Identification of Dates on which Spikes in Route Searches Occurred and Discussion of the Factors

Taku MORIYAMA [a*], Masashi KUWANO [b], Mio HOSOE [c]

[a,b,c] *Dept. of Management of Social Systems and Civil Eng., Tottori University, Tottori, 680-8552, Japan*
[a] *E-mail: moriyama@tottori-u.ac.jp*
[b] *E-mail: kuwano@tottori-u.ac.jp*
[c] *E-mail: d19t4003b@edu.tottori-u.ac.jp*

Abstract: Transportation operators are required to respond to daily and spikes in demand. To estimate spikes in transportation demand, travel preference information such as trip planner log data can be an effective alternative to actual travel data (e.g., those obtained from smart cards). Here, we elucidate the relationship between spikes in the number of searches and transportation demand to clarify the correlation between search behavior and transportation demand. We applied a state-space model to operator-version trip planner historical data from October 1, 2018, to September 31, 2019, and extracted the days whereon spikes in searches occurred from the error component excluding level and periodic components. Upon examining the factors that affected each spike day, we determined some factors responsible for the spikes in transportation demand. Thus, the occurrence of spikes in demand could be investigated from the search history in the trip planner.

Keywords: trip planner, log data, time series analysis, irregular fluctuations, state space model

1. INTRODUCTION

Transportation operators are required to respond to spikes in the number of users. Congestion on buses owing to spikes in total users reduces the levels of user satisfaction. Furthermore, if passengers are left behind owing to situations such as overcapacity, the operator misses a direct opportunity to profit. Therefore, in transportation planning, it is important to elucidate the mechanism of spikes in users. Possible causes of sudden fluctuations include weather conditions (Khattak and De Palma, 1997; Cools et al., 2010), holidays, and events such as the New Year Holiday period (Pereira et al., 2015). Ohler et al. (2017) developed a neural network model for predicting total bus passengers considering holidays and weather conditions. However, factors that predominantly cause spikes in the number of users are subtle, even at later dates.

Recently, advancement and spread of ICT is remarkable, and big data on detailed traffic behavior can be continuously obtained from sources such as smart cards, mobile phone traces, detailed vehicle location data, and social media. There are many empirical studies on use of smart card data, and Pelletier et al. (2011) gave a review on the studies and summarized the advantages and disadvantages. Calabrese et al. (2013) provided a methodology for extracting mobility information from mobile phone trace data and conducted a case study in the Boston Metropolitan Area. Tang and Thakuriah (2012)

* Corresponding author.

analyzed the impact of the introduction of the Chicago transit authority's real-time transit information system that was made possible by the widespread use of automatic vehicle location (AVL) technology by using a linear mixed-effects model. Considering intelligent transportation systems that have attracted significant attention as an example of the use of big data in the transportation sector, Iliopoulou and Kepaptoglou (2019) reviewed and discussed applications such as strategic level planning, transit assignment, origin-destination (O-D), and transfer inference. Social media data is used for analyzing human movements and trip purposes (Rashidi et al., 2017). Location-based social network (LBSN) data for transportation planning has recently attracted interest (Laman et al., 2019). Yang et al. (2020) used LBSN data obtained in the city of Nanjing in China for trip generation modeling and suggested the relevance between human activity and land use.

Zannat and Choudhury (2019) analyzed advantages and disadvantages of smart card data, mobile data, AVL data, and other forms of transportation big data that can be used for public transport planning, and proposed some tasks for investigation. General Transit Feed Specification (GTFS), a standard bus information format created by the Tri-County Metropolitan Transportation District of Oregon (Tri-Met) in partnership with Google, is widely known and used for combining several data (Rodnyansky, 2018). A dataset containing GTFS for 25 cities is available in a researcher-friendly format by Kujala et al. (2018). In Japan, the proprietary format GTFS-JP developed by the Ministry of Land, Infrastructure, Transport and Tourism in March 2017 is widely promoted.

Martin et al. (2005) analyzed the trip planner history data on the Montreal Transit Commission and the Laval Transit Commission websites and observed a strong correlation between the number of web users and the actual number of transit users for each O-D. However, a problem with trip planner log data is that some transportation users do not use the search, and conversely, not all searchers use transportation, which may lead to bias. They indicated that web users can be tourists, non-resident business travelers, and other users that are not fully identified by traditional O-D surveys.

