Dynamic Origin-Destination Matrix Estimation and Prediction for Real-Time Traffic Management Systems

K. Ashok and M.E. Ben-Akiva
Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract
This paper provides a framework for real-time estimation/prediction of time-dependent Origin-Destination(O-D) matrices. The problem is formulated as a Kalman Filter where the state vector consists of deviations of O-D flows from prior estimates based on historical data. Based on the set of link counts obtained at the end of each interval, the O-D flow deviations predicted for that interval are modified. This process makes use of information about travel times and path choice fractions for vehicles that are already on the network. Apart from generating estimates of O-D matrices for the *current* time interval, the model also has facility to predict O-D matrices corresponding to *future* departure intervals and update O-D matrices corresponding to *past* departure intervals.

As a case study, the performance of the model is evaluated using data from the Massachusetts Turnpike. Initial results are favorable and suggest that the filtering technique is robust.

1 Introduction

Increasing attention has been paid in recent years to dynamic traffic management systems that include dynamic driver information and adaptive traffic control. Critical to the success of these systems is the presence of a real-time dynamic traffic assignment (DTA) model. (See Ben-Akiva et al.[2] for discussion and presentation of a real-time DTA.) A key input to the DTA is a time-dependent Origin-Destination matrix for the network. It may be mentioned that these matrices are very different from the matrices used in transportation planning applications for estimation of which, there exists a diverse body of literature. (See [1],[4],[5] for a review and bibliography.)

Since obtaining these matrices directly (for example from 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. Research in this arena can be broadly divided into two groups – pertaining to "closed" networks[1] and pertaining to general networks. Several estimators have been proposed for the former ([7],[9],[10],etc), usually for a single intersection or a freeway segment. In most practical situations

[1] A closed network implies that complete information is available on the entry and exit counts of the network at all points in time.

however, it is very difficult to envisage availability of complete information at all exit and entry locations. Also, most of these techniques do not allow for explicit modeling of measurement error – which may be an important requirement considering the heavy emphasis that dynamic traffic management systems place upon accurate sensor information. For general networks, Cascetta et al.[4] provide an extension of the static O-D estimation framework. Though this model cannot be used online in the real-time context because it does not have any predictive or updating element, it can be used as an off-line tool for creating a set of "historical estimates" of O-D flows for the network (as will be seen later). Finally, an alternative approach based on Kalman Filtering has been suggested by Okutani[11]. The state vector has been taken to be the vector of unknown O-D flows. A major problem with the specification is the autoregressive formulation of the transition equation which is described in the next section.

This paper describes a framework for real-time updating and prediction of these matrices for general networks. Section 2 discusses the mathematical formulation of the problem and the proposed estimation/prediction technique. Section 3 describes a case study where the model was implemented using data for the Massachusetts Turnpike. The paper concludes with a brief discussion of the major issues involved and directions for future research.

2 Model Formulation

2.1 The Idea of Deviations

Previous work on formulation of the O-D matrix prediction problem as a Kalman Filter by Okutani[11] uses for the transition equation, a simple autoregressive formulation for the O-D flows. An autoregressive process can only capture temporal interdependencies among O-D flows. Such a model does not represent structural information about trip patterns. The pattern of O-D trips is a function of spatial and temporal distribution of activities as well as the 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 travel demand 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 state vectors in the Kalman Filtering methodology. The transition and measurement equations of the filter shall be expressed in terms of deviations from historical estimates of O-D flows. Thus the estimation and prediction process would have indirectly taken into account all the experience gained over many prior estimations and would hence be richer in its structural content.

The above idea of deviations also overcomes another difficulty that was recognized

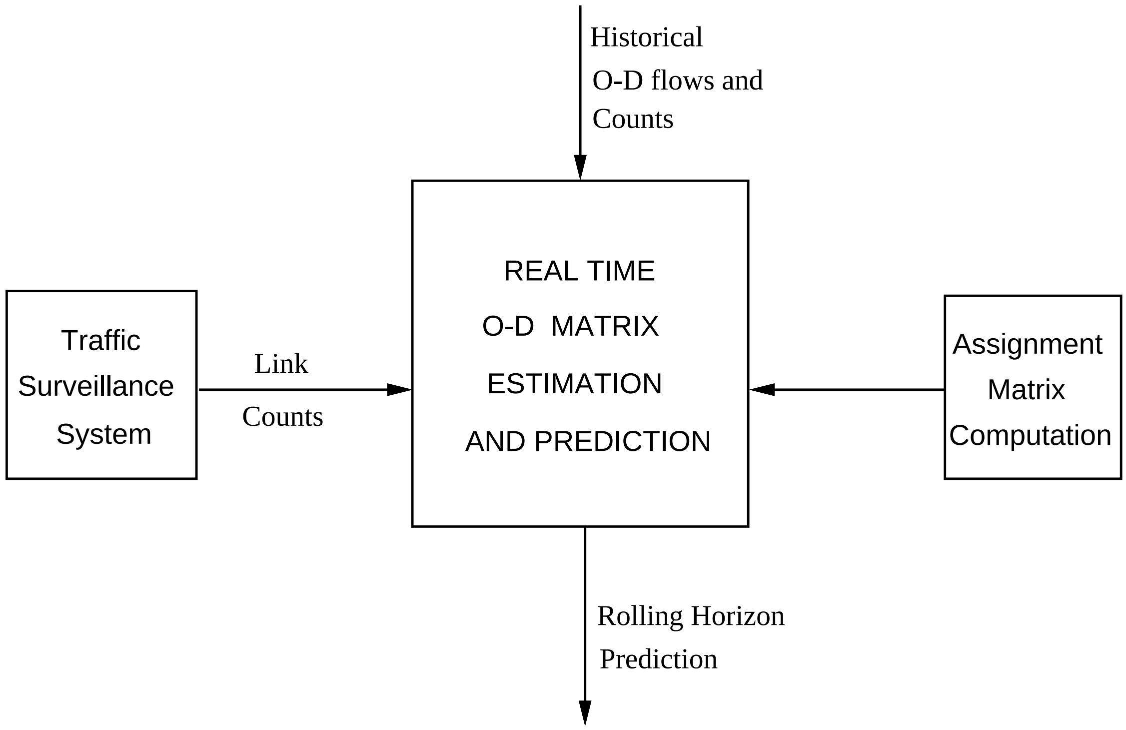


Figure 1: Real-time Interactions

by Okutani. A normal distribution for traffic variables (such as counts and 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.

2.2 Overview of Updating Methodology

Following the arguments advanced above, the state vector was defined in terms of deviations of O-D flows from historically computed estimates. Figure 1 shows the various inputs and outputs into the real-time O-D estimation module. At the end of each interval h, this module obtains historical O-D flow values from a database. The module also gets a vector of link counts for that interval from the surveillance system. In addition, it gets estimates of "assignment" matrices discussed in more detail in later sections[2]. The assignment matrices in turn have to be known in order to set up the measurement equation for the Kalman Filtering technique. By comparison of the link counts that were measured by the surveillance system with the counts obtaining by assigning the estimated O-D flows (for intervals $...h-1,h$) to the network, the O-D estimation module updates ("filters") these estimates. Predictions are then generated

[2]In simple terms, an assignment matrix maps a set of Origin-Destination matrices into a set of link flows.

for intervals $h+1, h+2, ...$ and the process continues.

