifying Accuracy (Fiscal 2014)

No.	Seat type	Relative error	Deviation
1	Non-reserved seat	7.7 %	30.2 %
2	Reserved seat	3.9 %	11.1 %
3	Overall	5.9 %	22.1 %

Relative error was higher than the target of 5% for non-reserved seats and lower than that target for reserved seats. And with non-reserved seats, deviation exceeded 30%, showing that relative error and deviation need to be improved on for the simulation to be used in actual work.

2.2 Improvement of Algorithm (Fiscal 2015)

In this study, we set an objective of improving estimation accuracy through the following two efforts.

- 1) Examine how to deal with the issues in the fiscal 2014 study covered later and implement those possible in the simulator.
- 2) In addition to the above, extract causes of lowered accuracy, study how to deal with those, and implement those possible in the simulator.

2.2.1 Onboard Survey

In order to improve reliability of results verification in improving the algorithm, we greatly increased the number of train runs surveyed over those of the previous year, taking into account items such as trains to be surveyed, type of train, interval between train arrival times, and number of runs on each line.

<Survey period>

Five days from November 30 to December 4, 2015

<Number of trains surveyed>

A total of 103 trains were surveyed. The number of trains where measurements were taken and passenger volume are shown by line in Fig. 3, 4, and 5.

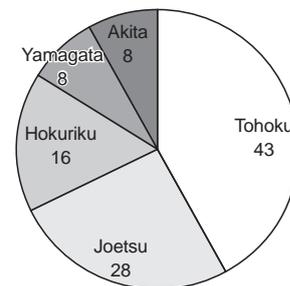


Fig. 3 Number of Train Runs per Line

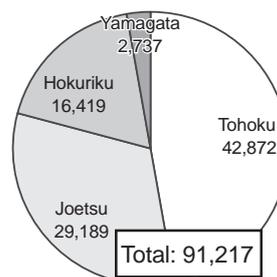


Fig. 4 Measured passenger volume (non-reserved seats)

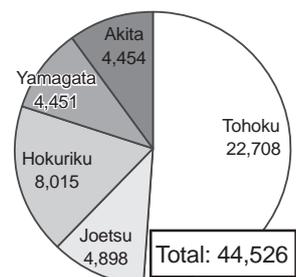


Fig. 5 Measured passenger volume (reserved seats)

2.2.2 Issues and Response

In judging passenger route where non-reserved seat is used, basically “train arrival time” and “passenger exiting time” are used to judge that “the passenger route was by the train that arrived closest to the exiting time” (Fig. 6).

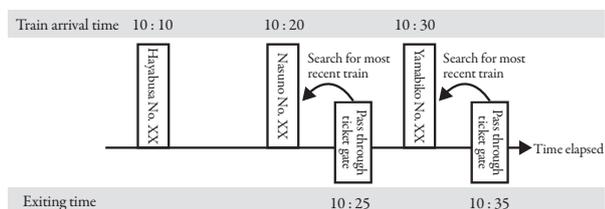


Fig. 6 Image of Judging Passenger Route

However, in stations such as Ueno where the distance from platform to ticket gate is far, the interval between train arrival time and passenger exiting time is longer, and the effect of that cannot be ignored particularly in sections where interval between trains is short. For that reason, we introduced in some stations a mechanism of changing “threshold between train arrival time and passenger exiting time” taken into account when judging passenger route. Additionally, we considered proportionally dividing passenger volume between trains for those with short intervals and arrival times close to each other. However, dividing at a uniform ratio does not necessarily lead to improved accuracy, and it may rather lead to reduced accuracy, so we put off responding to that and left it as an issue for the future.

It also came to light during the study that there were cases where, with some combinations of ticket types, double counting occurs and passenger route was judged incorrectly. We thus improved a total of 15 algorithms by means such as removing invalid combinations from input data and implemented those in the simulator.

2.2.3 Verification of Accuracy

Table 3 and 4 show the averages of relative error and deviation in the individual patterns. For reference, we compared those with passenger volume reports.

Table 3 Results of Verifying Accuracy: Relative Error (Fiscal 2015)

No	Pattern	Non-reserved seat	Reserved seat
1	Influence of transport disorder removed	6.9 %	4.9 %
2	No. 1 and tabulated by passenger volume report sections	6.1 %	4.9 %
3	No. 2 and two lines with ticket gate data	4.9 %	3.6 %
Reference	Passenger volume report	7.4 %	2.6 %

Table 4 Results of Verifying Accuracy: Deviation (Fiscal 2015)

No	Pattern	Non-reserved seat	Reserved seat
1	Influence of transport disorder removed	29.6 %	13.9 %
2	No. 1 and tabulated by passenger volume report sections	20.4 %	13.9 %
3	No. 2 and two lines with ticket gate data	8.9 %	9.0 %
Reference	Passenger volume report	28.8 %	6.7 %

(1) Target achievement status

We set targets of relative error of less than 5% and deviation of less than 10% based on the opinions of personnel in charge of transport planning. For lines where ticket gate data is available, targets were achieved for both relative error and deviation were in tabulation of passenger volume report sections spanning many sections between stations.

For non-reserved seats, relative error in all patterns and deviation other than in the No. 1 pattern were greater than in passenger volume reports.

(2) Issues

For reserved seats, however, results greater than those of passenger volume reports could not be achieved in any patterns of relative error and deviation. This is assumed to be because the effect of referencing ticket sales data and ticket type data from the onboard ticket inspection system is large with passenger volume reports.

