Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows

K. Ashok

Principal

Marketing and Planning Systems

1100 Winter Street, Waltham, MA 02451

Tel: (781) 890-2228 * Fax: (781) 890-6719

E-mail: ashok@mit.edu

M.E. Ben-Akiva

Edmund K. Turner Professor of Civil and Environmental Engineering

Director, Intelligent Transportation Systems Program

Massachusetts Institute of Technology

Room 1-181, 77 Massachusetts Avenue, Cambridge, MA 02139

Tel: (617) 253-5324 * Fax: (617) 253-0082

E-mail: mba@mit.edu

Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows

K. Ashok
Marketing and Planning Systems
1100 Winter Street, Waltham, MA 02451

M.E. Ben-Akiva
Massachusetts Institute of Technology
Room 1-181, 77 Massachusetts Avenue, Cambridge, MA 02139

Abstract
This paper presents a new suite of models for the estimation and prediction of time-dependent Origin-Destination (O-D) matrices. The key contribution of the proposed approach is the explicit modeling and estimation of the dynamic mapping (the *assignment* matrix) between time-dependent O-D flows and link volumes. The assignment matrix depends upon underlying travel times and route choice fractions in the network. Since the travel times and route choice fractions are not known with certainty, the assignment matrix is prone to error. The proposed approach provides a systematic way of modeling this uncertainty to address both the *offline* and *real-time* versions of the O-D estimation/prediction problem. Preliminary empirical results indicate that generalized models with a stochastic assignment matrix could provide better results compared to conventional models with a fixed matrix.

There is increasing emphasis on Dynamic Traffic Assignment (DTA) models for analysis and management of urban and suburban congestion problems (Ben-Akiva (1985), Ben-Akiva et al.(1994)). A key input to a DTA is a set of time-dependent Origin-Destination matrices. Since obtaining these matrices directly (for example from traffic surveys) is extremely difficult and costly, the usual procedure is to estimate these indirectly from the traffic volumes they induce on the links of the network.

A DTA may be used in real-time within a traffic management system or as an offline tool for assessment of network performance. Correspondingly, we distinguish between two types of dynamic O-D flow estimation problems – the *offline* estimation problem and the *real-time* estimation and prediction problem. The former, as the name indicates, is carried out offline and involves estimation of a set of dynamic O-D matrices given a time-series of link volumes (and potentially other information such as travel times, historical O-D flows, etc). The latter operates in tandem with a DTA within a real-time traffic management system. An additional issue of relevance in the real-time problem is *prediction* of future O-D flows. Prediction of future flows is important for generation of optimal traffic control and route-guidance strategies.

Most research on Dynamic O-D Estimation and Prediction to date has focused on "closed" networks, i.e., where complete information is available on the entry and exit traffic counts of the network during each measurement interval (Bell (1991), Chang and Tao (1996), Chang and Wu (1994), Cremer and Keller (1987), Nihan and Davis (1987), Van der Zijpp (1996)). These approaches suffer from the limitation that all entry and exit counts should be known – an unrealistic scenario in a general urban traffic network. For general networks, Cascetta et al.(1993) provide an extension of the static O-D estimation framework for the offline problem. An enhanced model with predictive ability was developed by Inaudi et al.(1994). In this model, the "sequential" model of Cascetta et al. is used for estimation. Values thus estimated are then used to generate predictions by a separate "filtering" approach that combines historical and estimated O-D information using the concept of "deviations" proposed by Ashok and Ben-Akiva (1993) (see below). For the real-time problem, a state-space modeling approach has been suggested by Okutani (1987). The state vector has been taken to be the vector of unknown O-D flows. However, the auto-regressive formulation of the transition equation is simplistic and only captures temporal interdependencies between O-D flows. Finally, another approach based on state-space modeling with the state-vector defined in terms of O-D flow *deviations* was presented by Ashok and

Ben-Akiva (1993). This addresses the problem with the auto-regressive specification in Okutani's work. Kachroo et al. (1995) extended this approach to account for serial correlation of errors in the autoregressive formulation.

The most critical issue in O-D matrix estimation – whether static or dynamic – is the mapping between the observed link flows and the unobserved O-D flows. This mapping is accomplished by means of an *assignment* matrix. In the dynamic problem, the assignment matrix depends on link and path travel times and traveler route-choice fractions – all of which are time-varying. More importantly, these quantities are at best imperfectly observed, thereby introducing errors into the O-D estimation process. As we show later, using these erroneous assignment matrices yields biased and inconsistent estimates of the O-D flows. None of the approaches described above take explicit cognizance of an erroneous assignment matrix or model the uncertainty in the underlying dynamics that generate these matrices. Two approaches that have attempted to do so, but have limitations, are those by Chang and Wu (1994) and Bell (1991). The former use a stochastic assignment matrix in their Kalman Filter based approach. They use a starting value for the assignment matrix for the first interval and a random walk transition equation for subsequent intervals. A weakness of the model is that measurement equations for the assignment fractions are not specified for subsequent intervals. Moreover, their technique is only valid for closed networks. Bell (1991) uses "travel time dispersion" parameters to model an assignment matrix but these parameters are assumed to be constant.

In this paper, we propose different approaches that introduce stochasticity of the assignment matrix into the formulation. Furthermore, we show how conventional formulations of the dynamic O-D estimation/prediction problem for both the offline and the real-time case can be modified for general networks. We organize the paper as follows. Sections 2 and 3 discuss the role of the assignment matrix in more detail. Section 4 outlines different procedures for explicit modeling of errors in these matrices. Sections 5 and 6 describe the consequent modifications to the conventional offline and real-time formulations. Case studies that evaluate the various models are described in the following section. We conclude with recommendations for future research.

1 Basic Concepts

Consider a period of length $\mathcal{T}$ divided into equal intervals of length H. Consider a network with n_{LK} links and n_{OD} O-D pairs. It is assumed that n_l of these links are equipped with detectors. Denote by x_{rh} the number of vehicles between the rth O-D pair that left their origin in interval h and by $\mathbf{x}_h$ the corresponding $(n_{OD} * 1)$ vector. Further denote by y_{lh} the observed traffic counts crossing detector l during interval h and by $\mathbf{y}_h$ the corresponding $(n_l * 1)$ vector.

The *fundamental* relationship between traffic counts and O-D flows can then be stated as follows:

$$y_{lh} = \sum_{p=h-p'}^{h} \sum_{r=1}^{n_{OD}} a_{lh}^{rp} x_{rp} + v_{lh} \tag{1}$$

where a_{lh}^{rp} is the fraction of the rth OD flow that departed its origin during interval p and crossed detector l during interval h. v_{lh} is the measurement error while p' is the maximum number of time intervals required to travel between any O-D pair of the network. In matrix form the above equation reduces to:

$$\mathbf{y}_h = \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p + \mathbf{v}_h \tag{2}$$

where the matrix $\mathbf{a}_h^p$ is an $(n_l * n_{OD})$ *assignment* matrix of contributions of $\mathbf{x}_p$ to $\mathbf{y}_h$ and $\mathbf{v}_h$ is the vector of measurement errors.

The interpretation of the above equation is straightforward. The flow across any detector during interval h is comprised of contributions from O-D flow vectors corresponding to departures during h,h-1,...,h-p'. The assignment matrix consists of the proportions of these O-D flows that constitute the link flow. The error term reflects the possibility of counting errors.

We next examine the assignment matrices $\mathbf{a}_h^p$ in further detail.