Borole et al. (2013) discussed the implementation of a dynamic trip planner (Shaller, 2002) based on real-time data from a GPS-equipped vehicle tracking system. A trip planner is a web system that creates a transit itinerary based on an origin and destination. This dynamic trip planner predicts bus delays and traffic congestion to enable users select a rapid and significantly convenient method to travel. Herein, transportation information is shared beyond the boundaries of operators, particularly in large cities in developed countries, and trip planners that consider routes and real-time congestion information by multiple operators are available. However, in Japan and several countries, often only limited and static trip planners provided by each transport operator are available as means to search bus routes. Sierpiński et al. (2016) devised a new trip planner that considers users' preferences and conducted a case study for a part of Silesian Voivodeship. The study explained that the devised trip planner can be user-friendly and make the transport system sustainable.

Roosmalen (2019) developed a prediction model for the number of passengers on the basis of trip planner log data and by incorporating machine learning and discussed its accuracy. However, the relationship between the number of searches as a proxy variable for transportation demand and the actual number of transportation users is not completely clear.

From Martin et al. (2005), because trip planners are used for unplanned trips differing from daily trips such as commuting to work and school, their log data may be useful for detecting spikes in demand. In addition, if a spike in demand for transportation occurs and the number of people aiming to use the service is smaller than capacity, most users of the planner will be considered as passengers. In contrast, if the number of people aiming to use the service unexpectedly exceeds capacity, some passengers will not be included in the data. The total users that were denied boarding because capacity was exceeded

and those who used other modes of transportation or cancelled their trip owing to congestion are not included in the actual number of transportation users.

This study aims to contribute the convenience and the sustainability of transportation system through establishing a methodology for analyzing the factors of spikes in demand. We focus on the log data of a trip planner to clarify the mechanism. If we know the occurrence of the spikes in advance, we can increase in the number of buses and propose alternative means of transportation. We propose a method for extracting the days whereon the number of searches suddenly increases from the search log data and infer the actual transportation demand on those days. In this study, the number of instances of trip planner usage is considered as time series data, and the error component obtained from the state-space model is used to extract days of spike in demand. Several studies have applied state-space models to traffic behavior analysis (Chang and Wu, 1994; Montero et al., 2015; Rrecaj and Todorova, 2018). Although this method is not novel, to the best of our knowledge, there are no cases of model application to search log data. We consider the factors that caused the spike in the number of searches on the days extracted by the analysis and infer if a spike in demand actually occurred on that day.

2. OVERVIEW OF DATA USED AND BASIC AGGREGATION

2.1 Overview of Data Used

In this study, we analyze the total trip planner searches and the total bus users for Shinki Bus that operates mainly in the southern sector of Hyogo Prefecture, Japan. The Shinki Bus Group operates

Figure 1. Target area for analysis

Table 1. Functions of Shinki Bus NAVI and their descriptions

Function	Details
Route finding	Displays route, time, and fare
Timetable display	Displays the timetable for a bus stop
Boarding point map display	Displays a map showing the location of bus stops
Bus position display	Displays the current position of the bus
Route display	Displays the bus stops for buses on the specified route on a map

Figure 2. Condition input screen and route guidance screen of Shinki Bus NAVI (in Japanese)

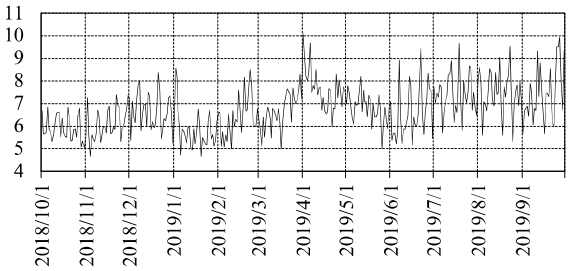

Figure 3. Number of searches (thousands)

