ESTIMATION OF DYNAMIC ASSIGNMENT
MATRICES AND OD DEMANDS USING ADAPTIVE
KALMAN FILTERING

Shou-Ren Hu
Assistant Professor, Department Of Transportation Management, Tamkang University,
Taipei, Taiwan

Samer M. Madanat
Associate Professor, Department of Civil and Environmental Engineering,
University of California, 114 McLaughlin Hall, Berkeley, CA 94720

James V. Krogmeier
Associate Professor, School of Electrical and Computer Engineering, Purdue
University, 338 Materials and Electrical Engineering Building, West Lafayette,
IN 47907

Srinivas Peeta
Associate Professor, School of Civil Engineering, Purdue University, 1284 Civil
Engineering Building, West Lafayette, IN 47907

The purpose of this research was to develop a dynamic model for the on-line
estimation and prediction of freeway users’ origin-destination (OD) matrices. In this
paper, we present a Kalman Filtering algorithm that uses time-varying assignment
matrices generated by using a mesoscopic traffic simulator. The use of a traffic simulator
to predict time-varying travel time model parameters was shown to be promising for the
determination of dynamic OD matrices for a freeway system. Moreover, the issues of
using time-varying model parameters, effects of incorporating different sources of
measurements and the use of adaptive estimation are addressed and investigated in this
research.

Key words: adaptive filters, Kalman filtering, optimal estimation, origin-destination
demands, traffic simulator.

INTRODUCTION

The problem of dynamic estimation and prediction of transportation network users’
orIGIN-DESTINATION MATRICES AND OD DEMANDS USING ADAPTIVE
KALMAN FILTERING

Shou-Ren Hu
Assistant Professor, Department Of Transportation Management, Tamkang University,
Taipei, Taiwan

Samer M. Madanat
Associate Professor, Department of Civil and Environmental Engineering,
University of California, 114 McLaughlin Hall, Berkeley, CA 94720

James V. Krogmeier
Associate Professor, School of Electrical and Computer Engineering, Purdue
University, 338 Materials and Electrical Engineering Building, West Lafayette,
IN 47907

Srinivas Peeta
Associate Professor, School of Civil Engineering, Purdue University, 1284 Civil
Engineering Building, West Lafayette, IN 47907

The purpose of this research was to develop a dynamic model for the on-line
estimation and prediction of freeway users’ origin-destination (OD) matrices. In this
paper, we present a Kalman Filtering algorithm that uses time-varying assignment
matrices generated by using a mesoscopic traffic simulator. The use of a traffic simulator
to predict time-varying travel time model parameters was shown to be promising for the
determination of dynamic OD matrices for a freeway system. Moreover, the issues of
using time-varying model parameters, effects of incorporating different sources of
measurements and the use of adaptive estimation are addressed and investigated in this
research.

Key words: adaptive filters, Kalman filtering, optimal estimation, origin-destination
demands, traffic simulator.

INTRODUCTION

The problem of dynamic estimation and prediction of transportation network users’
origin-destination matrices has received increasing attention in view of its applicability to on-
line traffic management systems. These time-dependent OD matrices are important inputs to
Advanced Traffic Management and Information Systems (ATMIS). Specifically, with the
information contained in the time-varying OD matrices, it is possible to project network users’
travel demands up to a time horizon of interest and pre-determine optimal control and routing
policies that achieve some desirable system-wide objectives. For instance, with the predictions
of time-varying OD demands, traffic control center is able to forecast future traffic conditions
and predict congestion so that appropriate control actions (e.g., ramp metering, variable message signing) can be determined, and effective traffic information can be provided to motorists. Therefore, an effective model for the dynamic estimation and prediction of transportation network users’ OD trip demands is crucial to advanced traffic management systems.

In general, long-term average OD trip demands are needed for transportation planning purposes (e.g., future network expansion and urban planning). On the other hand, short-term time-varying OD demands are important inputs to traffic control center for traffic management purposes. These OD trip demands are traditionally obtained by home-interview surveys and/or license plate surveys, which are highly expensive and time consuming. Another economic source of information to infer network OD demands is the automatically recorded link traffic counts. Since link traffic counts are the measurements of various OD flows using these links, the information contained in the measured link traffic counts is possible to infer the unknown OD demands. The topic of combining different sources of information to determine time-dependent OD demands has been investigated for the past decade (Ben-Akiva et al., 1985; Yang et al., 1992; Lo et al., 1996; van der Zijpp, 1997).

The estimation of network users’ OD demands involves two key issues: spatial issue of route choice and temporal issue of traffic dispersion. Specifically, for a given OD pair, there may exit more than one path between this OD pair. The issue of route choice needs to be investigated. On the other hand, link traffic counts measured during a certain time interval are composed of contributions from various OD flows corresponding to departures during prior time intervals; therefore the temporal dispersion of traffic movements needs to be studied. Depending on the structure of the target network, the estimation of networks users’ OD demands involves partially or fully these two issues. Table 1 shows these relationships.

