

Contributing Factors for Train Delays during Morning Rush Hour in Japanese Metropolitan Areas

Keigo Ohshima*, Kayoko Yamamoto

Graduate School of Informatics and Engineering, University of Electro-Communications, Tokyo, Japan

Email: *o1930022@edu.cc.uec.ac.jp

How to cite this paper: Ohshima, K. and Yamamoto, K. (2020) Contributing Factors for Train Delays during Morning Rush Hour in Japanese Metropolitan Areas. *Journal of Transportation Technologies*, 10, 154-168.

<https://doi.org/10.4236/jtts.2020.102010>

Received: February 22, 2020

Accepted: March 22, 2020

Published: March 25, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The present study aims to reveal the contributing factors for train delays in Tokyo metropolitan area by conducting statistical analyses, focusing on passenger trains, and using a variety of information by including data concerning train cars, stations, passengers, tracks and working timetables as explanatory variables. The present study conducted 2 types of statistical analyses including the standard multiple regression analysis and the logistic regression analysis by setting “average delay time” which indicates the quantitative conditions of delays, and “occurrence of delays” which indicates the qualitative condition, as objective variables. According to the results of the logistic regression analysis, the possibility of direct operations increasing the delay occurrence rate was quantitatively indicated. Therefore, direct operations are regarded as a contributing factor for train delays concerning metropolitan areas in recent years. Additionally, it was confirmed that the concentration of demand on terminal stations is also a contributing factor for train delays. On the other hand, it is certain that direct operations contribute to improving the convenience of passengers as well as the operational efficiency of train cars. Therefore, it would be ideal to resolve delays by easing the concentration of demands which may be accomplished by recommending off-peak commuting as well as adjustments to the working timetables.

Keywords

Train Delay, Morning Rush Hour, Train Line Network, Statistical Analysis, Standard Multiple Regression Analysis, Logistic Regression Analysis

1. Introduction

In the metropolitan areas of many countries, the commutable zones spread in

the suburbs according to the urbanization, and train lines are congested and delayed during rush hour. Especially in the case of Japan, many train lines in the metropolitan areas have intense commuter rush every morning, and a large number of train lines are delayed during rush hour. However, the frequency and time of such delays vary depending on the characteristics of each train line. According to the Ministry of Land, Infrastructure, Transport and Tourism [1], in Tokyo metropolitan area, the number of days where delay certificates were issued during the 20 weekdays in 2016 was a maximum of 19.1 days and a minimum of 1.4 days. Additionally, there is an increase in mutual direct operations between train lines in recent years, and further expansions of such operations can be expected in the future. Due to this increase, it has also become common for an incident caused in one location to affect the entire metropolitan area.

Furthermore, according to the Tokyo Metro Co., Ltd. [2], the number of passengers in specific stations in central Tokyo has increased due to mutual direct operations, and the congestion within station yards has become even more significant. Therefore, in order to improve the convenience of train lines network in metropolitan areas, it is essential to multilaterally analyze the contributing factors for delays with the characteristics of each train line in mind. On the other hand, quantitative analyses are extremely important in searching for the above contributing factors.

Therefore, focusing on passenger trains, the present study aims to reveal the contributing factors for train delays in Japanese metropolitan areas by conducting statistical analyses. The above contributing factors will be made clear using various data by adding information concerning train cars, stations, passengers, tracks and working timetables as explanatory variables. Additionally, by preparing data for both single train lines and entire direct operation sections, the above contributing factors can be identified based on the current conditions of metropolitan train networks.

2. Related Work

The present study will be categorized as a study related to the train delays in metropolitan areas. In this category, the preceding studies can be divided into two groups. The first one is the studies related to the modeling of passengers' behaviors, and the second one is related to the characteristics of train delays focusing on specific lines and train lines network. In Japan, because train lines network is tremendously complicated and the congestion becomes serious problem, there are many preceding studies in both these two groups. The following are representative examples of studies closely related to the present study.