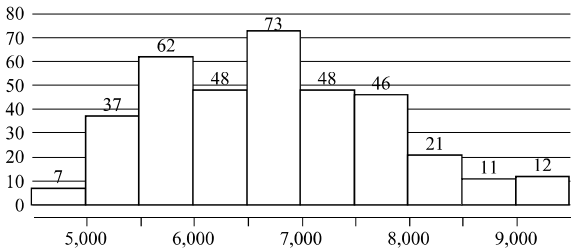

Figure 4. Frequency distribution of the number of searches

route buses primarily in the southern part of Hyogo Prefecture, and also provides express bus service from Kobe City and Himeji City to the Shikoku, Sanyo, and Sanin regions (see Hyodo et al., 2005 for the transportation usage rate in East Asian countries). In this study, we analyze the routes in the northern Himeji sector (Figure 1) from the log data of the operator version of the trip planner "Shinki Bus NAVI" for a one-year period—from October 1, 2018 to September 30, 2019. Figure 1 was processed and created using digital national land information (bus route and bus stop data).

From Table 1, Shinki Bus NAVI has five functions such as route search and timetable display. There are two ways to use the system: through a browser on a PC or smartphone and through a smartphone application. Shinki Bus NAVI is provided only in Japanese. Figure 2 shows the condition input screen and the route guidance screen if used via a PC browser. The name of the bus stop on the Shinki Bus route, the date and time of scheduled departure or arrival, and the method of specifying the time can be recorded. The log data of Shinki Bus NAVI include information on the date and time of system use, bus stop names, function used, and device used (e.g., Android, iPhone, PC, feature phone).

2.2 Basic Aggregation

Figure 3 shows the total number of uses of Shinki Bus NAVI in the northern Himeji sector over the target period (hereinafter referred to simply as the “number of searches”). There was a total of 2,463,749 uses in the year. There are approximately 6,750 searches daily with the highest number of searches recorded on Monday, April 1, 2019 (10,130) and the lowest on Sunday, January 20, 2019 (4,646). Figure 4 shows the frequency distribution of the number of searches. The average number of searches on weekdays (excluding holidays) is approximately 6,856, and the average number of the searches on Saturdays, Sundays, and holidays (hereinafter referred to as “holidays”) is approximately

6,535. The standard deviations are approximately 1,051 and 1,073.

Considering the combinations of origins and destinations in the northern Himeji sector, the average number of searches in the descending order of frequency was 138/day for "Yashiro Eigyosho to Himeji Station North Exit (hereafter, Himeji Station)"; 97.9/day for "Himeji Station to Yashiro Eigyosho"; 75.6/day for "Himeji Station to University of Hyogo School of Engineering"; 74.1/day for "Himeji Station to Shoshazan Ropeway"; and 59.8/day for "University of Hyogo Faculty of Environment and Human Studies to Himeji Station".

3. OVERVIEW OF THE STATE-SPACE MODEL

If $t = 1, 2, \cdots, T$ is a point in time, x_t is a state value, and y_t is an observation value, the linear-Gaussian state-space model is represented by the state equation in Equation (1) and the observation equation in Equation (2)

$$x_t = F_t x_{t-1} + G_t v_t \tag{1}$$

$$y_t = H_t x_t + w_t. \tag{2}$$

However, $\{v_t\}$ and $\{w_t\}$ are independent of each other and independently follow a Gaussian distribution with a mean of 0. In the state-space model, x_t at each point in time is obtained by applying the Kalman filter algorithm (Anderson and Moore, 1979). The predicted values of the states in the linear and Gaussian state-space model given by Equations (1) and (2) are obtained by Equation (3). The variance in this case is given by Equation (4)

$$x_{t|t-1} = F_t x_{t-1|t-1} \tag{3}$$

$$V_{t|t-1} = F_t V_{t-1|t-1} {F_t}^T + G_t Q_t {G_t}^T. \tag{4}$$

Herein, Q_t is the variance-covariance matrix of v_t and $x_{j|k}$ denotes the conditional expectation and is defined as follows

$$x_{j|k} := \mathbb{E}\left[x_j \middle| y_1, y_2, \dots, y_k\right].$$

$V_{j|k}$ is defined as follows

$$V_{j|k} := \mathbb{E}\left[\left(x_j - x_{j|k}\right)\left(x_j - x_{j|k}\right)^T\right].$$