<table>
<thead>
<tr>
<th></th>
<th>Intersections</th>
<th>Freeways</th>
<th>General Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial-Route Choice</td>
<td>×</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>Temporal-Traffic Dispersion</td>
<td>×</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

The state-of-the-art methods for the dynamic estimation and prediction of network OD matrices mainly focus on linear systems; i.e., turning proportions at intersections or OD demands for freeway systems. Some of these methods assume that the time taken to traverse a given OD pair is either small and can be ignored (Cremer and Keller, 1987; Nihan and Davis, 1987, 1989; Cascetta and Nguyen, 1988) or is equal to a fixed number of time intervals (Nihan and Hamed, 1992). Such assumptions are not suitable for freeways or general networks, especially in the presence of traffic congestion, where travel times are significant and time dependent (Ashok and Ben-Akiva, 1993). Bell (1991) incorporated the effect of traffic dispersion on the dynamic estimation of network OD demands. He proposed two recursive estimators that sequentially estimate the time-dependent platoon dispersion fractions and the corresponding OD proportions for a freeway system. Later research (Chang and Wu, 1994; van der Zijpp and Hamerslag, 1994) recognized the potential improvement of using time-varying model parameters in the dynamic estimation and prediction of network OD demands.

Ashok and Ben-Akiva (2000) presented two state-space models for real-time estimation / prediction of time-dependent OD flows. Instead of defining the state-vector as the OD flows
themselves, the first model defines the state-vector in terms of the deviations in OD demands, while the second model defines it as the deviations of departure rates from each origin and the shares headed to each destination. Preliminary test results demonstrate that such formulations make the real-time estimation process computationally tractable. Moreover, the second formulation yields better predictions with some loss of accuracy in filtered estimates.

This paper presents an adaptive Kalman Filtering algorithm for the dynamic estimation of freeway motorists’ OD demands using time-varying link traffic count information. This paper does not address the problem of estimation of OD matrices in networks that allow multiple routes. One important aspect of the proposed adaptive estimator is the use of a traffic simulator to predict time-varying travel times, that are used to compute time-varying assignment fractions. The proposed dynamic model was evaluated through simulation experiments due to the difficulty in acquiring real-world OD data and time-varying travel times. Preliminary test results show the capability of the proposed adaptive estimator for the dynamic estimation and prediction of freeway OD demands. The issues of using time-varying model parameters, effects of incorporating different sources of measurements and the use of adaptive estimation are discussed and investigated in the present research. The remainder of this paper is organized as follows. Section 2 provides the background information about this problem. Section 3 describes the methodology of the simulation-based adaptive Kalman Filtering algorithm. A set of simulation evaluation results and conclusions are presented in sections 4 and 5, respectively.

BACKGROUND

Problem Statement

Time-dependent OD matrices represent network users’ trip demands given network traffic conditions. For a freeway system, the demand for a given OD pair represents the number of vehicles entering the freeway segment from an upstream entrance and destined to a downstream exit. Time-varying link traffic counts (e.g., entrance, exit, and mainline section counts) are available for some links from traffic surveillance systems. The problem is to determine time-dependent OD matrices given discrete time series of entrance, exit, and mainline traffic counts for a freeway system. Figures 1 shows the observed freeway section traffic counts.

Basic Relations

To specify the problem, the following variables are defined:
\[ q_k(i) \]: the number of vehicles entering the freeway segment from entrance \( i \) during time interval \( k, i=1, 2, \ldots, N-1; \)
\[ y_k(j) \]: the number of vehicles leaving the freeway segment from exit \( j \) during time interval \( k, j=2, 3, \ldots, N; \)
\[ m_k(l) \]: the number of vehicles crossing the freeway mainline section \( l \) during time interval \( k, l=2, 3, \ldots, N-1; \)
$x_i(i,j)$: the number of vehicles entering the freeway segment from entrance $i$ during time interval $k$ that are destined to exit $j$;

$a'_i(i,j,d)$: the fraction of $x_i(i,j)$ that arrives at the measurement location $d$ (either exit $j$ or mainline section $l$) during time interval $k$, $r \leq k$.

FIGURE 1 Observed freeway section counts.

The state variables to be estimated and predicted in this research are the time-dependent OD flows, $x_i(i,j)$. Moreover, in view of platoon dispersion of traffic flow, the assignment fraction variables $a'_i(i,j,d)$ are defined to capture the temporal relationship between the time-dependent OD flows and the observed link traffic counts.

The basic relations in the dynamic traffic system are described below.

