計畫編號: NSC91-2415-H-009-002-  $\frac{91}{108}$  08 01 92 07 31

計畫主持人: 周幼珍

報告類型: 精簡報告

。<br>在前書 : 本計畫可公開查詢

行政院國家科學委員會專題研究計畫 成果報告

92 10 31

# Estimation of Dynamic Origin-Destination by State Space Model

dynamic origin-destination(O-D)

real-time traffic control

strategy

Gaussian Kalman filter models

Gibbs Sampler

observations

normality transition matrix

random matrix simultaneous state variable iterative estimation algorithm

#### **Abstract**

Dynamic origin-destination (O-D) pattern representing time-dependent trip demands from one place (origin) to another (destination) is one of the most essential input data for most traffic operational analyses. Historical studies assumed that the transition matrix is known or at least approximately known, which is unrealistic for a real world network. And due to the fact that the number of trips to a specific destination, y, is easy to obtain and the O-D variable, x (path flow based in this research), is not directly observable, a Gaussian state space model is formulated to describe the relationships of x and y, observation equations, and the dynamics of x , state equations with unknown transition matrix. Under the assumption of Gaussian noise terms in state space model, the distribution of random transition matrix F is derived. A solution algorithm combining Gibbs sampler and Kalman filter to tackle the problem of simultaneous estimation of F and  $x_t$  based on the latest available information is proposed. Real O-D data from Taipei Rapid Transit is used to verify the presented model and solution method. Preliminary results are generally satisfactory, showing that also in the unknown transition matrix case, significant estimates are achieved.

 $\mathbf{matrix}$ 

**Keywords:** Time-varying origin-destination matrices, Gaussian state space model, Kalman filter, Gibbs sampler.

## **1 Introduction**

 Due to the high cost of origin-destination O-D data collection in the highway, traffic engineers and researchers have been seeking estimation methods to derive the valuable O-D flow information from less expensive traffic data, mainly, link traffic counts of surveillance systems. The existing studies on estimating time-dependent O-D matrices with time-series of link flows were classified into two categories by [1], i.e. dynamic traffic assignment DTA based and non-DAT based. Namely, the assignment based studies naturally employ the DTA concept to formulate the interrelation between O-D flows and link flows so as to establish a set of observation equations. e.g. [2]; [3]; [1]; [4]; [5] Computation of the assignment matrix is a difficult task, for the elements in the matrix depend on mapping time-varying path flows to link flows. The non-DAT based studies manage to formulate observation equations without the DTA notion but assume that complete entering and leaving flows at origin and destination nodes are available from traffic counts. These methods were successfully implemented against some special networks for which the entry and exit counts are provided. e.g. [6]; [7]; [8]; [9]; [10]; [11] All of the above approaches were also divided into two groups in another way, pertaining to closed networks and pertaining to general networks, by [4]. A closed network possesses some similar meaning with non-DTA based method but does not strongly imply that the assignment matrix is not used in the solution algorithm. In this paper, based on the previous work of [12], extensions to consider lagged effects and both forward and backward filtering are formulated. Especially, without prior information of transition matrix, another very hard work is calculating assignment matrix. We propose an algorithm to simultaneously estimate O-D matrices and transition matrix. Performance of the proposed approach is evaluated by the observed passenger counts data from the Nangkang line of Taipei rapid transit systems.

# **2 Model Specifications**

### **2.1 Bsaic formulations**

The standard state space model is coupled with two parts: transition equations and measurement equations, and is illustrated as the following form.

$$
x_{t} = Fx_{t-1} + u_{t},
$$
  
\n
$$
y_{t} = Hx_{t} + v_{t}, t = 1, 2, ..., n
$$
\n(1)