2 Parameterizing the Assignment Matrix

Let each O-D pair r be connected by a set of paths $\mathcal{K}_r$. Assume that there exist in total K paths between the n_{OD} O-D pairs in the network i.e. $K = \|\mathcal{K}_1 \cup \mathcal{K}_2 \cup ... \cup \mathcal{K}_{n_{OD}}\|$. Each path $k = 1, 2, .., K$ corresponds to a unique O-D pair. Denote by L_k the set of links comprised in path k. Finally, denote by $\mathcal{F}_h^k$ the flow along path k departing the origin in interval h. Thus during any interval h, the following relationship holds:

$$x_{rh} = \Sigma_{k \in \mathcal{K}_r} \mathcal{F}_h^k \tag{3}$$

Let q_{kh} denote the fraction of travelers corresponding to O-D pair r and departure interval h that choose path k, with $\Sigma_{k \in \mathcal{K}_r} q_{kh} = 1 \;\; \forall \;\; r, h$. We then get the following relationship between O-D and path flows:

$$\mathcal{F}_h^k = x_{rh} q_{kh} \tag{4}$$

Recognizing that the link flows are comprised of contributions from many different path flows, Equation (1) can be restated in terms of path flows, as shown by Cascetta et al.(1993):

$$y_{lh} = \Sigma_{p=h-p'}^{h} \Sigma_{k=1}^{K} \alpha_{lh}^{kp} \mathcal{F}_p^k + v_{lh} \tag{5}$$

where α_{lh}^{kp} defines a mapping between path and link flows and is defined as the contribution of the kth path flow departing the origin during interval p towards the flow across detector l during interval h. Note that in conventional *static* O-D matrix estimation, this would be either one or zero; a matrix of these fractions is the familiar link-path incidence matrix.

Finally, as shown by Cascetta et al.(1993), a simple manipulation of equations (1), (4) and (5) yields the following expression for the assignment matrix:

$$a_{lh}^{rp} = \Sigma_{k: k \in \mathcal{K}_r} \alpha_{lh}^{kp} q_{kp} \tag{6}$$

Analytical expressions for the link-path incidence fractions can be obtained using information about link travel times. In addition to these travel times however, an assumption about movement of vehicles through the network is required. For example, Cascetta et al. derive expressions based on the assumption that vehicles within a group (k, p) (henceforth referred to as a *packet*) are uniformly comprised within the departure duration T and stay within this interval as they move across the network. In other words, vehicles within a packet are uniformly distributed between the leader and the last follower over a span of time T. This assumption can be easily relaxed to permit the effects of "stretching" and "squeezing" of packets as they traverse the network. Such effects could be significant for example, if trip durations are relatively large or travel time variations across successive time-intervals are significant. Following this relaxation, the link-path incidence fractions would be given by the following expression:

$$
\begin{aligned}
\alpha_{lh}^{kp} &= 1 && \textit{if } (h-1)T < \eta_{1l}^{kp} < \eta_{2l}^{kp} < hT \\
&= (hT - \eta_{1l}^{kp})/(\eta_{2l}^{kp} - \eta_{1l}^{kp}) && \textit{if } (h-1)T < \eta_{1l}^{kp} < hT < \eta_{2l}^{kp} \\
&= T/(\eta_{2l}^{kp} - \eta_{1l}^{kp}) && \textit{if } \eta_{1l}^{kp} < (h-1)T < hT < \eta_{2l}^{kp} \\
&= (\eta_{2l}^{kp} - (h-1)T)/(\eta_{2l}^{kp} - \eta_{1l}^{kp}) && \textit{if } \eta_{1l}^{kp} < (h-1)T < \eta_{2l}^{kp} < hT \\
&= 0 && \textit{otherwise}
\end{aligned}
\tag{7}
$$

where η_{1l}^{kp} and η_{2l}^{kp} represent the crossing times of the first and last vehicle in the packet (k, p) at detector l. To use the above relationship, one would in addition have to know the departure times of the first and last vehicles. A convenient assumption might be to have the first depart at the beginning of a departure interval and the last at the end.

Travel times (or more typically, speeds) can be obtained either from a traffic surveillance system (e.g. sensors on the roadway, video cameras, probe vehicles) or a DTA model. In addition to link and path travel times, information about the path choice fractions q_{kh} is required in order to apply equation (6). One way of obtaining these is by using discrete choice models that utilize information about generalized costs along different paths during each interval. An example of such a model is provided by Cascetta et al.(1996).

To summarize, computation of the assignment matrix is highly complicated. Moreover, the estimates obtained from application of equations (6) and (7) may suffer from errors on several fronts

- Travel times obtained from the surveillance system are subject to measurement error due to, for example, sensor malfunction.
- The assumption of uniform distribution of vehicles within a packet might be invalid under certain situations, e.g., an incident.
- Choice fractions obtained from route choice models might be erroneous because of inaccuracies either in the model coefficients or in the data.
- True departure times of first and last vehicles within a packet are unknown.

Finally, there could be scenarios in which some (or all) travel times might be entirely unobserved (are endogenous). These are explored in greater detail next.

3 Endogeneity in the Assignment Matrix

We turn our attention now to the case of erroneous travel times and/or route choice fractions that result in an imperfect assignment matrix. In such a case, we can write:

$$\mathbf{a}_h^{p\bullet} = \mathbf{a}_h^p + \nu_h^p \tag{8}$$

where $\mathbf{a}_h^{p\bullet}$ denotes the erroneous assignment matrix and ν_h^p a random error. Equation (2) becomes

$$\begin{aligned}\mathbf{y}_h &= \Sigma_{p=h-p'}^{h} \mathbf{a}_h^{p\bullet} \mathbf{x}_p - \Sigma_{p=h-p'}^{h} \nu_h^p \mathbf{x}_h + \mathbf{v}_h & (9)\\ \mathbf{y}_h &= \Sigma_{p=h-p'}^{h} \mathbf{a}_h^{p\bullet} \mathbf{x}_p + \tilde{\mathbf{v}}_h & (10)\end{aligned}$$

where the new error term $\tilde{\mathbf{v}}_h = \Sigma_{p=h-p'}^{h} \mathbf{v}_h - \nu_h^p \mathbf{x}_h$. Clearly, $\mathbf{a}_h^{p\bullet}$ and $\tilde{\mathbf{v}}_h$ are correlated because of Equation (8). This constitutes a standard error-in-variables problem in econometrics (See for example, Greene (1993)). Application of a least squares based estimator (such as a Kalman Filter) to (10) in this situation, yields biased and inconsistent estimates for $\mathbf{x}_h$. Depending on the level of uncertainty in the assignment matrix, the extent of bias might be significant.

The impact of this inconsistency could be even more severe if the O-D flows and assignment matrix were to be obtained by an iterative scheme. This could be the case, for example, if the travel times were entirely unobserved and the assignment fractions were obtained by applying the following series of steps: (a) Load a preliminary set of O-D flows to a traffic simulator. (b) Use the observed assignment matrix to recompute the O-D flows. (c) Repeat (a) using updated O-D flows. In such a situation, the assignment fractions are endogenous and indirectly depend on the O-D flows. To see this dependence, we first notice that link travel times depend directly upon link flows. The latter is related to path flows (since a link flow is essentially a weighted combination of several path flows using that link). Finally, the path flows are related to the O-D flows through equation (4). Thus, equation (2) could be expressed as:

$$\mathbf{y}_h = \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p (\mathbf{x}_h, \mathbf{x}_{h-1}, ..., \mathbf{x}_{h-p'}) \mathbf{x}_p + \mathbf{v}_h \tag{11}$$

In this situation, even if convergence in O-D flows and assignment fractions were to be attained (which in itself is far from guaranteed), the final estimates could be highly biased since errors in the assignment matrix are not explicitly accounted for, during each iteration. It is likely, of course, that these iterations may reduce the error in the assignment matrix.

This brings us to the central theme of this paper. In the following sections, we describe two approaches that explicitly take into account the stochasticity of the assignment matrix and accordingly modify the formulations developed in previous research.

4 Modeling a Stochastic Assignment Matrix

In the first approach, we envisage adding randomness to the assignment matrix by means of additional equations of the following form:

$$\mu(\mathbf{a}_h^{p\bullet}) = \mu(\mathbf{a}_h^{p}) + \mathbf{u}_h^{p} \tag{12}$$

In the above equations, $\mathbf{a}_h^{p\bullet}$ is the value of the assignment matrix computed from the equations (6) and (7) that use measured or estimated travel times and route choice fractions. $\mu(.)$ is an operator that re-arranges the elements of a $(m * n)$ matrix into a $(mn * 1)$ vector. The error term $\mathbf{u}_h^{p}$ reflects the fact that the estimate $\mathbf{a}_h^{p\bullet}$ is subject to error. As mentioned earlier, this error could arise from two sources – (a) because of imperfect measurements/estimates of travel times and route-choice fractions and (b) because of incorrectness of assumptions involved in equations (7).

