Asian Transport Studies, Volume 2, Issue 4 (2013), 411-420

Statistical Analysis on Multivariate Expressway Time Series Traffic Under the Different Toll Policies

Makoto TSUKAI [a], Shinichi INOUE [b], Masashi KUWANO [c], Makoto OKUMURA [d]

[a,c] *Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, 739-8527 Japan*
[a] *E-mail: mtukai@hiroshima-u.ac.jp*
[b] *Port and Airport Technology Office, Kanto Region Bureau of Land and Transportation, 2-1-4, Hashimoto-cho, Kanagawa-ku, Yokohama, 221-0053 Japan;*
E-mail: inoue-s83ab@pa.ktr.mlit.go.jp
[c] *E-mail: kuwano@hiroshima-u.ac.jp*
[d] *Center for Northeast Asian Studies, Tohoku University, 41, Kawauchi, Aoba-ku, Sendai, 980-8576 Japan; E-mail: mokmr@cneas.tohoku.ac.jp*

Abstract: Recently, various toll policies and social experiments have been conducted for ETC users of Japanese expressways. While the vast number of reports refers to the change of traffic for each section, the influence of toll policies has rarely been evaluated from the statistical viewpoint, due to the lack of an adequate model to deal with the multivariate time series including many peaks. The purpose of this study is to clarify the influence of toll policies on the traffic levels on expressways by using Independent Component Analysis (ICA). The continuously observed traffic data at several different sites on the expressway were decomposed into several independent series by using mixing coefficients (i.e. weight parameters) to convert the independent series into the observed series. ICA isolates the common and stable fluctuations in several traffic series in relation to toll policies. Moreover, the estimated independent series under the different toll policies were statistically compared to clarify whether the significant difference between the independent series occurs, or not.

Key Words: *Independent component analysis, Kolmogorov-Smirnov test, Toll policy, Continuously-observed traffic*

1. INTRODUCTION

1.1 Background and Conventional Studies

In Japan, the local expressway networks in each region have been interconnected by the year 2000. The inter-regional accessibility has improved substantially in recent years. Accordingly, the user environment relevant to the toll policies has remarkably changed. In terms of hardware environment, number of Electronic Toll Collection (ETC) system users has rapidly increased due to several discounting policies applied to ETC users. Subsequently, ETC user shares in observed traffic became 70% or more in 2009. The increase of ETC users has enabled toll policy manager to adequately control the toll levels corresponding to the characteristics of demand, considering regional or target specific conditions. After long discussions about the expressway management policy from 2000 to 2005, the expressway tolls have been discounted for each car type, or setting the discounting periods in a day, or setting the temporal social experiment.

It is very important to set the appropriate future toll levels based on the comparison of the different types of policies. Considering the huge investment in expressways up to now, less

effective toll discounting policy should not be allowed to continue considering the need to reinvigorate regional economic activities. For the effective use and management of expressway network, the effect of toll policies in the past and their persistency (how long the discounting policy continue to increase the traffic) should be carefully analyzed.

Iida and Takayama (1981) studied the characteristics of traffic variation on expressway considering season, day of the week and weather. Using analysis of variance, or chi-squared test, the periodic variation or interdependency among the decomposed variations was identified. Sugie et al. (2006) empirically studied the effect of discount toll policy at Hanshin expressway. This study analyzed that ETC data aggregated with driver's personal attribute and reported that the trip generation patterns of drivers during holiday and weekday, or trip frequency had changed before and after the discounting policy was in effect. There are some studies about the influence of toll policies on observed traffic but the number of papers is limited, because of its inherent difficulties. While the purpose of analysis is to isolate the influence of toll policy from the other factors, the observed traffic include temporal shift and changes in traffic generation. Therefore, the observed data series would be "contaminated" by factors other than toll policies, including seasonal fluctuations and economic conditions. As a result, the influence of toll policies on traffic is often evaluated using a simple comparison of the total or the average traffic between the periods before and after the new policies introduced.

The statistical approach to analyze the multivariate continuously observed traffic data is developed in some ways. Iryo et al. (2007) proposed a structural model for expressway traffic, which include short term and long term variation of traffic. Through the application of this model, it is clarified that Inoue et al. (2006) studied the multiple series by using Fourier analysis, or factor analysis (FA). In this study, the regional touring trip is analyzed by using the multivariate time series data observed at the Honsyu-Shikoku connecting bridges. The common fluctuation feature of traffic is extracted by using FA, and then Fourier analysis is applied to the estimated series.