Assignment Proportion Constraints. The arrival times for $x_i(i,j)$ are distributed among time intervals $r$, $r+1$, . . . , $r+u$, where $u$ is the maximum number of time intervals taken to travel between any OD pair of the freeway segment. Theoretically, it follows that:

$$0 \leq a'_i(i,j,d) \leq 1; 1 \leq i < j \leq N; r \leq k$$

$$\sum_{k=r}^{r+u} a'_i(i,j,d) = 1; 1 \leq i < j \leq N$$

Link Flow Conservation. The entrance traffic counts are given by:

$$q_k(i) = \sum_{j=i}^{N} x_k(i,j)$$

The exit traffic counts during time interval $k$ are composed of contributions from all upstream OD flows corresponding to departures during time intervals $k$, $k-1$, . . . , $k-u$. In other words, there exits a "time lag" effect when distributing the OD flows to the corresponding exit traffic flows. Mathematically, this relationship can be expressed as follows:

$$y_k(j) = \sum_{r=k-u}^{k} \sum_{i=1}^{j-1} a'_k(i,j,d)x_r(i,j)$$

Similarly, the freeway mainline section traffic counts are given by:
Equations (3) through (5) provide the relationships between the dynamic OD flows and the observed link traffic counts. Let \( z_k \) be an \( m \)-vector of link traffic counts measured at time interval \( k \), \( A_k \) and \( X_k \) represent the \((m \times ns)\) augmented \(^\dagger\) assignment matrix and the \( ns \)-vector of augmented OD flows at time interval \( k \), respectively. The above equations can be simplified as a single measurement equation as follows:

\[
z_k = [q_k, y_k, m_k]^T = A_k X_k
\]

(6)

This measurement equation plays an important role in the determination of dynamic OD flows. Specifically, if \( z_k \) and \( A_k \) are known in each time interval, it is possible to infer the unknown OD flows, \( X_k \) through various estimators.

**METHODOLOGY**

**System Description**

The dynamic traffic system is represented by a linear, finite-dimensional stochastic system as depicted in Figure 2.

![Diagram](source: Anderson and Moore, 1979)

FIGURE 2 Basic dynamic model (Source: Anderson and Moore, 1979).

This system can be described by the following state-space equations:

\[
x_{k+1} = \sum_{r=k-p+1}^{k} f_k^r x_r + w_k
\]

(7)

\[
z_k = \sum_{r=k-u}^{k} a_k^r x_r + v_k
\]

(8)

where \( x_{k+1} \) is an \( n \)-vector of OD flows at time \( k+1 \), and \( z_k \) is an \( m \)-vector of link traffic counts measured at time \( k \) \((m \leq n)\). \( f_k^r \) is the \((n \times n)\) transition matrix which describes the effects of previous OD flows \( x_r \) on current OD flows \( x_{k+1} \), and \( a_k^r \) is the \((m \times n)\) assignment matrix whose

---

\(^\dagger\)To capture the temporal relationship of the dynamic OD flows and the observed link traffic counts, an augmented OD flow vector \((ns \times 1)\) which includes the OD flow vectors of all prior intervals up to \( s \) is introduced, where \( n \) is the number of OD pairs to be estimated at each time interval and \( s \) will be given later.
entries specify the contributions of OD flows $x_r$ to $z_k$. $w_k$ and $v_k$ are the n-vector input (random) and m-vector output (measurement) noise processes, respectively. These noise processes are assumed to be individually white noise and gaussian processes with zero means and known covariance matrices, i.e., $Q_k$ and $R_k$. It is also assumed that $w_k$ and $v_k$ are independent processes. Moreover, the initial state $x_0$ is assumed to be a gaussian random variable of known mean and covariance, i.e., $\bar{x}_0$ and $p_0$. Furthermore, equation (7) is termed the transition equation and is formulated as an autoregressive model of order $p$. Equation (8) is termed the measurement equation which describes the temporal relationships between the dynamic OD flows and the observed link traffic counts, and $u$ is the maximum number of time intervals required to travel between any OD pair of the entire freeway segment.

Given the above dynamic system, the aim of the filtering problem is to find an estimate of $x_k$ using measurements $z_0, z_1, \ldots, z_k$. Mathematically, we seek to recursively estimate the system state variables at each time interval:

$$\hat{x}_{k|k-1} = E[x_k | z_0, z_1, \ldots, z_{k-1}] = E[x_k | Z_{k-1}]$$

and their corresponding error covariance matrices:

$$\sum_{k|k-1} = E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T]$$

$$\sum_{k|k} = E[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T]$$

where $\hat{x}_{k|k-1}$ is the state estimate given observations up to time $k-1$ and is termed the "one-step ahead" prediction, while $\hat{x}_{k|k}$ is the state estimate given observations up to time $k$ and is termed the "filtered" estimate (Anderson and Moore, 1979).

