

# Contributing Factors for Delays during the Morning Commute Hours and the Impact of the Spread of COVID-19 for Metropolitan Train Lines in Japan

Keigo Ohshima, Kayoko Yamamoto\*

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

Email: \*kayoko.yamamoto@uec.ac.jp

**How to cite this paper:** Ohshima, K. and Yamamoto, K. (2021) Contributing Factors for Delays during the Morning Commute Hours and the Impact of the Spread of COVID-19 for Metropolitan Train Lines in Japan. *Journal of Transportation Technologies*, 11, 519-544.

<https://doi.org/10.4236/jtts.2021.114033>

**Received:** July 18, 2021

**Accepted:** August 28, 2021

**Published:** August 31, 2021

Copyright © 2021 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 conduct 2 types of statistical analysis to reveal the impact of the spread of COVID-19 on train delays by comparing the potential contributing factors before, during and after the outbreak of the virus in the metropolitan train lines in Japan. First of all, the result of the present study clearly revealed the changes in contributing factors for train delays caused by the spread of COVID-19. Specifically, the contributing factors for train delays changed due to the decrease of passengers by the effect of the outbreak of the virus. Additionally, though large terminal stations were considered to be a major contributing factor in causing and increasing train delays in the past, this was not the case after the spread of COVID-19. Therefore, under such conditions, it is more effective to make improvements in small to medium stations and tracks rather than terminal stations. Furthermore, as the decrease in passengers also decreased train delays in commuter lines going to the suburbs due to the spread of COVID-19, the contributing factor for such lines is the excessive number of passengers. Therefore, as for countermeasures for train delays after the effects of COVID-19, it is necessary to disperse passengers in order to avoid passengers concentrating in the same time zones and train lines.

## Keywords

Train Delay, Morning Rush Hour, Train Line Network, COVID-19, Statistical Analysis, Standard Multiple Regression Analysis, Logistic Regression Analysis

---

## 1. Introduction

Regarding many train lines in the metropolitan area in Japan, the commuting

rush hour in the morning can be very heavy with delays in many train lines. However, the frequency and length of train delays vary from one train line to another. According to the Ministry of Land, Infrastructure, Transport and Tourism [1], the number of days in which train delays certificates were issued during a period of 20 weekdays in 2016 was a maximum of 19.1 days and a minimum of 1.4 days. Additionally, through services between train lines have been increasing in recent years and a further increase is also expected. Therefore, it is not uncommon for an issue that occurred in one location to affect the entire metropolitan area. Furthermore, according to the Tokyo Metro Co., Ltd. [2], through services has caused an increase in passengers at specific central stations which has significantly intensified the congestion within stations. Therefore, in order to improve the convenience of railway networks within the metropolitan area, the contributing factors for train delays must be analyzed and improved based on the characteristics of each train line.

On the one hand, COVID-19 is an infectious disease that was first erupted in Wuhan, China, in December 2019, and spread across the entire globe in 2020. In order to prevent its spread, the Japanese government declared a state of emergency based on the Act on Special Measures for Pandemic Influenza and New Infectious Diseases Preparedness and Response for Tokyo Metropolis and Saitama, Chiba, Kanagawa, Osaka, Hyogo and Fukuoka Prefectures for the first time on April 7th, 2020. However, the infections continued even after the removal of the state of emergency. Additionally, the number of people taking trains decreased due to the spread of COVID-19. According to the East Japan Railway Company [3], their train operating revenue for April 2020 compared with the previous year was 50.5% for commuter passes and 24.0% in total including non-commuter passes. In this way, the spread of COVID-19 has affected the trend of people taking trains and this change is ongoing as the virus continues to spread.

On the other hand, quantitative analysis to search for contributing factors is extremely important in order to discover potential contributing factors. The present study aims to conduct statistical analysis to reveal the impact of the spread of COVID-19 on train delays by comparing the potential contributing factors before, during and after the outbreak of the virus. Specifically, contributing factors can be identified from a wider range of elements by adding train cars, stations, tracks and timetables as explanatory variables. Moreover, by preparing data for both single train line sections as well as through service train line sections, contributing factors for train delays can be clarified based on the recent conditions of metropolitan railway networks. Accordingly, the present study will clarify whether the spread of COVID-19 had effects on train delays, by comparing the contributing factors before, during and after the outbreak of COVID-19.

## 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.

Regarding the studies related to the modeling of passengers' behaviors, Uematsu *et al.* [4], Landex *et al.* [5], Dollevoet *et al.* [6], Jian *et al.* [7], Iwakura *et al.* [8], Kobayashi *et al.* [9] [10], Kunimatsu *et al.* [11], Corman [12] and König *et al.* [13] developed their original simulation models and conducted train delay simulations. Additionally, Kanai *et al.* [14] proposed an optimal train delays management from passengers' viewpoints considering the whole railway network. Börjesson *et al.* [15] investigated how passengers on long-distance trains value unexpected delays relative to scheduled travel time and travel cost. Sato *et al.* [16] formulated the timetable rescheduling problem as a Mixed Integer Programming (MIP) problem and introduced a timetable rescheduling algorithm that outputs a rescheduling plan minimizing further inconvenience to the passengers caused by the disruption. Robenek *et al.* [17] 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. Li *et al.* [18] analyzed passengers' alternative choices and the corresponding influence mechanism with train delay in detail. Xu *et al.* [19] 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). Yap *et al.* [20] propose a supervised learning approach to predict multiple types of disruption occurrence at different stations of a public transport network and measure the impacts on passenger delays.

Regarding the studies related to the characteristics of train delays focusing on specific lines and train lines networks, in Japan, Kariyazaki *et al.* [21] [22] [23] [24], Yamamura [25] [26], Miyazaki *et al.* [27], Kobayashi *et al.* [28] and Ohshima *et al.* [29] conducted current state analyses of train delays and its expansions in order to clearly grasp the actual conditions in the metropolitan area. The above studies also identified the characteristics of train delays using data related to urban railways in the metropolitan area. In other countries, Goverde [30] and Corman *et al.* [31] 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.* [32] 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.* [33] 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.* [34] conducted statistical analyses to examine the effects of accident cause, type of track, and derailment speed in the U.S. Bergström *et al.* [35] addressed the lack of reliability within the Swedish rail network by identifying passenger train delay distributions. Markovića *et al.* [36] proposed machine learn-

ing models that capture the relation between passenger train arrival delays and various characteristics of a railway system in Serbia. Wen *et al.* [37] conducted statistical analysis on primary delays in the Wuhan-Guangzhou high-speed railway (HSR). Mussanov *et al.* [38] described the train delays performance of different train types under combinations of structured and flexible operations on single-track railway lines in North America. Onet *et al.* [39] developed a data-driven Train Delay Prediction System (TDPS) for the Italian railway network which exploits the most recent big data technologies, learning algorithms and statistical tools. Arshad *et al.* [40] presented the prediction of train delay in Indian Railways through machine learning techniques. Wang *et al.* [41] collected and analyzed a three-month dataset of weather, train delay and train schedule records in order to understand the patterns of delays and to predict delay time of the Beijing-Guangzhou railway. Huang *et al.* [42] [43] developed Deep Learning (DL) approach to predict train delays in the railway networks in China and the Netherlands. Huang *et al.* [44] proposed a hybrid model to predict the main consequences of disruptions and disturbances during train operations, targeting Wuhan-Guangzhou and Xiamen-Shenzhen HSR lines. Mohd *et al.* [45] developed a machine learning model to predict the delay in the arrival of trains combining previous train delay data and weather data in India.

While the above are representative studies related to the present study, except for Ohshima *et al.* [30], there are few studies that have quantitatively revealed the contributing factors for train delays in the metropolitan area despite it being a significant issue that greatly impacts the daily lives of many people. This is because train networks, especially in large metropolitan areas in any country, are complex and contributing factors for train delays differ according to the train line or station. Therefore, the present study has a profound significance as it proposes strategies to reduce train delays by identifying the contributing factors. Additionally, in comparison with the aforementioned preceding studies, referring to the result of Ohshima *et al.* [30], the first originality of the present study is that quantitative analysis is conducted regarding the potential contributing factors for delays that generally occur in the train lines within the metropolitan area in Japan by conducting statistical analyses after collecting and processing multiple open data concerning various train lines and train delays. The second point of originality is that a plan to deal with the changes in contributing factors for train delays is proposed by comparing those before, during and after the outbreak of COVID-19 in order to identify the impact of train delays caused by the virus.