In order to clarify the characteristics of multivariate series, independent component analysis (ICA) was developed in signal processing of communication engineering (Hyvarinenn et al., 2001). ICA was originally developed as an algorithm for blind source separation, which deals with multivariate series of sounds, including voices of several persons. The problem here is how to separate the individual voice from the recorded (i.e. mixture) signals, without using additional information. Along with the progress of studies in its theoretical background and more effective algorithms, ICA has been applied to various multivariate series (Akaho, 2002), such as climate change (Ilin et al., 2006) or financial time series (Cheung and Xu, 2001). The important application for engineering field is "fault detection", which will separate the outlier or abnormal series from the mixed multivariate series. Zhang and Zhang (2010) demonstrated another algorithm to detect a fault using the following steps: 1) Learning (estimating) by using normally processed observations to extract "normal" independent series; 2) Confidence intervals of "normal" independent series, and the mixed signals are estimated; 3) Target observations which could include the "fault" process are compared with the estimated confidence intervals; 4) Detect whether faults processes exist or not. In this procedure, the performance of the ICA algorithm in finding the normal or reference series is important. Wand and Shi (2010) proposed a new algorithm to detect the waste water processing, even if the observed series follows a skewed or non-Gaussian distribution.

1.2 Purpose of the Study

The purpose of this study is to statistically analyze the influence of past toll policies on the daily observed series of traffic in order to gain information for setting the toll level in the future. By using Independent Component Analysis (i.e., ICA), continuously observed traffic time series data are decomposed into independent series using mixing coefficients, in order to clarify the significant difference of the distribution of independent series under the different toll policies and by seasonal factors. In addition, the independent series and the mixing coefficients are analyzed by the statistical test, and the characteristics of the traffic at each observation spot are clarified.

The sections in this paper are organized as follows. Section 2 shows the formulation of the model. Section 3 shows the summary of data. Section 4 shows the result of the estimated model. Section 5 shows the summary and conclusions.

2. MODEL SPECIFICATION

2.1 Independent Component Analysis

Independent Component Analysis is one of multivariate analysis which is a statistical approach to summarize the multivariate into the principal, or fewer variables. Let $x_i(t)$ be observed traffic data at site i, $s_j(t)$ be a j th independent series which will correspond to the terms of toll policies or seasonal factors, and be a mixing coefficient. The basic model of ICA is,

$$x_i(t) = \sum_{j=1}^{k} a_{ij} s_j(t), \quad (i = 1, ..., n) \tag{1}$$

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2}$$

where $\mathbf{x}$, $\mathbf{A}$ and $\mathbf{s}$ are vector representations of $x_i(t)$, a_{ij} and $s_j(t)$, respectively.

In equation (2), $s_j(t)$ are combined by the mixing coefficient a_{ij}, and a_{ij} is estimated so as to maximize the independence among the series of $s_j(t)$. Note that in applying the ICA model, the average of each series over the period should be subtracted, respectively, to decrease the calculation load, and the number of independent series should be less than that of observed series. In equation (1) the order of independent series are not fixed during the ICA estimation procedure, because the permutation in independent and plus / minus manipulation among $s_j(t)$ and a_{ij} can be allowed. Therefore, the obtained series is permutated according to the order of variance contribution to the original series, and the plus/ minus manipulation is applied in order to get many positive coefficients on table, for ease in the interpretation of each series.

In ICA, as well as factor analysis (FA), mixing coefficient and independent series are simultaneously estimated. Hyvarinenn (1997) proposed a FastICA algorithm in which an independence of the series are measured by a kurtosis (i.e, the forth moment of the series). An advantage of ICA over conventional PCA or FA is that no assumption is made for the statistical distribution of unobserved factor in ICA, while PCA or FA estimate the unobserved factor by using the second moment information. Another advantage in empirical analysis is that ICA can capture the outlier in the estimated factor, something that is not clearly captured in PCA and FA. Such differences stem from the implicit assumption for the unobserved series.

