ASSIGNMENT-MATRIX-FREE DYNAMIC ESTIMATION OF ORIGIN-DESTINATION MATRICES

Rattaphol PUEBOOBPAPHAN1 and Takashi NAKATSUJI2

1Assistant Professor, School of Transportation Engineering, Suranaree University of Technology (111 University Avenue, Muang, Nakhonratchasima, 30000, Thailand) 
E-mail: rattaphol@sut.ac.th

2Member of JSCE, Professor, Graduate School of Engineering, Hokkaido University (Kita 13, Nishi 8, Kita-ku, Sapporo, 060-8628, Japan) 
E-mail: naka@eng.hokudai.ac.jp

This paper presents a new approach for dynamic estimation of origin-destination (OD) matrices based on Unscented Kalman Filter (UKF). The new approach, differs from most of the conventional approaches, does not assume linearity of the original nonlinear relationship between OD parameters and measurements. Instead, the nonlinearity is retained by representing this relationship using dynamic traffic simulation. This eliminates the need to compute the assignment matrix and offers some degrees of flexibility in selecting the traffic simulation model and incorporating speed as additional measurement variables. Preliminary results demonstrate the potential of the proposed approach and the contribution from using speed data.

Key Words : origin-destination matrix, OD flows, dynamic estimation, unscented kalman filter

1. INTRODUCTION

Time-varying origin-destination (OD) matrix is an essential input for on-line traffic control and management systems. Several models have been proposed for estimating dynamic OD parameters using traffic counts. The most crucial part of this problem is how to map the unobserved time varying OD parameters with the observed time-varying traffic counts. There are several approaches ranging from a model of linear relationship between entry and exit flows in the pioneering works of1),2), to the more complex form of using neural network 3). The most common mapping function employed so far is by the use of assignment matrices4),5),6).

There are some critical issues when applying conventional assignment matrix scheme as a mapping function. First, it is very difficult to determine the rational assignment matrix explicitly, particularly in determining the time-dependent link use proportions. This issue arises due to the difficulty of obtaining the time-dependent travel times from all origins to all observation stations. The next issue is the functional form of the mapping equation. Since in congested situation, travel times greatly depend on the path flows which also depend on OD flows5),7), the assignment parameters thus indirectly depend on OD flows. This dependence makes the mapping equation becomes implicitly nonlinear equation. Most of the previous researches, however, consider the mapping function as a linear function by treating the assignment matrix as just a coefficient of the equation. The mean and error covariance of the estimated OD parameters are thus biased as the dependence between the assignment matrix and the OD parameters are completely ignored8).

In this paper, we propose an innovative method for dynamic OD estimation in response to the above issues. The new method, differs from the conventional methods, does not assume linearity of the original nonlinear relationship between OD parameters and traffic counts. This direct nonlinear consideration is accomplished by means of a new filtering technique called Unscented Kalman Filter (UKF) and dynamic traffic simulation model. Natural constraints of the OD estimation problem are also considered explicitly.

The remaining of this paper is organized as follows. Section 2 reviews related literature. Section 3 presents the basic concepts of the dynamic OD estimation in which the issues stated above are discussed explicitly. Since the UKF is relatively new
2. LITERATURE REVIEW

The problem of estimation of time-varying OD from traffic counts has long been a subject of research up until to date. In the pioneering works of Cremer and Keller\(^1\),\(^2\), the time-series of entry and exit flows were employed to transform the under-determined static problem to the over-determined dynamic problem. The linear relationship between entry and exit flows was applied in these models. In subsequent works, Cremer and Keller\(^9\) and Nihan and Davis\(^10\) applied this idea and presented a family of recursive estimation methods including the Kalman filter (KF). This simple formulation between entry and exit flows is, however, only suitable for intersections or small networks because they assume the travel time required for a vehicle to traverse the network is either negligible or equal to a specified number of time intervals.

To make the model suitable for larger network or congestion, more sophisticated models that relax the above assumption, by considering the effect of platoon dispersion and allow the distribution of travel time to span over a number of time intervals, have been considered\(^11\),\(^12\). However, because of dynamic nature, the exact relationship between the OD parameters and traffic counts is very complicated and difficult to write in a closed analytical function. To reduce the complexity, many studies have used assignment proportions to map between the OD parameters and traffic counts\(^13\),\(^14\),\(^15\),\(^16\). The OD parameters may be set as OD proportions\(^5\),\(^11\),\(^13\),\(^14\),\(^15\),\(^16\), OD flows\(^5\),\(^12\),\(^13\) or deviation of OD flows\(^7\),\(^17\).