### **3. Framework and Method**

#### **3.1. Framework and Process**

In Section 4, data concerning train lines that are explanatory variables and data concerning train delays that are explained variables are collected and processed in order to conduct statistical analyses. Next, in Section 5, statistical analyses are conducted based on the data collected and processed in the previous section.

Furthermore, in Section 6, contributing factors for train delays are evaluated and discussed based on the results of statistical analyses in the previous section. Moreover, Kariyazaki *et al.* [22] revealed that the number of train delays certificates issued is higher on weekday mornings comparing with weekends. Therefore, the present study will only target weekday mornings.

In order to reveal the impact of COVID-19 on train delays, the present study will analyze 2 time periods: Before and after it was affected by COVID-19. First, data from June 2018 will be analyzed to clarify the contributing factors for train delays before the spread of COVID-19. The reason for selecting this month to represent the period before COVID-19 is that it is not affected by any holidays as there are no national holidays in June, and that it had comparatively low rainfall during the rainy season. Next, data from 2020 were analyzed in order to reveal the contributing factors for train delays caused by COVID-19. 2020 was divided into smaller periods to analyze and identify the changes in impact according to the progression of the spread of infections.

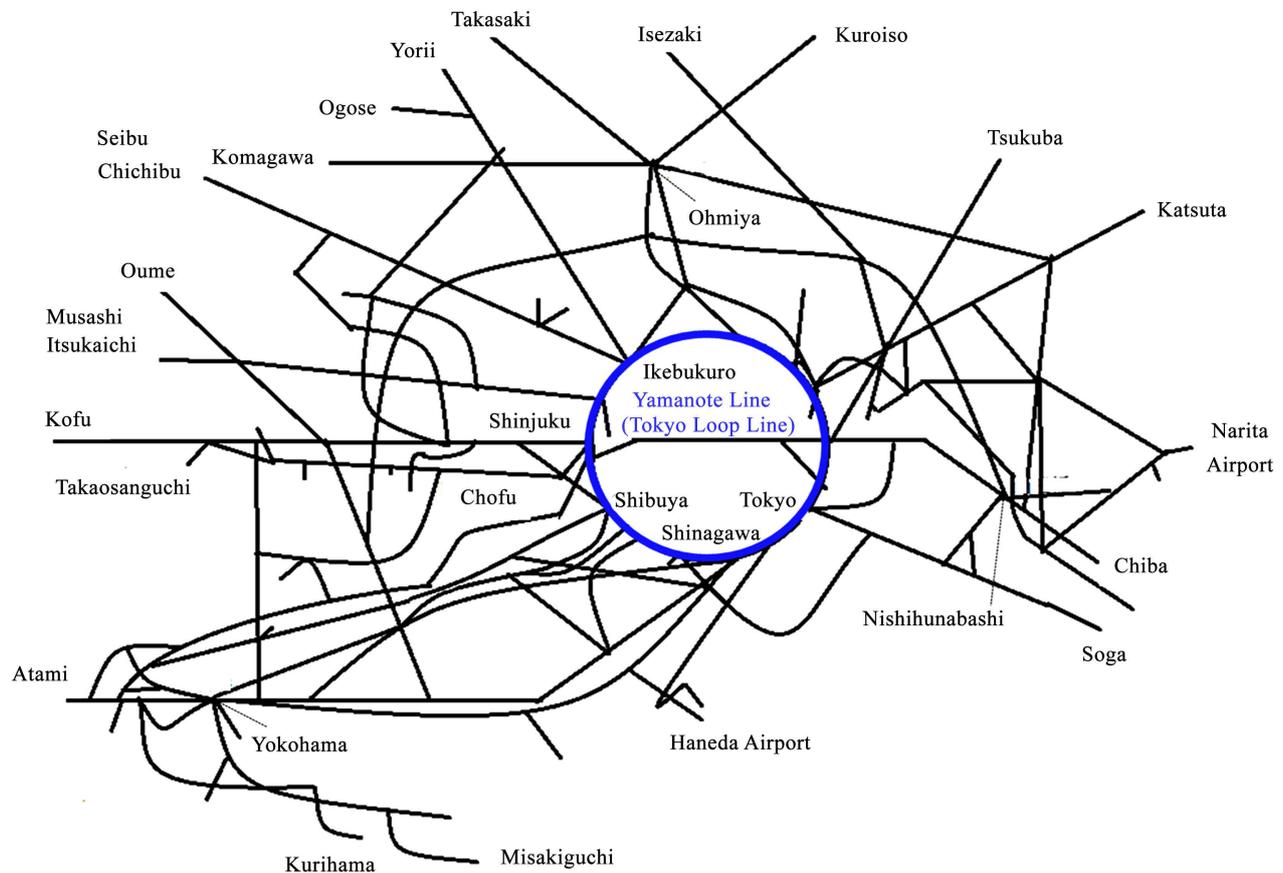
### 3.2. Method

In order to quantitatively grasp contributing factors for train delays, the present study will conduct 2 types of statistical analysis: standard multiple regression analysis and logistic regression analysis. The explained variable for the former is the “average delay time”, which indicates the quantitative condition of delays, while that for the latter is the “number of days that a delay occurred”, which indicates the qualitative condition of delays. Additionally, the stepwise method employing the Akaike Information Criterion (AIC) [46] will be used as a variable selection method.

### 3.3. Target Train Lines

In the present study, the 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. The Tokyo metropolitan area consists of six prefectures such as Tokyo Metropolis and Kanagawa, Chiba, Saitama, Yamanashi, Tochigi, Gunma and Ibaragi Prefectures. Thus, in the 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 the Tokyo metropolitan area. However, in the Tokyo metropolitan area, as train lines network is tremendously complicated, it is difficult to display all 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 Metropolis, and most of train lines are radially extended from the sub-centers such as Shinjuku, Shibuya, Ikebukuro and Shinagawa to the suburban areas.



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

## 4. Collection and Processing of Data

### 4.1. Collection of Train Delay Data

For train delay data, train delay certificates made public on the website of each company were recorded and used. **Figure 2** shows a sample of train delay certificate that describes date of delay occurrence, time period of delay, train line of delay occurrence, and maximum delay time. The website of each company was regularly visited, and the delay times displayed on the train delay certificates released on the respective days were recorded.

If there was no train delay certificate issued on a certain day, the delay time for that day was recorded as zero. Though the recording time used for delays was between the first train to 10 am, the records were made from the first train to 9 am for some companies that did not issue train delay certificates until 10 am. If there were multiple train delay certificates for the corresponding time period, the longest delay is recorded as the delay time. Since many of the major railway companies provide train delay certificates combining main and branch train lines, train line data were also calculated by combining multiple train lines accordingly. Additionally, in cases where train delay certificates were separately provided, one from before major stations and the other starting from major sta-

tions, the data for the 2 zones were combined. **Figure 3** shows sample preprocessed data concerning the average delay time for standard multiple regression analysis, and the number of days that a delay occurred for logistic regression analysis.

### 4.2. Explanatory Variables

In order to consider the effect of train cars, stations, passengers, tracks and working

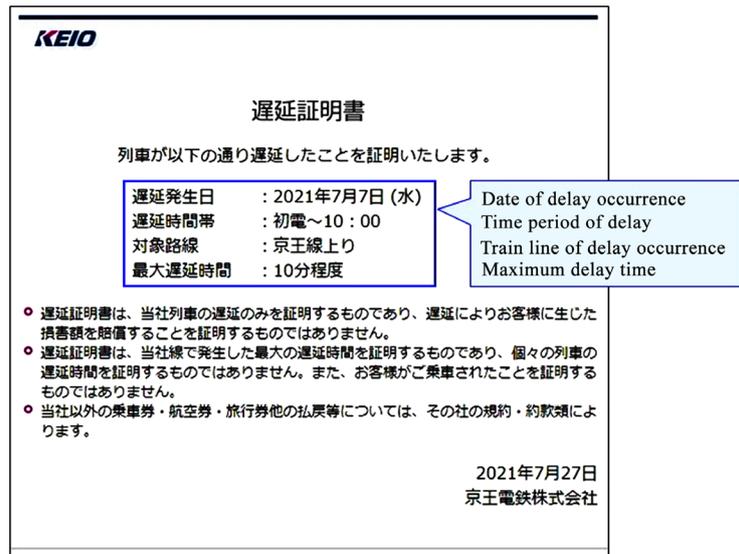


Figure 2. Sample of train delay certificate.