In PCA or FA, the information of data distribution used in the analysis is at most the second moment, while in ICA, at least the fourth moment. Since the higher moment of distribution is more sensitive to outlier, i.e., a sample located around the edge of distribution (Wang, et al., 2006) can be expected to produce a better estimate.

In this study, the FastICA algorithm is applied to observed traffic series in order to extract the independent characteristics of day to day distribution of traffic corresponding to the short-term concentrations of traffic due to consecutive holidays in January, May, and August.

2.2 Kolmogorov-Smirnov test

An important difference between "fault detection" in the designed system and the observation for social phenomenon is that it is difficult to define the "normal" or "reference" state in the observed traffic. In order to avoid such a problem, the Kolmogorov-Smirnov Test (i.e., KS test) is applied to check the identity of probability density in the couple of sample series (Feller, 1948). The KS test is applied for each couple of an independent series to clarify a significant difference in the probability density distribution of those two series under the different toll policies. The KS test does not assume the specific probability density function as the population distribution. In other words, KS test is a comprehensive test for all moments. The null hypothesis and alternate hypothesis of KS test are as follows:

H_0 : Two sets of samples follow the identical probability density distribution

H_1 : Two sets of samples follow the different probability density distribution

and the KS test statistics is

$$\chi_0^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} \tag{3}$$

where, n_1 and n_2 are the number of bands for each series. When $n_1, n_2 > 40$, a test statistic χ_0^2 approximately follows a chi-square distribution with two degrees of freedom.

The test statistics is a function of the value D, which is defined as the maximum distance between the empirical cumulative distribution functions for the different observation periods. Let $F(s)$ be the probability distribution function of the period s, and $G(s')$ be the probability distribution function of the other period. Then, D is defined as,

$$D = \sup_s \left| F(s) - G(s') \right| \tag{4}$$

FastICA was applied to ten series of day-to-day traffic to estimate ten independent series with the mixing coefficients. A set of mixing coefficients corresponding to each independent series were tested for their significance (Shimizu et al. 2006). This test enables us to specify the statistical significance on the observed traffic. By observing the estimated independent series, the peak periods of each series showing the characteristics of the concentrated demand is identified.

3. DATA AND TOLL POLICIES IN JAPANEASE EXPRSSWAY

In this study, the continuously-observed daily traffic series at the three bridges connecting Honsyu and Shikoku are analyzed. The observation period is 2,469 days, from January 1,

2002 to October 31, 2008. The locations of ten cross-sections are as follows: seven sections on the Nishi-Seto expressway, a section on the Seto-Chuou expressway, and two sections on the Kobe-Awaji-Naruto expressway. Although the observed traffic was recorded across five types of vehicles, the study used the total amount of traffic.

Although many toll policies were applied during the observation period, we focus on two policies because the major policy changes applied for all cross-sections are twice as follows:

1) Discounting toll policy for all sections started from July 1, 2003, and
2) Discounting toll policy for ETC user started from July 1, 2005.

As discussed in 2.2, we do not assume the "normal" state among the observed period. Instead, some comparisons between the different terms are made. In order to compare whether each independent series has changed among the periods under the different toll policies or not, the observed data series devided into several subseries as follows. Considering the independent series are divided into four periods such as follows:

(a) January 1, 2002～June 30, 2003 (547 days)
(b) July 1, 2003～June 30, 2005 (731 days)
(c) July 1, 2005～June 30, 2007 (730 days)
(d) July 1, 2007～October 31, 2008 (488 days)

Though the toll policies did not change between periods (c) and (d), period (d) was set to approximately equalize the number of days at each period. Using the KS test, a significant difference of the probability density distributions between the all combinations of sub-period (a) to (d) was clarified. When the significant difference is detected, we can deduce that the observed traffic under the deferent toll policies had changed.

4. MIXING COEFFICIENTS AND INDEPENDENT SERIES

ICA is applied to the ten observed series in order to estimate the independent series with the mixing coefficients. The result of the statistical test for the mixing coefficients is shown in Table 1. In this test, we adjusted the significance level to 0.5% (=5%/10) by the Bonferoni Collection standard in order to keep the size of test for a set of coefficients to individual series to 5%. By observing the temporal characteristics of the estimated series, we identified the peak periods of each independent series. Each series are named based on the peaks appearing in each of the independent series. The summary of these series is shown in Table 2 with skewness and kurtosis. Skewness and kurtosis are the characteristics of probability distribution, such as asymmetricity around the average and the thickness of both sides of distribution "edge", respectively. Note that the each independent series is standardized with 0 mean and unit variance. From table 2, we can understand that skewness and kurtosis corresponds to the peak frequency and the range of distribution. Among the independent series some of characteristic series as s_1, s_3, s_4, s_7, and s_8 out of ten independent series are shown in figure 1. Note that the subscript of an independent series indicates the descending share of the observed series calculated from the mixing coefficients.