Before proceeding into further details, it is worthwhile to formalize the approach.

2.3 Preliminary Definitions

Consider a period of length $\mathcal{T}$ divided into equal intervals of length T. The network is represented as a directed graph that includes a set of consecutively numbered nodes $\mathcal{N}$ and a set of numbered links $\mathcal{L}$. The network is assumed to have n_{LK} links and n_{OD} O-D pairs. It is assumed that n_l of these links are equipped with counting stations. Each O-D pair r is connected by a set of paths K_r. A path $k \in K_r$ is an ordered set of links $L_{k,r}$.

2.4 The Transition Equation

Denote by x_{rh} the number of vehicles between the rth O-D pair that left their origin in interval h and by x_{rh}^H the corresponding best historical estimate. Further, let the corresponding $(n_{OD} * 1)$ vectors of all O-D flows be given by $\mathbf{x}_h$ and $\mathbf{x}_h^H$ respectively. To set up the transition equation, it is hypothesized that the deviations in O-D flows from a historical base value at period h can be related to the deviations in O-D flows of previous time periods by the autoregressive form

$$x_{r,h+1} - x_{r,h+1}^H = \sum_{p=h-q'}^{h} \sum_{r'=1}^{n_{OD}} f_{rh}^{r'p} (x_{r'p} - x_{r'p}^H) + w_{rh} \tag{1}$$

where the coefficients $f_{rh}^{r'p}$ describe the effect of the deviation $(x_{r'p} - x_{r'p}^H)$ on the deviation $(x_{r,h+1} - x_{r,h+1}^H)$ and w_{rh} is a random error. q' is the number of lagged O-D flow deviations[3] assumed to affect the O-D flow deviation in interval $h+1$.

An autoregressive process is a commonly used structure for representation of time-series data. In equation (1), it models the temporal relationship among deviations in O-D flows. It captures correlation over time among deviations which arise from unobserved factors that are correlated over time. Such factors include weather conditions, special events, temporary changes in the transportation network, etc. Also, the model in the above form is highly general and assumes dependence of deviations corresponding to one O-D pair on deviations corresponding to other O-D pairs in prior periods. In practical applications (for example in the freeway case study in this paper) this may be unnecessarily general and relationships between deviations across different O-D pairs may be safely ignored.

In matrix form, the equation can be expressed as:

$$\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^H = \Sigma_{p=h-q'}^{h} \mathbf{f}_h^p (\mathbf{x}_p - \mathbf{x}_p^H) + \mathbf{w}_h \tag{2}$$

where $\mathbf{f}_h^p$ is an $(n_{OD} * n_{OD})$ matrix of effects of $(\mathbf{x}_p$ - $\mathbf{x}_p^H)$ on $(\mathbf{x}_{h+1} - \mathbf{x}_{h+1}^H)$ and $\mathbf{w}_h$

[3]excluding the deviation in the interval h

an $(n_{OD} * 1)$ vector of random errors. The following assumptions are made about the error vectors:

1. $E[\mathbf{w}_h] = 0$
2. $E[\mathbf{w}_h \mathbf{w}_l'] = \mathbf{Q}_h \delta_{hl}$ where $\delta_{hl} = 1$ if h=l and 0 o.w. $\forall h, l$ and $\mathbf{Q}_h$ is an $(n_{OD} * n_{OD})$ variance-covariance matrix.

The above formulation differs from that of Okutani[11] who used an autoregressive process of the form:

$$\mathbf{x}_{h+1} = \Sigma_{p=h-q'}^{h} \mathbf{f}_h^p \mathbf{x}_p + \mathbf{w}_h \tag{3}$$

As described in Section 2.1, this formulation seems to be inappropriate. It is purely statistical and given the high variability in traffic patterns over short time intervals, the noise factor in the above equation is likely to dominate.

Computation of the matrix $\mathbf{f}_h^p$ involves estimating linear regression models for each O-D pair and for each interval. If one makes the assumption that the structure of the autoregressive process remains constant with respect to h, the values of the matrix $\mathbf{f}_h^p$ would only depend on the difference $(h-p)$ and not on individual values of h and p. To simplify the problem further, it may be reasonable to assume a diagonal structure for the matrix – thereby ignoring interdependencies across O-D flows. The value of the parameter q' could be obtained from statistical significance tests on the regression coefficients for various lags.

The error covariance matrix for the transition equation could be easily calculated from the residuals of the above regressions. Again, on the assumption that the structure of the autocorrelation remains constant, the set of matrices $\{Q_{rh}\}$ would reduce to one matrix $\{Q_r\}$.

2.5 The Measurement Equation

Denote by y_{lh} the observed traffic counts on link l during interval h and by $\mathbf{y}_h$ the corresponding $(n_l * 1)$ vector. Then the measurement equation can be written as:

$$y_{lh} = \sum_{p=h-p'}^{h} \sum_{r=1}^{n_{OD}} a_{lh}^{rp} x_{rp} + v_{lh} \tag{4}$$

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

$$\mathbf{y}_h = \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p + \mathbf{v}_h \tag{5}$$

where the matrix $\mathbf{a}_h^p$ is an $(n_l * n_{OD})$ matrix[4] of contributions of $\mathbf{x}_p$ to $\mathbf{y}_h$ and $\mathbf{v}_h$ is the vector of measurement errors. Assume that

[4]The matrices $\mathbf{a}_h^p$ are the "assignment" matrices alluded to in earlier sections.

1. $E[\mathbf{v}_h] = 0$

2. $E[\mathbf{v}_h \mathbf{v}'_l] = \mathbf{R}_h \delta_{hl}$ where $\delta_{hl} = 1$ if h=l and 0 o.w. $\forall h, l$ and $\mathbf{R}_h$ is an $(n_l * n_l)$ variance-covariance matrix.

3. $E[\mathbf{w}_h \mathbf{v}'_l] = 0 \; \forall h, l$ i.e. Transition and measurement errors are uncorrelated.

The interpretation of the above equation is straightforward. The flow on any link 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 imperfect measurements.

Based on the arguments advanced in section 2.1 and in line with the formulation of the transition equation, one would like to define the state vector as a vector of deviations of O-D flows from the best historical estimates. Equation (5) can then be written as:

$$\mathbf{y}_h - \mathbf{y}_h^H = \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p (\mathbf{x}_p - \mathbf{x}_p^H) + \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p^H - \mathbf{y}_h^H + \mathbf{v}_h \qquad (6)$$

where $\mathbf{y}_h^H$ denotes the historical values of counts for time interval h. It should be noted that the matrices $\mathbf{y}_h^H$ and $\mathbf{x}_h^H$ are known matrices $\forall h$ and are obtained from a "database" created off-line.

Computation of the matrices $\mathbf{a}_h^p$ is a difficult exercise.[5] The fractions contained in these matrices depend on path fractions as well as the mapping of time-dependent path flows to link flows. One way of determining the former could be by means of discrete choice models. To estimate the latter, one would need in addition, knowledge about travel times. Travel times could be obtained from a traffic surveillance system (e.g. probe vehicles) or from a simulation model (See for example Ben-Akiva et al.[2]).