Equation (12) represents a straightforward approach of incorporating stochasticity in the assignment matrix. The disadvantage of this method is the additional computational load imposed by adding a large number of decision variables. In practical applications therefore, it may be necessary to prune the number of assignment fractions considered random in order to make the technique computationally tractable. For example, one might wish to add additional equations (12) only for assignment fractions corresponding to O-D pairs with high flows.

Deeper examination of the assignment matrix and its dependence on travel times and route-choice fractions as given by equations (6) and (7) suggests an alternative approach. Define the assignment matrix $\mathbf{a}_h^p$ by the following relationship:

$$\mathbf{a}_h^p = a(\mathbf{t}_h, \mathbf{t}_{h-1}, ..., \mathbf{t}_{h-p'}, \mathbf{q}_p) \tag{13}$$

where $\mathbf{t}_h$ denotes a $(n_{LK} * 1)$ vector of travel times for interval h, $\mathbf{q}_p$ a $(K * 1)$ vector of route choice fractions for departure time-interval p and the function $a(.)$ defined by equations (6) and (7). The above equation can be compactly represented as:

$$\mathbf{a}_h^p = a(\mathbf{T}_h, \mathbf{q}_p) \tag{14}$$

where $\mathbf{T}_h$ denotes the augmented vector $[\mathbf{t}_h \mathbf{t}_{h-1} ... \mathbf{t}_{h-p'}]'$. Note that $a(.)$ could also in general be a function of *future* travel times. This dependence comes from equation (7) as well as the fact that the $\mathbf{q}$'s could depend on future travel times. Our formulation captures errors in future travel times through the stochasticity of the $\mathbf{q}$'s (see below), and of course, the a's.

If we assume that equation (14) is exact, the only remaining sources of errors in the assignment matrix are those in $\mathbf{T}_h$ and $\mathbf{q}_p$. This provides the motivation for the second approach. Instead of directly dealing with the assignment matrix fractions as in the earlier approach, we work with the underlying travel times and route-choice fractions in the second. We therefore add additional equations of the following form:

$$\mathbf{T}_h^\bullet = \mathbf{T}_h + \lambda_h \tag{15}$$

$$\mathbf{q}_p^\bullet = \mathbf{q}_p + \psi_p \tag{16}$$

where $\mathbf{T}_h^\bullet$ and $\mathbf{q}_p^\bullet$ denote the measured/estimated values of $\mathbf{T}_h$ and $\mathbf{q}_p$ and λ_h and ψ_p represent vectors of random error terms.

Again for computational reasons, one might wish to treat some or all of the route choice fractions as fixed. Also note that if the route choice fractions $\mathbf{q}_p$ can be explicitly represented as a function of travel times $\mathbf{t}_p$ (for example using a logit formulation that uses additive link travel times to compute path travel times), this relationship could be substituted into equation (14) and equations (16) are rendered unnecessary. This technique was used in the case study (Section 7).

Both the approaches presented above have advantages and shortcomings. While the first approach is likely to be more computation intensive (due to the size of the assignment matrix), it could be useful in modeling situations where the assumptions behind equations (7) break down. For example, the assumption that vehicles within a packet stay uniformly distributed between a leader and follower might be violated on urban networks with traffic control signals. On the other hand, the second approach is more efficient because it benefits from the information contained in equation (14). It could be used in freeways or on moderately congested arterials. It could also be more useful when specific information is available on sensor errors, for example, when it is known that a particular sensor systematically overestimates or underestimates say travel speeds.

We again recognize that either method involves the addition of a large number of additional parameters to be estimated. While this is the unavoidable cost of greater model generality, it raises two important issues. The first pertains to the notion of "system observability" or identifiability of the parameters (A detailed discussion of the observability of models with a *fixed* assignment matrix can be found in the earlier paper by the authors, Ashok and Ben-Akiva (1993)). The second refers to whether the gains from the model generality could be negated by the increased complexity (and perhaps, necessary approximations) in the estimation procedure – this is particularly true of the real-time model that we discuss later. Regarding the first issue, we note that for each additional parameter to be estimated, we add (at least) one equation. In fact, as we will see in Section 6, for the real-time problem, we also add equations that attempt to capture the evolution of travel times or assignment fractions over time. If valid, these equations further bolster the observability of the system. The second issue could be particularly relevant for large problems and is difficult to address without further empirical evidence. In practical applications, we would recommend a judicious choice of decision variables, perhaps only estimating those that we hypothesize could have "large" errors or those that correspond to "important" (as measured by, say, heavy flows or large travel times) links.

In closing we might point out that the practicality and efficacy of these generalized models would improve with increased deployment of ITS. One way of looking at these models is that they provide a framework for combining and integrating multiple types, and sources, of measurements (link flows, travel times, speeds, etc) with different error characteristics. For example, instantaneous travel speeds could, in future, be obtained both from sensors and from probe vehicles with different degrees of error. Greater penetration of ITS would imply more and better data, thus the issue of model identifiability would be addressed even more satisfactorily.

5 Offline Estimation

The offline dynamic O-D estimation problem has been formulated by Cascetta et al.(1993) as an extension of static estimation. Since equation (2) typically consists of far more O-D pairs than number of measurements (link counts), additional information is required. Usually, this is provided using an *apriori* O-D matrix. This involves adding to equation (2) another equation of the following form:

$$\mathbf{x}_h^H = \mathbf{x}_h + \mathbf{w}_h \tag{17}$$

where $\mathbf{x}_h^H$ represents an apriori or historical matrix. The solution procedure involves minimization of a two part error function. The first part represents the deviation of the decision variables (the O-D flows) from the apriori values. The second part represents the deviation of estimated link volumes from the observed counts.

Cascetta et al. propose two types of estimation procedures – a *simultaneous* estimation procedure and a *sequential* estimation procedure. The former attempts to simultaneously solve for O-D flows corresponding to all time-intervals of interest. The latter attempts to solve for O-D flows one interval at a time utilizing the estimate from the prior interval as an apriori matrix. The tests conducted by Cascetta et al. indicate very small gains in precision from the simultaneous estimation over the sequential estimation. Because of its computational advantages however, the sequential procedure seems to be the preferred one for practical applications.

We present in this section, a modification of their sequential estimator that incorporates a stochastic assignment matrix. Consider the second of the two approaches suggested in Section 4. Since the sequential model only estimates O-D flows corresponding to one period and holds O-D flows corresponding to prior periods constant, equations (2), (15) and (16) are modified as follows:

$$\begin{aligned}
\mathbf{y}_h &= \Sigma_{p=h-p'}^{h-1} a(\hat{\mathbf{T}}_{h-1}, \mathbf{t}_h, \hat{\mathbf{q}}_p)\hat{\mathbf{x}}_p + a(\mathbf{t}_h, \mathbf{q}_h)\mathbf{x}_h + \mathbf{v}_h && (18)\\
\mathbf{t}_h^\bullet &= \mathbf{t}_h + \lambda_h && (19)\\
\mathbf{q}_h^\bullet &= \mathbf{q}_h + \psi_h && (20)
\end{aligned}$$

where $\hat{\mathbf{x}}_p$, $\hat{\mathbf{T}}_p$, and $\hat{\mathbf{q}}_p$ denote estimates from prior intervals that are fixed during interval h.

A GLS based solution would then involve minimization of the following error criterion for each interval:

$$[\hat{\mathbf{x}}_h, \hat{\mathbf{t}}_h, \hat{\mathbf{q}}_h] = arg\ min\ [(\mathbf{x}_h - \mathbf{x}_h^H)'\mathbf{W}_h^{-1}(\mathbf{x}_h - \mathbf{x}_h^H) + (\mathbf{y}_h - \mathbf{y}_h^\bullet)'\mathbf{V}_h^{-1}(\mathbf{y}_h - \mathbf{y}_h^\bullet)$$
$$+(\mathbf{t}_h - \mathbf{t}_h^\bullet)'\mathbf{\Lambda}_h^{-1}(\mathbf{t}_h - \mathbf{t}_h^\bullet) + (\mathbf{q}_h - \mathbf{q}_h^\bullet)'\mathbf{\Psi}_h^{-1}(\mathbf{q}_h - \mathbf{q}_h^\bullet)$$

where $\mathbf{y}_h^\bullet = \Sigma_{p=h-p'}^{h-1} a(\hat{\mathbf{T}}_{h-1}, \mathbf{t}_h, \hat{\mathbf{q}}_p)\hat{\mathbf{x}}_p + a(\mathbf{t}_h, \mathbf{q}_h)\mathbf{x}_h$. $\mathbf{W}_h$, $\mathbf{V}_h$, $\mathbf{\Lambda}_h$ and $\mathbf{\Psi}_h$ represent the covariances of the error terms $\mathbf{w}_h$, $\mathbf{v}_h$, λ_h and ψ_h in equations (17), (18), (19) and (20) respectively; their inverses reflect the degree of confidence placed on the various sources of information. The optimization would be subject to non-negativity constraints on $\mathbf{x}_h$, $\mathbf{t}_h$ and on the fact that $\Sigma_{k\in\mathcal{K}_r} q_{kp} = 1 \ \ \forall \ \ (r,p)$ and $0 \leq q_{kp} \leq 1 \ \forall \ (k,p)$.