The filtering of x_t from the obtained observations and state predictions is given by Equation (5)

$$x_{t|t} = x_{t|t-1} + K_t\left(y_t - H_t x_{t|t-1}\right). \tag{5}$$

The Kalman gain, K_t, is given by Equation (6)

$$K_t := V_{t|t-1} {H_t}^T (H_t V_{t|t-1} {H_t}^T + \sigma^2)^{-1}. \tag{6}$$

The variance of the filtering is given by Equation (7)

$$V_{t|t} = (I - K_t H_t) V_{t|t-1}. \tag{7}$$

After all states are obtained by prediction and filtering, smoothing is performed to correct the state values again, while going back through the time points in sequence. This smoothing is given by Equation (8)

$$x_{t|T} = x_{t|t} + A_t \left(x_{t+1|T} - x_{t+1|t} \right), \tag{8}$$

where

$$A_t := V_{t|t} {F_{t+1}}^T V_{t+1|t}^{-1}. \tag{9}$$

and the variance of smoothing is given by Equation (10)

$$V_{t|T} = V_{t|t} + A_t \left(V_{t+1|T} - V_{t+1|t} \right) {A_t}^T. \tag{10}$$

In this study, we use a local level model with the observation equation expressed by Equation (11) wherein a periodic variation element (r_t) is added

$$y_t = p_t + r_t + w_t. \tag{11}$$

Let p_t denote the probabilistic level component defined as follows

$$p_t = p_{t-1} + v_t. \tag{12}$$

Let the random variable sequences $\{v_t\}$ and $\{w_t\}$ be mutually independent and each independently follow the following Gaussian distribution

$$v_t \sim N(0, \tau^2), \qquad w_t \sim N(0, \sigma^2).$$

Periodic component r_t is assumed to have a 7-day periodicity as shown below

$$\sum_{i=0}^{6} r_{t-i} = z_t. \tag{13}$$

However, $\{z_t\}$ is assumed to independently follow the following Gaussian distribution

$$z_t \sim N(0, \eta^2).$$

4. EXTRACTION OF DAYS WITH A SPIKE IN SEARCHES AND DISCUSSION OF FACTORS

In this section, we describe the results of applying the local level model described in the previous section to the log data of Shinki Bus NAVI. After observing the level component and the periodic component obtained by this application, we focus on the error component wherein the "other factors" are mixed (w_t; hereafter referred to as the "error component"). Thereafter, days whereon a spike in searches suddenly occurred are defined on the basis of this error component, and the days with a spike in searches are obtained for one year, i.e., the data period concerned. We investigate the reason for the spike in the number of searches on each day whereon a spike in searches occurred and infer if a spike in demand for transportation services actually occurred.

4.1 Extraction of Days with a Spike in Searches by State-Space Model

Figure 5 shows the logarithmically transformed number of searches in one year. The level component, period component, and error component obtained by applying the state-space model to the search data are shown in Figures 6, 7, and 8, respectively.

Considering the level component in Figure 6, there is a considerable increase over the studied period. Upon examination on the basis of a detailed time period, the level component was found to be high from early December 2018 to early January 2019, from mid to late February 2019, and from late March to early April 2019. In addition, an increasing trend was observed from mid-September 2019 that can be due to users seeking information on the tax increase and fare hike in October 2019.

The periodic component shown in Figure 7 is aggregated for each day of the week, and the obtained average values are presented in Table 2. Table 2 presents that the periodic component increases from Monday to Friday and reaches its maximum value on Friday. The periodic component decreases from Friday to Sunday, reaching its minimum value on Sunday. On weekdays, it is expected that some number of people use the service for business.

Subsequently, the error component that cannot be explained by the level and periodic components (Figure 8) is used to identify dates whereon spikes in searches occurred and to investigate the factors.

4.2 Consideration Factors Accounting for Spikes in Searches

In this study, the day whereon the standardized error component exceeds the upper 95th percentile of the t-distribution with $364 = (365 - 1)$ degrees of freedom is defined as a day whereon a spike in searches occurred (hereafter, "spike day"). From the analysis, 20 days in the year were determined to be spike days.

