An Analysis of Day Traffic Volume able to correspond to Data update and Structural Change by Kalman Filter

Kuniaki Sasaki[a], Kota Nakazawa[b], Takashi Yamamoto[c], Hitoshi Iguchi[d]

[a] Department of Civil Engineering, University of Yamanashi, Kofu, 400-8511, Japan
[a] *E-mail: sasaki@yamanashi.ac.jp*
[b] Atsugi Construction Office, Central Nippon Expressway Company Limited, Atsugi, 243-0032, Japan; *E-mail: k.nakazawa.aa@c-nexco.co.jp*
[c,d] Planning Analysis Division, Nippon Expressway Research Institute Co. Ltd.
[c] *E-mail: t.yamamoto.ae@ri-nexco.co.jp*
[d] *E-mail: h.iguchi.aa@ri-nexco.co.jp*

Abstract: Due to the introduction of ITS devices, we can have got the massive updatable traffic data at low cost. The focus of this research is to analyze the massive updating traffic data that force heavy computational load. The state-space model based on the recurrent equation system is one of the solutions for massive data problem, because recurrent equation allows phased processing of the data and can update with relatively small computational load. We applied that model to the day traffic volume of an expressway in Japan. The model we applied includes the circulation factor of day of the week besides the factors of auto-regressive and trend. The result showed the effect of change of weekend toll charge on long term trend and on the day of the week factor numerically. The result showed that the prediction error for short term prediction would be reduced by data updating.

Keywords: State-Space model, Kalman Filter, Data Updating, Structural Change, Traffic Volume

1. INTRODUCTION

The traffic analyzing models such as discrete choice models and four-step travel demand analysis models assume a time-stable structure. However, the structure of the travel demand is not always time-stable. A big change in choice environment of travel would change the structure of demand. One of the good examples has happened in Japan: In March 2009, the weekend toll-expressway charging method was drastically changed and was restored to its former method in 2011. Before the implementation of the toll system change, the toll was charged based on the distance without ceiling. If one used an expressway for 100km, he has to pay 2700 yen ($30), and for 500km, he had to pay 12,000 yen ($132). From 2009 to 2011, the weekend toll was limited to 1000 yen ($11) maximum. Therefore, when one used expressway on weekend about 500km, he saved 11,000 yen ($121). This policy is said to have forced a drastic change in the mode and the date choices for long distance trip, because a 500km trip by the bullet train costs 15,000 yen. That is, it has affected tremendously the variation structure of the traffic volume. Any useful knowledge from this data of the change in the structure would provide us with a more precise perspective of the traffic volume variation in Japan because it became easy to change the toll by penetration of ETC.

Using the data collected by the electric systems such as ETC, however, has a problem of handling massive and updated data. The state-space model based on the recurrent equation system is one of the solutions for this problem, because recurrent equation allows phased processing of the data and can update with relatively small computational load. Moreover, the structural change of travel demand could be detected automatically by the analysis of massive data by state-space model, because this model does not assume fixed structure such as the assumption of ordinary statistical models. The most popular model of state-space model is Kalman filter model with linear recurrence equation system. The Kalman filter model with linear system would be suitable to analyze the day-to-day variation of travel volume on expressway in Japan, because the demand structure of traffic volume of the expressway is simpler than that of the ordinary roads. This Kalman filter approach is used, for example, to calculate a spacecraft location and to manage river flood (Pan and Ru-yih, 2004), as well as for econometrics (Mittnik, 1990, Hyndmana et. al, 2002). In the field of transportation analysis, Herrera and Bayen (2010) applied a kind of Kalman filter model to the state estimation of highway traffic.