In Table 1, all of mixing coefficients of s_1, s_2, s_4 and s_6 are shown to be significant.

Table 1. Mixing coefficients a_{ij} and the result of the statistical test

	S1		S2		S3		S4		S5	
<Nishi-Seto Expressway>										
Mukaishima-honsen to Mukaishima	1223.70	*	1288.16	*	734.98	*	502.54	*	355.77	*
Mukaishima to In-no-shima-kita	1255.34	*	1375.12	*	639.46	*	507.40	*	350.41	*
In-no-shima-kita to Ikuchijima-kita	955.49	*	1150.84	*	646.79	*	425.68	*	196.27	*
In-no-shima-minami to Ohmishima	832.49	*	1148.15	*	636.38	*	362.17	*	176.40	*
Ohmishima to Hakatajima	802.04	*	1104.15	*	510.54	*	404.47	*	164.64	*
Hakatajima to Ohshima	801.06	*	1101.58	*	509.40	*	544.19	*	177.95	*
Umashima to Imabarikita	798.45	*	1146.63	*	564.04	*	505.19	*	186.21	*
<Seto-chuoh Expressway>										
Kojima to Hitsuishi	2258.20	*	2250.36	*	906.27	*	862.20	*	1263.12	*
<Kobe-Awaji-Naruto Expressway>										
Tarumi to Awaji	5363.57	*	4296.92	*	2386.71	*	2048.08	*	1500.52	*
Awajishima-minami to Narutokita	4222.01	*	3028.46	*	1582.93	*	1525.49	*	1017.82	*

	S6		S7		S8		S9		S10	
<Nishi-Seto Expressway>										
Mukaishima-honsen to Mukaishima	425.89	*	288.36	*	933.98	*	261.68	*	44.49	
Mukaishima to In-no-shima-kita	369.24	*	284.78	*	924.83	*	136.60		21.13	
In-no-shima-kita to Ikuchijima-kita	325.69	*	147.97	*	576.87	*	-141.53	*	-25.96	
In-no-shima-minami to Ohmishima	436.89	*	97.67	*	313.81	*	235.16	*	71.94	
Ohmishima to Hakatajima	552.32	*	66.05		327.27	*	55.29		65.22	
Hakatajima to Ohshima	471.81	*	87.34	*	410.27	*	74.30		77.30	
Umashima to Imabarikita	466.22	*	119.07	*	428.70	*	108.59	*	298.17	*
<Seto-chuoh Expressway>										
Kojima to Hitsuishi	791.10	*	7.74		35.11		-15.61	*	-44.57	
<Kobe-Awaji-Naruto Expressway>										
Tarumi to Awaji	1489.69	*	1846.45	*	223.09		1229.94	*	243.27	
Awajishima-minami to Narutokita	1034.83	*	420.31	*	155.17		623.81	*	87.97	

*:adjusted significance at 99.5% level by Bonfferoni Correction

Table 2. Summary of the independent series

	Characteristics	Season, Location	Skewness	Kurtosis
S1	Peaks at the succesive holidays	New year, May and August	2.532	47.010
S2	Peaks at May holodays	May	4.162	94.317
S3	Distinct peak at May 5	May in 2006	2.740	20.275
S4	Peaks at August holidays	August	5.498	75.284
S5	Negative peaks	July 30, Dec. 30 in 2004, Sept. 5 in 2005	-1.562	24.117
S6	Peak at Newyear holodays	Newyear	6.161	66.322
S7	Temporal peak in summer vacation	Last week in July to first ween in August	2.001	14.712
S8	Increasing trend	Nishi-seto Expressway	0.168	-0.398
S9	Periodical peak	Sunday	0.862	1.492
S10	Illegal peak	Third Sunday of May	4.980	52.706

The temporal distributions of s_1, s_2, s_4 and s_6 have consecutive peaks of the corresponding holidays in every year. Especially, in s_4, the peaks of the August holidays after 2005 are higher than those of the previous years. Such the traffic change at August would be influenced

by the change in the toll policy introduced after July 1, 2005. Concerning the distribution of s_3 , the largest peak appears May 2, 2006. This peak would be influenced by the discounted toll policy introduced after April 1, 2006.