The dates determined to be spike days are presented in the first column of Table 3. The second and third columns of the table presents the number of searches on the day and the ranking of the number of searches for the year. The fourth and fifth columns show the difference between the number of searches on the previous day (hereinafter referred to as "difference with the previous day") and the ranking of the difference with the previous day for the year. Considering the number of searches and the ranking, the number of searches on the days considered to be a spike day is not necessarily high. In addition, considering the difference with the previous day and the ranking, the number of searches did

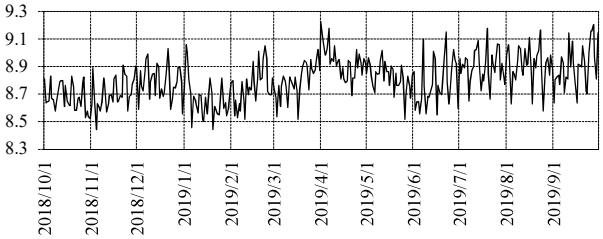

Figure 5. Logarithmically transformed number of searches

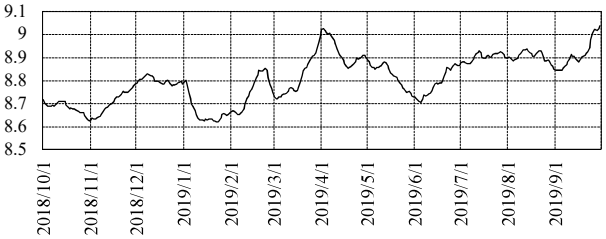

Figure 6. Level component

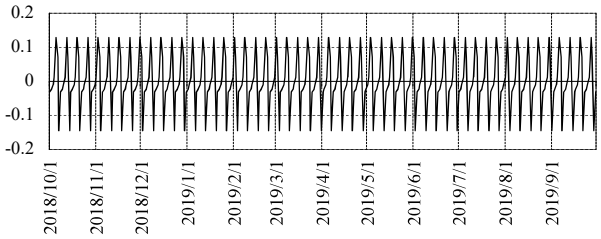

Figure 7. Periodic component

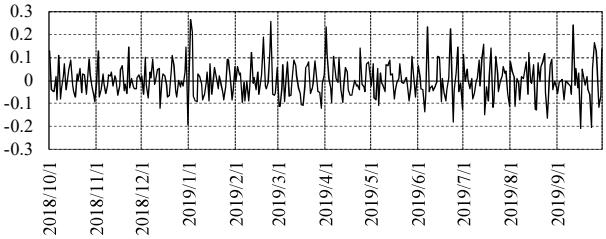

Figure 8. Error component

Table 2. Average values of the periodic component

Mon	Tue	Wed	Thu	Fri	Sat	Sun
−3.12	−2.37	−0.957	1.16	12.7	7.17	−14.6

not necessarily increase rapidly from the previous day on the days judged to be a spike day. This may be due to the fact that the day of the week that is a periodic component, has a strong influence on the number of searches. That is, by applying the state-space model, dates that are not evident from the number of searches or their increments are extracted as spike days. Accordingly, it is important to determine if a spike in demand actually occurred on the extracted spike days to determine the validity of the analysis results.

Therefore, we selected events in the suburbs of Hyogo Prefecture that may be related to the spike in the number of searches from websites and newspaper articles. The last column of Table 3 presents the events that may have triggered the search. From Table 3, poor weather conditions occurred on some of the days determined to be spike days. On Monday, October 1, 2018, a typhoon passed through the area, and on Friday, June 7, 2019, there was such heavy rainfall that evacuation advisories were issued in Hiroshima Prefecture and other areas. Herein, two major reasons can be given for the increase in the number of searches associated with poor weather conditions. First, the number of users who checked the current bus location increased because of the fear of schedule disruptions. This may have led to an increase in the number of searches although not in transportation demand. Second, the users who wanted to use the bus as an alternative means of transportation. In this case, demand for transportation can increase with the number of searches. In Japan, buses are often used as an alternative mode of transport by motorcyclists, cyclists, or walkers in bad weather (Ohmori et al., 2011). Noguchi (1999) analyzed the effect of weather on the use of buses as a means of transportation to and from railway stations, and observed that residents living in 1.5 km of a station select buses in bad weather. In Hamamatsu City, an additional "Rainy Bus" is operated in rainy weather to accommodate the rapid increase in the number of high school students and other users (Ohmori et al., 2011). In the northern