Solution of the Kalman Filtering Problem

Given the initial conditions, $\bar{X}_0$ and $P_0$, the Kalman filter recursively estimates the state variables by the following equations:

$$\sum_{k|k-1} = F_k \sum_{k-1|k-1} F_k^T + Q_k$$

$$K_k = \sum_{k|k-1} A_k^T [A_k \sum_{k|k-1} A_k^T + R_k]^{-1}$$

$$\sum_{k|k} = \sum_{k|k-1} - K_k A_k \sum_{k|k-1}$$

$$\hat{X}_{k|k-1} = F_k \hat{X}_{k-1|k-1}$$
\[ \hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k [z_k - A_k \hat{X}_{k|k-1}] \]  

where \( \hat{X}_{k|k-1} \) is the ns-vector of augmented "one-step ahead" OD flow predictions, and \( \hat{X}_{k|k} \) is the ns-vector of augmented "filtered" OD flow estimates. Since the transition and measurement equations given in (7) and (8) specify the dependence of the current state and measurement vectors on previous state vectors in more than one preceding interval, an augmented state vector that includes the state vectors of all prior intervals up to \( s = \max(p, u) \) is introduced. Here, \( p \) is the order of the transition equation, which is an autoregressive time series model; and \( u \) is the maximum number of time intervals required to travel between any OD pair of the entire freeway segment. \( F_k \) (ns\( \times \)ns) and \( A_k \) (m\( \times \)ns) are the augmented transition and assignment coefficient matrices, respectively. \( K_k \) (ns\( \times \)m) is the Kalman gain matrix.

The Adaptive Kalman Filtering Algorithm

The Kalman Filtering algorithm described in the previous section requires the knowledge of the assignment matrices \( A_k \), \( \forall k \). These model parameters in the OD flow estimation problem are in fact functions of travel times, since they capture the temporal dispersion of traffic flow. The adaptive Kalman Filtering algorithm is a sequential process that estimates the time-varying model parameters (travel times) and the corresponding dynamic OD flows. Specifically, the adaptive KF estimator uses "one step ahead" OD flow estimates (to be explained later) as feedback to the traffic simulator. At each time interval, with these predicted OD flows, the traffic simulator, which captures traffic realism predicts realistic travel times specifying the assignment fractions. Therefore, the Kalman filter is capable to provide desirable estimates of the unknown OD flows at each time interval. Figure 3 illustrates the proposed adaptive Kalman Filtering algorithm.
As shown in Figure 3, the proposed adaptive Kalman Filtering algorithm estimates the current state vector $\hat{X}_{k \mid k}$ by summing the following components: 1) "one-step ahead" predictions, $\hat{X}_{k \mid k-1}$, and 2) weighted predicted errors from link traffic counts, $e_k$. The Kalman gain matrix $K_k$ is used as a weight matrix to adjust the effect of the prediction errors in link traffic counts. At each time interval, the traffic surveillance system measures a vector of link traffic counts, $z_k$ and current link travel times, $t_k$. Given the current link travel times and historical OD data, the traffic simulator generates the predicted path travel times, $\hat{t}_k$ which are used for the specification of the assignment matrix, $\hat{A}_k$. We use predicted, rather than currently measured, link travel times because the former are closer to the travel times that will actually be experienced by motorists.

An additional feature of the algorithm is the following. In the case of non-recurrent traffic congestion, the predicted travel times based on historical OD data might be highly inaccurate. In such cases, the "one-step ahead" OD flow predictions, $\hat{X}_{k+1 \mid k}$ predicted by the Kalman filter can be fed back to the traffic simulator to predict the time-varying travel time model parameters, as indicated in Figure 3 by the dashed feedback arrow.

**CASE STUDIES**

The methodology developed in the previous section was evaluated using simulation experiments due to the difficulty in acquiring real-world OD data and time-varying travel time information. DYNASMART (Mahmassani et al., 1993), a mesoscopic traffic simulator, has been employed to implement the experiments.
Network Configuration

The test network consists of 8 traffic zones, 20 nodes and 38 directed arcs, with total distance of 6 miles. Each node represents either an origin / destination centroid or a dummy node that connects freeway mainline and entrance / exit ramps. All links are half mile except for the entrance / exit links, which are 0.25 mile. In the case studies, only east bound traffic is considered, thus the number of OD pair is reduced by half, i.e., 10 OD pairs are of interest. Figure 4 shows the configuration of the test network.

FIGURE 4 Test network configuration

Data Description

Given the initial time-varying OD demands and network description, a set of time-varying link traffic counts and travel time information were generated through the simulation experiments. It needs to be pointed out that the simulated link traffic counts and travel time information is collected after a certain simulation warm-up time period, usually it takes 30 minutes. The update frequency was 5 minutes, and the peak period of interest was 50 minutes, i.e., 10 time intervals. In the experiments, three demand levels with the network-loading ratio of 1:1.21:1.52 were generated to investigate the impacts of network loading on the performance of the Kalman Filtering algorithm. In each experimental case, 10 repetitions were made and the average performance was obtained for evaluation purposes.

Initiation

Before implementing the Kalman Filter, the transition equation needs to be specified. We calibrated the coefficient and error covariance matrices of the transition equation by using real-world historical OD data obtained from an early study (Ashok and Ben Akiva, 1993). We have also verified the necessary conditions of a stationary autoregressive model (Pindyck and Rubinfeld, 1991). Moreover, the initial conditions of the filter, $X_0$ and $P_0$ were taken from the same historical data set. This assumption does not seriously affect the performance of Kalman Filter, as the effect of the initial conditions disappears after a few filtering time intervals (Jazwinski, 1970).
Furthermore, since the assignment fractions in the measurement equation are functions of travel times, to specify the assignment fractions, we have obtained the means and variances of travel times at each time interval under the three demand levels. At each estimation time interval, in the case of adaptive estimation, the predicted travel times information given by the traffic simulator is used to specify the assignment matrices.