Based on the use of assignment proportions, the relationship between OD parameters (say OD flows) and traffic counts can be written compactly as shown in Equation (1).

\[
y(k) = \sum_{p-k-p}^{k} a_i(p)x(p) + \xi(k) \tag{1}
\]

where \(y(k)\) denotes the \((n_x \times 1)\) vector of \(y_i(k)\) which is the observed traffic count at detector station \(l\) during interval \(k\) and \(n_x\) is the total number of counting stations, \(x(p)\) denotes the \((n_{OD} \times 1)\) vector of \(x_i(p)\) which is the number of vehicles between OD pair \(ij\) that left their origin during interval \(p\) and \(n_{OD}\) is the total number of OD pairs, \(a_i(p)\) denotes the \((n_x \times n_{OD})\) assignment matrix of \(a_{ij}(p)\) which is the assignment proportion of the OD flow from \(i\) to \(j\) that departed its origin \(l\) during interval \(p\) and crossed the counting station on link \(l\) during interval \(k\), \(\xi(k)\) denotes the \((n_x \times 1)\) vector of \(\xi_i(k)\) which is the measurement error of counts at link \(l\) during interval \(k\), and \((p'+1)\) denotes the maximum number of time intervals required to travel between any OD pair.

Equation (1) implies that the flow across any detector station during interval \(k\) is comprised of the contributions from OD flows departed its origin during \(k\), \(k-1\), ..., \(k-p'\). The error term reflects the possibility of measurement errors from detector. It is noted that assignment matrix \(a_i(p)\) depends on the time-dependent path-choice fractions and the time-dependent link use proportions\(^5\),\(^17\). The former can be determined by means of discrete choice models that utilize the information of generalized costs along the paths. The latter is conventionally determined using the information from time-dependent travel times obtained from surveillance system or from simulation model. Fig. 1 illustrates the concept described in equation (1).

There are some critical issues when applying the assignment matrix to map between traffic counts and OD parameters. The first is the difficulty in observing time-varying travel times from all OD pairs to all observation stations. The second is on the nonlinearity of the original relationship between them. Similar criticism on this issue was also found in Balakrishna and Koutsopoulos\(^18\). The third is on the treatment of natural inequality and equality constraints in the estimation problem. Moreover, it is difficult to use other types of traffic measurements in the assignment mapping scheme. The measurement data used to be limited to traffic count data. The following sections explicitly discusses the above-mentioned issues and describe a new idea used in this paper.
3. BASIC CONCEPTS

Consider a network of $n_{OD}$ OD pairs and $n_l$ links, apart from them $n_l$ links are equipped with detectors. Based on our earlier discussion, the general nonlinear dynamic mapping between the unknown OD flows and the observed traffic counts can be written compactly as shown in Equation (2).

$$y(k) = h[x(k), \ldots, x(p), \ldots, x(k - p')] + E\left[\xi(k)\right]$$  \hspace{1cm} (2)

where $h[\cdot]$ denotes the $(n_l \times 1)$ vector of a nonlinear dynamic mapping function $h[\cdot]$ that describes dynamic traffic flow and path choice processes of vehicles on the network. In this section, we show that the statistics (mean and variance) of $y(k)$ calculated based on the simplified linear mapping Equation (1) are biased and are not the same as the statistics that would be obtained from the original nonlinear mapping Equation (2).

Assume that $z$ and $u$ are scalar Gaussian random variables (the same assumption used so far in standard KF and Extended Kalman Filter: EKF) with the following nonlinear relationship: $u = g(z)$. The mean value of $u$ can be obtained by taking the expectation of the nonlinear function, i.e. $\mu_u = E[u] = E[g(z)]$. However, the conventional approach in dynamic OD estimation simply transforms the expectation of $z$ through the nonlinear function to obtain the estimate of mean value of $u$, i.e. $\mu_u = g(E[z])$ which is not the same as $\mu_u = E[g(z)]$ (9). With the same logic to the real dynamic OD problem, the estimate value of traffic counts using the linear assignment mapping equation is just similar to performing $\hat{y}(k) = h[x(k), \ldots, x(p), \ldots, x(k - p')]$ which is not the same as performing

$$\hat{y}(k) = E[y(k)] = E[h[x(k), \ldots, x(p), \ldots, x(k - p')]]$$  \hspace{1cm} (3)

Therefore, the linear assignment matrix equation generates a bias to the OD estimation.

It is our objective to develop a new approach that applies directly the nonlinear dynamic mapping between the OD parameters and the measurements rather than using the biased linear assignment mapping equation.

As the purpose of the mapping function is to estimate the number of vehicles (or flows) that will cross at each observation station when loading a given set of OD flows onto the network, any sophisticated and reliable model that can perform this task can then be used directly as a new mapping equation. In this paper, we replace the conventional linear mapping function with the dynamic traffic simulation model to avoid the bias mentioned above.