The time series analysis of traffic flow has already started in 1980s. Okutani and Stephanedes (1984) applied Kalman Filter to the traffic flow analysis. Nanthawichit, C et. al (2003) applied Kalman Filter for update traffic state using updating data. Recently, the expansion of computational ability made the treatment of more complicated structure of time series model possible (Lafortune et. al, 1993, van D. Voort et. al, 1996, van Lint et. al, 2005). For example, there are some studies such as Jiang and Adeli (2005) which incorporated wavelet transform and neural network analysis with time series analysis for analyzing nonstationary feature of traffic flow. These studies focused on the micro variation on short period of traffic flow because it is hard to analyze long period for its computational reason. If the time series analysis is applied to the aggregated data in long interval such as day traffic volume or time traffic volume, the cyclical characteristics such as weekly or daily effects have to be considered. Besides, long term trend should be considered because the traffic volume of expressway is generally affected by the travel circumstance such as economic condition or toll system. One of the main purposes of this study is to confirm the applicability of state space model with cyclic effects for finding the time series characteristics of traffic demand for long term data. Moreover, when the toll charge system has drastically changed as stated in the previous paragraph, it is consequently expected that the driver's behavior has changed. The focus of this study is to apply the state-space model to the day traffic volume data of expressway in the long term including significant change of toll charge and to find the structural change of travel demand in that term.

2. MODELS

In this study, we apply the state-space model which can adjust the circulating effect that can reflect the constantly updated data and the structural changes (Kitagawa and Gersch, 1984). Some studies have applied the state-space model, the general form of the time series analysis, to the traffic volume analysis, however they rarely have used the long-term data for the analysis and consequently did not consider the circulating effects. In our study, we apply the state-space model to the long term data and find the circulation effect on the traffic volume. The basic definition of the model is shown in equation (1) and (2).

$$x_n = F x_{n-1} + G v_n \tag{1}$$
$$y_n = H x_n + \varepsilon_n \tag{2}$$

where,

x_n : unobservable traffic volume

y_n : traffic counts at the exit gate by the ETC

F, G, H : parameter matrices
v_n, ε_n : system error and measurement error, respectively

The equation (1) is the system model and the equation (2) is the measurement model. Because the system model is to describe the transition of the system, it is plausible that the periodic terms such as the effects of the time of the day, and the non-periodic terms such as the trend are included in the system model. The system equation is modified as the equation (3) considering those periodic and non-periodic factors.

$$y_n = t_n + s_n + p_n + \varepsilon_n \tag{3}$$

where,
ε_n : $N(0, \sigma^2)$ as white noise factor of the n[th] data
s_n : periodic term of the n[th] data
t_n : trend of traffic volume of the n[th] data
p_n : non-systematic noise such as the correlation between each data

The trend indicates the systematic longitudinal variation of the error term that are the change in a certain period as shown in the equation (4). The coefficient $c_i^{(k)}$ is a parameter defined by the dimension of difference.

$$t_n = \sum_{i=1}^{k} c_i^{(k)} t_{n-i} + v_{n1} \tag{4}$$

Periodic term in this week is a factor that varies on the traffic periodically during a day or a week, which is shown as equation (5). In equation (5), q denotes the period of the one cycle of the periodic factor. Because we are going to consider the effect of the day of the week on a day traffic volume, we fixed it to 7.

$$s_n = -\sum_{i=1}^{q-1} s_{n-i} + v_{n2} \tag{5}$$

where,
v_{n2} : $N(0, \tau_2^2)$ as random error term

We assume that the system noise consists of two types. One is the white noise and the other is the colored noise that follows the autoregressive process. The autoregressive process is defined as the equation (6) which expresses the colored noise of correlation. By considering autoregressive process we are going to extract the short-term trend from the variation of traffic. *m* denotes the dimension of auto-regression.

$$p_n = \sum_{i=1}^{m} a_i p_{n-i} + v_{n3} \tag{6}$$

where,
v_{n3} : $N(0, \tau_3^2)$ as white noise factor

This process builds the unobservable traffic volume (system state) x_n as a space vector which consists of the trend, the period and the auto-regressive components, as shown in equation (7).

$$x_n = [t_n, t_{n-1}, \cdots, t_{n-k+1}, s_n, s_{n-1}, \cdots, s_{n-q+2}, p_n, p_{n-1}, \cdots, p_{n-m+1}]^t \tag{7}$$

The parameter matrices, F, G, H are defined in the equation (8). F is the state transition matrix, and G denotes the input matrix of the system noise at system updating. H denotes the output matrix of the system state.

$$F = \begin{bmatrix} F_t & O & O \\ O & F_s & O \\ O & O & F_p \end{bmatrix} G = \begin{bmatrix} G_t & O & O \\ O & G_s & O \\ O & O & G_p \end{bmatrix} H^t = \begin{bmatrix} H_t^t \\ H_s^t \\ H_p^t \end{bmatrix} \quad (8)$$

Because the model system defined so far assumes normal distribution as error components, the average vector and covariance matrices are the sufficient statistics. Therefore these matrices enable the traffic volume estimation. The general algorithm for the state estimation for this type of model system is the Kalman filter.