The error covariance matrix $\mathbf{R}_h$ would be computed from historical data. From observing the differences $(\mathbf{y}_h - \Sigma_{p=h-p'}^{h} \hat{\mathbf{a}}_h^p \hat{\mathbf{x}}_p)$ over multiple time-intervals, the variance-covariance matrices $\mathbf{R}_h$ can be easily computed.

2.6 Transformation of variables

Equations (2) and (6) cannot be used directly as the Kalman filter equations. This is because the transition equation as given in (2) specifies dependence of the deviations in O-D flows for period $h+1$ on deviations of *more than one* preceding period. Likewise, (6) specifies the dependence of measurement deviations on O-D flow deviations of more than one preceding period. In order to be able to apply the Kalman filter, variables have to be transformed appropriately. The simplest solution is the standard technique of state augmentation – constructing an *augmented* state vector to include the state vector on all prior intervals up to $s = max(p', q')$.[6] Similar augmentation of the transition and assignment matrices yields after some algebra,

$$\mathcal{Y}_h = \mathbf{A}_h \mathcal{X}_h + \mathcal{B}_h + \mathbf{v}_h \qquad (7)$$

[5]See for example Cascetta et al.[4].
[6]The same technique was used by Okutani[11].

where

$\mathcal{Y}_h = \mathbf{y}_h - \mathbf{y}_h^H$
$\mathcal{X}_h = \mathbf{X}_h - \mathbf{X}_h^H$
$\mathcal{B}_h = \mathbf{A}_h \mathbf{X}_h^H - \mathbf{y}_h^H$ and

$$\mathcal{X}_{h+1} = \mathbf{\Phi}_h \mathcal{X}_h + \mathbf{W}_h \tag{8}$$

In the above equations, $\mathbf{X}_h$ and $\mathbf{X}_h^H$ are $(n_{OD}(s+1) * 1)$ augmented vectors; $\mathbf{A}_h$ is an augmented $(n_l * n_{OD}(s+1))$ matrix; $\mathbf{\Phi}_h$ is an augmented and appropriately modified $(n_{OD}(s+1) * n_{OD}(s+1))$ matrix and the error $\mathbf{W}_h$ is an $(n_{OD}(s+1) * 1)$ vector with zeros in the bottom $(n_{OD}s)$ cells and an $(n_{OD}(s+1) * n_{OD}(s+1))$ variance matrix $\mathcal{Q}_h$.

2.7 Filter Estimates

Assume that the initial state of the system $\mathcal{X}_0$ has mean $\bar{\mathcal{X}}_0$ and variance $\mathbf{P}_0$. Note that knowledge of $\mathcal{X}_0$ implies knowledge of $\mathbf{x}_0$, $\mathbf{x}_{-1}$, $\mathbf{x}_{-2}$, ... ,$\mathbf{x}_{-s}$ and the corresponding historical estimates. Then, using the assumptions made about the errors in sections 2.4 and 2.5, the following results can be stated.[7]

$$\mathbf{\Sigma}_{0|0} = \mathbf{P}_0$$

$$\mathbf{\Sigma}_{h|h-1} = \mathbf{\Phi}_{h-1} \mathbf{\Sigma}_{h-1|h-1} \mathbf{\Phi}'_{h-1} + \mathcal{Q}_{h-1}$$

$$\mathbf{K}_h = \mathbf{\Sigma}_{h|h-1} \mathbf{A}'_h (\mathbf{A}_h \mathbf{\Sigma}_{h|h-1} \mathbf{A}'_h + \mathbf{R}_h)^{-1}$$

$$\mathbf{\Sigma}_{h|h} = \mathbf{\Sigma}_{h|h-1} - \mathbf{K}_h \mathbf{A}_h \mathbf{\Sigma}_{h|h-1}$$

$$\hat{\mathcal{X}}_{0|0} = \bar{\mathcal{X}}_0$$

$$\hat{\mathcal{X}}_{h|h-1} = \mathbf{\Phi}_{h-1} \hat{\mathcal{X}}_{h-1|h-1}$$

$$\hat{\mathcal{X}}_{h|h} = \hat{\mathcal{X}}_{h|h-1} + \mathbf{K}_h (\mathcal{Y}_h - \mathbf{A}_h \hat{\mathcal{X}}_{h|h-1} - \mathcal{B}_{h-1})$$

$$h = 1, 2, ...$$

It can be proved[8] that the filter – as given by the above equations – produces the smallest Mean Square Error (MSE) covariance matrix for the state vector among a class of linear estimators. Under additional conditions of normalcy of the errors, the filter produces the smallest MSE estimate among a class of all estimators whether linear or non-linear. Moreover, the estimate is unbiased and orthogonal to its error.

To extend the model to k-step prediction, all that is required is to multiply the filtered vector by the $\mathbf{\Phi}$ matrix k times. What is obtained using the above formulae are the deviations; to obtain the O-D flows themselves, appropriate historical estimates have to be added. Moreover, one notes from the special structure of the state vector

[7] These equations comprise the solution to the standard linear discrete Kalman filtering problem. For a rigorous derivation of these equations, the reader is referred to Gelb[8].

[8] See for example Gelb[8].

that each flow is filtered $s+1$ times. In other words, all the vectors from $\hat{\mathcal{X}}_{h|h}$ to $\hat{\mathcal{X}}_{h+s|h+s}$ would contain some estimate of the deviation in the flow $\mathbf{x}_h$. By adding to these the vector $\hat{\mathbf{x}}_h^H$, different estimates of the actual O-D flow value $\mathbf{x}_h$ can be retrieved. Since the last estimate makes use of the most information, it is likely to be the most efficient.

Finally, a discussion of "observability"[9] which is a desirable property of such dynamic systems is in order here. Essentially, observability defines our ability to determine the initial state vector $\mathcal{X}_0$ uniquely from a set of measurements. Under conditions of non-observability, the effects of the initial estimates do not disappear with time and therefore it is critical to obtain accurate initial values. An analogous situation may be found in conventional (static or dynamic) matrix estimation where one typically starts with an apriori matrix and uses traffic counts to modify the apriori estimates. The apriori matrix is necessitated by the fact that the number of observations (traffic counts) is typically much less than the number of O-D pairs. The apriori matrix in effect increases the number of observations. However, the estimates obtained would always depend on the apriori information provided. In the model proposed here, the initial state vector $\mathcal{X}_0$ and the transition equation provide information similar to that provided by the apriori matrix in conventional estimation by exploiting temporal interdependencies among the O-D flows. The nice feature however of real-time updating is that under conditions of observability, the influence of the initial value of the state vector would disappear with time.

The most obvious and critical factor affecting observability in our problem is the ratio (n_l/n_{OD}). For a given number of O-D flows, measuring more counts increases the chances of observability being satisfied. The degree of linkage between O-D flows and counts is another factor. An extreme example of this arises when an entire column of the assignment matrix is zero implying that a particular O-D flow never gets measured.[10] A third consideration is the degree of linkage between O-D flows over time. Again, an extreme example of this arises when the transition matrix is fully populated by zeros.