Additionally, since in general the simulation model provides not only the traffic counts but also average speeds or travel times, they can be considered as the additional measurement variables to further provide information to the estimator (3). Similar idea on using other types of measurements was also discussed in Balakrishna and Koutsopoulos (10) in order to estimate OD flows on off-line basis. Our next objective is therefore on the investigation of the benefit that can be gained from utilizing speed data in addition to traffic counts, in order to estimate OD flows on on-line basis. Later we refer to the measurement variables as either only traffic counts or traffic counts with speed data.

Unfortunately, because of the simulation nature, the new method which uses the simulation model directly as the dynamic mapping has no definite closed analytical form and standard KF method cannot be applied. This also prevents us from using the most popular technique to deal with nonlinear system such as the EKF because of the derivation of Jacobian matrix is extremely difficult. The new method is thus based on the new filtering technique called UKF. This new technique offers a number of unique properties that allow the use of any complex function such as the simulation model or even the black-box process to be used in the new estimation method.

As the UKF is relatively new and plays an important role in the new estimation method, in the next section we provide the fundamental concepts of this filter.

4. UNSCENTED KALMAN FILTER

Similar to the EKF, state distribution in the UKF is represented by a Gaussian random variable (GRV). However, the UKF applies a deterministic sampling technique, called Unscented Transformation (UT), to first represent the state distribution by a number of carefully chosen sample points (typically called “sigma points”). These sigma points completely capture the true mean and covariance of the original variable. These sigma points are then propagated through the nonlinear transformation so a set of transformed sigma points can be obtained. These transformed points capture the mean and covariance of the transformed variable accurately to the second order Taylor series expansion for any nonlinearity.

The unscented transformation is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation. Assume a $(L \times 1)$ vector of random variables $x$ to be propagated through a nonlinear function, $y = f(x)$. Note that the time index $k$ is dropped in order to simplify the notation. Further assume that the mean and error covariance of $x$ are as $\hat{x}$ and $P$, respectively.
(L×(L×1)) matrix $X$ of (L×1) sigma vectors $X_i$ is drawn as follows:

$$X_0 = \hat{x}$$

$$X_i = \hat{x} + \sqrt{[(L + \lambda)\mathbf{P}]}_i, (i = 1,\ldots,L)$$

$$X_i = \hat{x} - \sqrt{[(L + \lambda)\mathbf{P}]}_{i+L}, (i = L+1,\ldots,2L)$$

where $\lambda = \alpha^2 (L + \kappa) - L$ is a scaling parameter of the sigma points. Constant $\alpha$ determines the spread of sigma points around the mean $\hat{x}$ and is usually set to a small value, i.e. $0.0001 \leq \alpha \leq 1$. Constant $\kappa$ is a second scaling parameter. $\sqrt{[(L + \lambda)\mathbf{P}]}_i$ is the $i$th column of the matrix square root (e.g., lower-triangular Cholesky factorization). Propagating these sigma vectors through the nonlinear function, we can obtain:

$$Y_i = f(X_i), \quad (i = 0,\ldots,2L)$$

(5)

Once the sigma vectors $Y_i$ are calculated, the mean and covariance of $y$ can be approximated using a weighted sample mean and covariance of these sigma vectors.

$$\bar{y} \approx \sum_{i=0}^{2L} W^w_i Y_i$$

$$\mathbf{P}_y \approx \sum_{i=0}^{2L} W^v_i (Y_i - \bar{y})(Y_i - \bar{y})^T$$

(6)

(7)

The weight $W_i$ can be calculated as:

$$W^w_i = \frac{\lambda}{L + \lambda}$$

$$W^v_i = \frac{\lambda}{L + \lambda} + 1 - \alpha^2 + \beta$$

$$W^w_i = W^v_i = \frac{1}{2(L + \lambda)}, \quad i = 1,\ldots,2L$$

(8)

where $\beta$ is used to incorporate prior knowledge of the distribution of $x$. In this paper, as there is no prior knowledge on the distribution and scaling, we set $\kappa$ and $\beta$ equal to zero. Based on our preliminary test, the value of $\alpha$ is set equal to 0.1. Readers interested in this topic may consult \cite{8,20,21,22,23} for more detail.

5. PROBLEM FORMULATION

In this section, we formulate the problem of dynamic estimation of OD matrices under the framework of UKF and with the dynamic traffic simulation model replacing the conventional mapping function by assignment matrices. The model consists of two basic elements, transition and measurement equations. Moreover, the method to deal with natural constraints of the problem is presented.