Regarding the studies related to the modeling of passengers' behaviors, in Japan, Uematsu *et al.* (2009) [3] analyzed the causes for delays and developed a simulation system that analyzes the delay occurrence and influence mechanism using an agent model. More specifically, they modeled the movement of passengers and developed a system that simulates delays due to concentration of de-

mand which is caused by passenger movements. Based on this result, Iwakura *et al.* (2013) [4] developed a multi-agent simulation model to analyze train knock-on delay in the busiest line in Tokyo metropolitan area. Additionally, Kobayashi *et al.* (2016) [5] estimated train boarding door choice model for knock-on urban train delay simulation, considering the volumes of boarding and disembarking passengers due to passenger demand and structure design of each station such as number of stairs and location. Kanai *et al.* (2011) [6] proposed an optimal delay management from passengers' viewpoints considering the whole railway network. Kunimatsu *et al.* (2012) [7] developed a microsimulation system to simulate both train operation and passengers' train choice behavior. They applied this system to an actual railway line in a metropolitan area and evaluated two train schedules by calculating the generalized cost, which reflects each passenger's disutility based on his or her experience. Sato *et al.* (2013) [8] formulated the timetable rescheduling problem as a Mixed Integer Programming (MIP) problem, and introduced a timetable rescheduling algorithm which outputs a rescheduling plan minimizing further inconvenience to the passengers caused by the disruption.

In other countries, Landex *et al.* (2010) [9] simulated the disturbances and modelling of expected train passenger delays in Denmark. Börjesson *et al.* (2011) [10] investigated how passengers on long-distance trains value unexpected delays relative to scheduled travel time and travel cost in Sweden. Dollevoet *et al.* (2011) [11] proposed a model where passengers' rerouting is incorporated in the delay management process in the Netherlands. Jian *et al.* (2012) [12] proposed a simulation model to investigate the relationship between train delays and passenger delays and to predict the dynamic passenger distribution in a large-scale rail transit network in Shanghai City, China. Robenek *et al.* (2016) [13] analyzed and improved the current planning process of the passenger railway service in light of the recent railway market changes, in order to introduce the passenger centric train timetabling problem in Switzerland. Li *et al.* (2016) [14] analyzed a passenger's alternative choices and the corresponding influence mechanism with train delay in detail, taking up the Shanghai URT system in China. Xu *et al.* (2018) [15] proposed the last train delay management especially for serious effect on transfer passengers' regular trips, using bi-objective mixed integer programming (MIP) model and genetic algorithm (GA), and taking up the Beijing subway network in China as a target.

Regarding the studies related to the characteristics of train delays focusing on specific lines and train lines network, in Japan, Kariyazaki *et al.* (2010, 2011) [16] [17] used various data to identify the current condition of delays of metropolitan train lines and conducted a present data analysis of the occurrence and influence of delays. They also presented characteristics of delays using the operation history of Tokyo subways (Tokyo Metro) in addition to the operating and stopping time between direct operation sections of private railways and Tokyo subways. Based on these studies, Kariyazaki *et al.* (2013) [18] analyzed urban railway de-

lays and estimated the effects of the countermeasures adopting a cellular automaton model. Kariyazaki *et al.* (2015) [19] also formulated a train operation simulation model, which reproduces the behavior of train operation under the knock-on delay, taking into account the interaction between the trains. Yamamura *et al.* (2014) [20] conducted a delay occurrence factor analysis using past train operation data of Tokyo subways, and provided countermeasures by identifying contributing factors for train delays. Miyazaki *et al.* (2014) [21] analyzed train delay data and conducted a present data analysis by taking the characteristics of each train line into consideration. They developed a simulation model that re-enacts train operations during rush hour and considered the impact of early train departures and the installation of platform screen doors.

In other countries, Goverde (2010) [22] and Corman *et al.* (2012) [23] computed the propagation of initial delays over a periodic railway timetable, and the domino effect of secondary delays over the entire network in the Netherlands. Dingler *et al.* (2010) [24] determined the cause of train delays making extensive use of a simulation tool known as rail traffic controller (RTC) in the United States (U. S.). Cule *et al.* (2011) [25] adapted and applied the state-of-the-art techniques for mining frequent episodes to the specific problem, in order to reveal the hidden patterns of trains passing under the knock-on delay in Belgium. Liu *et al.* (2012) [26] conducted statistical analyses to examine the effects of accident cause, type of track, and derailment speed in the U.S. Bergström *et al.* (2013) [27] addressed the lack of reliability within the Swedish rail network by identifying passenger train delay distributions. Markovića *et al.* (2015) [28] proposed machine learning models that capture the relation between passenger train arrival delays and various characteristics of a railway system in Serbia. Wen *et al.* (2017) [29] conducted statistical analysis on primary delays in Wuhan–Guangzhou high-speed railway (HSR). They also investigated the primary delays including delay causes, delay frequencies, delays’ temporal and spatial occurrences, affected number of trains and delay recovery patterns. Mussanov *et al.* (2017) [30] described the delay performance of different train types under combinations of structured and flexible operations on single-track railway lines in North America.