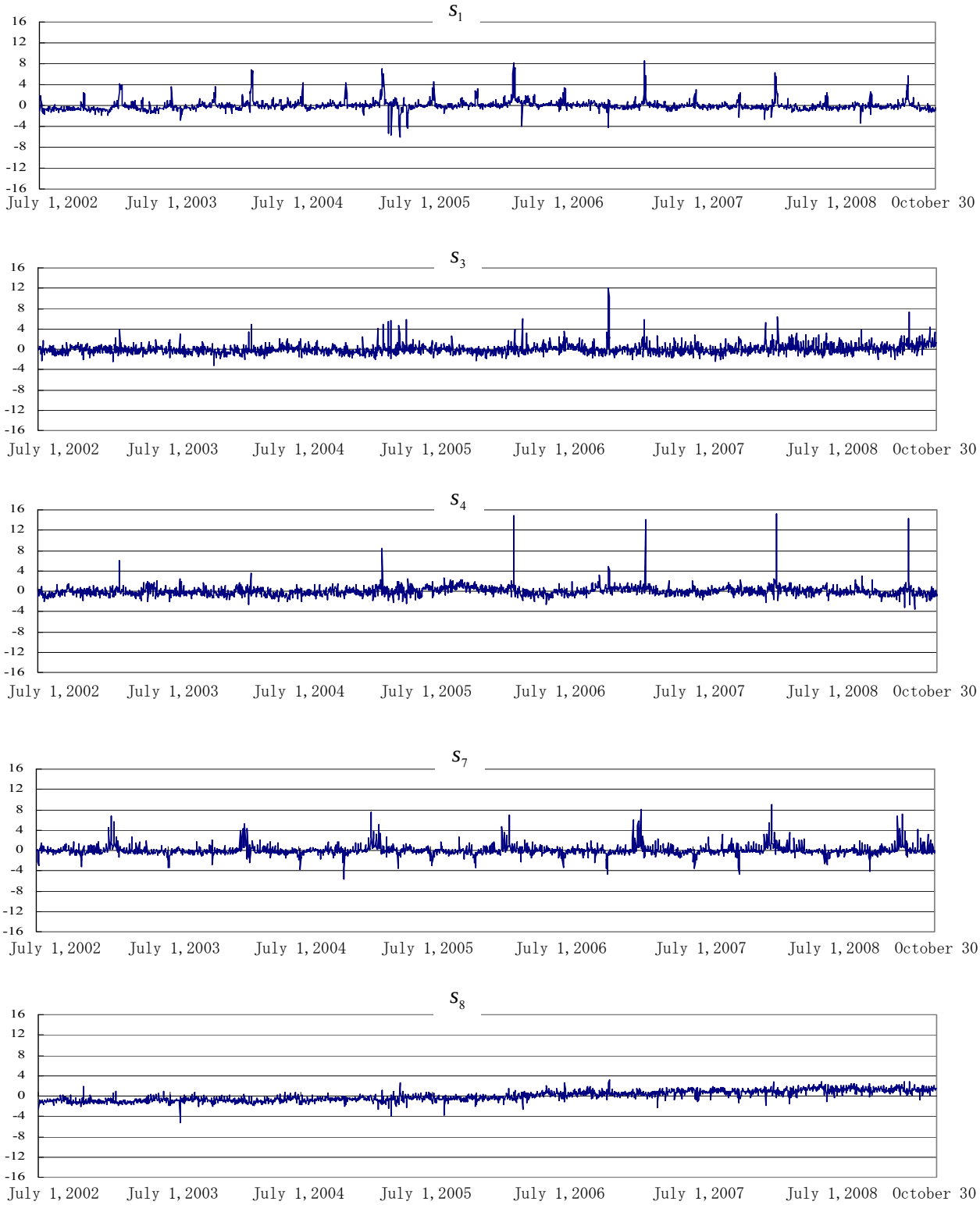

Figure 1. Independent series (S1, S3, S4, S7, and S8)

On the other hand, the mixing coefficients of s_7 are significant over almost all the cross-sections of Nishi-Seto expressway except Ohmishima to Hakatajima and Kobe-Awaji-Naruto expressway. The distribution of s_7 shows the consecutive peaks from the fourth week of July to first week of August, therefore the peak would correspond to the demand in summer vacation.