Average delay time									
Tokaido line	0	10	20	30	20	0	0	0	10
Yokosuka-Soubu line	0	0	0	20	61	10	0	10	0
Keihintohoku-Negishi line	0	0	20	10	10	10	0	0	0
Yokohama line	0	0	10	0	10	0	0	0	0
Naibu line	0	0	10	0	10	0	0	0	0
Yamate line	0	0	10	10	0	0	0	0	0
Chuo line	10	0	0	0	0	10	0	0	10
Chuo-Soubu line	0	0	0	10	20	0	0	0	0
Ome line	10	0	0	0	10	0	0	0	0
Itsukaishi line	0	0	0	0	10	0	0	0	0
Utsunomiya line	0	0	0	10	20	0	20	0	0
Takasaki line	0	0	0	10	10	0	0	0	0
Saikyo-Kawagoe line	10	0	0	10	20	10	0	0	0
Joban rapid line	10	0	0	20	20	0	0	0	0
Number of days that a delay occurred									
Train lines	2020/4/8	2020/4/9	2020/4/10	2020/4/13	2020/4/14	2020/4/15	2020/4/16	2020/4/17	2020/4/20
Tokaido line	0	1	1	1	1	0	0	0	1
Yokosuka-Soubu line	0	0	0	1	1	1	0	1	0
Keihintohoku-Negishi line	0	0	1	1	1	1	0	0	0
Yokohama line	0	0	1	0	1	0	0	0	0
Naibu line	0	0	1	0	1	0	0	0	0
Yamate line	0	0	1	1	0	0	0	0	0
Chuo line	1	0	0	0	0	1	0	0	1
Chuo-Soubu line	0	0	0	1	1	0	0	0	0
Ome line	1	0	0	0	1	0	0	0	0
Itsukaishi line	0	0	0	0	1	0	0	0	0
Utsunomiya line	0	0	0	1	1	0	1	0	0
Takasaki line	0	0	0	1	1	0	0	0	0
Saikyo-Kawagoe line	1	0	0	1	1	1	0	0	0
Joban rapid line	1	0	0	1	1	0	0	0	0

Figure 3. Sample preprocessed data for average delay time and number of days that a delay occurred.

timetables in 2018 and 2020 (My LINE Tokyo Timetable in 2018 and 2020) [47] [48] on train operations, based on the results of Ohshima *et al.* [30], 10 explanatory variables shown in **Table 1** will be selected. **Table 1** enumerates these explanatory variables together with the data sources, and **Figure 4** shows sample data for explanatory variables.

Train lines	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Tokaido line	104.6	4	633.5	1844	19	66780	4	2	21	2	12	45
Yokosuka-Soubu line	112.5	7	536.4	1864	19	64150	6	3	28	2	15	50
Keihintohoku-Negishi line	81.2	1	123.8	1480	26	71250	4	1	46	2	31	94
Yokohama line	42.6	1	99.8	1184	19	37110	3	2	20	2	8	12
Naibu line	35.5	0	35.5	888	25	40860	6	1	26	2	9	14
Yamate line	34.5	0	34.5	1628	23	58990	8	1	29	2	26	104
Chuo line	134.1	4	304.9	1480	30	81000	7	3	43	2	16	52
Chuo-Soubu line	60.2	2	138.8	1480	26	75230	5	1	39	2	23	66
Ome line	18.5	2	85.8	1340.71	17	29836	4	1	13	1.92973	2	6
Itsukaishi line	11.1	2	55.5	888	6	7070	2	1	7	1	1	3
Utsunomiya line	163.5	2	375.6	1844	14	36580	5	2	34	2	11	37
Takasaki line	105	3	453.3	1844	14	42300	5	2	25	2	4	13
Saikyo-Kawagoe line	67.5	3	115.9	1471.58	19	51050	6	2	24	1.601481	10	41
Joban rapid line	123.3	1	208	2044.84	19	59640	5	1	25	2	12	31

**Figure 4.** Sample data for explanatory variables.

**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 [49] [50].
2) Number of stations	Kotsu Shinbunsha: My LINE Tokyo Timetable [47] [48].
3) Number of transported passengers per hour during peak hours (unit: people/hour)	Ministry of Land, Infrastructure, Transport and Tourism: Statistics information on congestion rates [49] [50].
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 [47] [48].
6) Average number of tracks	Kotsu Shinbunsha: My LINE Tokyo Timetable [47] [48].
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 [49] [50].
8) Number of trains according to type	Kotsu Shinbunsha: My LINE Tokyo Timetable [47] [48].
9) Number of through service train lines	Kotsu Shinbunsha: My LINE Tokyo Timetable [47] [48].
10) Length of through service sections (unit: km)	Kotsu Shinbunsha: My LINE Tokyo Timetable [47] [48].
11) Number of connecting operation stations	Station maps from the website of each railway company.
12) Length of lines with connecting operation (unit: km)	Station maps from the website of each railway company.

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 variable relates to the train's car indicating the capacity of a train, calculated by dividing the transportation capacity per hour within the most congested section during peak hours by the number of operating trains (7) per hour.

2) Number of stations:

This variable represents the number of stations on the target train lines.

3) Number of transported passengers per hour during peak hours (unit: people/hour):

This variable indicates the number of passengers on a train in the most congested sections during peak hours.

4) Number of stairs and escalators in terminal stations:

This variable relates to the stations and it is the number of stairs and escalators on the platform of the terminal station. Stations from each train line with the highest number of incoming/outgoing passengers were selected as terminal stations.

5) Length of train lines (unit: km):

This variable relates to tracks indicating the length of train lines. For the distance, working kilometers are used.

6) Average number of tracks:

This variable indicates the number of tracks. For example, an entire train line that has double tracks is counted as 2, while a train line with quadruple tracks on one half and double tracks on the other are counted as 3.

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

This variable concerns the timetables indicating the number of operating trains per hour in the most congested sections during peak hours.

8) Number of trains according to type:

This variable relates to the timetables which indicate the number of each train type, such as "rapid", "express" and "commuter express", operated during the target time slot.

9) Number of through service train lines:

This variable represents the number of train lines with through services operated by multiple trains on corresponding train lines. Through service sections are mentioned in the following section.

10) Length of through service sections (unit: km):

This variable indicates the length of the entire through service sections including the relevant train lines. The length of train lines with no through service will be the same length calculated in (5).

11) Number of connecting operation stations:

This variable indicates the number of stations on the relevant train line in which passengers can transfer to other lines. Transfers between train lines that are included in the same train line group are not counted as connecting stations on data.

#### 12) Length of lines with connecting operation (unit: km):

This variable indicates the total number of connecting train lines for each station on the relevant train line. The criteria for transfers are the same as the number of connecting stations in (11). Additionally, if transfers can be made to multiple lines included in the same train line group from the same station, they will be compiled.

### 4.3. Setting Connecting Operation Sections

In order to consider the through services that have been increasing within the metropolitan area in recent years, explanatory variables concerning the entire through service sections were adopted. The criterion for through service section was “the sections in which a train that runs on the relevant train line during the target period runs to the end of the target period”, and the through service section were set based on the My Line Tokyo Timetables [47] [48].

### 4.4. Dividing of the Analysis Period

In the present study, 2020 was divided into smaller periods to analyze and identify the changes in the contributing factors for train delays according to the progression of the spread of COVID-19. The analysis periods were divided as shown below based on events that greatly impacted the general public such as the declaration of the state of emergency in Tokyo Metropolis as well as the start and discontinuation of the Go To Travel Campaign by the Japanese government. The Go To Travel Campaign was an initiative to provide subsidies equivalent to a maximum of 50% towards travel products and accommodations. In the present study, this campaign will be considered as a Japanese government initiative to encourage people to go out, and its start and discontinuation date will be used for the dividing of analysis periods. **Table 2** presents each period and its details in the present study.