Comparing with the preceding studies in the related fields mentioned above, focusing on passenger trains, the present study will demonstrate the originality by conducting statistical analyses of various kinds of data concerning train cars, stations, passengers, tracks, and working timetables with many train lines in metropolitan areas, in addition to conducting quantitative analyses of potential contributing factors for train delays. Furthermore, the present study also demonstrates the usefulness to clearly grasp the degree of effect of each contributing factors for train delays by conducting statistical analyses of the above data with many train lines. Accordingly, based on the analysis results, the present study can provide the detailed information to the countermeasures against the train delays in Japanese metropolitan area which has complicated train lines

network and serious congestion.

3. Framework and Method

3.1. Framework and Process

In Section 4, the data of train lines which is an explanatory variable, and the data of delays which is an explained variable are gathered and processed for use in statistical analyses. Next, in Section 5, statistical analyses are conducted based on the data gathered and processed in Section 4 and the potential causes for delays are discussed.

3.2. Method

In order to quantitatively grasp contributing factors for train delays, the present study will conduct 2 types of statistical analyses: the standard multiple regression analysis and the logistic regression analysis. Regarding objective variables, the former will be “average delay time” which indicate the quantitative situation of delays, while the latter will be “number of days with the occurrence of delays” which indicate the qualitative situation of delays.

Additionally, as the combination method for explanatory variables, there are 3 types including the all-possible regression method, the variable specification method, and the sequential selection method. As there are many explanatory variables to reveal the contributing factors for train delays in Japanese metropolitan areas making it difficult to select the best variable, the present study will use the sequential selection method. Additionally, this method has 3 types including the forward selection method, the backward elimination method, and the stepwise method. Among these, the stepwise method will be used in the present study, as it has the highest possibility of obtaining efficient variable combination, and has been the most used method in preceding studies in the related fields. Furthermore, the stepwise method adopting the Akaike Information Criteria (AIC) [31] is superior in having a clear process to select the appropriate variables based on a constant standard.

3.3. Target Train Lines

In the present study, Tokyo metropolitan area, which is the largest metropolitan area in Japan and has tremendously complicated train lines network and serious congestion, is selected as a target. Tokyo metropolitan area consists of six prefectures such as Tokyo Metropolis, and Kanagawa, Chiba, Saitama, Yamanashi, Tochigi, Gunma and Ibaragi Prefectures. Thus, in Tokyo metropolitan area, the range of train lines is very huge, it is necessary to grasp the outlines of the target train lines selected in the present study. Therefore, **Figure 1** describes the schematic diagram of the target train lines.

As shown in **Figure 1**, the present study targets 55 train lines of 17 railway companies in Tokyo metropolitan area. However, in Tokyo metropolitan area, as train lines network is tremendously complicated, it is difficult to display all

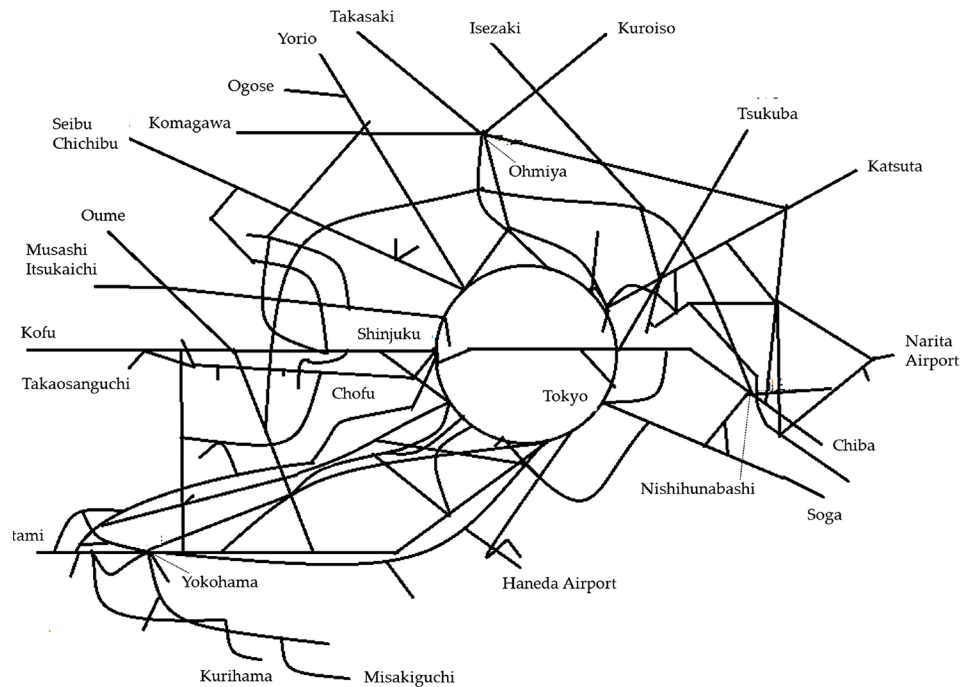


Figure 1. Schematic diagram of the target train lines in Tokyo metropolitan area.