The extrapolation of the traffic volume estimates

$$\left.\begin{aligned} x_{n|n-1} &= F x_{n-1|n-1} \\ V_{n|n-1} &= F V_{n-1|n-1} F^t + G Q G^t \end{aligned}\right\} \quad (9)$$

Kalman gain and the estimated traffic volume with additional observation

$$\left.\begin{aligned} K_n &= V_{n|n-1} H^t (H V_{n|n-1} H^t + R)^{-1} \\ x_{n|n} &= x_{n|n-1} + K_n (y_n - H x_{n|n-1}) \\ V_{n|n} &= V_{n|n-1} - K_n H V_{n|n-1} \end{aligned}\right\} \quad (10)$$

Where

$$Q = \begin{bmatrix} \tau_1^2 & & 0 \\ & \tau_2^2 & \\ 0 & & \tau_3^2 \end{bmatrix}, \; R = \sigma^2$$

The Kalman filter repeats the two processes. One is the extrapolating process of the traffic volume that estimates the prior distribution using traffic volume in the previous period. The other process is the traffic volume estimate process that updates the posterior probability distribution of the traffic volume by observing the current traffic volume. By using the Kalman filter with these two steps, we can estimate the conditional distribution of the traffic volume when the additional observation is available. In the proposed model system, estimating the parameters are the variance parameters of each system noise and measurement error , and the parameters of autoregressive model . These parameters are estimated by the maximum likelihood estimation using measurement data. After parameter estimation, we conduct the smoothing process to estimate all the past traffic volume again, using the equation (11). The smoothed estimation would be more precise because it is based on the whole measurement data.

Smoothing

$$\left.\begin{aligned} A_n &= V_{n|n} F^t V_{n+1|n}^{-1} \\ x_{n|N} &= x_{n|n} + A_n (x_{n+1|N} - x_{n+1|n}) \\ V_{n|N} &= V_{n|n} + A_n (V_{n+1|N} - V_{n+1|n}) A_n^t \end{aligned}\right\} \quad (11)$$

We cannot use the process explained in the previous section for the long term prediction because we do not have the necessary measurement data.
Thus, only the state-estimate extrapolation process is applied repeatedly as equation (12) if we

try to apply this model to the long term prediction.

$$\left.\begin{aligned} x_{n+i|n} &= F x_{n+i-1|n} \\ V_{n+i|n} &= F V_{n+i-1|n} F^t + G Q G^t \end{aligned}\right\} \qquad (12)$$

The second equation in (12) allows us to expect an increase in the variance of predicted distribution. This increase of variance would be vanished by adding the measurement data and applying the filter process.

3. EMPIRICAL STUDY

In practice, we applied to the above-proposed model the time series data of an expressway traffic volume taken from one pair of interchange provided by three expressway companies in Japan. The data was of the one-day traffic volume between Kawaguchiko and Hachioji in Chuo Expressway taken by the ETC (Electric Toll Collection) from 2005 to 2009. Kawaguchiko is a interchange to go to the Fuji Mountain, the most popular tourism area in Japan, and Hachioji is a city of the west gate of Tokyo. The distance between them is about 70km. The traffic depends on the tourist behavior that is affected by the day of the week and the weather. Tourism depends on the travel trends. Also, the toll of the expressway on weekend was drastically changed between 2005 and 2009. We believe that the data we use is suitable for testing the applicability of the proposed model because the model can deal with the periodic adjustment. The assumption in applying the proposed model is that the traffic state on the expressway is a hypothetical variable measured by ETC transaction, though about 90% of the total traffic volume is ETC users.