Table 3. Number of searches on spike days and possible wide-area causes

(MM/DD/YYYY)	Number of searches	Rank	Difference with the previous day	Rank	Possible cause
10/01/2018	6,709	178.5	–	–	Impact of Typhoon No. 24
11/02/2018	7,286	111	1,702	15	–
11/22/2018	7,396	101	1,535	21	Rainfall
12/30/2018	6,600	199	-735	286	End-of-year
01/02/2019	8,583	19	2,325	6	Start-of-year
01/03/2019	8,164	38	-419	250	Start-of-year
02/19/2019	8,187	37	1,422	30	Rainfall
02/23/2019	8,512	23	730	91	–
02/24/2019	7,757	61	-755	288	–
04/01/2019	10,130	1	2,990	2	Start of academic year
04/24/2019	8,304	31	1,539	20	Rainfall
06/07/2019	8,945	13	3,165	1	Record-breaking rainfall
06/22/2019	9,436	8	1,534	22	Rainfall
06/27/2019	8,341	28	1,107	46	Day before the summit
07/14/2019	7,581	77	-1,316	328	Rainfall
07/19/2019	9,676	4	2,120	9	Rainfall
09/11/2019	9,364	10	2,664	4	Record-breaking rainfall
09/25/2019	9,531	6	1,387	31	Day before the World Cup
09/26/2019	9,524	7	-7	183	Day of the World Cup
09/30/2019	9,365	9	2,666	3	Day of the World Cup

Himeji sector, a rapidly increasing demand for transportation is considered to be attributable for the rapid increase in the number of searches owing to poor weather conditions.

In addition, a spike in searches occurred in the case of major events such as the G20 Summit and the Rugby World Cup and during the year-end and New Year holidays. Given that Osaka City, where the G20 Summit was held, was subject to large-scale traffic restrictions and that traffic congestion was expected congestion was expected in Hyogo Prefecture, which is included in the commuting area of Osaka City, it can be inferred that demand for bus transportation increased. The Rugby World Cup was also held at the Misaki Park Stadium in Kobe City. It is thought that demand for transportation occurred because a lot of people traveled by bus and train to the venue to watch the game. December 31 (New Year's Eve) and January 2 (the day after New Year's Day) are widely recognized as special days for traffic congestion because of the increased demand due to shopping and shrine visits.

In contrast, factors attributable to weather conditions or events occurred on November 2, 2018, February 23, 2019, and February 24, 2019 could not be identified. In this study, we excluded O-D pairs with an average number of searches per day of less than one and those with overlapping origins and destinations, and examined the remaining 745 O-D pairs that showed abrupt changes on the days that were judged to be spike days.

Table 4 presents the O-D pairs for which the average number of searches increased by more than 100 on spike days. The corresponding increase was observed only during the six days presented in

Table 4. Possible causes of localized spikes in searches on spike days

(MM/DD/YYYY)	Origin	Destination	Number of searches on the day	Average number of searches	Multiplying factor	Possible cause
11/02/2018	Himeji Station (North Exit)	University of Hyogo School of Engineering	461	75.6	6.10	University of Hyogo School of Engineering Festival
02/19/2019	Yashiro Eigyosho	Himeji Station (North Exit)	313	138	2.27	Rainfall
	Himeji Station (North Exit)	Imajuku (Himeji City)	385	42.9	8.98	
02/23/2019	Himeji Station (North Exit)	Imajuku (Himeji City)	218	42.9	5.08	The 104th National Examination for Pharmacists
	Himeji Station (North Exit)	Himeji Dokkyo University	178	48.1	3.70	
02/24/2019	Himeji Station (North Exit)	Imajuku (Himeji City)	162	42.9	3.78	The day before University of Hyogo first semester entrance examination
	Himeji Station (North Exit)	School of Human Science and Environment, University of Hyogo	254	43.1	5.89	
	Himeji Station (North Exit)	School of Engineering	369	75.6	4.88	
04/01/2019	Himeji Station (North Exit)	University of Hyogo School of Engineering	196	75.6	2.59	Start of academic year
07/14/2019	Yashiro Eigyosho	Himeji Station (North Exit)	284	138	2.06	Rainfall
	Himeji Station (North Exit)	School of Human Science and Environment, University of Hyogo	213	43.1	4.94	School of Human Science and Environment open campus
	Himeji Station (North Exit)	Yashiro Eigyosho	219	97.9	2.34	