train lines in a single figure. Therefore, **Figure 1** shows the schematic diagram of the target train lines excluding subway lines. As shown in **Figure 1**, the Yamanote Line (the Tokyo Loop Line) surrounds the central part of Tokyo Metropolitan, and most of train lines are radially extended from the sub-centers such as Shinjuku to the suburban areas.

4. Collection and Processing of Data

4.1. Collection of Delay Data

For data concerning delays, delay certificates that are available on the website of each railway company were used. The delay time displayed on the delay certificates was recorded and if there were no delay certificates, the delay time was recorded as being 0 minutes. Kariyazaki *et al.* (2010) [16] indicated that the delay certificates reveal that the number of delays were significantly higher on weekdays than weekends, and were especially high in the morning. Therefore, weekday mornings are set as the target in the present study. The specific time zone targeted to the present study was set from the first train to 10 am. Additionally, the target period was set to 21 weekdays in June 2018. Because, in Japan, the difference between the days of the week can be minimized as there are no holidays in June, and the effect of the weather can be eliminated due to the good weather around this time.

Regarding “average delay time” which is an explained variable, the largest value is 8.6 minutes, the smallest is 0 minutes, and the average value is 8.5 minutes in June 2018. For “number of days with the occurrence of delays”, the largest value was 21 days, the smallest value was 0 days, and the average value

was 12.1 days. In this way, during the target period, 1 out of the 55 target train lines, which was a short train line (length of 13 km) with no direct operation, had no delays. On the other hand, one train line had delays every day during the target period which was significantly more compared with other lines. This train line has 2 direct operations, the length of the direct operation section is 173.8 km, the transportation capacity per train during peak hours is 1372.7 people/train, and the number of stations is 70. Therefore, it can be said that this is a train line that conducts direct operations in addition to being large-scaled in the first place.

4.2. Explanatory Variables

In order to consider the effect of train cars, stations, passengers, tracks and working timetables (My LINE Tokyo Timetable) [32] on train operations, 10 explanatory variables shown in **Table 1** will be selected. **Table 1** enumerates these explanatory variables together with the data sources.

In the following part, the details of the explanatory variables shown in **Table 1** are explained.

1) Transportation capacity for each train during peak hours (unit: people/train)

This is a variable concerning the total passenger capacity of train cars and is the transportation capacity during peak hours in the most congested section divided by the number of operating trains per hour (7).

2) Number of stations

This is the number of stations on the target train lines.

3) Transported passengers per hour during peak hours (unit: people/hour)

This is a variable indicating the number of passengers on trains during peak hours in the most congested section.

Table 1. Sources of data for explanatory variables.

Explanatory variables	Data sources
1) Transportation capacity for each train during peak hours (unit: people/train)	Ministry of Land, Infrastructure, Transport and Tourism: Statistics information on congestion rates (2017) [33]
2) Number of stations	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]
3) Transported passengers per hour during peak hours (unit: people/hour)	Ministry of Land, Infrastructure, Transport and Tourism: Statistics information on congestion rates (2017) [33]
4) Number of stairs and escalators in terminal stations	Station yard map from the website of each railway company
5) Length of train lines (unit: km)	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]
6) The average number of tracks	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]
7) Number of operating trains per hour during peak hours (unit: train/time)	Ministry of Land, Infrastructure, Transport and Tourism: Statistics information on congestion rates (2017) [33]
8) Number of trains according to type	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]
9) Number of lines with direct operation	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]
10) Length of direct operation sections (unit: km)	Kotsu Shinbunsha: My LINE Tokyo Timetable (2018) [32]

4) Number of stairs and escalators in terminal stations

This is a variable concerning stations, which include the number of stairs and escalators on the platform of terminal stations. Stations of each line with the highest number of passengers on the platforms were selected as the terminal stations in principle.

5) Length of train lines (unit: km)

This is a variable concerning the length of each train line. Working kilometers were used as the value to indicate the length.

6) The average number of tracks

This is a variable concerning the number of tracks. For example, lines that are double tracks are counted as 2 and lines that are half quadruple tracks and half double tracks are counted as 3.