The time series data of the day-traffic volume corresponds to the measurement data y_k in the proposed state-space model. The variance of day-traffic volume is decomposed into three components, the trend, the periodic adjustment and the auto-regressive factor. We fixed the dimension of the auto-regression to seven and the dimension of the trend to one, which was the highest AIC through some trials of the parameter estimation. The parameter estimated by using the data is shown in the Table 1.

Table-1 estimated parameter of the model

Parameter	*estimates*
σ^2	2328.59
τ_1^2	1.05
τ_2^2	0.04
τ_3^2	67.31
a_1	0.60
a_2	-0.15
a_3	0.08
a_4	-0.07
a_5	0.06
a_6	0.03
a_7	0.05
Max. likelihood	-10471.00
AIC	20964.00

FIgure-1 shows the actual day-traffic volume, and the trend factor estimated by using the proposed model in the same graph. The trend is the expectation value of traffic volume in time series as not affected by the other factors. The trend factor in it demonstrates that the seasonal trend peaks in August.

Also, the first week of May, a series of holidays in Japan, shows a specific increase in the day-traffic volume. However, from April of 2009, the trend factor increased by about five hundred, including the first week of May. This increase is supposed to be caused by the drastic change of toll charge in weekend that had started on the end of March.

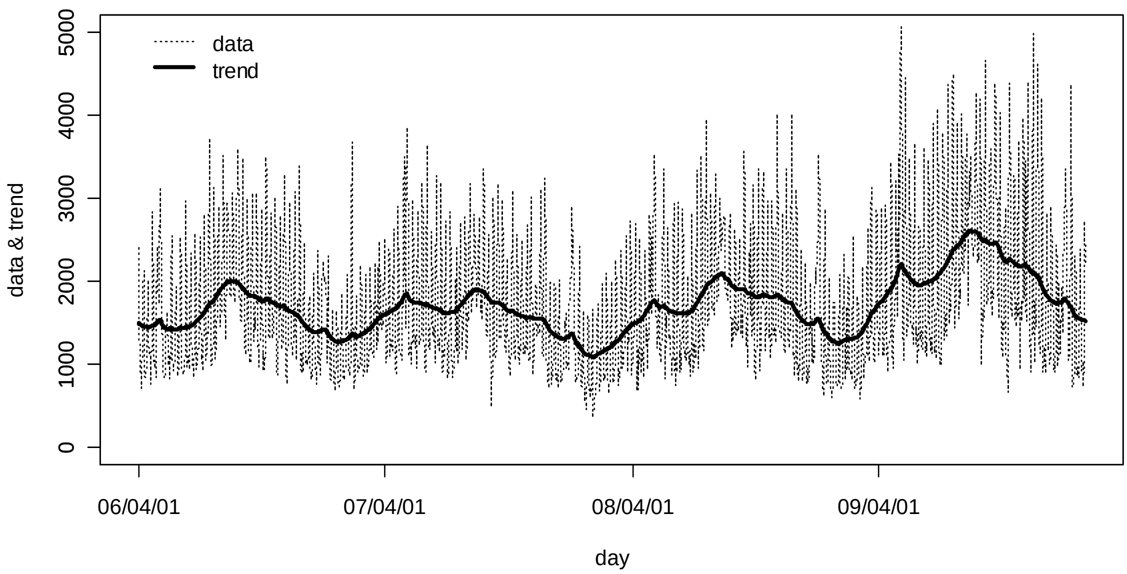

Figure-1 Time series of traffic volume and the trend factor

The effect by the day of the week (the week-day factor) is shown in Figure-2. The horizontal axis is the day and the vertical axis is the value of cyclical adjustment. Because there is more traffic volume on this expressway in weekends, this cyclical adjustment is high on weekends and low on weekdays. Therefore, we can see the upper level of this graph as weekends and the lower level of this graph as weekdays. The effect of the week-day factor makes about 1500 pcu difference between weekend and weekday. Furthermore, we can see that the cyclical adjustment itself is cyclical also in a year in this case: The differences of the cyclical adjustments tend to increase in the summer, because Kawaguchiko is a summer resort located in a mountain area. We think that the toll change on weekends caused about 350 pcu increase in the cyclical adjustment, because the cyclic factor on weekend increased and the cyclic factor on weekdays decreased. The track of the auto-regressive factor is shown in Figure-3. This factor indicates the systematic effect of error component that is related to the demand on the precedent days. Consequently, the auto-regressive factor mainly describes the traffic volume variation that occurs around national holidays, summer vacations and so on. The variance of auto-regressive factor increased after March 2009. That means the systematic component of unobserved factor increased after the toll change.