**Evaluation Results**

**General Performance of the KF Algorithm**

In theory, the KF algorithm provides the unbiased and linear minimum variance estimates when the model specifications and noise statistics are known. To demonstrate these desirable properties, we assume that the model parameters in terms of the time-varying assignment fractions are obtained from the post-simulation results; and the noise statistics in terms of the random and measurement error covariance matrices are obtained from the residual analysis. To evaluate the proposed Kalman Filtering algorithm, two goodness-of-fit performance indices are used: Root Mean Squared Errors (RMSE) and Chi-square ($\chi^2$) statistics, which are defined below:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K} (e_k - o_k)^2}{K}}$$  \hspace{1cm} (18)

$$\chi^2 = \sum_{k=1}^{K} \frac{(e_k - o_k)^2}{e_k}$$  \hspace{1cm} (19)

where $e_k$ and $o_k$ are the estimated and observed OD flows at time interval $k$, respectively, and $K$ is the total number of time intervals.

Table 2 shows the goodness-of-fit of the Kalman Filtering OD flow estimates under the low demand level. Furthermore, Tables 3 and 4 demonstrate the same information as in Table 2 for the medium and high demand levels, respectively. Note that the evaluation results shown in Tables 2 through 4 are obtained by using "true" time-varying model parameters (assignment fractions) given by the observed time-varying travel times after conducting the simulation experiments. From the evaluation results, several important findings are summarized:

(a) In the low and medium demand levels, most of the OD flow estimates given by the KF algorithm are statistically indifferent from the true values at the 95% level of significance.

(b) As travel demand increases to the high demand level, there were significant vehicles queuing at the second and third entrances due to highly traffic congestion. These vehicles might be held until next or later time intervals before entering the freeway segment. Recognizing this physical phenomena, the initial entry link traffic counts were adjusted accordingly to capture the holding vehicles. It was found that the KF algorithm still remains robust since 7 out of 10 OD flow estimates are still statistically indifferent from the observed OD flows. We found larger link traffic count prediction errors in the higher demand levels, which means the residuals in the measurement equations become large.
(c) The RMSEs and $\chi^2$ statistics for OD pair number 10 are both zero under the three demand levels since this OD pair becomes completely measurable after including the last entrance traffic counts information.

(d) We also found that as the travel demand increases, or the average speed decreases, the RMSEs for the OD flow estimates increase.

(e) Finally, in terms of the KF algorithm performance against the distance between the OD pair, we have found, in general, that the RMSE for an OD pair with longer distance is larger than that of a shorter OD pair. This is due to the larger traffic variations for a longer OD pair, which results in the higher prediction errors.

### TABLE 2 Overall performance for the goodness-of-fit under the low demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.05</td>
<td>1.95</td>
<td>3.85</td>
<td>4.28</td>
<td>2.79</td>
<td>5.85</td>
<td>6.57</td>
<td>3.89</td>
<td>3.89</td>
<td>0.00</td>
<td>35.12</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>2.26</td>
<td>4.87</td>
<td>14.3</td>
<td>3.70</td>
<td>3.24</td>
<td>7.33</td>
<td>14.9</td>
<td>65.6*</td>
<td>3.48</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### TABLE 3 Overall performance for the goodness-of-fit under the medium demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.92</td>
<td>3.55</td>
<td>4.10</td>
<td>5.44</td>
<td>3.11</td>
<td>8.11</td>
<td>6.95</td>
<td>4.02</td>
<td>4.02</td>
<td>0.00</td>
<td>42.22</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.05</td>
<td>17.6*</td>
<td>14.0</td>
<td>5.12</td>
<td>3.70</td>
<td>16.4</td>
<td>8.75</td>
<td>13.5</td>
<td>6.47</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### TABLE 4 Overall performance for the goodness-of-fit under the high demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>3.71</td>
<td>4.21</td>
<td>3.49</td>
<td>6.72</td>
<td>3.11</td>
<td>10.6</td>
<td>8.77</td>
<td>4.83</td>
<td>4.83</td>
<td>0.00</td>
<td>50.27</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>4.09</td>
<td>10.6</td>
<td>7.34</td>
<td>13.1</td>
<td>4.88</td>
<td>20.4*</td>
<td>23.7*</td>
<td>70.6*</td>
<td>7.74</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### The Effect of Using Time-varying Model Parameters

To investigate the effect of using time-dependent model parameters on the performance of the KF algorithm, we compare the predicted OD flows with a constant speed assumption to those with measured time-varying speeds under different demand levels. Table 5 compares the estimation errors in terms of the RMSEs for the OD flow estimates with constant and time-varying model parameters under the low demand level. The numbers shown in the first and second rows of Table 5 represent the RMSEs with constant model parameters, and the numbers given in the third row are the RMSEs with time-varying model parameters. The RMSEs shown in the second row are obtained by using the calibrated time-varying measurement error covariance matrices $R_k$, while the RMSEs in the first row are computed with fixed error covariance matrix $R, \forall k$. 