7) Number of operating trains per hour during peak hours (unit: train/time)

This is a variable concerning the working timetables indicating the number of operating trains per hour during peak hours in the most congested sections.

8) Number of trains according to type

This is a variable concerning the working timetables indicating the number of train types including “rapid”, “express” and “commuter express”, and operating in the target time zone.

9) Number of lines with direct operation

This is a variable indicating the number of lines with direct operations that involve multiple train lines. The direct operation sections are as explained below.

10) Length of direct operation sections (unit: km)

This is a variable indicating the total length of direct operation sections including the relevant train lines. Lines with no direct operations will have the same train line length as (5).

4.3. Setting Direct Operation Sections

In order to consider the recent increase of direct operations in metropolitan areas, the present study will adopt explanatory variables concerning the entire direct operation section. The standard for direct operation sections was set as “sections in which trains run on the applicable train line within the target period”, and the direct operation sections were set based on the My Line Tokyo Timetable [32].

5. Results and Discussion

In this section, R will be used to confirm the multicollinearity of explanatory variables, and reveal the contributing factors for train delays in Japanese metropolitan areas by conducting 2 types of statistical analyses. R is a programming language of statistics of open-source free software for statistics analysis. In the present study, using R, 2 types of statistical analyses including the standard multiple regression analysis and the logistic regression analysis by setting “average delay time” which indicates the quantitative conditions of delays, and “occur-

rence of delays” which indicates the qualitative condition, as objective variables. Additionally, 10 explanatory variables shown in **Table 1** are adopted in the above 2 types of statistical analyses.

5.1. Standard Multiple Regression Analysis

5.1.1. Variables Selection for the Standard Multiple Regression Analysis

As a result of using the stepwise method by increasing and decreasing variables in the standard multiple regression analysis, “transportation capacity for each train during peak hours”, “number of stairs and escalators in terminal stations”, “number of trains according to type”, and “length of direct operation sections” were selected as explanatory variables.

5.1.2. Evaluation and Discussion of the Standard Multiple Regression Analysis Results

Table 2 shows the result of the multiple regression analysis. The discussion of each explanatory variable is as shown below.

1) Transportation capacity for each train during peak hours

Trains with large transportation capacity are trains with more cars. If there are more train cars, the distribution of passengers becomes unbalanced. Therefore, even if the number of passengers is not extremely high, getting on and off the train may take longer, resulting in the train being delayed.

2) Number of stairs and escalators in terminal stations

If there are many stairs and escalators installed in the terminal station, it can be considered that the demand is concentrated on that station. Therefore, delays may be caused as getting on and off at the terminal takes time.

3) Number of trains according to type

Train lines with severe congestion issues tend to have fewer train types. Especially with many subway lines, the regression coefficient for the number of train types is under 0 as only local trains are operated.

4) Length of direct operation sections

As the frequency of accidents and trouble arising naturally increases when operating sections are longer, the average delay time also becomes longer.

Based on the information above, the average delay time is considered to increase

Table 2. Results of the standard multiple regression analysis.

	Regression coefficient	Standard error	t value	p value
Intercept	-3.994×10^{-16}	8.228×10^{-2}	0	1
Transportation capacity for each train during peak hours	0.3039	0.1312	2.317	0.024631
Number of stairs and escalators in terminal stations	0.3147	0.1092	2.881	0.005819
Number of trains according to type	-0.1686	8.721×10^{-2}	-1.934	0.058842
Length of direct operation sections	0.4017	0.1094	3.670	0.000589

due to the length of operation sections as well as the magnitude and concentration of the transportation demand.

5.2. Logistic Regression Analysis

5.2.1. Variables Selection for the Standard Multiple Regression Analysis

As a result of using the stepwise method by increasing and decreasing variables in the logistics regression analysis, “transported number of passengers per hour during peak hours”, “number of stairs and escalators in terminal stations”, “the average number of train lines”, “number of operating trains per hour during peak hours”, “number of trains according to type”, “number of lines with direct operation”, and “length of direct operation sections” were selected as explanatory variables. Comparing with the standard multiple regression analysis, the logistic regression analysis has 3 additional explanatory variables: “the average number of train lines”, “number of operating trains per hour during peak hours” and “number of lines with direct operation”.

5.2.2. Evaluation and Discussion of the Standard Multiple Regression Analysis Results

Table 3 shows the result of the logistic regression analysis. The discussion of each explanatory variable is as shown below.