In the context of observability, results from empirical studies (Section 3) conducted on the proposed model are encouraging. It was observed that for different values of initial starting conditions, the model produced identical filtered estimates. Another positive indication while testing the model was that it invariably succeeded in finding a gain matrix that moved the predicted estimates closer to the true values.

2.8 The Overall Framework

Figure 2 shows the overall framework for the dynamic O-D estimation problem. The framework consists of two major subcomponents. The lower box – which is of greater interest to us – is the real-time component while the upper box refers to the setting up of the historical database.

The lower portion of Figure 2 has been discussed in Section 2.2. As mentioned

[9] For a rigorous definition of observability, see for example Gelb[8].

[10] This might however be overcome if that O-D flow is related to other measurable O-D flows by means of a non-diagonal transition matrix.

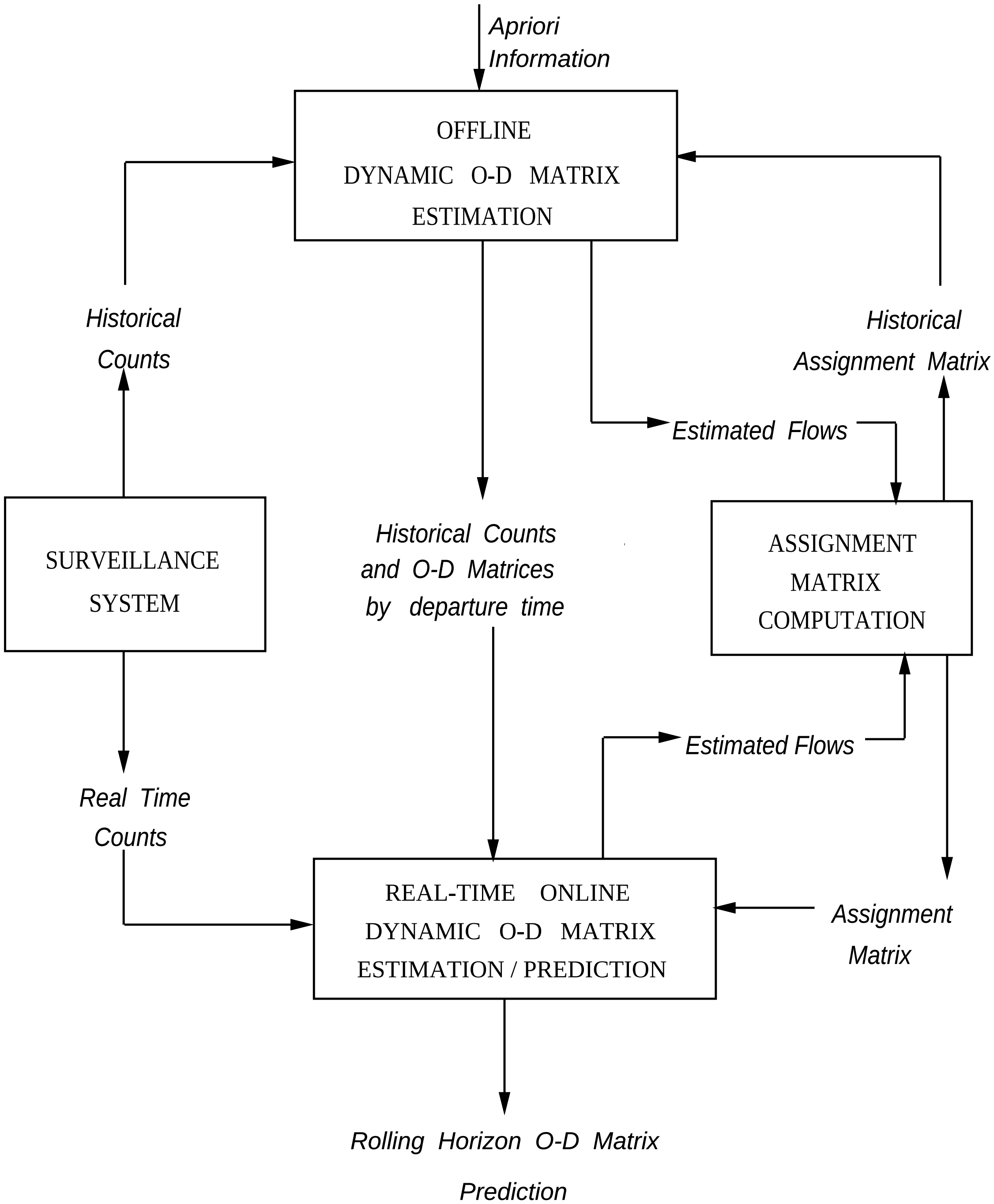


Figure 2: Overall Framework

earlier, this module interacts in real-time with the surveillance system, the historical database and the network analysis module which computes the assignment matrix. At the end of each updating interval, it accepts link counts from the surveillance system and retrieves from the historical database the historical O-D matrices and counts for that interval. There is however one difference between Figures 1 and 2. The lower part of Figure 2 has an arrow pointing from the O-D estimation module to the assignment matrix computation module. The purpose of this arrow is to allow for the possibility of feedback. This condition would arise because the assignment matrix is based on the (best) previously available O-D flow estimates.

To understand this possible feedback between the O-D estimation module and the network analysis module, assume that travel times in the network are not observable and hence have to be obtained by simulation. Consider a time instant t which corresponds to the end of the departure time-interval h and the beginning of departure time-interval $h + 1$. At this point in time, the O-D module gets a set of link counts for the departure interval h from the surveillance system. Further it gets a set of matrices $\mathbf{a}_h^h$, $\mathbf{a}_h^{h-1}$, $\mathbf{a}_h^{h-2}$, ... , $\mathbf{a}_h^{h-p'}$. Using these counts and assignment matrices and by calculating the Kalman gain matrix $\mathbf{K}_h$, the O-D estimation module computes filtered O-D flows (apart from smoothed prior estimates) for interval h. These filtered O-D flows may be quite different from the (one-step predicted) flows that were used to determine the assignment matrices in the first place. Hence, the newly filtered flow of interval h (and possibly also the newly smoothed flows of prior intervals) is reloaded again on to the network analysis module to get revised estimates of the assignment matrices. This cycle of filtering/smoothing–computation of assignment matrices–filtering/smoothing continues till convergence is reached. Once convergence is attained, a one-step prediction is performed to generate estimates of O-D flows for interval $h+1$. This is now passed to the assignment matrix computation module along with the filtered/smoothed flows of intervals h, $h-1$, ... , $h+1-p'$ to get preliminary estimates of assignment matrices for interval $h + 1$ and the process continues.

It is again emphasized that the above iterative process would only apply if the travel times are not measured directly or are liable to suffer from substantial imprecision. Otherwise, the simple analysis of section 2.2 holds satisfactorily. Further, one notes that the iterative approach to determining the assignment matrix implies an approximate solution to a non-linear Kalman Filter – the non-linearity arising from the dependence of the assignment matrix on the state-vector. While convergence cannot be guaranteed, empirical study (Section 3) indicates that the filtering procedure is fairly robust with respect to quality of assignment matrices; hence one would expect the quality of the estimate to get better with each iteration.