Table 4, and the total number of O-D pairs was 12. From the table, most identified pairs went from Himeji Station to high schools or universities. Referring to their websites and other sources, for example, a festival was held at the University of Hyogo on November 2, 2018, and a first semester entrance examination was held on February 24, 2019, also at the University of Hyogo. It is possible that many of the passengers heading to these events were high school and university students who did not have their own cars, possibly resulting in a spike in demand for transportation by bus.

5. CONCLUSION

This study examined the relationship between spikes in the number of trip planner searches and transportation demand by applying a state-space model to the log data of Shinki Bus NAVI. A stochastic level component and a cyclic component according to the day-of-week were set in the equation of state, and the days on which spikes occurred were defined on the basis of the value of the remaining error component. From the analysis, 20 spike days were extracted from the target year. Considering the factors of the spike in the total searches, we can determine plausible causes for all 20 days. We observed that sudden transportation demand occurs on spike days extracted by applying the state-space model, indicating a co-occurrence between spikes in searches and transportation demand. We did not extract all days whereon spikes in transportation demand occurred. In future research, it is necessary to obtain boarding and alighting history data for Shinki Bus users, extract all the days whereon spikes in demand occur to successfully determine if the number of searches suddenly increased on those days.

From this study, in the northern Himeji sector, large-scale events, poor weather conditions, and events at high schools and universities in the area can be factors for spikes in demand. Therefore, discovering factors that cause spikes in demand by applying a state-space model to search history data is useful for subsequent transportation planning. There were some local spikes in transportation demand owing to small-scale events and other factors; however, the total volumes in the entire northern Himeji sector were estimated not to be extreme. If the proposed state-space model is applied to the data of each route individually, we can estimate if spikes in demand occurred or not for each route. The clarification of the mechanism for the spikes in transportation demand at the scale of each route or each O-D pair can reveal important information for elucidating the mechanism of spikes in demand in an area.

Table 5 presents the number of O-D pairs that were searched for a multiple of five times or more than the average number of searches on each extracted day. This suggests that the transportation demand increases for several O-D pairs if the number of searches increases rapidly because of factors such as the Rugby World Cup, the beginning of the year, and poor weather conditions. Therefore, on the basis of the route, these events can cause a spike in transportation demand. In contrast, events that affect individual routes can include smaller and more localized events such as local festivals.

Therefore, because events of various scales can affect the demand of an individual route, adequate effort is required to elucidate the spikes in demand if the scope is narrowed. Therefore, it is important to analyze the spikes in transportation demand at various scales. Thus, comprehensive acquisition of event data of different scales and the establishment of an analysis method are necessary.

In addition, it is necessary to consider weather information and event conditions as exogenous variables to predict spikes in demand. On the basis of this, a model that incorporates exogenous variables can be developed.

Table 5. Number of usage pair on which demand suddenly increased on a spike day

(MM/DD/YYYY)	Number of pairs with spike in demand	(MM/DD/YYYY)	Number of pairs with spike in demand
10/01/2018	11	04/24/2019	19
11/02/2018	11	06/07/2019	14
11/22/2018	8	06/22/2019	15
12/30/2018	10	06/27/2019	9
01/02/2019	37	07/14/2019	8
01/03/2019	12	07/19/2019	24
02/19/2019	12	09/11/2019	31
02/23/2019	7	09/25/2019	21
02/24/2019	4	09/26/2019	30
04/01/2019	37	09/30/2019	10

ACKNOWLEDGEMENTS

We express our gratitude to Shinki Bus Co., Ltd. for providing the data for this study. This work was supported by *the Japan Society for the Promotion of Science, JSPS KAKENHI Grant Number JP20H02277*.

REFERENCES

Anderson, B.D., Moore, J.B. (1979) *Optimal Filtering*. Prentice-Hall, Englewood Cliffs.