Table 3. The result of KS test for s_4

	period (a)	period (b)	period (c)	period (d)
period (a)	0			
period (b)	51.22 **	0		
period (c)	39.13 **	4.74	0	
period (d)	21.02 **	10.76 **	7.77	0

Table 4. The result of KS test for s_8

	period (a)	period (b)	period (c)	period (d)
period (a)	0			
period (b)	151.82 **	0		
period (c)	852.4 **	797.51 **	0	
period (d)	897.87 **	922.85 **	293.48 **	0

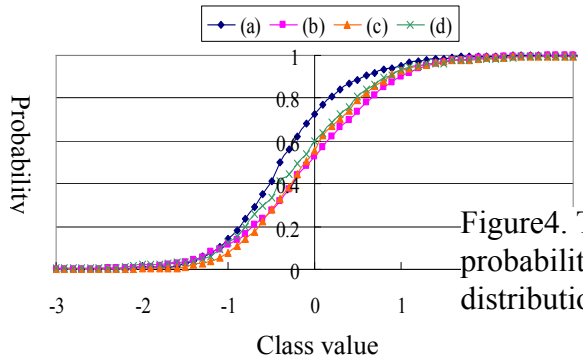

Figure4. The probability distributions

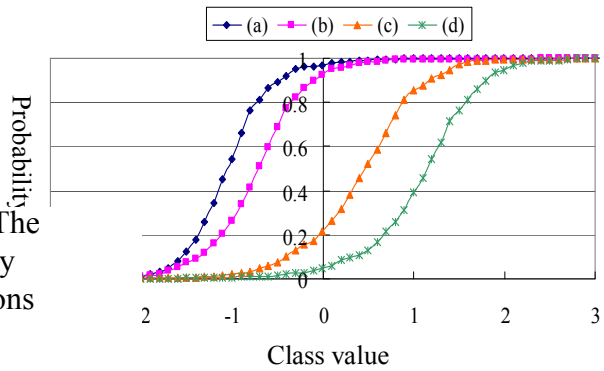

Figure 3. The probability distribu sub-period of s_4

. The probability distributions for sub-period of s_8

Note that the period is not outstanding at first glance of the original series.

The mixing coefficients of s_8 are significant only for Nishi-Seto expressway. The distribution of s_8 does not have a distinctive peak, but has the increasing trend in the whole observation term. The mixing coefficients of s_9 are significant for some sections in Nishi-Seto expressway and all sections of Kobe-Awaji-Naruto expressway. The distribution of s_9 shows a periodical peak every Sunday. The mixing coefficients of s_{10} are significant only for the section from Umajima to Imabari-kita in Nishi-Seto expressway. The distribution of s_{10} has the distinct peak at the third Sunday in May in every year.

As discussed in section3, each independent series are divided into four sub-periods corresponding with different toll policies, and applied KS test in order to test the significant difference among them. In this test, the significance level is set at 1%. As a result of this test, the combinations of sub-periods in s_1, s_2, s_5, s_6, s_8, s_9 and s_{10} were found significant. The probability distribution functions of these independent series are different for different toll policies. On the other hand, the combinations of sub-period in s_7 were not significant. Therefore, the temporal peak demand in summer vacation seems stable and unchanged in terms of probability density function. Concerning s_4, we observed another tendency for temporal demand change. For comparison, the results of the KS test for s_4 and s_8 are shown

in Table 3 and Table 4, respectively. In addition, the cumulative probability distributions of sub-periods for s_4 and s_8 are shown in Figure 3 and Figure 4.

The result of the KS test for s_4 shows a significant difference in the probability density distributions between period (d) and period (a), and (b). On the other hand, the difference found between period (d) and period (c) is not considered to be significant because the toll policy was not changed in those periods. The probability distributions of sub-periods for (b), (c), and (d) were skewed in positive area when compared with period (a). On the other hand, the distributions of these three periods were similar. The result of the KS test for s_8 shows that all probability density distributions are significantly different. The probability distributions become skewed in positive area which indicates that the observed traffic in Nishi-Seto expressway tends to increase as time passed.