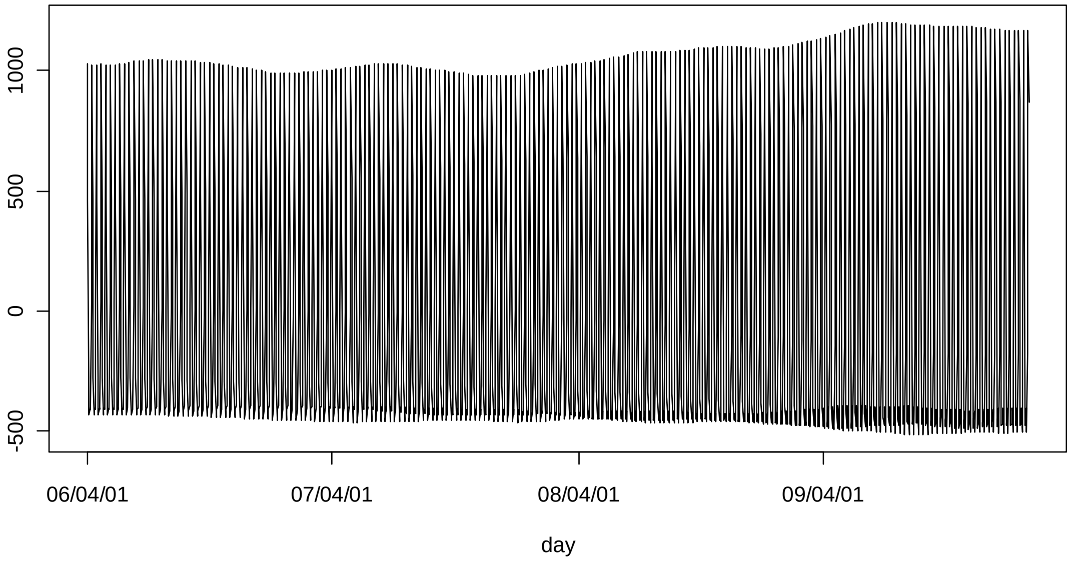

Figure-2 The series of cyclical (the day of week) effect

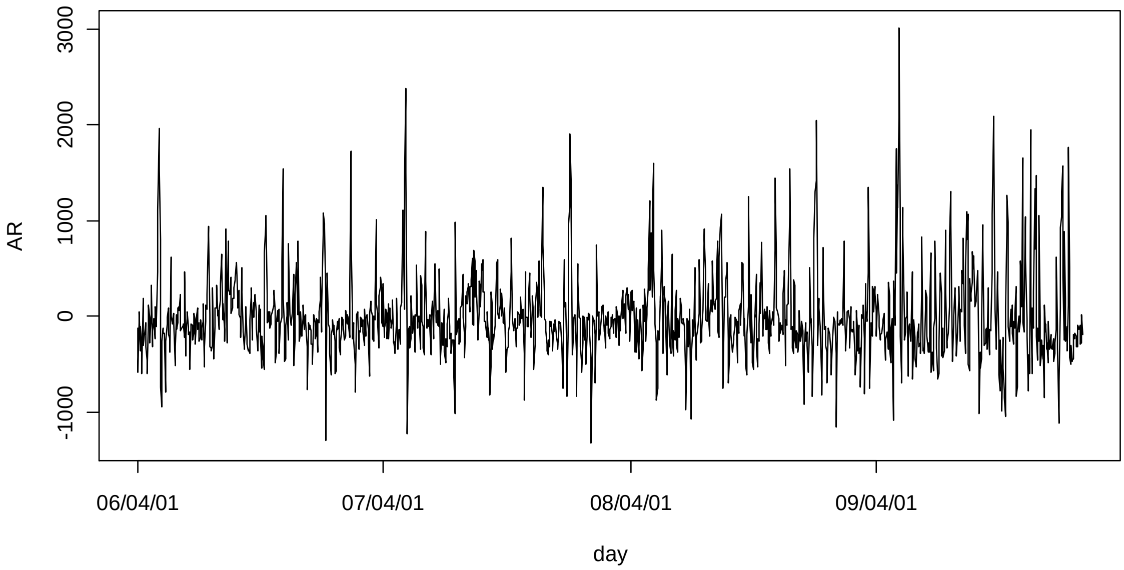

Figure-3 The series of autoregressive factor

The proposed model that uses the Kalman filter algorithm can interpolate when some measured data is missing and can extrapolate the prediction when long term prediction is necessary. For example, the Figure-5 indicates the prediction of the traffic volume from January to February 2008 that is extrapolated by the data till December 2007, while Figure-4 shows the prediction of the traffic volume in February 2008 by extrapolating by the data till January 2008. The model reduces the error in the prediction about the traffic- volume in February by adding one-month date. These Figures show the estimated state after the smoothing process. The estimated state after smoothing is mostly consistent with the actual traffic volume.