**Table 2.** List of analysis periods in 2020.

Number	Start	End	Events of the starting date
1	January 4th	April 7th	-
2	April 8th	May 25th	Declaration of the state of emergency in Tokyo Metropolis and Saitama, Chiba, Kanagawa, Osaka, Hyogo and Fukuoka Prefectures
3	May 26th	July 17th	Lifting of the state of emergency mentioned above
4	July 22nd	September 30th	Start of the Go To Travel Campaign
5	October 1st	December 14th	Start of the application of the Go To Travel Campaign for travels to and from Tokyo Metropolis
6	December 15th	December 28th	Announcement of temporary discontinuation of the Go To Travel Campaign

## 5. Results

### 5.1. Overview of the Results

This section will conduct analyses using the R language and examine the results. R is a programming language used in statistical analysis for free and open-source software. In this section, the explanatory variables selected as a result of Section 4.2, and the results of the regression analysis using the selected explanatory variables will be shown in the table. The items in the table of the results differ according to the type of analysis. In the standard multiple regression analysis, **Table 3** and **Tables 5-10** indicate the regression coefficient, standard error, t value, and p value of each explanatory variable. In the logistic regression analysis, **Table 4** and **Tables 11-16** indicate the regression coefficient, Exp (regression

**Table 3.** Results of the standard multiple regression analysis for June 2018.

	Regression coefficient	Standard error	t value	p value
Intercept	$6.609 \times 10^{-15}$	$2.676 \times 10^{-2}$	0	1
Number of through service train lines	$9.174 \times 10^{-2}$	$4.255 \times 10^{-2}$	2.156	$3.1298 \times 10^{-2}$
Length of through service sections	0.1210	$4.446 \times 10^{-2}$	2.722	$6.587 \times 10^{-3}$
Transportation capacity for each train during peak hours	0.1449	$4.284 \times 10^{-2}$	3.382	$7.43 \times 10^{-4}$
Number of stairs and escalators in terminal stations	$6.732 \times 10^{-2}$	$3.989 \times 10^{-2}$	1.686	$9.1761 \times 10^{-2}$
Number of trains according to type	$-6.319 \times 10^{-2}$	$3.242 \times 10^{-2}$	-1.989	$5.1538 \times 10^{-2}$
Number of connecting operation stations	-0.1336	$7.039 \times 10^{-2}$	-1.897	$5.8024 \times 10^{-2}$
Length of lines with connecting operation	0.2755	$7.618 \times 10^{-2}$	3.617	$3.11 \times 10^{-4}$

**Table 4.** Results of the logistic regression analysis for June 2018.

	Regression coefficient	Exp (regression coefficient)	Standard error	z value	p value
Intercept	-4.107	$1.645611 \times 10^{-2}$	0.9369	-4.384	$1.17 \times 10^{-5}$
Number of through service train lines	0.2299	1.25851362	$4.970 \times 10^{-2}$	4.626	$3.72 \times 10^{-6}$
Number of operating trains per hour during peak hours	$-8.150 \times 10^{-2}$	0.92173524	$1.618 \times 10^{-2}$	-5.036	$4.76 \times 10^{-7}$
Number of transported passengers per hour during peak hours	$2.882 \times 10^{-5}$	1.00002882	$6.651 \times 10^{-6}$	4.334	$1.47 \times 10^{-5}$
Number of stairs and escalators in terminal stations	0.2034	1.22552414	$7.381 \times 10^{-2}$	2.755	$5.87 \times 10^{-3}$
Number of stations	$-9.050 \times 10^{-3}$	0.99099086	$5.399 \times 10^{-3}$	-1.676	$9.367 \times 10^{-2}$
Average number of tracks	1.651	5.21247949	0.5042	3.274	$1.06 \times 10^{-3}$
Length of lines with connecting operation	$3.513 \times 10^{-2}$	1.03575200	$5.269 \times 10^{-3}$	6.667	$2.60 \times 10^{-11}$

coefficient), standard error, z value and p value. As suggested in the tables, the larger the absolute value of the regression coefficient is, the greater the impact its explanatory variable will have on the explained variable.

### 5.2. Results for before the Spread of COVID-19 (June 2018)

In this section, **Table 3** and **Table 4** show the results before the spread of COVID-19 (June 2018). In the standard multiple regression analysis for June 2018, the regression coefficient for the length of lines with connecting operation was exceptionally high, while the regression coefficient for the number of connecting operation stations was largely negative. This is because the train delay time in commuter lines connecting to large terminal stations that go out towards the suburbs is longer than that of the central lines that connect with many stations. The greater regression coefficient for the transportation capacity for each train during peak hours demonstrated the large-scale impact on the delay times of commuter lines, which tend to have more cars, going out to the suburbs. Additionally, the regression coefficient for the length of through service section was also high. This is because the delay times are long for the lines that connect with other commuter lines through the train lines in the city center. From the above, before the spread of COVID-19, delay times were longer in commuter lines going towards the suburbs and lines that run through services with many other train lines.

In regards to the logistic regression analysis for June 2018, the regression coefficient for the average number of tracks was especially high, and the regression coefficient for the number of stairs and escalators in terminal stations was also high. Tracks were accordingly quadrupled, indicating that the commuter lines with large-scale terminal stations going out to suburbs have a high delay probability. Additionally, the regression coefficient for the number of through service train lines was also high, and this suggests that such lines with many train lines have a high delay probability. From the above, before the spread of COVID-19, the delay probability was high for commuter lines going towards the suburbs and train lines with through services with many other train lines.

### 5.3. Results for after the Spread of COVID-19 (2020)

In this section, **Tables 5-16** show the results after the spread of COVID-19 (2020). Each period of 2020 is marked by the number indicated in **Table 2** of this section. In order to grasp the condition of the period in which COVID-19 spread the most, the results for Period 2 when the state of emergency was declared will be introduced first. Next, the results for the entire year of 2020 are compared in order to grasp the change in contributing factors for train delays due to the progression of the spread of COVID-19.

In the standard multiple regression analysis for during the state of emergency (Period 2), the regression coefficients for the length of train lines and length of through service sections were largely positive, while that of the number of trains

**Table 5.** Results of the standard multiple regression analysis for Period 1.

	Regression coefficient	Standard error	t value	p value
Intercept	$-7.301 \times 10^{-15}$	$1.569 \times 10^{-2}$	0	1
Length of train lines	0.1310	$3.038 \times 10^{-2}$	4.313	$1.66 \times 10^{-5}$
Number of connecting operation stations	$8.809 \times 10^{-2}$	$2.822 \times 10^{-2}$	3.121	0.00182
Length of lines with connecting operation	$7.644 \times 10^{-2}$	$3.547 \times 10^{-2}$	2.155	0.03125
Number of operating trains per hour during peak hours	-0.1066	$2.727 \times 10^{-2}$	-3.908	$9.49 \times 10^{-5}$
Number of transported passengers per hour during peak hours	0.1809	$2.828 \times 10^{-2}$	6.396	$1.80 \times 10^{-10}$
Number of trains according to type	$-8.957 \times 10^{-2}$	$2.541 \times 10^{-2}$	-3.524	0.00043
Average number of tracks	$3.357 \times 10^{-2}$	$1.916 \times 10^{-2}$	1.752	0.07986
Number of connecting operation stations	-0.1422	$4.499 \times 10^{-2}$	-3.160	0.00159
Length of lines with connecting operation	0.2248	$4.564 \times 10^{-2}$	4.926	$8.77 \times 10^{-7}$

**Table 6.** Results of the standard multiple regression analysis for Period 2.

	Regression coefficient	Standard error	t value	p value
Intercept	$-2.562 \times 10^{-15}$	$2.355 \times 10^{-2}$	0	1
Length of train lines	0.1701	$3.805 \times 10^{-2}$	4.471	$8.30 \times 10^{-6}$
Length of through service sections	0.1346	$3.440 \times 10^{-2}$	3.911	$9.56 \times 10^{-5}$
Number of transported passengers per hour during peak hours	$7.718 \times 10^{-2}$	$2.717 \times 10^{-2}$	2.841	$4.56 \times 10^{-3}$
Number of each train type	-0.1167	$2.722 \times 10^{-2}$	-4.286	$1.92 \times 10^{-5}$