Following a similar vein, the first approach suggested in Section 4 would involve minimization of the following error criterion:

$$\begin{aligned}[\hat{\mathbf{x}}_h, \hat{\mathbf{a}}_p^h] = arg\ min\ [(\mathbf{x}_h - \mathbf{x}_h^H)'\mathbf{W}_h^{-1}(\mathbf{x}_h - \mathbf{x}_h^H) \\ +(\mathbf{y}_h - \mathbf{y}_h^\bullet)'\mathbf{V}_h^{-1}(\mathbf{y}_h - \mathbf{y}_h^\bullet) \\ +\Sigma_{p=h-p'}^{h}(\mu(\mathbf{a}_p^h) - \mu(\mathbf{a}_h^{p\bullet}))'(\mathbf{U}_h^p)^{-1}(\mu(\mathbf{a}_p^h) - \mu(\mathbf{a}_h^{p\bullet}))]\end{aligned} \tag{21}$$

where $\mathbf{U}_h^p$ is the covariance matrix for the error term $\mathbf{u}_h^p$ in equation (12). $\mathbf{a}_h^{p\bullet}$ represents the value of the assignment matrix computed from the equations (6) and (7) that use measured or estimated travel times and route choice fractions. Again standard constraints would apply on $\mathbf{x}_h$ and $\mathbf{a}_h^p$.

6 Real-Time Estimation and Prediction

All of the discussion in the preceding sections can be extended to real-time models. Previous work by the authors (Ashok and Ben-Akiva (1993, 2000)) has shown that the O-D estimation and prediction problem can be formulated as a state-space model. The main idea behind this approach is the definition of the state-vector in terms of the *deviations* of O-D flows from apriori flows (usually the best historical estimates) instead of the O-D flows themselves.

It is worth reiterating here, the rationale behind the use of deviations in the transition equation of the state space model. An autoregressive process based on the O-D flows as defined by Okutani (1987) can only capture temporal interdependencies among O-D flows. Such a model does not represent any structural information about trip patterns. The pattern of O-D trips is a function of spatial and temporal distribution of activities as well as characteristics of the transportation system. It is highly unlikely therefore that a simple autoregressive process would be able to capture the complex structure of activities that result in the spatial and temporal pattern of trip making.

Suppose that O-D matrices have been estimated from historical data for several previous days or months. These already estimated O-D matrices subsume a wealth of information about the relationships that affect trip making and about their variation over space and time. One simple way then of incorporating structural relationships is to include all the prior estimation into the real-time O-D estimation problem. The simplest way to do this is to use *deviations* of O-D flows from best available historical estimates instead of the actual flows themselves as unknown variables. Thus the estimation and prediction process would have indirectly taken into account all the experience gained over many prior estimations and would be richer in its structural content.

The idea of deviations overcomes another difficulty that was recognized by Okutani. A normal distribution for traffic variables (such as O-D flows) is a useful property for available statistical tools such as the Kalman Filtering technique used by Okutani. However, the traffic flow variables used by Okutani have skewed distributions whereas the corresponding deviations would have symmetric distributions and hence be more amenable to approximation by a normal distribution.

Given these arguments, the transition equation may be stated as follows:

$$\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^{H} = \Sigma_{p=h-q'}^{h} \mathbf{f}_{h}^{p} (\mathbf{x}_p - \mathbf{x}_{p}^{H}) + \mathbf{w}_{h} \tag{22}$$

where $\mathbf{f}_h^p$ is an $(n_{OD} * n_{OD})$ matrix of effects of ($\mathbf{x}_p$ - $\mathbf{x}_p^H$) on $(\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^H)$, q' the degree of the autoregressive process and $\mathbf{w}_h$ an $(n_{OD} * 1)$ vector of random errors. The basic measurement equation is still given by (2). In *deviation* form, this equation can be given by

$$\mathbf{y}_{h} - \mathbf{y}_{h}^{H} = \Sigma_{p=h-p'}^{h} \mathbf{a}_{h}^{p} (\mathbf{x}_p - \mathbf{x}_{p}^{H}) + \mathbf{v}_{h} \tag{23}$$

where $\mathbf{y}_h^H = \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p^H$. The error terms are assumed to satisfy standard assumptions of independence, zero mean and zero serial correlation (For details, refer to Ashok and Ben-Akiva (1993)).

The transition equation as given in (22) specifies dependence of the deviations in O-D flows for period $h+1$ on deviations of more than one preceding period. Likewise, (2) specifies the dependence of measurements on O-D flows of more than one preceding period. A common technique to overcome this problem is to construct an augmented state comprising of O-D flow deviations from multiple time-intervals as in Ashok and Ben-Akiva (1993) or Okutani (1987). However, this procedure results in vastly increasing the size of the state vector potentially rendering a real-time application of the model to large and congested networks infeasible. An alternate approximate procedure was accordingly proposed and implemented in Ashok and Ben-Akiva (2000). The approximation is based on the conjecture that most of the information about an O-D flow is likely to be provided the first time it is counted. If this were true, O-D flows corresponding to prior departure intervals could be held constant at their prior estimated values and only the flows for the current departure interval need to be estimated (Note that this is analogous in philosophy to the sequential offline model described earlier). The transition and measurement equations for this model can be stated as follows:

$$\mathbf{y}_h - \mathbf{y}_h^H = \mathbf{a}_h^h(\mathbf{x}_h - \mathbf{x}_h^H) + \mathbf{b}_h + \mathbf{v}_h \tag{24}$$

$$\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^H = \mathbf{f}_h^h(\mathbf{x}_h - \mathbf{x}_h^H) + \mathbf{c}_h + \mathbf{w}_h \tag{25}$$

where
$\mathbf{b}_h = \Sigma_{p=h-p'}^{h-1}\mathbf{a}_h^p(\hat{\mathbf{x}}_p - \mathbf{x}_p^H)$
$\mathbf{c}_h = \Sigma_{p=h-q'}^{h-1}\mathbf{f}_h^p(\hat{\mathbf{x}}_p - \mathbf{x}_p^H)$ and $\hat{\mathbf{x}}_p$ is an estimate of $\mathbf{x}_p$.

In order to extend this framework to incorporate the ideas described in Section 4, we need additional measurement and transition equations. For the second approach, we augment the state by the travel times (or speeds) and route choice fractions. Equations (19) and (20) are used as additional measurement equations. We hypothesize that the transition equations for the travel times (or speeds) and route choice fractions follow the relationship shown below:

$$t_{ih+1} = \frac{t_{ih+1}^H}{t_{ih}^H} t_{ih} + \gamma_{ih}^1 \tag{26}$$

$$q_{ih+1} = \frac{q^H_{ih+1}}{q^H_{ih}} q_{ih} + \gamma^2_{ih} \tag{27}$$

where t_{ih} and q_{ih} denote the ith travel time and route choice fraction respectively. The conjecture underlying these equations is that the ratio of interval-over-interval travel times (or route-choice fractions) is stable on a day-over-day basis. The superscript H on the variables indicates historical values. γ^1_{ih} and γ^2_{ih} allow for errors in these equations. In matrix form, these equations can be represented as follows:

$$\mathbf{t}_{h+1} = \mathbf{M}_h \mathbf{t}_h + \mathbf{\Gamma}^1_h \tag{28}$$

$$\mathbf{q}_{h+1} = \mathbf{N}_h \mathbf{q}_h + \mathbf{\Gamma}^2_h \tag{29}$$

where $\mathbf{M}_h$ and $\mathbf{N}_h$ are diagonal matrices with the ith diagonal element given by (t^H_{ih+1}/t^H_{ih}) and (q^H_{ih+1}/q^H_{ih}) respectively. $\mathbf{\Gamma}^1_h$ and $\mathbf{\Gamma}^2_h$ represent random vectors of errors. Thus the complete specification for this approach would involve Equations (25), (28) and (29) as transition equations and Equations (24), (19) and (20) as measurement equations.