[Translation Details]

(c) The RMSEs and $\chi^2$ statistics for OD pair number 10 are both zero under the three demand levels since this OD pair becomes completely measurable after including the last entrance traffic counts information.

(d) We also found that as the travel demand increases, or the average speed decreases, the RMSEs for the OD flow estimates increase.

(e) Finally, in terms of the KF algorithm performance against the distance between the OD pair, we have found, in general, that the RMSE for an OD pair with longer distance is larger than that of a shorter OD pair. This is due to the larger traffic variations for a longer OD pair, which results in the higher prediction errors.

### TABLE 2 Overall performance for the goodness-of-fit under the low demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.05</td>
<td>1.95</td>
<td>3.85</td>
<td>4.28</td>
<td>2.79</td>
<td>5.85</td>
<td>6.57</td>
<td>3.89</td>
<td>3.89</td>
<td>0.00</td>
<td>35.12</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>2.26</td>
<td>4.87</td>
<td>14.3</td>
<td>3.70</td>
<td>3.24</td>
<td>7.33</td>
<td>14.9</td>
<td>65.6*</td>
<td>3.48</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### TABLE 3 Overall performance for the goodness-of-fit under the medium demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.92</td>
<td>3.55</td>
<td>4.10</td>
<td>5.44</td>
<td>3.11</td>
<td>8.11</td>
<td>6.95</td>
<td>4.02</td>
<td>4.02</td>
<td>0.00</td>
<td>42.22</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.05</td>
<td>17.6*</td>
<td>14.0</td>
<td>5.12</td>
<td>3.70</td>
<td>16.4</td>
<td>8.75</td>
<td>13.5</td>
<td>6.47</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### TABLE 4 Overall performance for the goodness-of-fit under the high demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>3.71</td>
<td>4.21</td>
<td>3.49</td>
<td>6.72</td>
<td>3.11</td>
<td>10.6</td>
<td>8.77</td>
<td>4.83</td>
<td>4.83</td>
<td>0.00</td>
<td>50.27</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>4.09</td>
<td>10.6</td>
<td>7.34</td>
<td>13.1</td>
<td>4.88</td>
<td>20.4*</td>
<td>23.7*</td>
<td>70.6*</td>
<td>7.74</td>
<td>0.00</td>
<td>-</td>
</tr>
</tbody>
</table>

*$\chi^2 > \chi^2_{.95,9} = 16.69$

### The Effect of Using Time-varying Model Parameters

To investigate the effect of using time-dependent model parameters on the performance of the KF algorithm, we compare the predicted OD flows with a constant speed assumption to those with measured time-varying speeds under different demand levels. Table 5 compares the estimation errors in terms of the RMSEs for the OD flow estimates with constant and time-varying model parameters under the low demand level. The numbers shown in the first and second rows of Table 5 represent the RMSEs with constant model parameters, and the numbers given in the third row are the RMSEs with time-varying model parameters. The RMSEs shown in the second row are obtained by using the calibrated time-varying measurement error covariance matrices $R_k$, while the RMSEs in the first row are computed with fixed error covariance matrix $R, \forall k$. 

TABLE 5 RMSEs with constant and time-varying model parameters under the low demand

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>∑</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.94</td>
<td>4.81</td>
<td>6.39</td>
<td>11.2</td>
<td>3.03</td>
<td>6.90</td>
<td>5.30</td>
<td>3.05</td>
<td>3.05</td>
<td>0.00</td>
<td>47.6</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.11</td>
<td>3.32</td>
<td>5.11</td>
<td>7.75</td>
<td>2.92</td>
<td>9.19</td>
<td>7.92</td>
<td>3.48</td>
<td>3.48</td>
<td>0.00</td>
<td>46.3</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.05</td>
<td>1.95</td>
<td>3.85</td>
<td>4.28</td>
<td>2.79</td>
<td>5.85</td>
<td>6.57</td>
<td>3.89</td>
<td>3.89</td>
<td>0.00</td>
<td>35.1</td>
</tr>
</tbody>
</table>

<sup>a</sup>RMSEs with constant model parameters and fixed error covariance matrix
<sup>b</sup>RMSEs with constant model parameters and time-varying error covariance matrices
<sup>c</sup>RMSEs with time-dependent model parameters and error covariance matrices

Tables 6 and 7 show the similar information under the medium and high demand levels, respectively. As can be seen in the three tables, the estimation error difference increases as the demand level increases. Moreover, the OD flow estimates with time-varying model parameters always outperform those with constant model parameter assumption. However, with the measurement error covariance matrices being calibrated, those OD flow estimates with constant parameters become accurate. The OD flow estimates are relatively close to those with time-varying parameters. These findings are important since we can always gain improvements in the OD flow estimation accuracy by calibrating the measurement error covariance matrices. These results also reflect the statement by various researchers (Nihan and Davis, 1987; Ashok and Ben-Akiva, 1993) that the KF procedure is fairly robust with respect to inaccuracies in the assignment matrices or system design parameters, under the assumption of known noise statistics.