**Table 7.** Results of the standard multiple regression analysis for Period 3.

	Regression coefficient	Standard error	t value	p value
Intercept	$-5.511 \times 10^{-15}$	$1.939 \times 10^{-2}$	0	1
Length of train lines	0.22887	$3.795 \times 10^{-2}$	6.027	$1.94 \times 10^{-9}$
Number of connecting operation stations	$6.740 \times 10^{-2}$	$3.372 \times 10^{-2}$	1.998	0.045790
Length of through service sections	$6.379 \times 10^{-2}$	$4.145 \times 10^{-2}$	1.539	0.123939
Number of transported passengers per hour during peak hours	$8.846 \times 10^{-2}$	$2.793 \times 10^{-2}$	3.167	0.001559
Number of trains according to type	-0.1383	$2.790 \times 10^{-2}$	-4.956	$7.72 \times 10^{-7}$
Number of connecting operation stations	-0.1562	$5.367 \times 10^{-2}$	-2.909	0.003658
Length of lines with connecting operation	0.1894	$5.469 \times 10^{-2}$	3.463	0.000544

**Table 8.** Results of the standard multiple regression analysis for Period 4.

	Regression coefficient	Standard error	t value	p value
Intercept	$-2.556 \times 10^{-15}$	$1.837 \times 10^{-2}$	0	1
Length of train lines	0.1557	$3.203 \times 10^{-2}$	4.861	$1.24 \times 10^{-6}$
Length of through service sections	0.2304	$2.957 \times 10^{-2}$	7.792	$9.64 \times 10^{-15}$
Number of operating trains per hour during peak hours	$-8.277 \times 10^{-2}$	$2.964 \times 10^{-2}$	-2.793	0.00527
Number of transported passengers per hour during peak hours	0.1529	$3.874 \times 10^{-2}$	3.946	$8.17 \times 10^{-5}$
Number of stairs and escalators in terminal stations	$-7.450 \times 10^{-2}$	$2.875 \times 10^{-2}$	-2.591	0.00962
Number of trains according to type	-0.1309	$2.577 \times 10^{-2}$	-5.081	$4.04 \times 10^{-7}$
Number of connecting operation stations	-0.1155	$5.112 \times 10^{-2}$	-2.258	0.02400
Length of lines with connecting operation	0.1614	$5.229 \times 10^{-2}$	3.086	0.00205

**Table 9.** Results of the standard multiple regression analysis for Period 5.

	Regression coefficient	Standard error	t value	p value
Intercept	$-1.176 \times 10^{-15}$	$1.706 \times 10^{-2}$	0	1
Length of train lines	0.2802	$3.318 \times 10^{-2}$	8.446	$<2 \times 10^{-16}$
Number of through service train lines	$8.866 \times 10^{-2}$	$3.008 \times 10^{-2}$	2.947	0.00324
Length of through service sections	$6.417 \times 10^{-2}$	$3.693 \times 10^{-2}$	1.738	0.08238
Number of operating trains per hour during peak hours	-0.1209	$2.800 \times 10^{-2}$	-4.317	$1.63 \times 10^{-5}$
Number of transported passengers per hour during peak hours	0.2075	$3.622 \times 10^{-2}$	5.728	$1.12 \times 10^{-8}$
Number of stairs and escalators in terminal stations	$-7.447 \times 10^{-2}$	$2.708 \times 10^{-2}$	-2.750	0.00600
Number of trains according to type	-0.1604	$2.778 \times 10^{-2}$	-5.773	$8.65 \times 10^{-9}$
Number of connecting operation stations	-0.2714	$4.819 \times 10^{-2}$	-5.633	$1.95 \times 10^{-8}$
Length of lines with connecting operation	0.2655	$4.922 \times 10^{-2}$	5.395	$7.44 \times 10^{-8}$

**Table 10.** Results of the standard multiple regression analysis for Period 6.

	Regression coefficient	Standard error	t value	p value
Intercept	$-1.324 \times 10^{-15}$	$3.788 \times 10^{-2}$	0	1
Length of train lines	0.4182	$8.025 \times 10^{-2}$	5.211	$2.67 \times 10^{-7}$
Length of through service sections	0.1261	$6.239 \times 10^{-2}$	2.021	0.043741
Number of transported passengers per hour during peak hours	0.1147	$4.530 \times 10^{-2}$	2.532	0.011635
Number of stations	-0.2328	$6.063 \times 10^{-2}$	-3.840	0.000138

**Table 11.** Results of the logistic regression analysis for Period 1.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p value
Intercept	-3.794	$2.251302 \times 10^{-2}$	0.5975	-6.350	$2.16 \times 10^{-10}$
Number of through service train lines	0.1348	1.14434786	$3.501 \times 10^{-2}$	3.851	$1.18 \times 10^{-4}$
Length of through service sections	$1.117 \times 10^{-3}$	1.00111781	$5.345 \times 10^{-4}$	2.090	$3.6599 \times 10^{-2}$
Transportation capacity for each train during peak hours	$7.189 \times 10^{-4}$	1.00071912	$1.692 \times 10^{-4}$	4.249	$2.15 \times 10^{-5}$
Number of transported passengers per hour during peak hours	$6.059 \times 10^{-6}$	1.00000606	$3.466 \times 10^{-6}$	1.748	$8.0410 \times 10^{-2}$
Number of stairs and escalators in terminal stations	$6.553 \times 10^{-2}$	1.06772353	$3.811 \times 10^{-2}$	1.719	$8.5536 \times 10^{-2}$
Number of stations	$-9.290 \times 10^{-3}$	0.99075273	$2.897 \times 10^{-3}$	-3.207	$1.340 \times 10^{-3}$
Average number of tracks	0.7348	2.08508685	0.2930	2.508	$1.2141 \times 10^{-2}$
Length of lines with connecting operation	$2.178 \times 10^{-2}$	1.02201926	$2.410 \times 10^{-3}$	9.036	$<2 \times 10^{-16}$

**Table 12.** Results of the logistic regression analysis for Period 2.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p value
Intercept	-1.9347811	0.1444559	0.9305960	-2.079	$3.7610 \times 10^{-3}$
Length of train lines	$1.48908 \times 10^{-2}$	1.0150022	$2.5325 \times 10^{-3}$	5.880	$4.11 \times 10^{-9}$
Number of through service train lines	0.1787198	1.1956856	$5.55136 \times 10^{-2}$	3.219	$1.285 \times 10^{-3}$
Transportation capacity for each train during peak hours	$9.385 \times 10^{-4}$	1.0009389	$2.863 \times 10^{-4}$	3.278	$1.046 \times 10^{-3}$
Number of operating trains per hour during peak hours	$3.71886 \times 10^{-2}$	1.0378887	$1.58399 \times 10^{-2}$	2.348	$1.8886 \times 10^{-2}$
Number of trains according to type	-0.2703075	0.7631448	$7.21054 \times 10^{-2}$	-3.749	$1.78 \times 10^{-4}$
Average number of tracks	-1.4075520	0.2447417	0.5196405	-2.709	$6.755 \times 10^{-3}$
Number of connecting operation stations	$2.69567 \times 10^{-2}$	1.0273234	$1.41192 \times 10^{-2}$	1.909	$5.6233 \times 10^{-2}$

**Table 13.** Results of the logistic regression analysis for Period 3.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p value
Intercept	-3.9398350	$1.945142 \times 10^{-2}$	0.2772774	-14.209	$<2 \times 10^{-16}$

**Continued**

Length of train lines	$8.0389 \times 10^{-3}$	1.00807135	$2.4639 \times 10^{-3}$	3.263	$1.10 \times 10^{-3}$
Number of through service train lines	0.1462204	1.15745131	$5.74517 \times 10^{-2}$	2.545	$1.092 \times 10^{-2}$
Length of through service sections	$1.8572 \times 10^{-3}$	1.00185888	$8.337 \times 10^{-4}$	2.228	$2.590 \times 10^{-2}$
Transportation capacity for each train during peak hours	$1.2679 \times 10^{-3}$	1.00126873	$2.336 \times 10^{-4}$	5.427	$5.72 \times 10^{-8}$
Number of stairs and escalators in terminal stations	$7.02043 \times 10^{-2}$	1.07272733	$4.67139 \times 10^{-2}$	1.503	0.13288
Average number of tracks	-0.1895380	0.82734129	$4.80941 \times 10^{-2}$	-3.941	$8.11 \times 10^{-5}$
Number of connecting operation stations	$2.97481 \times 10^{-2}$	1.03019500	$9.7559 \times 10^{-3}$	3.049	$2.29 \times 10^{-3}$