As mentioned earlier, the box in the upper part of Figure 2 refers to the setting up of the historical database of O-D flows. These matrices would be estimated using multiple days of data (for example could be averaged by day of week) and would be updated daily or weekly. A database of historical counts would also be obtained and stored. This component requires as input historical counts from a surveillance system and the assignment matrices corresponding to each interval. The estimation and updating of the historical O-D matrices could be performed using, for example, one of the least squares based methods described in Cascetta et al.[4]. The O-D flow matrices are obtained by minimizing a two part error function reflecting errors in apriori values and in the counts obtained by loading the apriori values onto the network. Again, computation of the assignment matrices could be an iterative process. In that case, one would start with an initial estimate of the O-D flows and compute the assignment matrices. This would be followed by recomputation of O-D flows and the process would continue till convergence.

3 Case Study

3.1 Data Description

Data on traffic movements was available for 3 days for the the Massachusetts turnpike, a toll freeway stretching from the New York State Border to Weston (Route 128) – a distance of about 120 miles. There are 15 entry/exit ramps; hence, the network has 210 possible entry/exit or O-D pairs. However, in the analysis, only East Bound traffic was considered; hence the number of O-D pairs was reduced by half. Information was available on entry and exit ramps of each vehicle and entry[11] and exit times.

3.2 Preparing Filter Inputs

Since the data was in disaggregate form, it had to be aggregated to obtain time-dependent traffic counts for each of the 14 links in the network. This aggregation was carried out by assuming an average vehicle speed of 55 mph and using this speed to calculate the entry time of each vehicle on each link.[12] Hypothetical counting stations were assumed to be located at the entrance to each link on the network. The values of the assignment matrices and O-D flows were computed such that the relationship

$$y_{lh} = \sum_{p=h-p'}^{h} \sum_{r=1}^{n_{OD}} a_{lh}^{rp} x_{rp} \tag{9}$$

held exactly. Since the network was linear, no path choice model was required and the movement of each vehicle could be tracked using the constant speed assumption.

The data for the third day (of the three days for which data was available) was chosen for implementing the Kalman Filter. The historical database of O-D flows and counts was created from the data of the first day. Again, because complete information was available, true historical values could be computed. The matrices $\mathbf{f}_h^p$ of autoregressive coefficients were computed by simple Ordinary Least Squares regressions using the deviation of O-D flow values of the second day from those of the first. Thus, the calibration of the transition equation coefficient matrices used only historical information. Two assumptions were made in the process of obtaining these matrices. Firstly, it was assumed that the autoregressive structure remained constant over the whole day so that one could have enough observations in one day to

[11]Entry times were not considered accurate enough, hence they had to be back-calculated from exit times assuming an average speed. More on the speed assumption later.

[12]This assumption of a uniform average speed for all vehicles at all times is unrealistic. However, for the purposes of implementing and evaluating the proposed model, all that is needed is a set of "reasonable" O-D matrices, counts and assignment matrices consistent with each other. The issue of whether the speeds assumed for generating counts is realistic is not directly relevant.

estimate all the parameters. Secondly, it was assumed that a flow between O-D pair r for a period was related only to rth O-D flows of prior intervals. The covariance of the transition equation error was retrieved from these regressions as explained in earlier sections. An autoregressive process of order 4 (i.e. $q' = 4$) provided the best fit. Because of the exactness of the estimated quantities in equation (9), there was no measurement error in the problem.

3.3 Application of the filter

There were 105 unknown flows to be estimated in each 15-minute departure interval. However, due to the transformation of variables, there were also flows corresponding to s prior intervals to be estimated. In this case study, p' was set equal to 8 since the maximum time taken to traverse the network was about 120 minutes ($= 8 * 15$). Since q' was only 4, the value of s was 8. Hence the number of unknown O-D flows to be estimated during each interval equaled 945.

The obvious choices for the initial estimates for the state-vector and its initial covariance matrix were the corresponding historical values. Besides the filtered flows, one step, two step and three step predictions were also computed.

3.4 Results

Figures 3 and 4 show results for the estimation scenario presented for one O-D pair. The flows are in terms of the number of vehicles departing their origin during each 15 minute interval. Note that Figure 3 shows the *first* filtered flow – because of transformation of variables, each flow is filtered 9 (i.e. $s+1$) times. It is apparent from Figure 3 that the filtered estimates are significantly closer to the true values than the corresponding historical values. Also, it can be seen that the quality of the predictions deteriorates progressively and that the predicted estimates tend to converge to the historical values as the prediction time-step is increased. This is to be expected given the autoregressive formulation because for every one-step-ahead prediction, the deviation (in the prior interval) is effectively multiplied by a fraction. The deviations are themselves small given the limited variability in traffic flow over the three days. Multiplying them repeatedly by a fraction reduces them still further. Adding them to the historical values – which are much larger in magnitude in comparison – hence yields estimates that do not differ by much from the historical values.

In the above implementation, there is very little variability in the traffic demand over the three days of analysis and hence, the historical flows themselves provide a good approximation to the true values. To test the performance of the filter in replicating the true values with poor historical information, another implementation was carried out using very poor historical values.[13] Figure 5 shows results for one O-D pair. It is seen that the filtering procedure is fairly robust and the quality of the filtered estimates does not seem to be very sensitive to historical information. However, the predicted estimates – though significantly better than the historical values – leave room for improvement.

[13] These actually corresponded to historical values for the evening peak.