Also, in this study, ticket gate data could be obtained for all stations on just two lines, the Tohoku Shinkansen and Joetsu Shinkansen. In terms of number of train runs, this makes up only about 70% of the total data. This is because there are stations without automatic ticket gates on the Akita and Yamagata Shinkansen and ticket gates data cannot be obtained from the five stations on the new section of the Hokuriku Line operated by JR-West. However, it may be possible to overcome those issues if the proposal for the future covered in the next chapter is achieved. If all ticket gate data can be obtained for the Hokuriku Shinkansen, estimation can be done with ticket gate data of 80 to 90% of all trains.

3 Proposal for Future Practical Use

3.1 System Composition in Practical Use

We examined the system composition when the simulator is put into practical use. Fig. 7 shows that configuration of the current system.

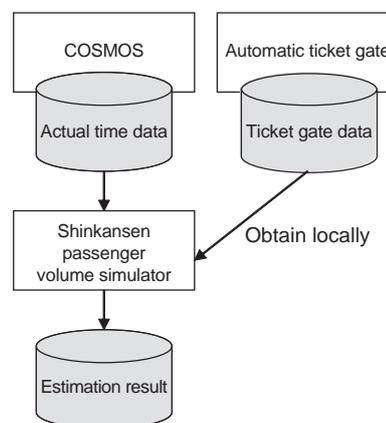


Fig. 7 Composition of Current System

The current simulator uses actual time data results obtained from the Computerized Safety, Maintenance and Operation Systems of Shinkansen (COSMOS) and ticket gate data obtained locally from automatic ticket gates to estimate passenger volume of individual trains. Fig. 8 shows the proposed configuration of the system in practical use.

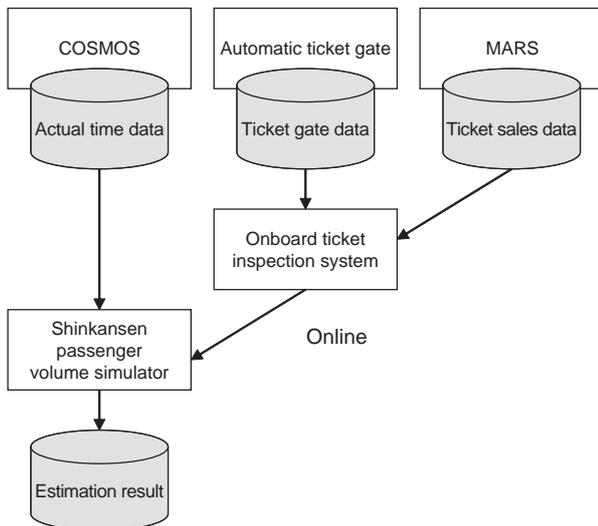


Fig. 8 Proposed Composition of System in Practical Use

In forming the system, the following benefits can be obtained by linking data online via the existing onboard ticket inspection system or linking as a subsystem within the onboard ticket inspection system.

- With the onboard ticket inspection system, data of automatic ticket gates of Shinkansen stations operated by JR East is linked, so there is no need to build a new mechanism for data linkage with stations.
- Automatic ticket gate data is gathered from Hokuriku Shinkansen stations in the JR-West area as well, so accuracy of estimation for that Shinkansen line is improved. The same applies for the Hokkaido Shinkansen, which was opened later.
- With the system, ticket sales data from MARS also is linked, so estimation accuracy for reserved seats will probably approach accuracy of passenger volume reports.

3.2 Other Improvement Proposals

Currently, the majority of stations on the Yamagata and Akita Shinkansen do not have automatic ticket gates, so ticket gate data can only be obtained for some stations, causing a decline in simulator accuracy.

However, simplified Suica ticket gates have been set up and utilized at some of those stations, and it is known that those are linked with the Suica ID management system. If IC log information that is anticipated in the future to be obtainable from those simplified ticket gates can be used, improvement in simulator accuracy can be expected.

4 Conclusion

In this paper, the estimation method proposed by the authors demonstrated accuracy greater than that of passenger volume reports for non-reserved seats in tabulation by passenger volume report sections. In that, passenger volume including OD by data obtainable from existing equipment and data sources is estimated without installing sensors and the like. Therefore, it can be said to be effective in busy times, especially when trains are congested with standing passengers.

As an outlook for the future, estimation accuracy could possibly be improved further by building a system that collects ticket gate data and accumulating and analyzing that data on a day-to-day basis. Additionally, predictions for the future will be possible by combining external data such as that for weather and events. In that way, venues for utilizing data may expand beyond transport business to a variety of areas including services, such as proving information to customers. Into the future, we will work to create further value for the Shinkansen and contribute to its advancement through studies on utilization of data.

Reference:

- 1) Tsutomu Suzuki, Minoru Kimura, Kenji Koyo, *Shinkansen Jidokaisatsuki Data wo Katsuyoshita Diagram Sakusei Shien System no Kaibatsu* [in Japanese]
- 2) Yuu Shikoda, Kazutaka Ito, Sei Sakairi, *Shinkansen Ryokaku Simulator nikansuru Kiso Kenkyu* [in Japanese]
- 3) Kazutaka Ito, Yuu Shikoda, Sei Sakairi, Fumiaki Mashiba, "Shinkansen Josha Jinin System no Kaihatsu" [in Japanese], *Proceedings of 52nd Railway Cybernetics Symposium*, Article No. 203 (November 2015)
- 4) Kazutaka Ito, *Shinkansen Josha Ninzu Simulator no Suikei Seido Kojo nikansuru Kenkyu*
- 5) East Japan Railway Company, In-house communications training basic course "Unsou Keikaku/Unkou Kanri Dai 2 Sho Unsou Juyou Yosoku 2-2 Kotsuryo no Chousa" [in Japanese]