For convenience, all the variables and subscripts in (1) are described to fit the terminology of transit O-D estimation. Let  $x_t = (x_{t1}, x_{t2},..., x_{tp})'$  be the O-D transition vector at time period *t*,  $t = 1$ , *2,..., n,* and  $x_{ij}$  denote the number of passengers traveled at time period *t* of O-D pair *j, j = 1,* 2,..., p;  $y_t = (y_{t1}, y_{t2},..., y_{tq})'$  is an observed vector at time interval *t*, and the elements of  $y_t$  are the counts of transmitted messages at all *q* stations; *F* is a  $p \times p$  random matrix; *u* and *v* are the noise

terms with independent  $N_p(0, \Sigma)$  and independent  $N_q(0, \Gamma)$  respectively. For considering the

lagged effects, it is necessary to distinguish the formulation between entry measurement equations and exit measurement equations in (1). Then the entry measurement equations can be written as

$$
y_{ot} = H_o x_t + v_{ot}, t = 1, 2, \dots, n
$$
 (2)

, where  $y_{ot} = (y_{ot1}, y_{ot2},..., y_{otq})'$  is an observed vector of entry counts at time interval *t*; the

elements of  $y_{ot}$  are the entry passengers at all *q* stations;  $H_o$  describe the *q*×*p* origin-OD pair incident matrix;  $v_{\alpha t}$  are the noise terms with independent  $N_a(0, \Gamma_a)$ . The exit measurement equations are rewritten as

$$
y_{dt} = \sum_{t'=1}^{t_{\text{max}}} H_{dt'} x_{t-(t'-1)} + v_{dt}, t = 1, 2, ..., n
$$
 (3)

, where  $y_{dt} = (y_{dt1}, y_{dt2},..., y_{dtq})'$  is an observed exit vector at time interval *t*; the elements of

 $y_{dt}$  are the counts of exit passengers at all *q* stations;  $H_{dt}$ <sup>'</sup> are the *q*×*p* destination-OD pair incident matrices considering  $t'$ ,  $t' = 1, 2, \dots, t_{\text{max}}$ , lagged effects.  $t_{\text{max}}$  is the maximal travel time among *p* O-D pairs. Waiting time and walking time of intra station should be included in calculating *t'*.  $v_{dt}$  are the noise terms with independent  $N_q(0, \Gamma_d)$ .

### **2.2** *F* **matrix and Gibbs sampler**

For simplicity, following concept and expressions are translated by means of standard state space model, i.e. (1). Just as mentioned, transition matrix in this paper is assumed to be a random matrix without any historical information of transition vectors. Then, we focus on the discussion of the transition equations in (1). If the transition equations in (1) are rewritten as

$$
x'_{t} = x'_{t-1}F' + u'_{t}, t = 1, 2, ..., n
$$
\n<sup>(4)</sup>

Let

$$
X_{n} = \begin{bmatrix} x_{1}' \\ M \\ x_{n}' \end{bmatrix}, X_{n-1} = \begin{bmatrix} x_{0}' \\ M \\ x_{n-1}' \end{bmatrix}, U = \begin{bmatrix} u_{1}' \\ M \\ u_{n}' \end{bmatrix}
$$
 (5)

then we have

$$
X_n = X_{n-1}F' + U \tag{6}
$$

We shall further suppose that the parameterization in terms of *F*' is so chosen such that it is appropriate to take *F*' as locally uniform,  $P(F') \propto \text{constant}$ . Consider the  $p \times p$  symmetric matrix

$$
S(F') = \left\{ S_{ij} \left( F'_i, F'_j \right) \right\} \tag{7}
$$

, where

$$
S_{ij}\left(F'_i, F'_j\right) = \sum_{t=1}^n u_{ti} u_{tj}
$$
  
=  $(X_{n(i)} - X_{n-1}F'_i)'(X_{n(j)} - X_{n-1}F'_j)$ 

;  $X_{n(i)}$  is the i-th column vector of  $X_n$ ;  $F'_i$  is the i-th column vector of  $F'$ . Let

$$
\hat{F}'_i = (X'_{n-1}X_{n-1})^{-1}X'_{n-1}X_{n(i)}
$$
\n(8)

be the least square estimate of  $F_i'$ ,  $i = 1, 2,..., p$ . Consequently, we get

$$
S(F') = A + (F' - \hat{F}')' X'_{n-1} X_{n-1} (F' - \hat{F}')
$$
\n(9)

and  $A = \{a_{ij}\}\$ is the *p*×*p* matrix with its elements

$$
a_{ij} = (X_{n(i)} - X_{n-1} \hat{F}_i)'(X_{n(j)} - X_{n-1} \hat{F}_j') \tag{10}
$$