(1) Transition equation

As the temporal relationship between the state variables is difficult to define explicitly, auto-regressive and random walk processes are the two forms of the transition equation that were used in most of the previous researches. We also follow the same idea. The transition equation is written as shown below:

$$b(k + 1) = b(k) + \zeta(k)$$

(9)

where $b(k)$ denotes the $(n_{OD} \times 1)$ vector of state variables $b_y(k)$ which is the fraction of flow originated at $i$ during interval $k$ that will destine at $j$. $\zeta(k)$ denotes a vector of modelling errors which is assumed to be Guassian white noise with zero mean and variance $\Phi(k)$. In addition, state variables $b_y(k)$ must satisfy the following inequality and equality constraints:

$$\sum_{ij} b_y(k) = 1, \quad \forall i = 1 \text{ to } n_o$$

$$0 \leq b_y(k) \leq 1, \quad \forall ij$$

(10)

(11)

where $n_o$ is the number of origins. The algorithm to deal with the above constraints in the estimation algorithm will be discussed later in this section. We note that this model is limited by the assumption that all origin flows, $o_i(k)$ are given and considered as inputs to the model. In some situations, these origin flows may not be available due to detector failure. One way to deal with this problem is to estimate both the origin flows and the OD proportions simultaneously\cite{17}. The inclusion of the origin flows as the additional state variables is, however, left as a topic for future research.

(2) Measurement equation

The measurement equation in this paper differs from the other previous researches because it retains the nonlinearity of this equation which is accomplished by means of dynamic traffic simulation model. We repeat again the general form of the measurement equation with some modification to facilitate the above transition equation.

$$y(k) = h[b(k), b(k - 1), \ldots, b(k - p'), u(k), \ldots, u(k - p')] + \xi(k)$$

(12)

where $y(k)$ denotes a vector of measurements (either only traffic counts or counts with speed data), $h[\cdot]$ denotes, as before, the dynamic mapping function which in this case is the dynamic traffic simulation model, $u(k)$ denotes a vector of input variables which are origin flows in this case, $\xi(k)$ denotes, exactly same as before, a vector of measurement errors which is assumed to be Guassian white noise with zero mean and covariance $\Psi(k)$ and is not correlated with $\xi(k)$.
that corresponds to each time interval \( p = (k - p') \) to \( k \), we simply load the estimated OD flows \( (x_{ij}^p) = o_i \times b_{ij} \) that corresponds to each time interval \( p \) onto the network according to the traffic simulation model \( h[\cdot] \) and record the measurements during the last interval \( p = k \). Because we update the traffic condition on the network every time interval based on the latest estimate of OD flows, the initial traffic condition at the beginning of the interval \( p = (k - p') \) therefore includes all the contributions from the previous estimate OD flows, \( \hat{x}(k - p' - 1), \hat{x}(k - p' - 2), x(k - p' - 3), \ldots \).

Equation (12) implies that the measurements corresponding to interval \( k \) provide information not only to \( b(k) \) but also to \( b(k - 1), b(k - 2), \ldots, b(k - p') \). This suggests that each OD proportions shall be estimated multiple times in order to fully exploit the information. We can perform this using the same method suggested by Okutani\(^{(12)}\) by augmenting the OD proportions from several previous intervals into the state vector. However, this greatly increases the number of state variables to be estimated and may not be efficient for on-line estimation. Ashok and Ben-Akiva\(^{(17)}\) have investigated the approximation of this form by estimating each OD flow only once and holding constant after their first estimate.

The implementation of this approximate model using any simulation model \( h[\cdot] \) is made as follows. In this case, we simply load the current estimate of the OD flows at time \( k \) onto the network according to the simulation model \( h[\cdot] \) and record the measurements during this interval. Similar to the previous, the initial traffic condition at the beginning of the interval \( k \) includes all effects from the previous OD flows, \( \hat{x}(k - 1), \hat{x}(k - 2), \ldots, \hat{x}(k - p') \).

(3) Imposing constraints

As generally known, the estimate of OD proportions using any type of KF algorithm may not satisfy their natural constraints shown in Equations (10) and (11). The method to deal with equality and inequality constraints used in this study is based on the active set method of\(^{(24,25)}\). We summarize the algorithms for both equality and inequality constraints below. To simplify the formulation, the index of time interval \( k \) is shown only if necessary. The equality and inequality constraints respectively are rewritten as:

\[
Db = 1 \quad \text{and} \quad D' b \geq 0 \quad (14)
\]

where \( D \) and \( D' \) are respectively known \( (n_o \times n_{OD}) \) and \( (n_{OD} \times n_{OD}) \) coefficient matrices, \( 1 \) and \( 0 \) are respectively \( (n_o \times 1) \) and \( (n_{OD} \times 1) \) vectors of equality and inequality constraints. The elements in coefficient matrices \( D \) and \( D' \) are properly set as either unity or zero so that the constraints are equivalent to Equations (10) and (11). For illustration purpose, equation (14) can be written for the network in Fig. 1 as follows:

\[
\begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
b_4 \\
b_5
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
h_1 \\
h_2 \\
h_3 \\
h_4
\end{bmatrix}
\geq
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\quad (15)
\]