Similarly for the first approach, we augment the state with the assignment fractions. Equation (12) now represents the additional measurement equation. We hypothesize that the transition equation obeys the following relationship:

$$\mu(\mathbf{a}^{p+1}_{h+1}) = \Upsilon_h \mu(\mathbf{a}^p_h) + \mathbf{\Gamma}^3_h \tag{30}$$

where Υ_h is a diagonal matrix with the ith element given by $[\mu(\mathbf{a}^{p+1^H}_{h+1})]_i/[\mu(\mathbf{a}^{p^H}_h)]_i$, the superscript H again denoting the historical value. $\mathbf{\Gamma}^3_h$ as before denotes the error vector. A complete specification of this approach therefore involves Equations (25) and (30) as transition equations and Equations (24) and (12) as measurement equations.

Regardless of the approach used, we notice that the measurement equation system is non-linear in the state-vector. A popular way to tackle the problem of non-linear estimation in a state-space framework has been to use the *Extended* Kalman Filter (EKF) algorithm (Refer to Appendix). In our problem, this involves a first-order Taylor linearization of the measurement equation about the best available estimate of the state vector. The resulting update equations for the filter closely resemble those of the conventional *linear* Kalman Filter. The estimates obtained from the

EKF could be improved by performing successive iterations of linearization and re-estimation leading to an *Iterated* EKF. A detailed discussion of non-linear filtering techniques may be found in Gelb (1974).

7 Case Studies

7.1 Data Description

The first dataset covered a 5.2 mile stretch of I-880 (NorthBound) near Hayward, California. This section had 4 on-ramps and 5 off-ramps with 20 O-D pairs. Ten minute detector data on traffic volumes and average speeds was available at 10 detector locations for a 2.5 hour morning peak period. Data over seven days was available. The first six days were used to generate inputs to the various models while the last day was used as a test-bed for both the offline and real-time models. Because of the linear nature of the network, there was no route choice in the problem. The maximum travel time between any O-D pair in the network was about 9 minutes. Congestion level was heavy during certain time intervals with speeds reaching 15-20 mph at some locations.

The second dataset was a 32 km freeway encircling the city of Amsterdam with 20 entrance and exit ramps. In this dataset, there exist two routes between each O-D pair – clockwise and counterclockwise. Moreover, in this dataset, we used a combination of actual and synthetic data to gain additional insight into model performance. After removing sensors with large errors from the dataset, information on observed average speeds and link counts was available at 65 locations over one minute intervals for one day. This information was aggregated over 15 minute intervals for the morning peak. Synthetic data on O-D flows and speeds was generated for an additional day by repeatedly applying Equations (22) and (28) using the first day estimates as the historical database. Details on the data generation process and the calibration of the transition error covariances may be found in Ashok (1996). There were 291 O-D flows to be estimated in each interval.

7.2 Preparation of Inputs

7.2.1 Offline Estimation

The offline models were tested on the I-880 data only. For these models, two apriori matrices were used, i.e., two sets of equations (17) were used with correspondingly two quadratic terms in the error criterion with different weighting factors. The first was obtained from the estimation results of previous days i.e. the apriori matrix for interval h corresponded to the estimated matrix for interval h for a previous day. The second matrix that was used corresponded to the estimate obtained on the same day in interval $h-1$. The covariance matrices for both the error terms were calibrated from the residuals for the previous days.

For stochastic assignment matrix modeling using the second approach, two sets of travel speeds were used. The first corresponded to the estimation result for the previous period $(h-1)$. The second corresponded to the values measured by the detectors during the current period h. Again, error variances for both error terms were calibrated from residuals for previous days.

Some elaboration about using speeds as the fundamental entities instead of travel times in our models is in order here. All the equations described in previous sections were written in terms of speeds rather than travel times. The structure of the equations was exactly the same. A potential problem with using speeds (as suggested by a reviewer) is that in congested traffic, small errors in speed estimates could lead to large errors in travel times. Since α_{lh}^{kp} is computed from travel times (equation (7)), this could affect the quality of the estimation process.

For the first approach, again, two sets of assignment matrix fractions were used. The first corresponded to the estimation result for the previous period $(h-1)$ while the second to the values computed using equations (7) and the measured speeds during the current period h. As before, error variances were calibrated from residuals for previous days. For the studies in this paper, 39 of the 106 assignment fractions were estimated along with the O-D flows. The remaining fractions corresponded to O-D pairs with very low flows and hence their randomness was not considered.

Since no vehicle stayed on the network over more than two time-intervals, $p' = 1$ i.e. during any interval p, only $\mathbf{a}_p^{p-1}$ and $\mathbf{a}_p^{p}$ needed to be estimated. Because no route choice existed for the I-880 dataset, the elements of $\mathbf{a}_p^{p-1}$ had to obey the following relationship:

$$a_{lp}^{rp-1} = 1 - a_{lp-1}^{rp-1} \tag{31}$$

The above relationship had to be satisfied for all (l, r) pairs for which $l \in L_k$ for any $k \in K_r$. We enforced this constraint in the following manner. For any interval h, we estimated the assignment fractions a_{lh}^{rh}. We then computed a_{lh+1}^{rh} using (31) and moved to interval $h + 1$.

7.2.2 Real-Time Estimation and Prediction

For the I-880 data, historical O-D matrices were created using the offline methods described earlier. The transition matrices $\mathbf{f}_h^p$ were obtained by simple OLS regressions on actual deviations from historical estimates. Two assumptions were made in the process of obtaining these matrices. Firstly, it was assumed that the autoregressive structure remained constant over the peak period. Secondly, it was assumed that a flow between O-D pair r for a period was related *only* to the rth O-D flows of prior intervals. The variances of the transition equation errors was approximated from the residuals of these regressions. An autoregressive process of order four provided the best fit. The matrices $\mathbf{M}_h$, $\mathbf{N}_h$ and Υ_h were directly computed from historical estimates. The error variances for the measurement equation were the same as for the offline methods described earlier. Constraints (31) were enforced exactly as before. In this context, we note that the EKF procedure does not guarantee that the estimated assignment fractions lie between zero and unity as required. We therefore truncated each negatively estimated fraction to zero and each fraction greater than unity to one. Our empirical results indicated that these were extremely rare occurrences.

For the Amsterdam Beltway, the historical database was obtained from the first day using the Cascetta et al. sequential model (with a GLS estimator) on the observed speeds and link counts over 15-minute estimation intervals. For route-choice, a simple logit model with path travel time as the only attribute for each route was used. Details on calibration of this model may be found in Ashok (1996). Note that the framework presented in this paper does not preclude use of more complicated route-choice models such as those in Cascetta (1996). The general testing procedure is shown in Figure 1. Error was added to the counts and speeds using the equations:

$$Measured\_counts = True\_counts * (1 - \delta_{cts} + 2 * \delta_{cts} * U) \tag{32}$$

$$Measured\_speeds = True\_speeds * (1 - \delta_{spd} + 2 * \delta_{spd} * U) \tag{33}$$