1) Transported number of passengers per hour during peak hours

The regression coefficient, which is extremely close to 0, indicates that it does not directly cause delays, while the change in the transported number of passengers affects the occurrence rate of delays.

2) Number of stairs and escalators in terminal stations

As with the discussion of the standard multiple regression analysis, the

Table 3. Results of the logistic regression analysis.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p-value
Intercept	-3.486	0.0306098	0.9067	-3.845	0.000120
Transportation capacity for each train during peak hours	3.372×10^{-5}	1.0000337	6.499×10^{-6}	5.189	2.12×10^{-7}
Number of stairs and escalators in terminal stations	0.2666	1.3054729	6.814×10^{-2}	3.912	9.16×10^{-5}
The average number of train lines	1.175	3.2379618	0.4972	2.363	0.018113
Number of operating trains per hour during peak hours	-4.529×10^{-2}	0.9557222	1.694×10^{-2}	-2.673	0.007523
Number of trains according to type	-0.2101	0.8105155	0.4972	2.363	0.018113
Number of lines with direct operation	0.1892	1.2082781	6.591×10^{-2}	2.871	0.004098
Length of direct operation sections	1.652×10^{-3}	1.0016530	9.982×10^{-4}	1.655	0.098000

concentration of demand in terminal stations is considered to cause an increase in the occurrence rate of delays.

3) The average number of train lines

The Saikyo-Kawagoe line, which had a small value compared to other lines due to the single track section within the Kawagoe line, had delays occurring every day during the target period. Additionally, most lines have double tracks and there is no significant difference among the values of each train line. Based on the above, the regression coefficient increased.

4) Number of operating trains per hour during peak hours

Delays become less frequent when the number of operating trains increases as the demand per train becomes smaller. On the other hand, as the intervals between trains become smaller when the number of trains increases, delays could occur as trains will be more likely to slow down as it gets closer to a train in front. As a result, the 2 above events cancel each other out and the regression coefficient becomes closer to 0.

5) Number of trains according to type

As with the discussion of the standard multiple regression analysis, train lines with few train types such as subway lines are more easily congested and have a higher occurrence rate of delays.

6) Number of lines with direct operation

Most lines conducting direct operations pass through the city center. As delays occur more frequently in city centers with high demand, train lines that conduct direct operations passing through the city center have more delays as the delays of trains in front affect the trains behind.

7) Length of direct operation sections

Tough the occurrence rate of accidents and trouble becomes higher when operation sections become longer, the regression coefficient value of the explanatory variable for the length of operation sections was low as such occurrence rate is still smaller compared to small-scale delays due to congestion. Additionally, as the p-value is high, the length of direct operation sections does not directly affect the occurrence rate of delays.

Based on the information presented above, the concentration of demand in addition to the number of trains with direct operations is highly correlated to the occurrence rate of delays. Therefore, it can be suggested the possibility that the increase in direct operations affects the occurrence of delays.

6. Conclusions

The present study revealed the contributing factors for train delays in Tokyo metropolitan area, Japan by conducting statistical analyses, focusing on passenger trains. More specifically, the above factors were grasped using various information including data concerning train cars, stations, passengers, tracks and working timetables as explanatory variables. Additionally, by preparing data for both single train lines and entire direct operation sections, the above contribut-

ing factors according to current conditions of metropolitan train networks were identified.

The present study conducted 2 types of statistical analyses including the standard multiple regression analysis and the logistic regression analysis by setting “average delay time” which indicates the quantitative conditions of delays, and “occurrence of delays” which indicates the qualitative condition, as objective variables. Regarding the comparison between the 2 types of statistical analysis results, the logistic regression analysis had 3 added explanatory variables: “the average number of train lines”, “number of operating trains per hour during peak hours”, and “number of lines with direct operation”.

According to the results of the logistic regression analysis, the possibility of direct operations increasing the delay occurrence rate was quantitatively indicated. Therefore, direct operations are regarded as a contributing factor for train delays concerning Tokyo metropolitan areas in recent years. Additionally, it was confirmed that the concentration of demand on terminal stations is also a contributing factor for train delays. On the other hand, it is certain that direct operations contribute to improving the convenience of passengers as well as the operational efficiency of train cars. However, as direct operations make it possible for passengers to arrive at their destination without transferring at terminal stations, direct operations can also be expected to ease the concentration of demand. Therefore, it would be ideal to resolve delays by easing the concentration of demands which may be accomplished by recommending off-peak commuting as well as adjustments to the working timetables.