**Table 14.** Results of the logistic regression analysis for Period 4.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p value
Intercept	-2.653	$7.045479 \times 10^{-2}$	0.5076	-5.226	$1.73 \times 10^{-7}$
Length of train lines	$1.253 \times 10^{-2}$	1.01260505	$1.785 \times 10^{-3}$	7.019	$2.24 \times 10^{-12}$
Number of through service train lines	0.1630	1.17703050	$4.255 \times 10^{-2}$	3.831	$1.28 \times 10^{-4}$
Transportation capacity for each train during peak hours	$6.368 \times 10^{-4}$	1.00063701	$3.477 \times 10^{-4}$	1.831	$6.7027 \times 10^{-2}$
Number of operating trains per hour during peak hours	$-4.641 \times 10^{-2}$	0.95465366	$2.159 \times 10^{-2}$	-2.149	$3.1596 \times 10^{-2}$
Number of transported passengers per hour during peak hours	$1.704 \times 10^{-5}$	1.00001704	$8.356 \times 10^{-6}$	2.039	$4.1426 \times 10^{-2}$
Number of trains according to type	-0.1800	0.83524205	$4.973 \times 10^{-2}$	-3.620	$2.94 \times 10^{-4}$
Length of lines with connecting operation	$9.702 \times 10^{-3}$	1.00974951	$3.046 \times 10^{-3}$	3.185	$1.449 \times 10^{-3}$

**Table 15.** Results of the logistic regression analysis for Period 5.

	Regression coefficient	Exp (regression coefficient)	Standard error	z value	p value
Intercept	-1.445	0.2356813	0.1891	-7.642	$2.14 \times 10^{-14}$
Length of train lines	$1.524 \times 10^{-2}$	1.0153522	$1.628 \times 10^{-3}$	9.359	$<2 \times 10^{-16}$
Number of through service train lines	0.1229	1.1307447	$3.776 \times 10^{-2}$	3.254	$1.14 \times 10^{-3}$
Number of operating trains per hour during peak hours	$-9.642 \times 10^{-2}$	0.9080841	$1.451 \times 10^{-2}$	-6.644	$3.06 \times 10^{-11}$

**Continued**

Number of transported passengers per hour during peak hours	$3.636 \times 10^{-5}$	1.0000364	$4.527 \times 10^{-6}$	8.031	$9.69 \times 10^{-16}$
Number of trains according to type	$-8.789 \times 10^{-2}$	0.9158592	$4.312 \times 10^{-2}$	-2.038	$4.150 \times 10^{-2}$
Number of connecting operation station	$-4.418 \times 10^{-2}$	0.9567818	$2.183 \times 10^{-2}$	-2.023	$4.303 \times 10^{-2}$
Length of lines with connecting operation	$1.936 \times 10^{-2}$	1.0195453	$6.278 \times 10^{-3}$	3.083	$2.05 \times 10^{-3}$

**Table 16.** Results of the logistic regression analysis for Period 6.

	Regression coefficient	exp (regression coefficient)	Standard error	z value	p value
Intercept	-1.4713748	0.2296096	1.2810231	-1.149	0.260724
Length of train lines	$2.79134 \times 10^{-2}$	1.0283066	$4.5895 \times 10^{-3}$	6.082	$1.19 \times 10^{-9}$
Number of through service train lines	0.2285692	1.2568005	$9.31264 \times 10^{-2}$	2.454	$1.4112 \times 10^{-2}$
Transportation capacity for each train during peak hours	$1.3681 \times 10^{-3}$	1.0013691	$4.473 \times 10^{-4}$	3.058	$2.225 \times 10^{-3}$
Number of operating trains per hour during peak hours	$4.40979 \times 10^{-2}$	1.0450847	$2.49337 \times 10^{-2}$	1.769	$7.6959 \times 10^{-2}$
Number of trains according to type	-0.3882041	0.6782739	0.1082648	-3.586	$3.36 \times 10^{-4}$
Average number of tracks	-1.4782268	0.2280417	0.7316531	-2.020	$4.3343 \times 10^{-2}$
Number of connecting operation stations	$-5.01086 \times 10^{-2}$	0.9511261	$2.45271 \times 10^{-2}$	-2.043	$4.1053 \times 10^{-2}$

according to type were largely negative. The high regression coefficients for the length of train lines and the length of through service sections show that the delay times increase in proportion with the length of the train operation zones. Additionally, the large negative regression coefficient for the number of trains according to type suggests the delay time is especially short for commuter lines where many types of trains run to the suburbs during commuting hours. From the above, delay times during the declaration of the state of emergency were long for lines with long operation zones, while it was not so long for commuter lines traveling towards the suburbs.

On the other hand, in the logistic regression analysis for Period 2, the regression coefficients for the average number of tracks and the number of trains according to type were largely negative. This suggests that quadruple tracks were accordingly developed, indicating that the delay probability especially on commuter lines with many types of trains traveling to the suburbs was low. Additionally, the regression coefficient for the number of through service train lines

is comparatively high. This is because the delay probability is low for older train lines that travel through the downtown area among the train lines without through services. From the above, the delay probability during the declaration of the state of emergency was low for commuter lines going towards the suburbs and lines going through downtown areas.

## **6. Evaluation and Discussion**

### **6.1. Results for before the Spread of COVID-19 (June 2018)**

When comparing the results of the standard multiple regression analysis and the logistic regression analysis for June 2018, the common finding was that train delays were frequent in commuter lines traveling to the suburbs. Additionally, the number of through service train lines was selected as a variable for both analyses, and the fact that their regression coefficients were high suggests that the high number of through service train lines significantly impacted both the delay probability as well as the delay times. The regression coefficient for the length of lines with connecting operation was high in the standard multiple regression analysis, while it was relatively low in the logistic regression analysis, and the number of connecting operation stations was not selected. In contrast, the regression coefficient for the number of stairs and escalators in terminal stations was high in the logistic regression analysis. Therefore, while the delay probability had the most impact from large-scale stations, delay times are also affected by the connectivity of entire train lines to railway networks.

From the above, before the spread of COVID-19, the high number of through service train lines affected both the delay probability and the delay time. The delay probability was affected by the scale of the largest stations, while the delay time was affected by the connectivity of the entire train lines to other railway networks.

### **6.2. Results for after the Spread of COVID-19 (2020)**

#### **6.2.1. Results for the Standard Multiple Regression Analysis**

The results of the standard multiple regression analysis for the whole year of 2020 are compared. First, the number of selected variables of the analyses for Periods 2 and 6 were low compared with other periods. This is because the decrease in the number of passengers during these 2 periods reduced the contributing factors for train delays. Additionally, the selected variables gradually increased from Period 3 to Period 5, with Period 5 having the same number of variables as Period 1. This indicates that the contributing factors for train delays increased to the condition similar to before the spread of COVID-19 due to the implementation of the Go To Travel Campaign where people started to return to the streets feeling less hesitant to go out. Additionally, the regression coefficients for the length of train lines and the length of through service sections for Period 6 were high as with Period 2. This shows that the length of train operation zones affected the delay times during Period 6 just as during the state of emergency.

From the above, while the contributing factors for train delays had begun to return to their original state before COVID-19 with people feeling less hesitant to go out due to the Go To Travel Campaign implemented after the state of emergency, the decision to temporarily discontinue the Go To Travel Campaign caused the situation to return to how it was during the state of emergency. This suggests that the decision to temporarily discontinue the Go To Travel Campaign affected the use of trains almost as much as the first state of emergency did.

### **6.2.2. Results for the Logistic Multiple Regression Analysis**

The results of the logistic regression analysis for the whole year of 2020 are compared. First, the regression coefficient for the average number of tracks was positive during Period 1 but turned negative during Periods 2 and 6. From this, while commuter lines with quadruple tracks traveling to the suburbs were a contributing factor to the increase in the delay probability before the progression in the spread of COVID-19, they had become contributing factors to decrease the delay probability due to the state of emergency and the discontinuation of the Go To Travel Campaign.