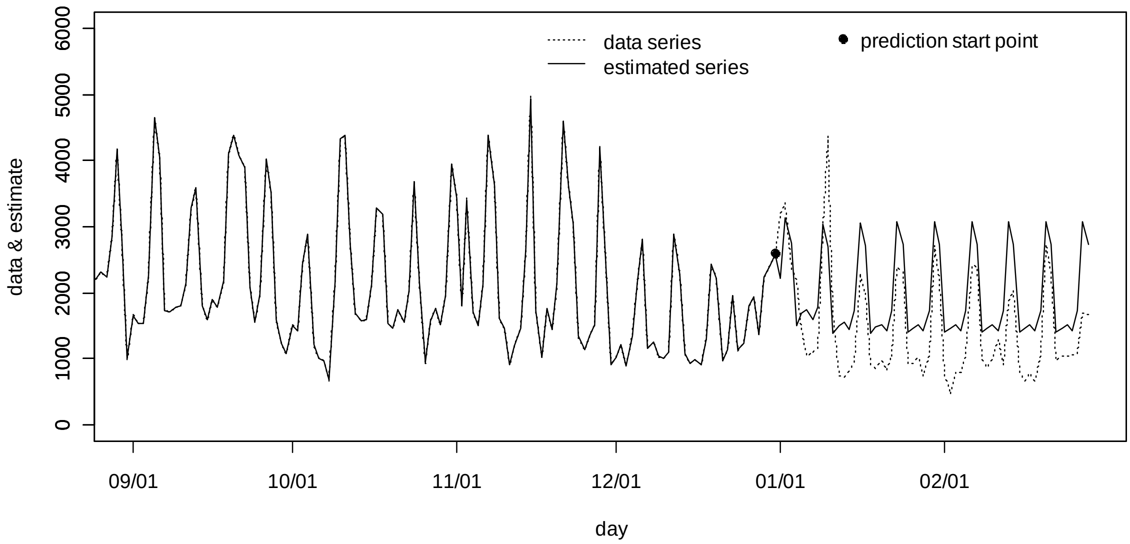

Figure-4 Extrapolating traffic volume with the data till 12/31/2008

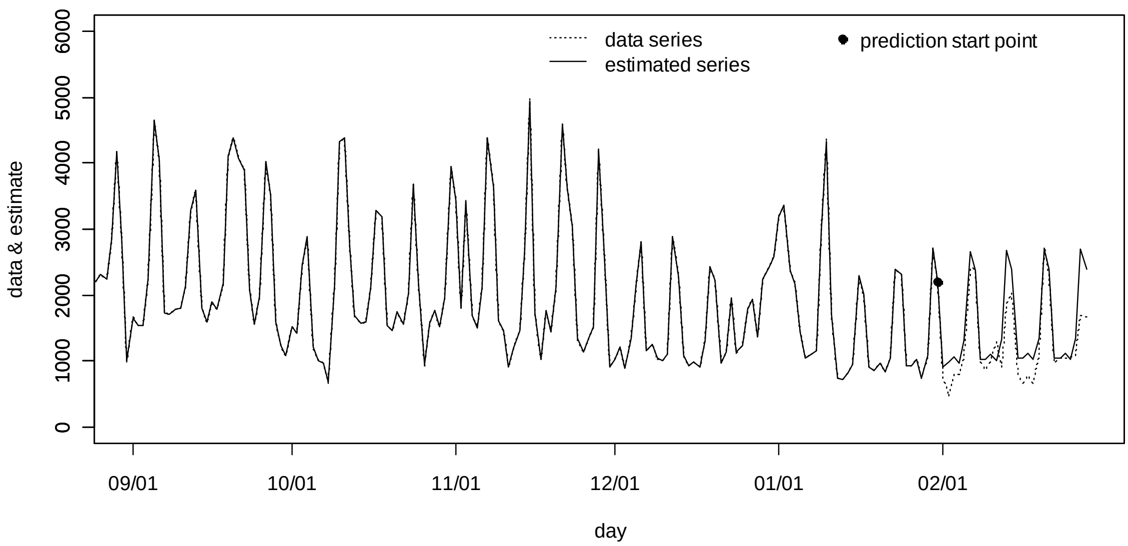

Figure-5 Extrapolating traffic volume with the data till 01/31/2009

One of the parameters estimated in this model is the variance of the error term of measurement equation. This variance can be figured out by the data on each time point. Therefore, we calculated it and showed it in the Figure-6. Though the variance is large at the initial stage, it is stable after about a couple of month. That is, the minimum period of data we need to get more precise