In the future, it will be necessary to prepare more explanatory variables and further consider the characteristics of each train line in analyses. Additionally, as the data in the present study were gathered when the weather was good, the effect of weather is not sufficiently reflected in the data. The effect of such contributing factors must be considered by extending the period for collecting data. In this way, it is a task for future research projects to improve analytical accuracy.

Acknowledgements

We gratefully acknowledge the work of past and present members of our laboratory. We wish to thank the anonymous reviewers for comments on earlier version of this paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Ministry of Land, Infrastructure, Transport and Tourism: Starting the Visualization of Train Delay [Internet]. <http://www.mlit.go.jp/common/001215328.pdf>
- [2] Tokyo Metro Co. Ltd. (2014) Ranking of the Average Daily Number of Passengers of Each Station in Fiscal Year 2013.

- https://www.tokyometro.jp/corporate/enterprise/passenger_rail/transportation/passengers/2013.html
- [3] Uematsu, S. and Iwakura, S. (2009) A Multi-Agent Simulation Model for Estimating Knock-On Delay of Tokyo Metropolitan Railway. *Proceedings of Infrastructure Planning*, **40**, 19-23.
- [4] Iwakura, S., Takahashi, I. and Morichi, S. (2013) A Multi Agent Simulation Model for Estimating Knock-On Train Delays under High-Frequency Urban Rail Operation. *Transport Policy Studies Review*, **15**, 31-40.
- [5] Kobayashi, W. and Iwakura, S. (2016) Development of Train Boarding Door Choice Model for Knock-On Urban Train Delay Analysis. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **72**, I1067-I1074. https://doi.org/10.2208/jscejipm.72.I_1067
- [6] Kanai, S., Shingo, K., Harada, S. and Tomii, N. (2011) An Optimal Delay Management Algorithm from Passengers' Viewpoints Considering the Whole Railway Network. *Journal of Rail Transport Planning & Management*, **1**, 25-37. <https://doi.org/10.1016/j.jrtpm.2011.09.003>
- [7] Kunimatsu, T., Hirai, C. and Tomii, N. (2012) Train Timetable Evaluation from the Viewpoint of Passengers by Microsimulation of Train Operation and Passenger Flow. *Electrical Engineering in Japan*, **181**, 51-62. <https://doi.org/10.1002/eej.21264>
- [8] Sato, K., Tamura, K. and Tomii, N. (2013) A MIP-Based Timetable Rescheduling Formulation and Algorithm Minimizing Further Inconvenience to Passengers. *Journal of Rail Transport Planning & Management*, **3**, 38-53. <https://doi.org/10.1016/j.jrtpm.2013.10.007>
- [9] Landex, A. and Nielsen, O.A. (2010) Simulation of Disturbances and Modelling of Expected Train Passenger Delays. In: Hansen, I.A., Ed., *Timetable Planning and Information Quality*, WIT Press, Boston, 85-93.
- [10] Börjesson, M. and Eliasson, J. (2011) On the Use of "Average Delay" as a Measure of Train Reliability. *Transportation Research Part A: Policy and Practice*, **45**, 171-184. <https://doi.org/10.1016/j.tra.2010.12.002>
- [11] Dollevoet, T., Huisman, D., Schmidt, M. and Schöbel, A. (2011) Delay Management with Re-Routing of Passengers. *Transportation Science*, **46**, 74-89. <https://doi.org/10.1287/trsc.1110.0375>
- [12] Jiang, Z., Li, F., Xu, R. and Gao, P. (2012) A Simulation Model for Estimating Train and Passenger Delays in Large-Scale Rail Transit Networks. *Journal of Central South University*, **19**, 3603-3613. <https://doi.org/10.1007/s11771-012-1448-9>
- [13] Robenek, T., Maknoon, Y., Azadeh, S.S., Chen, J. and Bierlairea, M. (2016) Passenger Centric Train Timetabling Problem. *Transportation Research Part B: Methodological*, **89**, 107-126. <https://doi.org/10.1016/j.trb.2016.04.003>
- [14] Li, W. and Zhu, W. (2016) A Dynamic Simulation Model of Passenger Flow Distribution on Schedule-Based Rail Transit Networks with Train Delays. *Journal of Traffic and Transportation Engineering*, **3**, 364-373. <https://doi.org/10.1016/j.jtte.2015.09.009>
- [15] Xu, W., Zhao, P. and Ning, L. (2018) Last Train Delay Management in Urban Rail Transit Network: Bi-Objective MIP Model and Genetic Algorithm. *KSCE Journal of Civil Engineering*, **22**, 1436-1445. <https://doi.org/10.1007/s12205-017-1786-0>
- [16] Kariyazaki, K., Hibino, N. and Morichi, S. (2010) Study on Mechanism of Worsening Punctuality in Urban Railway Services. *Infrastructure Planning Review*, **27**, 871-879. <https://doi.org/10.2208/journalip.27.871>