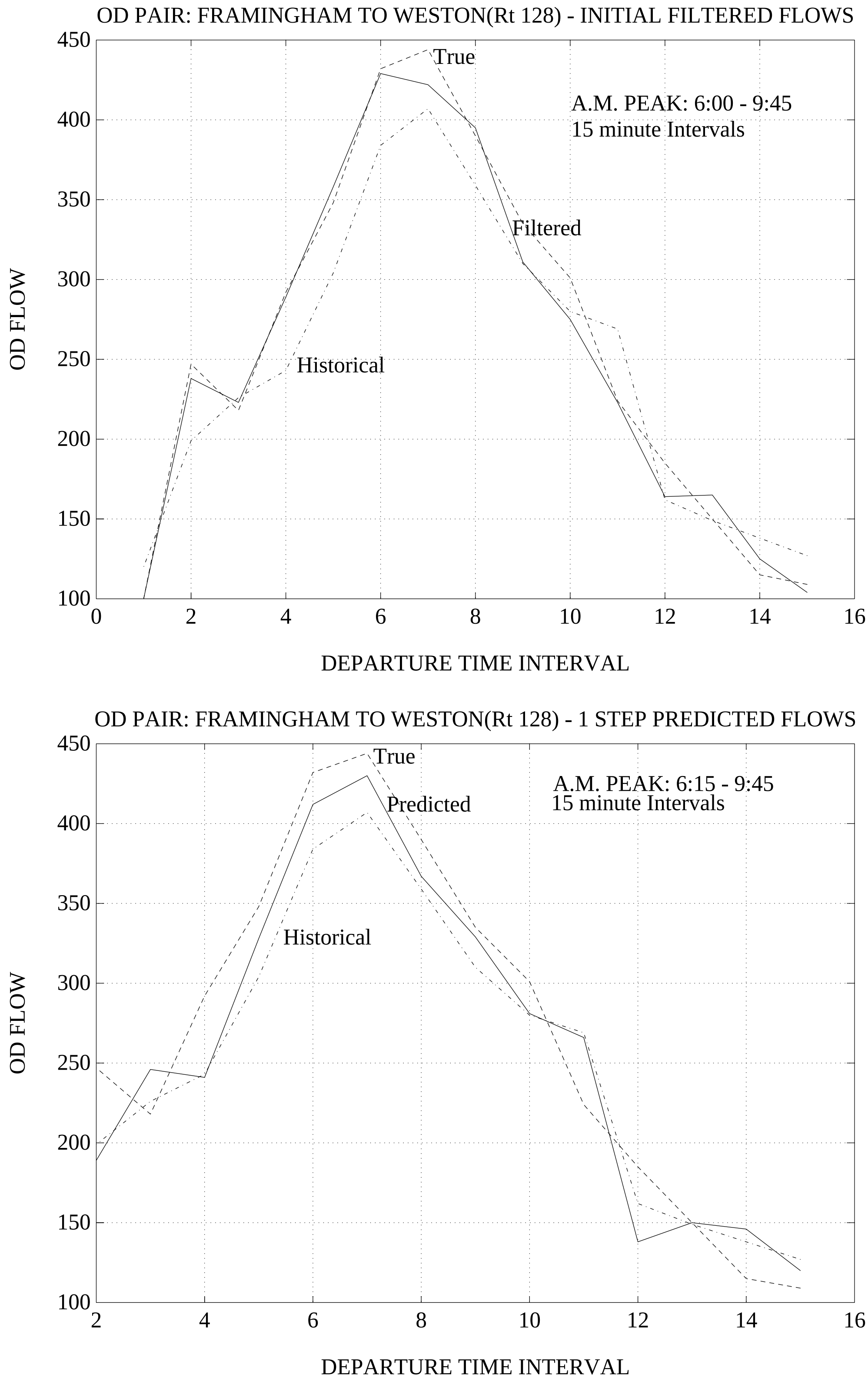
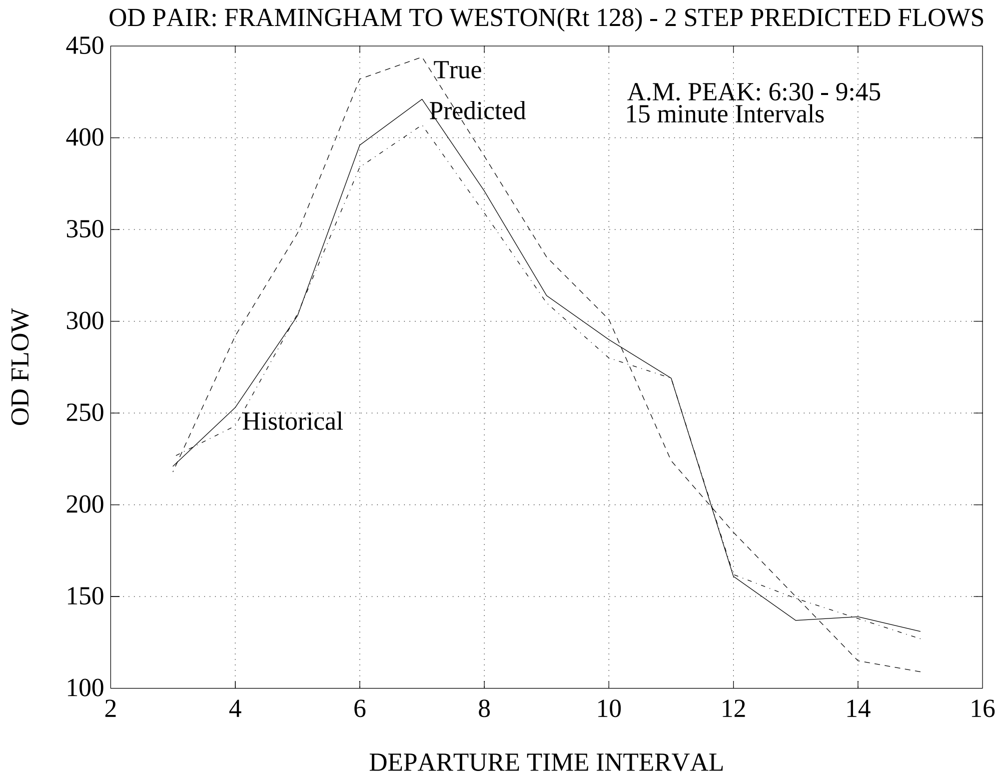