result is about 60 days to 90 days. The more the data is used to estimate, the smaller the variance became. However, the variance increased discontinuously after 『t(March 2009)t』 again. This increase suggests that the demand structure of this expressway itself has changed so that we cannot expect no change when we apply the same model to the post-change situation. Nonetheless, we can expect this increased variance to be readjusted in this model in a few weeks by this graph.

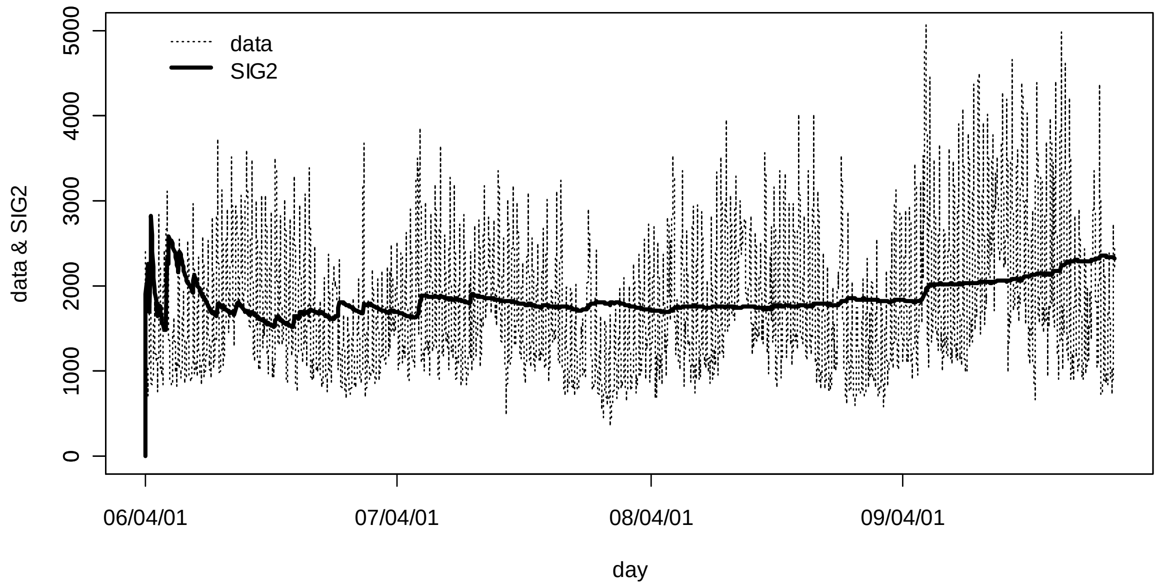

Figure-6 The expectation of measurement error estimated by the data until each time point