- [17] Kariyazaki, K., Hibino, N. and Morichi, S. (2011) Simulation Analysis of Daily Service Delay Focusing on Train Headway. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **67**, 1001-1010. https://doi.org/10.2208/jscejpm.67.67_I_1001
- [18] Kariyazaki, K., Hibino, N. and Morichi, S. (2013) Simulation Analysis of Train Operation to Recover Knock-On Delay Earlier. *Asian Transport Studies*, **2**, 284-294.
- [19] Kariyazaki, K., Hibino, N. and Morichi, S. (2015) Simulation Analysis of Train Operation to Recover Knock-On Delay under High-Frequency Intervals. *Case Studies on Transport Policy*, **3**, 92-98. <https://doi.org/10.1016/j.cstp.2014.07.007>
- [20] Yamamura, A. (2014) Delay Reduction Measures and Their Effects in Dense Transportation Operation Using Train Traffic Record Data. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **70**, 44-55. <https://doi.org/10.2208/jscejpm.70.44>
- [21] Miyazaki, K., Hibino, N. and Morichi, S. (2014) Analysis of Train Delay in Urban Railway Services Based on Characteristics of Each Line. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **70**, 477-486. https://doi.org/10.2208/jscejpm.70.I_477
- [22] Goverde R.M.P. (2010) A Delay Propagation Algorithm for Large-Scale Railway Traffic Networks. *Transportation Research Part C: Emerging Technologies*, **18**, 269-287. <https://doi.org/10.1016/j.trc.2010.01.002>
- [23] Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M. (2012) Optimal Inter-Area Coordination of Train Rescheduling Decisions. *Transportation Research Part E: Logistics and Transportation Review*, **48**, 71-88. <https://doi.org/10.1016/j.tre.2011.05.002>
- [24] Dingler, M., Koenig, A., Sogin, S. and Barkan, C.P.L. (2010) Determining the Causes of Train Delay. *Proceedings of the 2010 Annual AREMA Conference*, Orlando, 29 August-1 September 2010, 14 p.
- [25] Cule, B., Goethals, B., Tassenoy, S. and Verboven, S. (2011) Mining Train Delays. *10th International Symposium on Advances in Intelligent Data Analysis X*, Porto, 29-31 October 2011, 113-124. https://doi.org/10.1007/978-3-642-24800-9_13
- [26] Liu, X., Saat, M.R. and Barkan, C.P.L. (2012) Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates. *Journal of the Transportation Research Board*, **2289**, 154-163. <https://doi.org/10.3141/2289-20>
- [27] Bergström, A. and Krüger, A. (2013) Modeling Passenger Train Delay Distributions: Evidence and Implications. Working Papers in Transport Economics 2013: 3, CTS-Centre for Transport Studies Stockholm (KTH and VTI), 31 p. <https://www.diva-portal.org/smash/get/diva2:774278/FULLTEXT01.pdf>
- [28] Markovića, N., Milinkovićb, S., Tikhonovc, K.S. and Schonfelda, P. (2015) Analyzing Passenger Train Arrival Delays with Support Vector Regression. *Transportation Research Part C: Emerging Technologies*, **56**, 251-262. <https://doi.org/10.1016/j.trc.2015.04.004>
- [29] Wen, C., Li, Z., Lessan, J., Fu, L., Huang, P. and Jiang, C. (2017) Statistical Investigation on Train Primary Delay Based on Real Records: Evidence from Wuhan-Guangzhou HSR. *International Journal of Rail Transportation*, **5**, 170-189. <https://doi.org/10.1080/23248378.2017.1307144>
- [30] Mussanov, D., Nishino, N. and Dick, C.T. (2017) Delay Performance of Different Train Types under Combinations of Structured and Flexible Operations on Single-Track Railway Lines in North America. *Proceedings of the 7th International Conference on Railway Operations Modelling and Analysis*, Lille, 4-7 April 2017,

759-776.

- [31] Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
<https://doi.org/10.1109/TAC.1974.1100705>
- [32] Kotsushibunsha (2018) My Line Tokyo Timetable (June, 2018).
- [33] Ministry of Land, Infrastructure, Transport and Tourism: Statistics Information Related to Congestion Rate Data [Internet].
<http://www.mlit.go.jp/common/001245932.pdf>