Figure 3: Typical Model Performance: Filtered and One-Step Predicted Flows



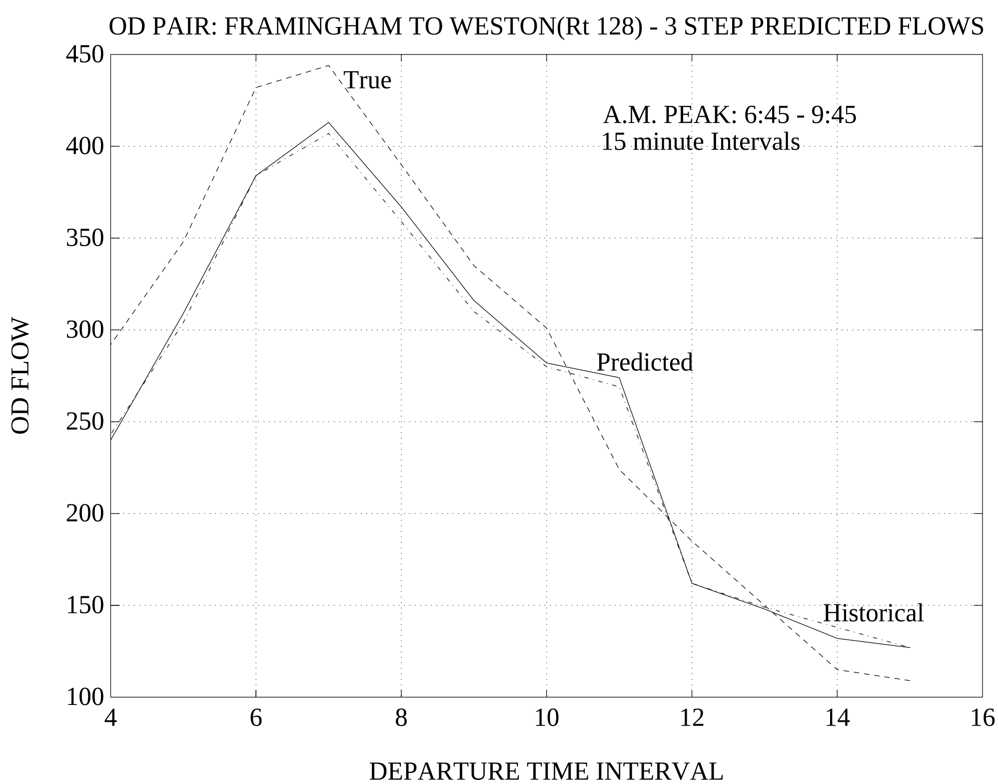


Figure 4: Two Step and Three Step Predictions

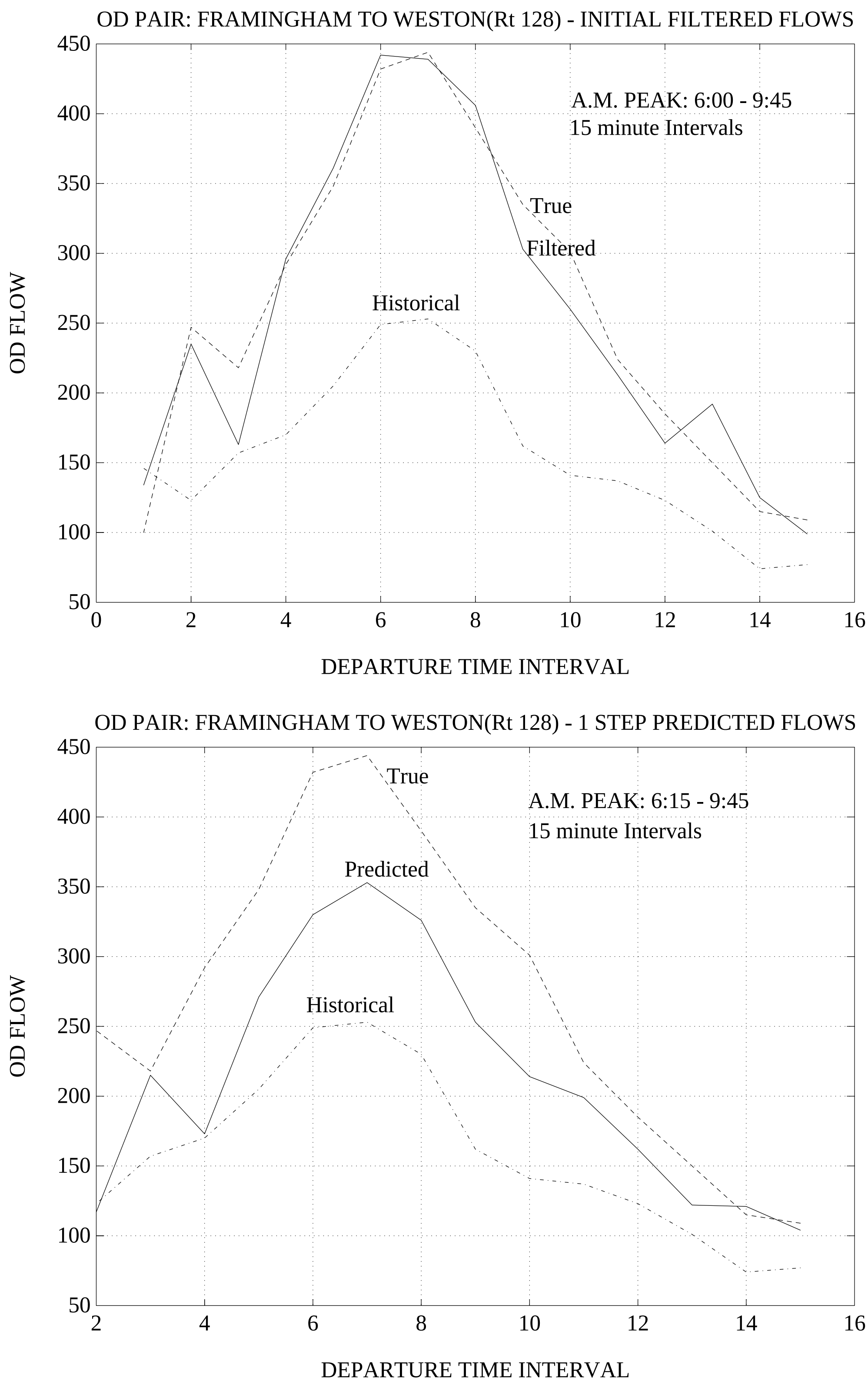


Figure 5: Filter Estimates and One-step Predictions with poor historical information

		MODEL	*HISTORICAL*
RMS Error	Filtered	5.3449	8.7015
	1-Step Predicted	9.1061	8.9501
	2-Step Predicted	8.7558	9.1146
	3-Step Predicted	9.3090	9.2490
RMSN Error	Filtered	0.2905	0.4729
	1-Step Predicted	0.4832	0.4750
	2-Step Predicted	0.4539	0.4725
	3-Step Predicted	0.4708	0.4678

Table 1: Root Mean Square and Normalized Root Mean Square Error Values

All the results presented thus far were for a specific O-D pair. The following two statistics were computed to get a better overall comparison of the historical estimates and those estimated by the model.

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 summation is over all O-D pairs and all intervals for which analysis was carried out.

Table 1 gives values of the RMS and RMSN errors. It is seen that the filtered estimates computed by the model have much lower errors. The 1-Step predictions are however on the whole slightly worse – the difference is not significant. 2-Step predictions made by the model are better than corresponding historical values while there is very little difference between 3-step predictions and historical values. This pattern may be attributed to the fact that there is very little variability in traffic patterns over the three days; hence the historical values provide good fit to the true values. It was however observed that the predictions perform better than their historical counterparts for O-D pairs with relatively high flows. Tables 2 and 3 show the values of RMS and RMSN errors with only the O-D pairs corresponding to flows more than 100 and 150 units taken into account.

The errors were also computed for the case with poor historical information. Table 4 shows that in this case, 1-Step and 2-Step predictions are significantly better than historical values but 3-Step predictions again do not show much difference from the historical values.

In this case study, further filtering of an O-D vector (beyond the first measurement) does not convey any additional information because the measurements are exact.

The most critical real-time input to the O-D estimation process is the assignment matrix. It is hence important that the sensitivity of the filter to imperfections in the assignment matrix be understood. In order to investigate this, the elements of the assignment matrix were perturbed using the formula:

		MODEL	*HISTORICAL*
Flows ≥ 100	Filtered	15.0397	38.6223
	1-Step Predicted	39.0104	39.4572
	2-Step Predicted	36.8199	40.1758
	3-Step Predicted	39.8751	40.4197
Flows ≥ 150	Filtered	15.1953	44.8590
	1-Step Predicted	44.4626	45.2173
	2-Step Predicted	42.1367	46.0628
	3-Step Predicted	43.8698	44.6862

Table 2: Root Mean Square Error Values for high demand pairs

		MODEL	*HISTORICAL*
Flows ≥ 100	Filtered	0.0585	0.1501
	1-Step Predicted	0.1547	0.1556
	2-Step Predicted	0.1453	0.1577
	3-Step Predicted	0.1598	0.1615
Flows ≥ 150	Filtered	0.0475	0.1403
	1-Step Predicted	0.1454	0.1469
	2-Step Predicted	0.1352	0.1485
	3-Step Predicted	0.1495	0.1515

Table 3: Root Mean Square Normalized Error Values for high demand pairs

$$a_{new} = a_{correct}[(1-\delta) + U * 2\delta]$$

for different values of δ. U is a random number drawn from a uniform distribution between zero and one. The performance of the filtered estimates for different values of δ is shown in Figure 6. It can be seen that for the range $\delta < 0.33$, the filtered estimates remain superior to the historical values indicating that the proposed filtering procedure is fairly robust with respect to inaccuracies in the assignment matrices[14].