4. SUMMARY

This study applied the state-space model with the Kalman filter to the day traffic volume of the expressway. Two purpose mainly motivated us to apply this type of model. One is to detect the change in the demand structure of expressway caused by the drastic change of the toll charge. Because the common statistical models assume a stable demand structure, no structural change can be detected automatically. Another purpose is to develop a model that can apply the updating data to the prediction of the traffic demand. Though the recent progress in the measurement system allows us to use the updating data, the data size becomes too big to analyze using the limited calculation resources. That is why we applied the recurrent equation model to reduce the calculation load. The model identified three factors as the sources of the variation of traffic volume, the trend, the week-day factor and the autoregressive factor. We found these three factors changed significantly after the toll charge in 2009 in Japan. After the toll change, the difference of the traffic volume between weekday and weekend increased and the trend of the traffic volume turned into increasing, while the variance of the auto-regressive factor increased.

The advantage of this study is that it uses only the daily traffic volume as the necessary data. Also, the effect of these factors can be measured as numeric change. Since this approach is less load of computation and finds the structural change automatically, it will become more important in the ITS era which yields massive and purposeless data. This study has assumed a linear structure of the state transition for simplicity on one expressway traffic volume. In applying this model to the network traffic, we need to introduce more complicated state transition system because of the complex characteristics of traffic. Nevertheless, we do conclude that this study showed the first step of updating method of the travel demand prediction by using the traffic volume.

ACKNOWLEDGEMENTS

This study was supported by three expressway companies in Japan. They provided not only data for analyzing but useful comments. We would like to express our deepest gratitude to the companies.

REFERENCES

van Der Voort, M., Dougherty, M. and Watson S. (1996) Combining Kohonen maps with ARIMA time series models to forecast traffic flow, Transportation. Research Part C, Vol. 4, pp.307-318

Herrera, J and A. Bayen (2010) Incorporation of lagrangian measurements in freeway traffic state estimation, Transportation research Part B, Vol. 44, pp.460-481

Hyndmana, R. J., A. B Koehlerb, R. D Snydera, S. Grosea (2002) A state space framework for automatic forecasting using exponential smoothing methods, International J. of Forecasting, Vol.18, pp. 439-454

Jiang, X. and H. Adeli (2005) Dynamic wavelet neural network model for traffic flow forecasting, Journal of Transportation Engineering, vol. 131, pp.771-779

Kitagawa, G. and W. Gersch, (1984) A Smoothness Priors - State Space Modeling of Time Series with Trend and Seasonality, J. of the American Statistical Association, Vol.79, pp.378-389

Lafortune, S., R. Sengupta, D. E. Kaufman and R. L. Smith (1993) Dynamic system-optimal traffic assignment using a state space model, Transportation Research Part B, Vol. 27, pp. 451-472

van Lint, J.W.C., S.P. Hoogendoorn and H.J. van Zuylen (2005) Accurate freeway travel time prediction with state-space neural networks under missing data, Transportation. Research Part C, Vo. 13, pp. 347-369

Mittnik, S. (1990) Macroeconomic forecasting experience with balanced state space models, International Journal of Forecasting, Vol. 6, pp.337-348

Nanthawichit, C, T. Nakatsuji and H. Suzuki (2003) Application of Probe Vehicle Data for Real-Time Traffic State Estimation and Short-Term Travel Time Prediction on a Freeway, Paper presented ad 82nd TRB annual meeting, Washington DC

Okutani, I. and Y. J. Stephanedes (1984) Dynamic prediction of traffic volume through Kalman filtering theory, Transportation Research Part B, 18(1), pp.1-11

Pan, Tsung-yi and Ru-yih Wang (2004) State space neural networks for short term rainfall-runoff forecasting, Journal of Hydrology, Vol. 297, pp. 34-50