Additionally, the regression coefficient for the number of trains according to type was largely negative during Period 2 but gradually moved towards zero from Period 3 to Period 5. This is because the delay probability increased as people were feeling less hesitant to go out with the implementation of the Go To Travel Campaign, and the use of commuter lines heading towards the suburbs gradually increased. However, the regression coefficient for the number of trains according to type in Period 6 was more largely negative than that of Period 2. This demonstrates that the delay probability had decreased due to the discontinuation of the Go To Travel Campaign which again led to a decrease in train use.

From the above, while the passengers of commuter lines traveling to the suburbs have increased to the point where the delay probability also increased, the passengers decreased to the point where delay probability also decreased due to the spread of COVID-19. Additionally, the decrease in the delay probability for commuter lines running towards the suburbs especially in Periods 2 and 6 show that the discontinuation of the Go To Travel Campaign had the same effect as the state of emergency on the use and the delay of such lines.

### **6.2.3. Comparison of the Results for the Standard Multiple Regression Analysis and the Logistic Regression Analysis**

The results of the standard multiple regression analysis and logistic regression analysis for the whole year of 2020 are compared. First, in contrast with the large decrease in the number of selected variables in the standard multiple regression analysis for Periods 2 and 6, the number of selected variables in the logistic regression analysis had no major changes throughout the year. This indicates that there are many contributing factors that affected the delay probability but the spread of COVID-19 did not cause any major changes. However, as the number

of passengers affected the length of delay times, it was evident that the contributing factors that increased delay times were reduced, when the difference between train lines were lessened due to the spread of COVID-19.

Additionally, while the regression coefficients for variables representing the length of operation zones such as the length of train lines and the length of through service sections increased in the standard multiple regression analysis, the regression coefficients for such variables hardly increased or were not selected in the logistic regression analysis. Therefore, it can be presumed that the length of train operation zones does not greatly impact the delay probability but greatly affects the length of delay times. From the above, though the contributing factors for train delays changed to the ones representing operation zones as a result of the decrease in passengers due to the spread of COVID-19, it did not greatly affect the delay probability.

### **6.3. Comparison of the Results for before and after the Spread of COVID-19**

The results of June 2018 (before COVID-19) and each period of 2020 (after the spread of COVID-19) are compared. First, when comparing Period 1 with June 2018, the trend of selected variables and the regression coefficients are similar in both the standard multiple regression analysis and the logistic regression analysis. Therefore, before the declaration of the state of emergency with the progression of the spread, there was no impact large enough to change the contributing factors for train delays from before the spread of COVID-19.

Next, Period 2 is compared with June 2018. In the standard multiple regression analysis, the number of selected variables was significantly reduced and the contributing factors for train delays also decreased. Additionally, with the logistic regression analysis, the regression coefficient for the average number of tracks was positive in June 2018 but was negative in Period 2. Therefore, it is evident that train delays had decreased in commuter lines with quadruple tracks traveling to the suburbs after the spread of COVID-19.

When comparing the results of the standard multiple regression analysis for Period 3 with June 2018, the selected variables are slightly different but the trend of the regression coefficients was relatively similar. This suggests that the lifting of the state of emergency caused the contributing factors for train delays to become similar to that of before the spread of COVID-19. Additionally, the selected variables are different from the results of June 2018 in the logistic regression analysis. In other words, while the variables representing transportation scale per hour and connectivity of railway networks were selected for June 2018, the variables indicating the length of operation zones and concentration towards major stations were selected for Period 3. Hence, in this period, train delays caused by the transportation scale were reduced with contributing factors of the length of train lines and major stations being the cause of most delays.

When comparing the results of the standard multiple regression analysis for Period 4 with June 2018, the trend of regression coefficients was relatively close,

while the selected variables were slightly different as with the case in Period 3. However, the regression coefficient for the number of stairs and escalators in terminal stations was positive in June 2018 but was negative in Period 4. Though this variable was not selected in Period 3, train delays in lines with large terminal stations had reduced due to the spread of COVID-19. Additionally, regarding the logistic regression analysis, while the selected variables were slightly different, the results were similar to that of June 2018 and the train delay occurrences were similar to the condition before COVID-19. However, the number of stairs and escalators in terminal stations was not selected, and train delays caused by large terminal stations had decreased.

When comparing the results of the standard multiple regression analysis for Period 5 with June 2018, the regression coefficient for the number of stairs and escalators in terminal stations was negative as with the case in Period 4. This indicated that the decrease of train delays in large terminal stations had continued. Additionally, the number of variables selected in Period 5 had become comparatively high, and the contributing factors for train delays had increased as the spread of COVID-19 slowed down. In the logistic regression analysis, just as in Period 4, the variable for the number of stairs and escalators in terminal stations was not selected, and the train delays caused by large terminal stations had decreased. Moreover, the regression coefficient for the number of through service train lines had decreased compared with June 2018, and train delays in lines with through services to many other train lines had also decreased compared with before COVID-19.

When comparing the results of the standard multiple regression analysis for Period 6 with June 2018, as mentioned in Section 5.2, the number of selected variables had decreased just as in Period 2, and the contributing factors for train delays had also decreased due to the temporary discontinuation of the Go To Travel Campaign. Additionally, the regression coefficient for the length of train lines was especially high in Period 6, and most of the train delays were proportional to the length of operation zones during this period. In the logistic regression analysis, the regression coefficient for the average number of tracks turned from positive to negative as in Period 2, and the train delay occurrences in commuter lines traveling to the suburbs decreased due to the temporary discontinuation of the Go To Travel Campaign. Furthermore, while there were many occurrences and increase of train delays in commuter lines traveling to the suburbs before the spread of COVID-19, as mentioned earlier, train delays decreased in such lines after the spread of the virus and this was especially significant in Periods 2 and 6. Therefore, it can be said that the contributing factors for train delays had changed due to the declaration of the state of emergency. Specifically, the contributing factors that previously increased train delays reduced those after the declaration of the state of emergency. Additionally, similar changes were seen due to the temporary discontinuation of the Go To Travel Campaign.

From the above, the number of train delays stopped increasing in the aforementioned commuter lines traveling to the suburbs, and the contributing factors have changed due to the state of emergency and the suspension of the Go To Travel Campaign. Moreover, the occurrence and increase of train delays caused by large terminal stations significantly decreased due to the spread of COVID-19.

## 7. Conclusions

The present study conducted 2 types of statistical analysis to reveal the impact of the spread of COVID-19 on train delays by comparing the potential contributing factors before, during, and after the outbreak of the virus within the metropolitan area in Japan. The results of the present study revealed the changes in contributing factors for train delays caused by the spread of COVID-19. Especially during the state of emergency, as shown in Section 5.2, the occurrence and increase of train delays in commuter lines going towards the suburbs decreased, and those became greatly affected by the length of the operating zones of trains. Additionally, after the decision to temporarily discontinue the Go To Travel Campaign, changes in contributing factors for train delays similar to that of the state of emergency were seen as described in Sections 5.1 and 5.2. This indicates that train delays were reduced in commuter lines traveling to the suburbs during the declaration of the state of emergency as well as after the temporary discontinuation of the Go To Travel Campaign.

Additionally, during the state of emergency and after the temporary discontinuation of the Go To Travel Campaign, train delays caused by factors that contributed to congestion in the past decreased, and the impact that was dependent on the length of operation zones had become relatively significant. This suggested that the contributing factors for train delays changed due to the decrease of passengers by the effect of the outbreak of COVID-19. In particular, the state of emergency and the temporary discontinuation of the Go To Travel Campaign had an extremely significant effect on the changes in contributing factors for train delays. The number of passengers had decreased during this period, and there is a possibility that contributing factors for train delays changed when passengers in the Tokyo metropolitan area decreased. Moreover, as mentioned in Section 5.3, though large terminal stations were considered to be a major contributing factor in causing and increasing train delays in the past, this was not the case after the spread of COVID-19. Therefore, under such conditions, it is more effective to make improvements in small to medium stations and tracks rather than terminal stations in order to improve train delays.

Furthermore, as the decrease in passengers also decreased train delays in commuter lines going to the suburbs due to the spread of COVID-19, the contributing factor for such lines is the excessive number of passengers. This suggests that train delays can be reduced by the distribution of passengers. As for countermeasures for train delays after the effects of COVID-19, it is necessary to disperse passengers in order to avoid passengers concentrating in the same time

zones and train lines. From the above, regarding future measures for train delays as the social situation has changed due to the impact of COVID-19, the improvement on overall equipment and the prevention of extreme congestion are more effective than intensive investments in one location.