Assume that \( m \) of the \( n_{OD} \) inequality constraints are active at the solution, and denote by \( \overrightarrow{D'} \) and \( \overrightarrow{o} \) the \( m \) rows of \( \overrightarrow{D'} \) and \( \overrightarrow{o} \) corresponding to the active constraints. We can then augment the active set of inequality constraints with the equality constraints as follows:

\[
\overrightarrow{D} = \begin{bmatrix} D' \end{bmatrix}
\quad \text{and} \quad \overrightarrow{d} = \begin{bmatrix} 1 \\
\overrightarrow{o}
\end{bmatrix}
\quad (16)
\]

Using these new sets of constraints, we can apply the method developed in Simon and Chia\(^{(29)}\) to explicitly constrain the estimate of the OD proportions. The result from unconstrained estimates is further corrected based on the constrained adjustment algorithm as follows.

\[
\hat{b}(k) = \hat{b}(k) - W^{-1}(k) \overrightarrow{D}^T(k) \cdot \left( [\overrightarrow{D}(k) W^{-1}(k) \overrightarrow{D}^T(k)]^{-1} [\overrightarrow{D}(k) \hat{b}(k) - \overrightarrow{d}(k)] \right)
\quad (17)
\]

where \( \hat{b}(k) \) and \( \hat{b}(k) \) are respectively the constrained and unconstrained estimate of the OD proportions during interval \( k \), and \( W \) is a weighting matrix. By setting \( W = P^+ \) and \( W = I \), the algorithm is respectively equivalent to the maximum probability method and the mean square method\(^{(24)}\).

The overall estimation procedure can be summarized as follows. At the beginning of any OD estimation interval, we obtain the priori estimate of the OD proportions using Equation (9). This estimate is used to generate the amount of traffic between OD pair \( ij \) during the current estimation interval. Then the

331
UKF algorithm is applied to generate a set of sigma points of OD proportions (hence OD traffic volume), and feed as input to the dynamic traffic simulation, which will be described in the next section. The outputs of the simulation, such as traffic flow and average speed at any particular segment, are compared with the actual observation, and the resulting error can be determined. The UKF algorithm then corrects the priori estimate of OD proportions based on this error. Finally, we impose the constraints on the new estimation result based on Equations (14) (16), and (17). This new estimate is the final result and will be used in the transition equation during the next estimation interval. Based on this new formulation, the important features of the proposed approach can be summarized below:

- It retains the inherent nonlinear relation between OD parameters and measurement variables by formulating them in terms of dynamic traffic simulation model,
- The proposed UKF-based approach does not require any differential operations that were requisite in the derivation of Jacobian matrix of the EKF-based approach.
- It offers a way to incorporate speed data in the estimation algorithm.
- It is flexible in terms of traffic simulation model so any reliable simulation model can be used in this approach.
- The natural equality and inequality constraints are also imposed in the UKF filtering algorithm.

6. TRAFFIC SIMULATION MODEL

The simulation model used in our preliminary test is a multi-commodity macroscopic traffic model which is modified from the cell-transmission model and the discretization of the first-order macroscopic model based on the Godunov scheme. Different commodities represent the flows from different OD pairs. It is noted that any other reliable traffic simulation model can also be used in this estimation approach. In the case study, the network is a freeway stretch and there is only one route, therefore the problem reduces to only the dynamic traffic simulation without route choice process.

The road network is first discretized into small segments and the simulation interval is properly selected. Traffic flow is characterized by three aggregated variables: density ($\rho$), space mean speed ($v$), and flow rate ($q$). The general formulations for ordinary segment are given below:

$$D_i (\tau) = \begin{cases} Q_e (\rho_i (\tau)) & \text{if } \rho_i (\tau) \leq \rho_{c,i} (\tau) \\ q_{\max,i} (\tau) & \text{if } \rho_i (\tau) > \rho_{c,i} (\tau) \end{cases}$$

$$S_{i+1} (\tau) = \begin{cases} q_{\max,i+1} (\tau) & \text{if } \rho_{i+1} (\tau) \leq \rho_{c,i+1} (\tau) \\ Q_e (\rho_{i+1} (\tau)) & \text{if } \rho_{i+1} (\tau) > \rho_{c,i+1} (\tau) \end{cases}$$

$$q_i (\tau) = \min \{D_i (\tau), S_{i+1} (\tau)\}$$

$$\rho_i (\tau + 1) = \rho_i (\tau) + \frac{\Delta \tau}{\Delta x} (q_{i-1} (\tau) - q_i (\tau))$$