, i.e. A is proportional to the sample covariance matrix. From the general result in Gaussian model, the posterior distribution of  $F'$  is then

$$
p(F'/X) \propto |S(F')|^{-\frac{n}{2}}
$$
  
=  $|A + (F' - \hat{F}')'X'_{n-1}X_{n-1}(F' - \hat{F}')|^{-\frac{n}{2}}$  (11)

It is clear now that we need a sampling scheme to generate conditional distributions of *F*′ and *X*  with

$$
x_t/F, x_{t-1}, \Sigma_u \sim N(Fx_{t-1}, \Sigma_u) \tag{12}
$$

$$
F'/X, \Sigma_u
$$
  
\sim [k(n, p)]<sup>-1</sup>|A<sup>(n-p)</sup>/2 $\left|X'_{n-1}X_{n-1}\right|^{p/2}$  $\left|A + FX'_{n-1}X_{n-1}F'\right|^{-n/2}$  (13)

The key to the above sampling scheme leads itself naturally to Gibbs sampler. Gibbs sampler is a technique for generating random variables from a distribution indirectly without deriving the

density. Given an arbitrary starting set of  $\{Z_1^{(0)}, Z_2^{(0)},..., Z_k^{(0)}\}$  $Z_1^{(0)}, Z_2^{(0)},..., Z_k^{(0)}\}$ , we draw

$$
Z_1^{(1)} \sim [Z_1 / Z_2^{(0)}, Z_3^{(0)},..., Z_k^{(0)}],
$$
  
\n
$$
Z_2^{(1)} \sim [Z_2 / Z_1^{(0)}, Z_3^{(0)},..., Z_k^{(0)}],
$$
  
\n
$$
Z_3^{(1)} \sim [Z_3 / Z_1^{(0)}, Z_2^{(0)}, Z_4^{(0)},..., Z_k^{(0)}],......
$$
  
\n
$$
Z_k^{(1)} \sim [Z_k / Z_1^{(0)}, Z_2^{(0)},..., Z_{k-1}^{(0)}].
$$

Thus each variable is visited in the natural order and k random generations forms an iteration. After m such iterations we have  $\{Z_1^{(m)}, Z_2^{(m)},..., Z_k^{(m)}\}$  $(m)$ 1 *m*  $Z_1^{(m)}$ ,  $Z_2^{(m)}$ , ...,  $Z_k^{(m)}$  }. Under mild conditions, the following results hold.[13]

GG 1 (Convergence).

visited infinitely often.

 ${Z_1^{(m)}, Z_2^{(m)},..., Z_k^{(m)}}$  $(m)$ 1 *m*  $Z_1^{(m)}, Z_2^{(m)},..., Z_k^{(m)}$  { $Z_1, Z_2,..., Z_k$ } and hence for each s,  $Z_s^{(m)}$  [ $Z_s$ ] as m ∞. In fact a slightly stronger result is proven. Rather than requiring that each variable be visited in repetitions of the natural order, convergence still follows any visiting scheme, provided that each variable is

#### GG 2 (Rate)*.*

Using the sup norm, rather than the L<sub>1</sub> norm, the joint density of  $\{Z_1^{(m)}, Z_2^{(m)},..., Z_k^{(m)}\}$  $(m)$ 1 *m*  $Z_1^{(m)}$ ,  $Z_2^{(m)}$ , ...,  $Z_k^{(m)}$ } converges to the true density at a geometric rate in m, under visiting in the natural order. A minor adjustment to the rate is required for an arbitrary visiting scheme.

GG 3 (Ergodic theorem).