Borole, N., Rout, D., Goel, N., Vedagiri, P., Mathew, T.V. (2013) Multimodal Public Transit Trip Planner with Real-time Transit Data. *Procedia-Social and Behavioral Sciences*, 104, 775-784.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti, C. (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301-313.

Chang, G.L., Wu, J. (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. *Transportation Research Part B: Methodological*, 28(2), 141-160.

Cools, M., Moons, E., Creemers, L., Wets, G. (2010) Changes in travel behavior in response to weather conditions. *Transportation Research Record*, 2157, 22-28.

Hyodo, T., Montalbo, C.M., Fujiwara, A., Soehodho, S. (2005) Urban travel behavior characteristics of 13 cities based on household interview survey data. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 23-38.

Iliopoulou, C., Kepaptsoglou, K. (2019) Combining ITS and optimization in public transportation planning: state of the art and future research paths. *European Transport Research Review*, 11, Article number:27.

Khattak, A.J., De Palma, A. (1997) The impact of adverse weather conditions on the propensity to change travel decisions: a survey of Brussels commuters. *Transportation Research Part A: Policy and Practice*, 31(3), 181-203.

Kujala, R., Weckström, C., Darst, R.K., Mladenović, M.N., Saramäki, J. (2018) A collection of public transport network data sets for 25 cities. *Scientific data*, 5, Article number:180089.

Laman, H., Yasmin, S., Eluru, N. (2019) Using location-based social network data for activity intensity analysis: A case study of New York City. *Journal of Transport and Land Use*, 12(1), 723-740.

Montero, L., Codina, E., Barceló, J. (2015) Dynamic od transit matrix estimation: formulation and model-building environment. In *Progress in Systems Engineering*. Advances in Intelligent Systems and Computing, 366, 347-353.

Noguchi, T. (1999) Modal choice of access trip to the railway station and users' evaluation: Comparison under the weather condition. *Papers on City Planning*, 34, 979-984.

Ohler F., Krempels K., Möbus S. (2017) Forecasting Public Transportation Capacity Utilisation Considering External Factors. In *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems*, 1, 300-311.

Ohmori, N., Harata, N., Sugitani, Y., Kunimoto, N. (2011) Travel behavior of university students on rainy days: A preliminary study on international comparison between east asian countries. *Journal of the Eastern Asia Society for Transportation Studies*. 9, 483-498.

Pelletier, M.-P., Trépanier, M., Morency, C. (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies,* 19(4), 557-568.

Pereira, F.C., Rodrigues, F., Ben-Akiva, M. (2015) Using data from the web to predict public transport arrivals under special events scenarios. *Journal of Intelligent Transportation Systems*, 19(3), 273-288.

Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, T.S. (2017) Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, 197-211.

Rodnyansky, S. (2018) Do it yourself: obtaining updated transit stop and route shapefiles in urban and nonurban areas. *Cityscape*, 20(1), 205-214.

Roosmalen, J.J. van (2019) Forecasting bus ridership with trip planner usage data : a machine learning application. Master thesis, University of. Twente, Enschede.

Rrecaj, A.A., Todorova, M.M. (2018) Estimating Short Time Interval Densities in a CTM-KF Model. *Advances in Science, Technology and Engineering Systems Journal*, 3(2), 85-89.

Shaller, B. (2002) *Effective Use of Transit Websites: A Synthesis of Transit Practice.* Washington, DC: Transportation Research Board.

Sierpiński, G., Staniek, M., Celiński, I. (2016) Travel Behavior Profiling Using a Trip Planner. *Transportation Research Procedia*, 14, 1743-1752.

Tang, L., Thakuriah, P. (2012) Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C: Emerging Technologies*, 22, 146-161.

Trépanier, M., Chapleau, R., Allard, B. (2005) Can trip planner log files analysis help in transit service planning? *Journal of Public Transportation*, 8(2), 79-103.

Yang, F., Li, L., Ding, F., Tan, H., Ran, B. (2020) A Data-Driven Approach to Trip Generation Modeling for Urban Residents and Non-local Travelers. *Sustainability*, 12(18), 7688

Zannat, K., Choudhury, C.F. (2019) Emerging Big Data Sources for Public Transport Planning: A Systematic Review on Current State of Art and Future Research Directions. *Journal of the Indian Institute of Science*, 99, 601-619.