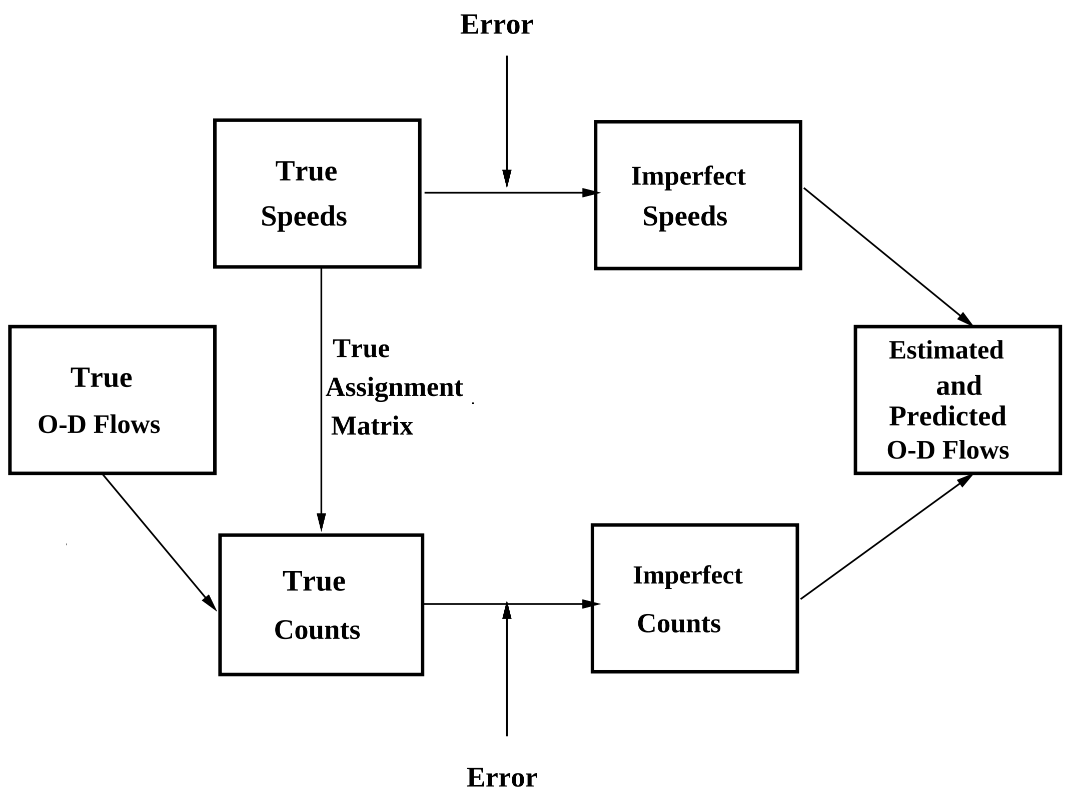

Figure 1: Testing procedure with synthetic data

Errors	*Off-Base*	*Off-Stoc-Spd*	*Off-Stoc-Assg*	*Historical*
RMS	26.9748	26.8931	26.7943	85.6206
RMSN	0.0254	0.0254	0.0253	0.0812

Table 1: Errors in Link Volumes Using Offline Models: I-880

where U is a random number drawn from a uniform distribution between zero and one. By assigning different values for δ_{cts} and δ_{spd}, the sensitivity of the model to errors in counts and speeds could be investigated. Measurement error variances could be exactly computed as a function of δ_{cts} and δ_{spd}. For each scenario, multiple runs were conducted.

7.3 Results

For evaluation of the results of various models, the following error measures were used.

1. Root Mean Square (RMS) Error $= \sqrt{\frac{\sum_i (x_i - \hat{x}_i)^2}{N}}$

2. Root Mean Square Error Normalized (RMSN) $= \frac{\sqrt{N \sum_i (x_i - \hat{x}_i)^2}}{\sum_i x_i}$

where the x and $\hat{x}$ represent the true and estimated O-D flows respectively and the summation is over all detectors and all intervals for which analysis was carried out.

For the I-880 dataset, the true O-D flows were unknown. We therefore compared the link counts obtained from re-assigning the estimated/predicted O-D flows with the measured link counts. This error measure should be interpreted cautiously since (a) the measured counts are themselves erroneous and (b) it is possible that even though the estimated link counts better match the measured link counts compared to models with a fixed assignment matrix that have fewer parameters, the estimated O-D flows may differ considerably from the true O-D flows.

7.3.1 Offline Estimation

Results for the offline models are shown in Table 1. Model *Off-Base* represents the basic (Cascetta et al.) sequential model that uses a deterministic assignment matrix. Model *Off-Stoc-Spd* estimates O-D flows along with speeds. Model *Off-Stoc-Assg* estimates O-D flows with assignment fractions. The last column shows the effect of

		Base-Appx	*Stoc-Spd*	*Stoc-Assg*	*Historical*
RMS Error	Filtered	27.6405	21.7885	21.3744	85.1615
	1-Step Predicted	79.4600	111.1101	72.4859	88.1974
	2-Step Predicted	77.4488	116.1716	67.2875	86.2065
	3-Step Predicted	102.3850	102.1264	74.9402	68.7522
RMSN Error	Filtered	0.0261	0.0205	0.0202	0.0808
	1-Step Predicted	0.0749	0.1062	0.0683	0.0832
	2-Step Predicted	0.0732	0.1126	0.0636	0.0815
	3-Step Predicted	0.0979	0.1012	0.0717	0.0657

Table 2: Errors in Link Volumes Using Real-Time Models: I-880

assigning historical O-D flows to the network. We observe that though all the models significantly outperform the historical values in terms of fit to link counts, there is almost no difference between the performance of Models *Off-Base*, *Off-Stoc-Spd* and *Off-Stoc-Assg*. An examination of the estimated O-D flows for Models *Off-Base*, *Off-Stoc-Spd* and *Off-Stoc-Assg* indicated very little difference as well. One explanation for these results could be that for a linear network (with no route-choice) where speeds do not change drastically interval-over-interval, the model is fairly insensitive to errors in the assignment matrix. Indeed, for Model *Off-Stoc-Spd*, an RMSN measure that compared estimated speeds with the measured ones indicated a difference of 13.7% between the two. It is interesting that this large difference in speeds does not seem to have translated into comparable differences in O-D and link flow estimates. Again, this might have to do with the linear structure of the network.

7.3.2 Real-Time Estimation and Prediction

Table 2 gives a comparison between the different real-time models tested for the I-880 dataset. Model *Base-Appx* is the basic real-time model with fixed assignment matrices (based on observed speeds) as represented by equations (24) and (25). Model *Stoc-Spd* denotes the real-time model with the state vector consisting of O-D flow deviations and speeds while Model *Stoc-Assg* has the O-D flow deviations augmented with the assignment fractions. Finally, results using the historical O-D flows have been shown for comparison.

We make several observations. Firstly, as in the offline case, all the three models show significantly better performance in filtering (estimation), compared to historical values. For the most part, one-step and two-step predictions are better as well. The most interesting find however is that unlike in the offline case, Models *Stoc-Spd* and *Stoc-Assg* significantly outperform Model *Base-Appx*. One reason could be that insofar as the real-time model works with deviations in O-D flows instead of the O-D flows themselves, it represents a different statistical model with different properties compared to the GLS based offline model. For example, the error terms in equation (25) are more amenable to a normal approximation than those in say equation (17). Also, unlike the offline methods, the transition equations in the real-time case provide an explicit modeling of the relationship between speeds and assignment fractions across time-intervals – thus the real-time models use more information. These equations also used historical speeds and assignment fractions – something the offline models did not. We also notice that Model *Stoc-Spd* performs worse than the others in prediction indicating perhaps the need for examining alternatives to equation (26).

For the Amsterdam Beltway, Figure 2 shows the degradation in the performance of *Base-Appx* as the error parameter δ_{spd} (equation (33)) is varied. (Recall that $\delta_{spd} = e$ implies a measurement error within $\pm e\%$ in speeds. For these graphs, we used a value of δ_{cts}=0.10.) Errors are stratified by size of O-D flow. Each bar shows the RMS/RMSN errors for a specific value of δ_{spd}. For comparison, errors in employing historical O-D flows (in other words, the difference between O-D flows on the first and second days) are shown in the last bar. It can be seen that while the model is fairly robust with respect to quality of speeds, the bias becomes significant for very large values of δ_{spd}. Figure 3 compares the fixed and stochastic assignment matrix models for $\delta_{cts} = 0.1$ and $\delta_{spd} = 0.8$. The first bar pertains to the model *Base-Appx* with true speeds, i.e., $\delta_{spd} = 0$. The second bar pertains to the application of *Base-Appx* with $\delta_{spd} = 0.8$. The third and fourth bars pertain to the *Stoc-Spd* and *Stoc-Asg* models respectively. Again, errors in historical estimates are shown for comparison. We see that *Stoc-Spd* performs extremely well – in fact, its performance almost rivals that of *Base-Appx* with true speeds. The same cannot be said, however, for *Stoc-Asg* which does no better than *Base-Appx*. This poor performance of *Stoc-Asg* relative to *Stoc-Spd* could be because of the following reasons:

- Speeds for the second day were generated by application of equation (28). Thus, the extra information provided by (28) was extremely valuable for model *Stoc-Spd* (since this was *true* information). Relatively, the information provided by

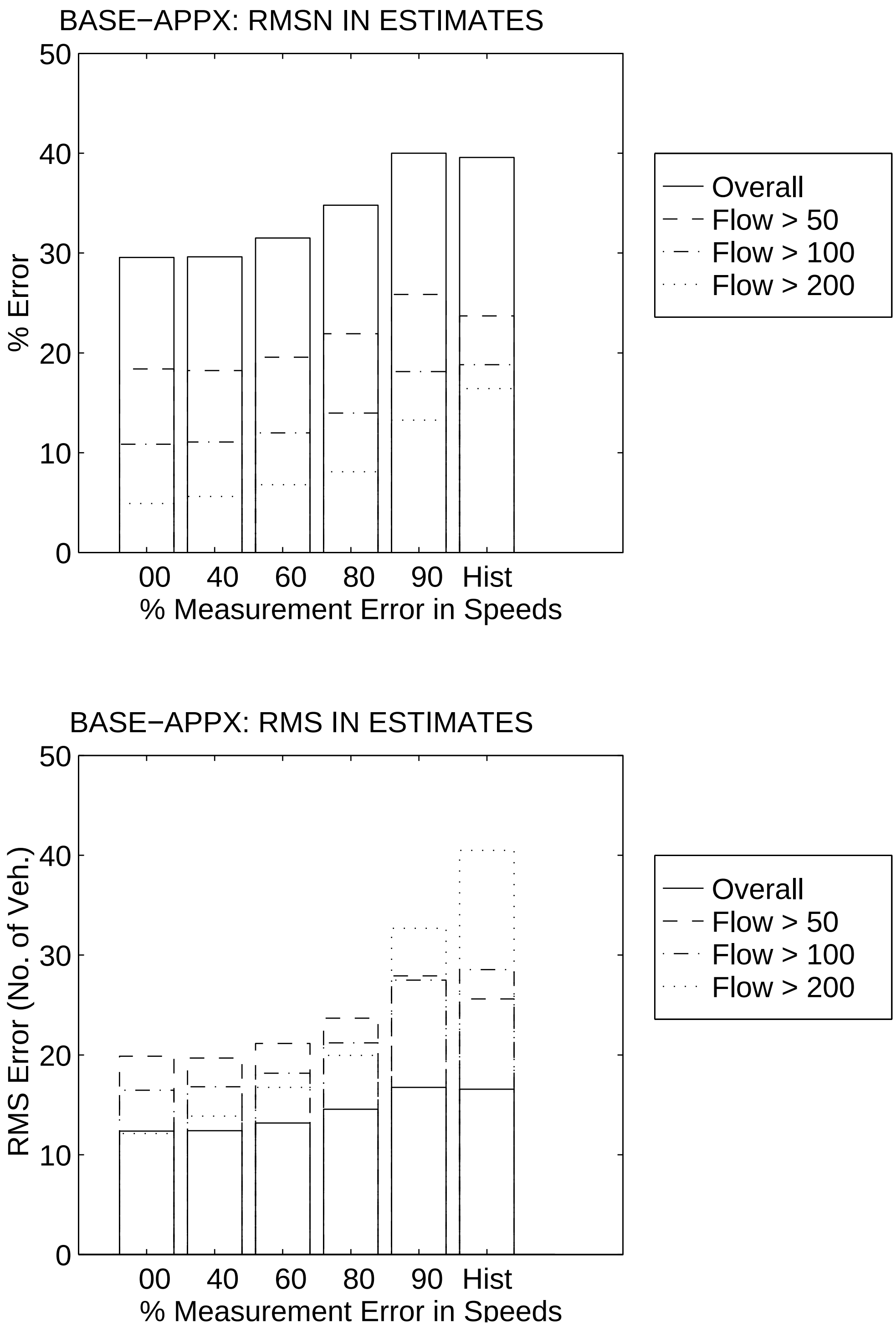

Figure 2: Model *Base-Appx* as a function of accuracy of speeds: Beltway

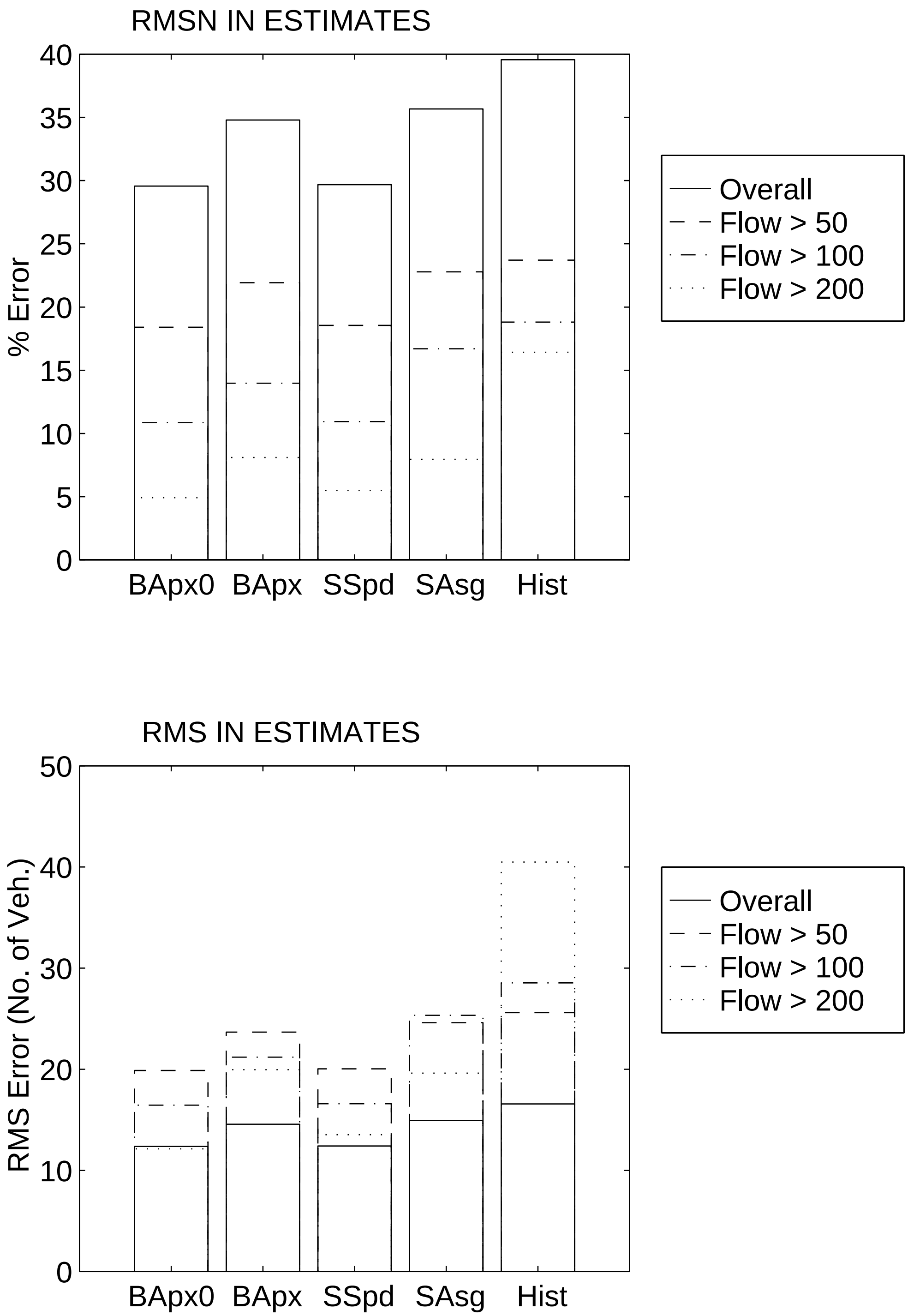

Figure 3: Fixed and Stochastic Assignment Matrix Models: Beltway

(28) for *Stoc-Asg* was not as useful since the transition equation (30) used by *Stoc-Asg* was not completely consistent with the data generation procedure.

- Because of computational reasons, all the assignment fractions could not be estimated. Only those corresponding to high O-D flows (234 in number) were considered stochastic.