For any measurable function T of  $Z_1, Z_2, ..., Z_k$  whose expectation exits

$$
\lim_{m\to\infty}\frac{1}{m}\sum_{l=1}^m T(Z_1^{(l)},Z_2^{(l)},...,Z_k^{(l)})\to E(T(Z_1,Z_2,...,Z_k))\,.
$$

As Gibbs sampling through *r* replications of the aforementioned *m* iterations produces *r* i.i.d. k tuples  $(Z_{1j}^{(m)}, Z_{2j}^{(m)},..., Z_{kj}^{(m)})(j = 1,2,...,r)$  $Z_{1j}^{(m)}, Z_{2j}^{(m)},..., Z_{kj}^{(m)}$ )(j = 1,2,..., r *m j*  $J_j^{(m)}, Z_{2j}^{(m)},..., Z_{kj}^{(m)}$   $(j = 1, 2,..., r)$ , which the proposed density estimate for  $[Z_s]$  having form

$$
[\hat{Z}_s] = \frac{1}{r} \sum_{j=1}^r [Z_s / Z_r^{(j)}, r \neq s].
$$
\n(14)

# **3 Solution Algorithm**

In the context of our state space model, samples from the conditional distributions by the sampling shceme mentioned in section 2.2 to implement the Gibbs sampler. Since the observation information is not used in the conditional distributions, the proposed algorithm combines the Kalman filter and Gibbs sampler and is briefly illustrated as the common algorithmic format.

• Step 1 (Initialization)

1. Use prior information to generate  $F^{(0)}$ 2.Given Σ, Γo, Γd 3.Given  $x0 \sim N(\mu 0, V0)$ 

• Step 2 ( Generate  $x_t^{(g)}$ ,  $t = 0, 1, 2, ..., n$  )

1.Generate  $x_0^{(g)}$  from N( $\mu$ 0,V0)

2. Generate  $x_1^{(g)}$  $x_1^{(g)}$  from  $x_1/x_0^{(g)}$ ,  $F^{(g)} \sim N(F^{(g)}x_0^{(g)}, \Sigma)$  $\left( x_1 \middle/ x_0^{(g)}, F^{(g)} \right) \sim N(F^{(g)} x_0^{(g)}, \Sigma)$ 

3.Use Kalman filter to filter  $x_t^{(g)}$ 

4. Repeat 2,3 for  $t = 2, 3, ..., n$ 

• Step 3 ( Generate  $F^{(g)}$  )

1.Calculate  $(X_{n(i)}^{(g)} - X_{n-1}^{(g)} \hat{F}'^{(g)}_{i})' (X_{n(i)}^{(g)} - X_{n-1}^{(g)} \hat{F}'^{(g)}_{i})$  ${a_{ii}^{(g)}}$  $(g) \hat{\mathbf{r}}$  $\prime$  $(g)$ 1  $(g)$  $(j)$  $(g) \hat{\mathbf{r}}$  $\prime$  $(g)$ 1  $(g)$  $(i)$  $(g)$  $(g)$   $(g)$ *g j g n g n j g i g n g n i g ij g ij g*  $a_{ii}^{(g)} = (X_{n(i)}^{(g)} - X_{n-1}^{(g)} \hat{F}'^{(g)}_{i})'(X_{n(i)}^{(g)} - X_{n-1}^{(g)} \hat{F}$  $A^{(g)} = \{a$  $=(X_{n(i)}^{(g)}-X_{n-1}^{(g)}\hat{F}'^{(g)}_{i})'(X_{n(i)}^{(g)}-X_{n-1}^{(g)}\hat{F}'_{i})$ =  $-1 \mathbf{r}_i \quad \mathbf{A}_{n(j)} - \mathbf{A}_{n-j}$ 

$$
\hat{F}'^{(g)}_i = (X'^{(g)}_{n-1}X^{(g)}_{n-1})^{-1}X'^{(g)}_{n-1}X^{(g)}_{n(i)}
$$

- 2. Calculate  $X_{n-1}^{\prime (g)} X_{n-1}^{(g)}$  $(g)$ 1 *g*  $X_{n-1}^{\prime (g)} X_{n-1}^{(g)}$
- 3. Generate w ~ Wishart ( $X_{n-1}^{\prime (g)} X_{n-1}^{(g)}$  $(g)$ 1 *g*  $X_{n-1}^{\prime (g)} X_{n-1}^{(g)}$ , n-p)

4.Generate

$$
Z = (z'_1, z'_2, ..., z'_p)'
$$
  
where  $z_k \sim \frac{iid}{N_p(0, A^{(g)})}$ 

- 5. Generate  $F'(g) = ((w1/2)')-1$  Z.
- Step 4 (Iteration)

Repeat Step 2, Step 3 m times.