TABLE 6 RMSEs with constant and time-varying model parameters under the medium demand

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>∑</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.90</td>
<td>5.74</td>
<td>10.7</td>
<td>16.4</td>
<td>6.47</td>
<td>16.4</td>
<td>18.7</td>
<td>5.02</td>
<td>5.02</td>
<td>0.00</td>
<td>88.3</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.57</td>
<td>5.30</td>
<td>6.28</td>
<td>11.1</td>
<td>2.95</td>
<td>6.84</td>
<td>6.57</td>
<td>2.24</td>
<td>2.24</td>
<td>0.00</td>
<td>46.1</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.92</td>
<td>3.55</td>
<td>4.10</td>
<td>5.44</td>
<td>3.11</td>
<td>8.11</td>
<td>6.95</td>
<td>4.02</td>
<td>4.02</td>
<td>0.00</td>
<td>42.2</td>
</tr>
</tbody>
</table>

TABLE 7 RMSEs with constant and time-varying model parameters under the high demand

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>∑</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10.1</td>
<td>11.4</td>
<td>11.0</td>
<td>25.5</td>
<td>5.53</td>
<td>39.7</td>
<td>43.9</td>
<td>14.8</td>
<td>14.8</td>
<td>0.00</td>
<td>177</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.17</td>
<td>6.66</td>
<td>5.65</td>
<td>11.5</td>
<td>4.80</td>
<td>13.6</td>
<td>16.4</td>
<td>3.66</td>
<td>3.66</td>
<td>0.00</td>
<td>68.1</td>
</tr>
<tr>
<td>RMSE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.71</td>
<td>4.21</td>
<td>3.49</td>
<td>6.72</td>
<td>3.11</td>
<td>10.6</td>
<td>8.77</td>
<td>4.83</td>
<td>4.83</td>
<td>0.00</td>
<td>50.3</td>
</tr>
</tbody>
</table>

The Effect of Using Different Sources of Measurements

Measurements obtained from various sources convey different information due to different degrees of measurement errors or model specification errors. For instance, the entry traffic counts provide error-free information since there is no "time lag" effect inherent in the entry traffic count measurements. On the other hand, mainline and exit traffic counts suffer from different degrees of errors due to the "time lag" effect. This section demonstrates the performance of the KF algorithm by incorporating different sources of measurements under
various demand levels in Figure 5. The first 4 traffic counts shown in this figure are the entrance traffic counts, while the rest are the exit traffic counts.

As demonstrated in Figure 5, the RMSEs monotonously decrease as more entrance traffic counts are incorporated. Further performance improvements can be achieved by including the first two exit traffic counts. Namely, the smallest RMSEs are found by incorporating the first 6 traffic counts’ information. It seems that the last two exit traffic counts convey less useful information. In view of these two exits’ locations, the exit traffic counts measured at these two exits suffer from significant "time lag" effects, which result in the inaccurate specification of assignment fractions. It is noted that the "time lag" effect has significant influence on the specification of model parameters (assignment fractions in our case) instead of measurement errors. It has also been reported (Lo et al., 1996) that the additional traffic count information might not necessarily help to improve the estimation accuracy due to wrongly specified model parameters. In summary, the entry traffic counts always provide useful information in determining time-varying OD flows in view of their error-free characteristics. On the other hand, some exit traffic counts may convey limited information.

FIGURE 5 The performance of the KF algorithm versus the number of measurements incorporated under different demand levels.

**Performance of the Adaptive KF Algorithm**

We have shown in the previous section that the quality of the model parameters plays an important role in the dynamic estimation of freeway OD flows. Specifically, accurate time-varying travel times are essential to compute the model parameters. To demonstrate the performance of the proposed adaptive KF estimator, we compare the OD flow estimates obtained by using historical travel time parameters to those given by the adaptive KF estimator. Table 8 shows the goodness-of-fit performance index, RMSE for the OD flow estimates based
on different types of model parameters under the low demand level. For comparison purposes, the RMSEs with measured true travel time parameters are also given in Table 8.

As can be seen in Table 8, when the travel demand is low, there is essentially no difference, in a statistical sense, among the OD flow estimates given by these three Kalman Filtering estimators. Tables 9 and 10 show the same information for the medium and high demand levels, respectively.