5. SUMMARY AND CONCLUSIONS

The application of Independent Component Analysis to continuously-observed expressway traffic data, allowed the separation of multiple traffic series into independent series and mixing coefficients in order to analyze the temporal characteristics of the observed traffic. The Komolgrov-Smirnov test was applied to each combination of independent series to determine the level of significance of the difference between the probability density distributions of four sub-periods that were set in relation to the historical application of various toll policies.

While some peaks were not so clear in the original observed traffic series, based on the results of Independent Component Analysis, it was found that the observation series can indeed be separated into characteristic independent series with the distinctive peaks, corresponding with the consecutive holiday demands in annual period, or the weekly period. Further, through the application of the Komolgrov-Smirnov test, almost all probability density distributions of the sub-period are evaluated to be significantly different, such that the observed traffic series are statistically different under the different toll policy.

In summary, the toll discounting policy introduced in 2003 had a great influence on the traffic, while the discounting policy introduced in 2005 did not greatly influence traffic levels. In relation to the peak period, Komolgrov-Smirnov test could pick up the significant difference between the sub-periods.

The remaining issue is that it is still necessary that an analytical procedure to avoid the arbitrariness in interpretation of temporal characteristics for independent series should be developed. Concerning the peak demand difference, more sensitive statistical tests to find the difference still likewise needs to be developed.

ACKNOWLEDGEMENTS

Our research is subsidized by EHRF in Japan (Express Highway Research Federation). The expressway traffic data is provided by the NEXCO Nishi-Nihon.

REFERENCES

Akaho, S. (2002) Conditionally independent component analysis for supervised feature extraction. *Neurocomputing*, 49, 139-150.

Ilin, A., Valpola, H., Oja, E. (2006) Exploratory analysis of climate data using source separation methods. *Neural Networks*, 19, 155-167.

Iida, T., Takayama, J. (1981) Statistical analysis on traffic variation on expressway. *Expressway and car*, 24, 22-32 (in Japanese).

Inoue, H., Tsukai, M., Okumura, M. (2003) Classification of traffic pattern by using cross-sectional traffic between Honsyu-Shikoku bridges. *Journal of Traffic Engineering in Japan*, 23, 217-220 (in Japanese).

Iryo, T., Iwanani, A., Asakura, Y. (2007) Analysis on weekly variation of inflow traffic in urban expressway. *Journal of Traffic Engineering in Japan*, 27, 173-176 (in Japanese).

Cheung, M., Xu, L. (2001) Independent component ordering in ICA time series analysis. *Neurocomputing*, 41, 145-152.

Feller, W. (1948) On the KOLOGOROV-SMIRNOV limit theorem for empirical distribution. *The Annals of Mathematical Statistics*, 19, 177-189

Hyvarinen, A., Oja, E., (1997) A fixed-point algorithm for independent component analysis. *Neural Computaion*, 9, 1483.

Hyvarinen, A., Karuhnenn, J., Oja, E., (2001) Independent component analysis. Wiley & Sons.

Shimizu, S., Hyvarinen, A., Kano, Yutaka., Hoyer, P., Kerminen, A. (2006) Testing significance of mixing and demixing coefficients in ICA. *Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA2006)*, 901-908, Charleston SC, USA, March 5-8.

Sugie, I., Yamamoto, M., Kanno, H., Yoshioka, M. (2006) Empirical analysis on social experiment in toll policies of expressway using ETC data. *Journal of Japan Society of Traffic Engineers in Japan*, 26,149-152 (in Japanese).

Wang, L., Shi, H. (2010) Multivariate statistical process monitoring using an improved independent component analysis. *Chemical Engineering Research and Design*, 88, 403-414.

Wang, Z., Wang, J., Calhoun, V., Rao, H., Detre, J., Childress, A. (2006) Strategies for reducing large fMRI data sets for independent component analysis. *Magnetic Resonance Imaging*, 24, 591-596.

Zhang, Y., Zhang, Y. (2010) Fault detection of non-Gaussian processes based on modified independent component analysis. *Chemical Engineering Science*, 65, 4630-4639.