The spread of COVID-19 continues even as the present study is advanced, and there may be further changes in contributing factors for train delays depending on future developments and trends. Therefore, train delay trends must continue to be closely observed until the pandemic comes to a complete end. Additionally, the decrease in passengers and changes in the use of trains due to COVID-19 have also occurred in other areas outside of the Tokyo metropolitan area. Accordingly, before, during, and after the outbreak of the virus, contributing factors for train delays peculiar to the Tokyo metropolitan area can be identified by comparing with other areas.

### Acknowledgements

We gratefully acknowledge the work of past and present members of our laboratory. We wish to thank the anonymous reviewers for comments on the 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 (2019) Starting the Visualization of Train Delay. <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.tokyo-metro.jp/corporate/enterprise/passenger\\_rail/transportation/passengers/2013.html](https://www.tokyo-metro.jp/corporate/enterprise/passenger_rail/transportation/passengers/2013.html)
- [3] East Japan Railway Company (2021) Changes in Year-on-Year Rate of Revenue of Railway Service in Fiscal Year 2020. <https://www.jreast.co.jp/investor/monthly/pdf/report.pdf>
- [4] 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.
- [5] 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*, WIT Press Publishing, Southampton, 85-93.
- [6] Dollevoet, T., Huisman, D., Schmidt, M. and Schöbel, A. (2011) Delay Management with ReRouting of Passengers. *Transportation Science*, **46**, 74-89. <https://doi.org/10.1287/trsc.1110.0375>
- [7] 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>

- [8] 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.
- [9] 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**, I\_1067-I\_1074. [https://doi.org/10.2208/jscejipm.72.I\\_1067](https://doi.org/10.2208/jscejipm.72.I_1067)
- [10] Kobayashi, W. and Iwakura, S. (2019) Agent-Based Model for Assessing Techniques to Reduce Knock-on Urban Train Delay. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **75**, 273-288. <https://doi.org/10.2208/jscejipm.75.273>
- [11] 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>
- [12] Corman, T. (2020) Interactions and Equilibrium between Rescheduling Train Traffic and Routing Passengers in Microscopic Delay Management: A Game Theoretical Study. *Transportation Science*, **54**, 785-822. <https://doi.org/10.1287/trsc.2020.0979>
- [13] König, E. and Schön, C. (2021) Railway Delay Management with Passenger Rerouting Considering Train Capacity Constraints. *European Journal of Operational Research*, **288**, 450-465. <https://doi.org/10.1016/j.ejor.2020.05.055>
- [14] 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>
- [15] 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>
- [16] 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>
- [17] Robenek, T., Maknoon, Y., Azadeh, S.S., Chen, J. and Bierlaire, 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>
- [18] 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>
- [19] Xu, W., Zhao, P. and Ning, L. (2018) Last Train Delay Management in Urban Rail Transit Network: Bi-Objective MIP Model and Genetic Algorithm. *KSCSE Journal of Civil Engineering*, **22**, 1436-1445. <https://doi.org/10.1007/s12205-017-1786-0>
- [20] Yap, M. and Cats, O. (2020) Predicting Disruptions and Their Passenger Delay Impacts for Public Transport Stops. *Transportation*, **48**, 1703-1731. <https://doi.org/10.1007/s11116-020-10109-9>
- [21] 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>
- [22] 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**, 67\_I\_1001-67\_I\_1010.  
[https://doi.org/10.2208/jscejipm.67.67\\_I\\_1001](https://doi.org/10.2208/jscejipm.67.67_I_1001)
- [23] 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.
- [24] 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>
- [25] 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/jscejipm.70.44>
- [26] Yamamura, A. and Tomii, N. (2019) Train Traffic Simulation based on Analysis of Historical Train Traffic Records for Dense Railway Networks. *IEEJ Transactions on Industry Applications*, **139**, 206-214. <https://doi.org/10.1541/ieejias.139.206>
- [27] 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**, I\_477-I\_486.  
[https://doi.org/10.2208/jscejipm.70.I\\_477](https://doi.org/10.2208/jscejipm.70.I_477)
- [28] Kobayashi, W., Fukuda, W. and Iwakura, S. (2020) Economic Evaluation for Delay and Punctuality of Urban Rail Transit Based on Scheduling Approach. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, **76**, 236-250. [https://doi.org/10.2208/jscejipm.76.3\\_236](https://doi.org/10.2208/jscejipm.76.3_236)
- [29] 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>
- [30] 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>
- [31] 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>
- [32] 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, August 29-September 1 2010, 14 p.
- [33] Cule, B., Goethals, B., Tassenoy, S. and Verboven, S. (2011) Mining Train Delays. *International Symposium on Intelligent Data Analysis*, Porto, 29-31 October 2011, 113-124. [https://doi.org/10.1007/978-3-642-24800-9\\_13](https://doi.org/10.1007/978-3-642-24800-9_13)
- [34] Liu, X., Saat, M.R. and Barkan, C.P.L. (2012) Analysis of Causes of Major Train Derailment and Their Effect on Accident Rates. *Transportation Research Record*, **2289**, 154-163. <https://doi.org/10.3141/2289-20>
- [35] Bergström, A. and Krüger, A. (2013) Modeling Passenger Train Delay Distributions: Evidence and Implications. *Proceedings of the 5th International Symposium on Transportation Network Reliability*, Hong Kong, 18-19 December 2012, 31 p.
- [36] 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>

- [37] 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>
- [38] 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.
- [39] Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N. and Anguita, D. (2018) Train Delay Prediction Systems: A Big Data Analytics Perspective. *Big Data Research*, **11**, 54-64. <https://doi.org/10.1016/j.bdr.2017.05.002>
- [40] Arshad, M. and Ahmed, M. (2019) Prediction of Train Delay in Indian Railways through Machine Learning Techniques. *International Journal of Computer Sciences and Engineering*, **7**, 405-411. <https://doi.org/10.26438/ijcse/v7i2.405411>
- [41] Wang, P. and Zhang, Q. (2019) Train Delay Analysis and Prediction Based on Big Data Fusion. *Transportation Safety and Environment*, **1**, 79-88. <https://doi.org/10.1093/tse/tdy001>
- [42] Huang, P., Wen, C., Fu, L., Peng, Q. and Tang, Y. (2020) A Deep Learning Approach for Multi-Attribute Data: A Study of Train Delay Prediction in Railway Systems. *Information Science*, **516**, 234-253. <https://doi.org/10.1016/j.ins.2019.12.053>
- [43] Huang, P., Li, Z., Wen, C., Lessan, J., Corman, F. and Fu, L. (2021) Modeling Train Timetables as Images: A Cost-Sensitive Deep Learning Framework for Delay Propagation Pattern Recognition. *Expert Systems with Applications*, **177**, Article ID: 114996. <https://doi.org/10.1016/j.eswa.2021.114996>
- [44] Huang, P., Lessan, J., Wen, C., Peng, Q., Fu, L., Li, L. and Xu, X. (2020) A Bayesian Network Model to Predict the Effects of Interruptions on Train Operations. *Transportation Research Part C: Emerging Technologies*, **114**, 338-358. <https://doi.org/10.1016/j.trc.2020.02.021>
- [45] Mohd, A. and Muqem, A. (2021) Train Delay Estimation in Indian Railways by Including Weather Factors through Machine Learning Techniques. *Recent Advances in Computer Science and Communications*, **14**, 1300-1307. <https://doi.org/10.2174/2666255813666190912095739>
- [46] 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>
- [47] Kotsushibunsha (2018) My Line Tokyo Timetable (June, 2018).
- [48] Kotsushibunsha (2020) My Line Tokyo Timetable (June, 2020).
- [49] Ministry of Land, Infrastructure, Transport and Tourism (2017) Statistics Information Related to Congestion Rate Data. <https://www.mlit.go.jp/common/001245351.pdf>
- [50] Ministry of Land, Infrastructure, Transport and Tourism (2019) Statistics Information Related to Congestion Rate Data. <https://www.mlit.go.jp/statistics/details/content/001365148.pdf>