TABLE 8 Comparison of the OD flow estimates with different types of model parameters under the low demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE⁠&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.88</td>
<td>2.68</td>
<td>5.08</td>
<td>5.84</td>
<td>2.51</td>
<td>5.41</td>
<td>4.57</td>
<td>2.83</td>
<td>2.83</td>
<td>0.00</td>
<td>34.6</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.85</td>
<td>3.07</td>
<td>5.51</td>
<td>7.29</td>
<td>2.19</td>
<td>3.82</td>
<td>4.15</td>
<td>3.19</td>
<td>3.19</td>
<td>0.00</td>
<td>35.3</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.05</td>
<td>1.95</td>
<td>3.85</td>
<td>4.28</td>
<td>2.79</td>
<td>5.85</td>
<td>6.57</td>
<td>3.89</td>
<td>3.89</td>
<td>0.00</td>
<td>35.1</td>
</tr>
</tbody>
</table>

RMSEs of the OD flow estimates with historical model parameters
RMSEs of the OD flow estimates given by the adaptive KF estimator
RMSEs of the OD flow estimates with measured travel time parameters

TABLE 9 Comparison of the OD flow estimates with different types of model parameters under the medium demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE⁠&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.82</td>
<td>4.53</td>
<td>5.11</td>
<td>9.21</td>
<td>2.61</td>
<td>10.4</td>
<td>9.78</td>
<td>8.57</td>
<td>8.57</td>
<td>0.00</td>
<td>60.6</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.10</td>
<td>4.28</td>
<td>5.26</td>
<td>8.17</td>
<td>3.33</td>
<td>6.99</td>
<td>5.35</td>
<td>2.93</td>
<td>2.93</td>
<td>0.00</td>
<td>43.3</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.92</td>
<td>3.55</td>
<td>4.10</td>
<td>5.44</td>
<td>3.11</td>
<td>8.11</td>
<td>6.95</td>
<td>4.02</td>
<td>4.02</td>
<td>0.00</td>
<td>42.2</td>
</tr>
</tbody>
</table>

TABLE 10 Comparison of the OD flow estimates with different types of model parameters under the high demand level

<table>
<thead>
<tr>
<th>OD #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE⁠&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7.69</td>
<td>6.45</td>
<td>8.42</td>
<td>12.4</td>
<td>3.49</td>
<td>19.4</td>
<td>12.5</td>
<td>10.4</td>
<td>10.4</td>
<td>0.00</td>
<td>91.2</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.34</td>
<td>3.25</td>
<td>6.18</td>
<td>7.05</td>
<td>3.61</td>
<td>10.2</td>
<td>10.8</td>
<td>7.23</td>
<td>7.23</td>
<td>0.00</td>
<td>60.9</td>
</tr>
<tr>
<td>RMSE⁠&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.71</td>
<td>4.21</td>
<td>3.49</td>
<td>6.72</td>
<td>3.11</td>
<td>10.6</td>
<td>8.77</td>
<td>4.83</td>
<td>4.83</td>
<td>0.00</td>
<td>50.3</td>
</tr>
</tbody>
</table>

It can be seen in Table 9 that under the medium demand level, the adaptive KF estimator remains robust with respect to the model parameters, while the quality of the OD flow estimates with historical model parameters deteriorates. The difference in the estimation accuracy of the OD flow estimates becomes more significant for the high demand level, as shown in Table 10.

The above results demonstrate the superiority of the proposed adaptive KF estimator over a KF estimator that uses historical model parameters. The adaptive KF estimator uses "one step ahead" OD flow estimates as feedback to the traffic simulator. With these predicted time-varying OD flows, the traffic simulator, which captures traffic realism, predicts realistic travel times. Compared to the state-of-the-art adaptive estimators that predict the model parameters using statistical methods, the adaptive estimator presented in this research is capable to capture traffic flow realism and provides accurate OD flow estimates. More importantly, in
view of its real-time feedback feature, the proposed adaptive KF estimator is applicable to on-line traffic management systems.

CONCLUSIONS

This research has presented an adaptive Kalman Filtering algorithm for the dynamic estimation and prediction of freeway OD matrices. One major aspect of this research is the use of a traffic simulator as a travel time predictor to obtain time-varying model parameters, even though we recognize that using a simulator adds a random error in the assignment matrix, which introduces a bias in the estimation of OD flows. The proposed methodology has been shown to be promising for the dynamic estimation of freeway OD flows. Specifically, the OD flow estimates given by the adaptive KF estimator have been found, in a statistical sense, indifferent from the OD flow estimates with measured travel time parameters.

The proposed methodology is essentially applicable to linear networks, such as intersections and freeway networks. The dynamic estimation of OD flows from link traffic counts for general network (Cascetta et. al, 1993; Ashok and Ben-Akiva, 2000) is a more challenging problem. An effective model should simultaneously capture both the temporal issue of traffic dispersion and the spatial issue of route choice. This is a topic that is worthy of further investigation.

ACKNOWLEDGEMENTS

The authors would like to thank Hani Mahmassani for providing us with DYNASMART, the traffic simulator used in the simulation experiments. We would also like to thank Moshe Ben-Akiva who provided us with the OD data set for model evaluation purposes and K. Ashok for his inputs to the model development.

REFERENCES