# **4 Numerical results**

Data resources in this paper is from O-D counts of Nangkang Line provided by Taipei Rapid Transit Corperation. There are nine stations, between Taipei City Hall and Lungshan Temple, considered and each is both origin node and destination node except two terminal nodes. From the topology of nodes, the selected number of O-D pairs is fifteen for one direction. O-D counts of 5-minute time period from 6:30am to 9:30am are collected. Then we show the estimations of O-D pattern and the observed O-D data on Table 1.



Table 1. O-D Estimations vs. Real O-D Data







The results shown above are generally satisfactory with estimated O-D pattern fitting with real data. And the value of root mean square error (RMSE) is evaluated as 3. However, it is clear that for those O-D pairs with very small or even zero counts illustrated here are not good enough.

This might come from the assumption of Gaussian noise terms. Or due to the limitation of computer memory, the iteration number in Step 4 is not enough to reach the convergent state. So it reminds us two things at least. First, non-Gaussian assumption would be better theoretically. The second one is that the Monte Carlo-like simulation process to reach the convergent state is deserved to further research both in theoretical and numerical analysis.

# **5 Conclusions**

This paper provides a method of estimating time varying origin-destination matrices for the rapid transit system by using the Gaussian state space model with an unknown transition matrix. An algorithm illustrates the process that combines the Kalman filter and Gibbs sampler to simultaneously estimate origin-destination matrices and transition matrix. Performance of the proposed approach is evaluated by the real observed passenger counts data. Preliminary results are generally satisfactory, showing that also in the unknown transition matrix case, significant estimates could be obtained. However, non-Gaussian assumption in noise terms and the convergent behavior of MonteCarlo-like simulation process are deserved to futher study both in theoretical and numerical analysis. Extentions to prediction process by considering real-time updating information and the hierarchy-based estimation algorithm to relax the limitation on the entry-exit-count specific network are also valuable issues in future research.

### **References**

[1] Ashok, K. and Ben-Akiva, M. E., "Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows", Transp. Sci. 34, pp. 21-36, 2000.

[2] Bell, M. G. H., "The Real Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion", Transp. Res.25, pp. 115-125, 1991.

[3] Cascetta, E. and Nguyen, S., "A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts", Transp. Res. –B 22B, pp. 437-455, 1988.

[4] Cascetta, E., Inaudi, D. and Marquis, G., "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts", Transp. Sci. 27, pp. 363-373, 1993.

[5] Chang, G. L. and Wu, J., "Recursive Estimation of Time-Varying O-D Flows from Traffic Counts in Freeway Corridors", Transp. Res. –B 28B, pp. 141-160, 1994.

[6] Chang, G. L. and Tao, X., "An Integrated Model for Estimating Time-Varying Network Origin-Destination Distributions", Transp. Res. –A 33A, pp. 381-399, 1999.

[7] Cremer, M.and Keller, H., "A New Class of Dynamic Methods for Identification of Origin-Destination Flows", Transp. Res.21, pp. 117-132, 1987.

[8] Geman, S. and Geman, D., "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", IEEE, Trans. Pat. Anal. Mach. Intel. 6, 721-741 1984.

[9] Hu, S. R., "An Adaptive Kalman Filtering Algorithm for the Dynamic Estimation and Prediction of Freeway Origin-Destination Matrices", Ph.D. dissertation, Purdue University, 1996.

[10] Nihan, N. L. and Davis, G. A., "Recursive Estimation of Origin-Destination Matrices from Input/Output Counts", Transp. Res.21, pp. 149-163, 1987

[11] Okutani, I., "The Kalman Filtering Approach in Some Transportation and Traffic Problems", In Transportation and Traffic Theory, N. H. Gartner and N. H. M. Wilson (eds), Elsevier, New York, pp. 397-416, 1987.

[12] Wu, J., "A Real-Time Origin-Destination Matrix Updating Algorithm for On-Line Applications", Transp. Res. –B 31B, pp. 381-396, 1997.

Gibbs Sampler

 $\overline{\text{IEEE}}$