$$q_i^d (\tau) = \frac{\rho_i^d (\tau)}{\rho_i (\tau)} \cdot q_i (\tau)$$

$$\rho_i^d (\tau + 1) = \rho_i^d (\tau) + \frac{\Delta \tau}{\Delta x} (q_i^d (\tau) - q_i^d (\tau))$$

$$v_i (\tau) = V_e (\rho_i (\tau))$$

$$w_i (\tau) = \phi \cdot v_i (\tau) + (1 - \phi) \cdot v_{i+1} (\tau)$$

where $\tau$ denotes the index of simulation interval. If not stated explicitly, $Z(\cdot)$ denotes any variable $Z$ measured over the whole of segment $i$. $D(\cdot)$ and $S(\cdot)$ denote respectively the local demand and local supply, $\rho(\cdot)$ denotes the density, $q(\cdot)$ is the flow at the downstream boundary of segment $i$, $V(\cdot)$ is the space mean speed, $w(\cdot)$ is the time mean speed at the downstream boundary of segment $i$, $\rho^d (\cdot)$ and $q^d (\cdot)$ represent the partial density and flow of commodity $d$, $Q(\cdot)$ and $V(\cdot)$ are respectively the equilibrium $\rho-q$ and $\rho-v$ relationships, $q_{\max,i}$ is the capacity flow of segment $i$, $\rho_{c,i}$ is the critical density, $\Delta \tau$ is the simulation interval, $\Delta x$ is the length of segment $i$, $\phi$ is a weighting parameter for estimation of time mean speed.

Equations (18)–(25) constitute the multi-commodity traffic flow model for ordinary section. The formulations for the segment with on- and off-ramps differ from that of the ordinary segment in the calculation of boundary flow at on- and off-ramps. These equations can be used to recursively estimate density of each segment. Flow of each commodity is moved toward their destination as a result of computing and updating partial boundary flow and partial density shown in Equations (22) and (22). Equation (24) provides only an approximation of space mean speed at equilibrium condition and Equation (25) provides only an approximation of time mean speed. Moreover, as this model was derived from the first-order macroscopic model, it has all the same limitations as that of the first-order macroscopic model. Replacing this model with the more sophisticated and reliable model is also included in our research plan.
7. CASE STUDY

(1) Data description

The 24-hour traffic data on November 1, 1994, from Matsubara line of the Hanshin expressway in Japan was used to evaluate the proposed method. The traffic condition on that day was quite smooth for almost the whole day but also congested around 7-8 p.m. with the average speed of around 20 km/hr. The weather condition was clear and there was no accident. This data was selected because of two major reasons. First, the actual OD flow information is available from this database. Second, we tried to choose the network with only one feasible route so that we can be certain that the performance of the model is purely from the proposed algorithm and not contaminated with the route choice modeling error. The schematic layout of the road section is given in Fig. 2. The study section is 2-lane and 11.220 km length and was divided into 28 segments for the purpose of traffic simulation. This section has in total 15 feasible OD pairs. The information on actual OD volumes was collected by the Hanshin Expressway Company Limited for every one hour from all on- and all feasible off-ramps. 5-minute detector data on traffic volumes, time mean speed, and occupancy were available at the mainstream entrance, mainstream exit, and at the end of segments 4, 9, 13, 17, 20, and 25. The 5-minute detector data on traffic volumes at all on- and off-ramps were also available.

(2) Implementation

Since the OD volumes were recorded at the time the vehicles entered the on-ramp, there might be time lag between the recorded time and the time the vehicles passed at the entrance O1 of the study section. The actual OD data thus might also include an error depending on the size of the network and the traffic condition. In this study, however, we assume that the value of this error is small and can be neglected. As the actual OD data is available every 1 hour, the estimation interval of OD proportions was also set as 1 hour. In this case, the detector data (measurement data) were also averaged over 1 hour.

To start the algorithm, the initial value of the OD proportions and density of all segments are required. However, these initial values were not available from the data. Therefore the initial value of density was assumed to be 50 veh/km for all segments. The effect from initial value of density is expected to disappear after a number of simulation runs.

Two sets of initial OD proportions, Ib1 and Ib2, were considered in this study. The first set Ib1, represents the case of no prior information, assumes OD flows are equally distributed (all destinations from the same origin attract the same proportion of traffic). The second set Ib2, represents the case of more accurate information of the initial value, uses the average values of the actual OD proportions over 24 hrs as the initial values, considering the availability of historical data.

Because there is no information available on the statistics of the modelling noise, the covariance matrix Φ was assumed to be constant and equal to 0.0001 along the diagonal. The statistics of measurement noises were determined on a trial-and-error basis bearing in mind the possible ranges suggested by Cremer[30]. The measurement noises for both flows and speeds were then set as zero mean with standard deviation equal to 5 percent of the measured values.