- Round-off errors could have contributed to inaccuracies in the final results.

7.4 Major Findings

The following summarizes the major findings of this case study.

1. In offline estimation, the generalized models do not show much improvement over conventional models. As mentioned earlier, this could be because of the linear network structure for the I-880 dataset. All models tested in this study however show significant superiority over historical values.

2. In real-time estimation and prediction, the generalized models show potential for improved performance. The conclusion to be drawn from the case studies is that a good knowledge of the underlying process that describes temporal evolution of speeds or assignment fractions (both in terms of the structure of the transition equations, as well as in the error covariances) is needed in order for these more complicated models to perform better. In this context, an advantage of the stochastic speeds approach over the stochastic assignment fractions approach is that since speeds are directly observed, it could be easier to develop and calibrate more complicated transition equations for this approach. A reassuring result – particularly for the stochastic speeds approach – in applying these models is that the impact of nonlinearities in the measurement equation do not appear to be the source of biases.

3. The generalized models are computationally more intensive than the conventional models. In practical applications therefore, the importance of a careful choice of additional decision variables (travel times, route-choice fractions and assignment parameters) cannot be overemphasized.

4. Part of the reason for lack of a significant improvement in performance of the generalized models in some cases could be due to the presence of other error

sources. For example, the sequential offline method treats previous interval estimates as fixed or error-free. Similarly, the real-time approximate method does not allow for updating of O-D flows estimated during previous time intervals and moreover treats historical O-D flows as exogenous inputs. While these conditions can be relaxed (see for example, Ashok (1996)), the price to be paid would be the addition of a large number of parameters.

It is reiterated that for the reasons mentioned earlier, fit to counts is an imperfect performance criterion. The results presented for the I-880 case study therefore are only useful for identifying general trends. To reach definitive conclusions based on these would be premature.

8 Conclusion

In this paper, a class of generalized models have been presented for dynamic O-D matrix estimation and prediction. These models explicitly recognize the fact that the mapping between O-D flows and link volumes is subject to uncertainty. Two approaches have been developed to address this issue. Both have been evaluated using case studies based in part on actual data. Preliminary results indicate that the generalized models offer potential for improvement over conventional ones. Future research should focus on evaluating these models on large networks, on investigating further techniques for stochastic modeling of the assignment matrix, and on computational issues.

References

K. Ashok, "Estimation and Prediction of time-dependent Origin-Destination Flows," Ph. D. Dissertation, Center for Transportation Studies, Massachusetts Institute of Technology, September 1996.

K. Ashok and M. Ben-Akiva, "Dynamic O-D matrix estimation and prediction for Real-Time Traffic Management Systems," In C.F. Daganzo, editor, *Transportation and Traffic Theory*, 465-484 (1993).

K. Ashok and M. Ben-Akiva, "Alternate Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows," *Transportation Science* 34 (2000).

K. Ashok and M. Ben-Akiva, "Estimation and Prediction of Dynamic O-D Flows by Combining Different Types of Measurements," Working Paper, Center for Transportation Studies, Massachusetts Institute of Technology, 1997.

M.G.H. Bell, "The real time estimation of origin-destination flows in the presence of platoon dispersion," *Transportation Research* 25, 115-125 (1991).

M. Ben-Akiva, "Dynamic Network Equilibrium Research," *Transportation Research* 19A, 429-431 (1985).

M. Ben-Akiva, H.N. Koutsopoulos and A. Mukundan, "A Dynamic Traffic Model System," *IVHS Journal* 2, 1-19 (1994).

E. Cascetta, D. Inaudi, and G. Marquis, "Dynamic Estimators of Origin-Destination Matrices using Traffic Counts," *Transportation Science* 27, 363–373 (1993).

E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks," University of Napoli, 1996.

G.L. Chang and X. Tao, "Estimation of Dynamic O-D Distributions for Urban Networks," *Transportation and Traffic Theory*, (1996).

G.L. Chang and J. Wu, "Recursive estimation time-varying O-D flows from traffic counts in freeway corridors," *Transportation Research* 28 (1994).

C.K. Chui and G. Chen, Kalman Filtering with Real-Time Applications, Springer-

Verlag, 1991.

M. Cremer and H. Keller, "A new class of Dynamic methods for the identification of Origin-Destination Flows," *Transportation Research* 21 117–132 (1987).

A. Gelb, editor, Applied Optimal Estimation, M.I.T. Press, 1974.

W. Greene, Econometric Analysis, Maxwell Macmillan International Publishing Group, 1993.

D. Inaudi, E. Kroes, S. Manfredi, and S. Toffolo, "The DYNA on-line O-D Estimation and Prediction Model," First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems, December 1994.

P. Kachroo, A. Narayanan, and K. Ozbay, "Investigating the Use of Kalman Filtering Approaches for Dynamic Origin-Destination Trip Table Estimation," Center for Transportation Research, Virginia Tech, 1995.

N.L. Nihan and G.A. Davis, "Recursive estimation of Origin-Destination matrices from input/output counts," *Transportation Research* 21 149–163 (1987).

I. Okutani, "The Kalman Filtering Approach in Some Transportation and Traffic Problems," In N.H. Gartner and N.H.M. Wilson, editors, *Transportation and Traffic Theory*, 397–416 (1987).

N. van der Zijpp, "Dynamic Origin-Destination Matrix Estimation on Motorway Networks," Ph. D. Thesis, Department of Civil Engineering, Delft University of Technology, 1996.

APPENDIX

Consider the non-linear system given by the following two equations:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{f}_t(\mathbf{x}_t) + \mathbf{w}_t \\ \mathbf{y}_t &= \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t \end{aligned} \tag{34}$$

where $\mathbf{f}_t$ and $\mathbf{h}_t$ are vector-valued functions of the state-vector $\mathbf{x}_t$ and the errors $\mathbf{w}_t$ and $\mathbf{v}_t$ have covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$ respectively. Assume that the initial state is given by $\bar{\mathbf{x}}_0$ with variance-covariance matrix $\mathbf{P}_0$. The Extended Kalman Filter (EKF) is then given by:

$$
\begin{aligned}
\mathbf{\Sigma}_{0|0} &= \mathbf{P}_0 \\
\mathbf{\Sigma}_{t|t-1} &= \mathbf{A}_{t-1}\mathbf{\Sigma}_{t-1|t-1}\mathbf{A}'_{t-1} + \mathbf{Q}_{t-1} \\
\mathbf{K}_t &= \mathbf{\Sigma}_{t|t-1}\mathbf{C}'_t(\mathbf{C}_t\mathbf{\Sigma}_{t|t-1}\mathbf{C}'_t + \mathbf{R}_t)^{-1} \\
\mathbf{\Sigma}_{t|t} &= \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{C}_t\mathbf{\Sigma}_{t|t-1} \\
\hat{\mathbf{x}}_{0|0} &= \bar{\mathbf{x}}_0 \\
\hat{\mathbf{x}}_{t|t-1} &= \mathbf{f}_{t-1}(\hat{\mathbf{x}}_{t-1|t-1}) \\
\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}_t(\hat{\mathbf{x}}_{t|t-1})) \\
t &= 1, 2, \ldots
\end{aligned}
\tag{35}
$$

where

$$
\begin{aligned}
\mathbf{A}_t &= \frac{\partial \mathbf{f}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t}) \\
\mathbf{C}_t &= \frac{\partial \mathbf{h}_t}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t-1})
\end{aligned}
$$

There is a variation of the discrete-time EKF worth mentioning here. The update step of the EKF involves the linearization of the measurement equation about the present best estimate of $\mathbf{x}_t$, i.e. $\hat{\mathbf{x}}_{t|t-1}$. However, once this step is completed, a presumably superior estimate $\hat{\mathbf{x}}_{t|t}$ is available which could then be used to relinearize the measurement equation and redo the update step. These iterations could be repeated as many times as deemed necessary. The resulting EKF is termed the *Iterated* EKF.

Acknowledgements

We are thankful to Mark Hickman, at University of Arizona, and Hisham Noeimi and Karl Petty at the University of California, Berkeley for helping us obtain I-880 data. We are also indebted to Nanne Van der Zijpp at Delft University for providing access to the Amsterdam Beltway data. We also benefited from the insights of the anonymous referees of this paper.