The multi-commodity traffic flow model presented in section 6 was used as the traffic flow simulator in the estimation algorithm. Because the shortest segment length is 190 m, the simulation interval was set as 5 seconds in order to ensure the numerical stability of the traffic flow model.

Fig.2 Schematic layout of study section.
Traffic data during 7:00 – 9:00 AM were used to calibrate the speed-density diagram by minimizing the error between the observed and estimated traffic condition. This was done on a trial and error basis. The general form of speed-density relationship which has four parameters to calibrate, namely, \( v_f \), \( \rho_{jam} \), \( a \), and \( b \), is shown in Equation (26). The results of calibration are as: \( v_f = 100 \text{ km/hr}, \rho_{jam} = 120 \text{ veh/km}, a = 1.5, \) and \( b = 3.0 \).

\[
v = v_f \left( 1 - \frac{\rho}{\rho_{jam}} \right)^a \quad (26)
\]

In this preliminary test, we considered 13 levels of measurements. Segments 17, 20, and 25 were selected as the additional measurement locations because at these locations we can observe most of the major OD flows. The details of 13 measurement scenarios are described in Table 1.

(3) Results
The results are presented separately in two parts: (i) using only flow data and (ii) using both flow and speed data.

a) Using only flow data
In this part, we consider only the case of using only flow data as generally did in the previous researches. Fig. 3 – 5 show respectively some examples of the plots of estimated and actual time-varying OD proportions, the estimated and actual OD flows, and the estimated and actual link flows, for both \( Ib_1 \) and \( Ib_2 \) from a scenario F17_20. The performance of the model is evaluated based on root mean square error (RMSE) of OD flows as shown in Equation (27).

\[
\text{RMSE} = \sqrt{\frac{\sum (x_i - \hat{x}_i)^2}{N}}
\]

where \( x \) and \( \hat{x} \) represent respectively the actual and estimated OD flows, and the summation is over all OD pairs and estimation intervals. To be fair for both \( Ib_1 \) and \( Ib_2 \), the results of the first two intervals were excluded from the RMSE statistics. Fig. 6 shows the RMSE values of \( Ib_1 \) and \( Ib_2 \) in case of using only flow data. It can be seen from Fig. 3(a) that the estimation algorithm can track the actual OD proportions even in the case of arbitrarily or poor setting of initial value (\( Ib_1 \)). The results are quite similar for both \( Ib_1 \) and \( Ib_2 \) (similarity also found in Fig. 4 – 5). In case of \( Ib_1 \), the estimated OD proportions become close to their actual values after the first estimation interval. The effects from the initial values disappeared within a small number of initial runs. This result illustrates the potential of the estimation algorithm in tracking the true OD proportions even in the absence of accurate initial values.

From Fig. 4, the plots of the estimated and actual OD flows are densely along the diagonal line, illustrating the well performance of the estimation algorithm. There was small difference between Fig. 4(a) and 4(b). Better initial value (\( Ib_2 \)) seems to provide superior results of the small OD flows.

The proposed method also provides very excellent by-product estimate of link flows regardless of their initial values as can be seen from Fig. 5. The traffic simulation model embedded in the estimation algorithm perfectly reproduced the observed link flows.

### Table 1 Details of Measurement Setup.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Measurement Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>Using only off-ramp and exit flows</td>
</tr>
<tr>
<td>F17</td>
<td>Similar to “Base” with additional flow measurements from segment 17</td>
</tr>
<tr>
<td>F20</td>
<td>Similar to “Base” with additional flow measurements from segment 20</td>
</tr>
<tr>
<td>F25</td>
<td>Similar to “Base” with additional flow measurements from segment 25</td>
</tr>
<tr>
<td>F17_20</td>
<td>Similar to “Base” with additional flow measurements from segments 17 and 20</td>
</tr>
<tr>
<td>F17_25</td>
<td>Similar to “Base” with additional flow measurements from segments 17 and 25</td>
</tr>
<tr>
<td>F20_25</td>
<td>Similar to “Base” with additional flow measurements from segments 20 and 25</td>
</tr>
<tr>
<td>FS17</td>
<td>Similar to “F17” with additional speed measurements from segment 17</td>
</tr>
<tr>
<td>FS20</td>
<td>Similar to “F20” with additional speed measurements from segment 20</td>
</tr>
<tr>
<td>FS25</td>
<td>Similar to “F25” with additional speed measurements from segment 25</td>
</tr>
<tr>
<td>FS17_20</td>
<td>Similar to “F17_20” with additional speed measurements from segments 17 and 20</td>
</tr>
<tr>
<td>FS17_25</td>
<td>Similar to “F17_25” with additional speed measurements from segments 17 and 25</td>
</tr>
<tr>
<td>FS20_25</td>
<td>Similar to “F20_25” with additional speed measurements from segments 20 and 25</td>
</tr>
</tbody>
</table>
**Fig. 6** shows the reduction of RMSE values when more accurate initial value is available. The number of detector data has little effect on the performance of the model. This might be from the fact that the predicted link flows were already perfect, so there is no room to improve the results using only these link flows.

**b) Using both flow and speed data**

The above results illustrated the benefit from using more accurate initial value $I_b^2$ when using only flow data. This part further investigates the benefit that can be gained from using speed data in addition to flows under the same case. **Fig. 7** illustrates the improvement in quality of the estimated OD flows by using speed data in addition to link flows. **Fig. 9** shows the reduction in RMSE of OD flows when using both flow and speed data.

![Fig. 3](image1.png) Estimation results of OD proportions (F17_20).

![Fig. 4](image2.png) Estimated and actual OD flows (F17_20).

![Fig. 5](image3.png) Estimated and actual link flows (F17_20).
From Fig. 7, though most of the times the OD estimates in case of with and without speed data are very close to each other, there exist some differences between them at the period around 7:00PM – 8:00PM. By checking from detector data, we found that this time is the most peak of the day and there were likely the congestion between segments 17 and 20. Fig. 8 shows the increasing in observed occupancy and decreasing in speed between segments 17 and 20, illustrating the heavy traffic during that time. The result from Fig. 7 thus implies that speed data can improve the performance of the estimation if there is traffic congestion. In case of light traffic (all other time periods), the results are almost the same as in the case of using only flows.

The improvements in the quality of the estimated OD flows by using both flows and speeds are confirmed in Fig. 9. Note also here that the performance is the same for the case of F25 and FS25 because the traffic condition on this link was always in light condition. To summarize, utilizing speed data in addition to flows can reduce the RMSE of OD flows particularly when there is traffic congestion.

![Fig.6 RMSE of OD flows.](image1)

![Fig.7 Effect of using speed data.](image2)

![Fig.8 Observed occupancy and time mean speed.](image3)
8. CONCLUSION

In this paper, a new method for dynamic estimation of OD matrices has been proposed. This method uses dynamic traffic simulation model to replace the conventional linear mapping equation between OD parameters and measurement variables based on the assignment matrices. This new mapping is considered to be more direct and retains the nonlinear relationship between OD parameters and measurement variables. Because of nonlinearity and the difficulty in deriving Jacobian matrix from the traffic simulation model, the new approach applies a new filtering technique, UKF, as the estimation algorithm. The natural equality and inequality constraints are considered explicitly in this approach. In addition to using only traffic counts, the proposed approach offers a way to incorporate speed data to further provide more information in the estimation, which is more important in case of traffic congestion. Moreover, it also offers some degrees of flexibility in selecting the traffic simulation model so that any reliable simulation model can be used in this approach.

The preliminary results from the case study are encouraging and illustrate a promising future in this research direction. Two major findings from the experiment can be summarized as follows:

- The algorithm is quite robust to the initial value of OD proportions. This is a desirable property as we can be certain that the estimation result will converge and follow the same pattern after a few estimation intervals.
- Using speed data in addition to the flow data was found to improve the accuracy of the algorithm particularly during congested condition.

The above findings were based on an experiment with the database from the expressway Matsubara line. This database was chosen due to the limited availability of the actual OD data. As a selected network has only one route, it offers a more direct way to test the performance of the OD estimation algorithm by separating the problem of OD estimation from route choice determination. On the other hand, when dealing with a network with more routes such as urban arterial roads, the route choice algorithm may greatly affect the performance of the proposed algorithm. In addition, in arterial network, the number of unknown OD parameters can be extremely large compared with the number of detector stations, leading to a data scarcity problem. Moreover, unlike the uninterrupted nature of freeway network, there are many traffic characteristics of urban arterial network which can be difficult to describe by the traffic simulation model. Examples are sinks and sources, bus stops, on-street parkings, interaction with pedestrian, etc. All of these can complicate the simulation model and hence affect the performance of the proposed approach.

There are still some remaining aspects of the proposed estimation framework that need to be investigated. The first is to extensively validate the proposed approach under various situations. Second, the model that fully exploits traffic information by estimating each OD flow multiple times should be examined. Third is to extend the model to the case of simultaneously estimate of origin volumes and OD proportions. Fourth is to incorporate all available types of measurements from probes, AVI, ETC, etc. into the estimation and investigate the contribution from these measurements. Finally, investigation of the effect of using different traffic simulation models as well as the effect of route choice process on the performance of the proposed approach should also be conducted. Moreover, extension of the proposed framework to deal with the problem of data scarcity such as those in the urban arterial network is also suggested.

ACKNOWLEDGMENT: Detector data and OD data provided by the Hanshin Expressway Company Limited was greatly appreciated.
REFERENCES


(Received April 19, 2010)