3.5 Major Findings

The following are the major findings from this case study.

1. The filtered O-D flow values are substantially better (in the sense of being closer to the true values) compared to the corresponding historical values. Also, the filtering procedure seems to be fairly robust with respect to quality of historical information and the assignment matrices.

[14]Perturbation of the assignment matrix introduced measurement error in the problem – hitherto unconsidered. The variance-covariance matrix $\mathbf{R}$ was calibrated for each δ from historical data.

		MODEL	*HISTORICAL*
RMS Error	Filtered	14.6659	29.2744
	1-Step Predicted	24.3734	29.8121
	2-Step Predicted	28.2335	30.2335
	3-Step Predicted	30.4809	30.9269
RMSN Error	Filtered	0.7970	1.5909
	1-Step Predicted	1.2935	1.5821
	2-Step Predicted	1.4637	1.5674
	3-Step Predicted	1.5417	1.5642

Table 4: Error Values with poor historical information

2. The one-step and two-step predictions perform better than the corresponding historical values for O-D pairs with high O-D flows. If there is not much variability between the historical O-D flows and the actual flows during the period of analysis, the predictions are no better than the historical values. There is not much to be gained from three step predictions over the historical values. The predicted values tend to converge to the historical values with increasing prediction horizon.

4 Future Research

4.1 Application Issues

The computational costs of implementing the model seem to be a function of the following three parameters:

1. Number of O-D pairs.
2. Spatial distribution of the network.
3. Congestion level in the network.

The number of O-D pairs is an obvious parameter since it is directly related to the dimensionality of the unknown vector to be estimated. The spatial distribution of the network is important because the number of lagged intervals in the measurement equation depends upon the maximum travel time between any two points in the network. This in turn depends upon whether the network is "clustered" or "dispersed". For the same number of nodes and links therefore, one would expect the computational costs associated with a linear network (a freeway or arterial) to be much higher compared to an urban network of a few intersections.[15] The congestion level in the network is also related to this fact. High congestion levels in the network would increase the maximum travel times and hence increase the number of lags.

[15] Unless the latter experiences high congestion.

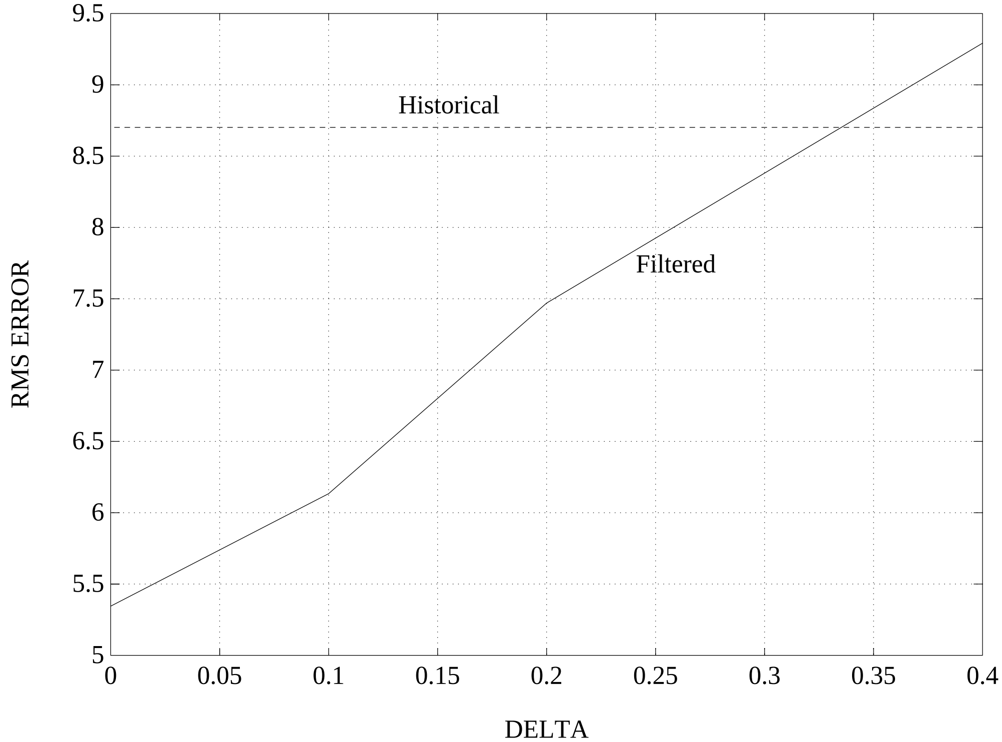


Figure 6: Filtering with imperfect assignment matrix

Another important issue is appropriate choice of estimation interval. Of primary relevance here is the time granularity required for the application that makes use of these matrices. Also, if the time intervals involved are very short, the predictability of the transition equation would be reduced since over very small intervals of time, fluctuations in traffic movements are essentially random. Very large estimation intervals on the other hand hold out the danger of masking the information contained in time-varying link counts. And finally, computational considerations are also important in choice of estimation interval because for small intervals, the number of lags to be considered (and hence the dimensionality of the state vector) would be high.

An appealing idea of reducing the computational costs associated with the model drastically would be to write the measurement equation as follows:

$$\mathbf{y}_h - \mathbf{y}_h^H = \sum_{p=h-p'}^{h-1} \mathbf{a}_h^p (\hat{\mathbf{x}}_p - \mathbf{x}_p^H) + \mathbf{a}_h^h (\mathbf{x}_h - \mathbf{x}_h^H) + \Sigma_{p=h-p'}^{h} \mathbf{a}_h^p \mathbf{x}_p^H - \mathbf{y}_h^H + \mathbf{v}_h \tag{10}$$

The above equation denotes an approximation to equation (6) since during each estimation interval, all O-D vectors corresponding to prior time intervals are held constant at their prior estimated values. Thus the O-D vector corresponding to each time interval would be filtered only once. Further empirical testing is required before drawing any conclusions.

The transition equation could be similarly rewritten for each estimation interval to consist of two distinct parts - one term representing a summation over all prior q' intervals and the other representing the interval of interest. It should be noted that as a result of this, the transition equation would also contain a constant term. Solutions to the filter problem with constants in both the measurement and transition equations are standard.[16]

4.2 Conclusion

In this paper, a new framework for O-D matrix updating and prediction in real-time has been presented. The principal innovation of the formulation is the definition of the state vector in terms of deviations of O-D flows from historical values. It is believed that this procedure has more theoretical justification. The performance of the model has been tested using data for the Massachusetts Turnpike. Initial results are encouraging and suggest that the filtering technique is quite robust. Further research should focus on the process of updating the historical database, on computational issues and on alternative formulations of the transition equation leading potentially to non-linear models.

References

[1] M. Ben-Akiva. Methods to combine different data sources and estimate Origin-Destination matrices. In N.H. Gartner and N.H.M. Wilson, editors, *International Symposium on Transportation and Traffic Theory*, pages 459–481. Elsevier Science Publishing Company Inc., 1987.

[2] M. Ben-Akiva, H.N. Koutsopoulos and A. Mukundan. A Dynamic Traffic Model System. Forthcoming in Proceedings of the Second International Capri Seminar on Urban Traffic Networks, 1992

[3] M. Ben-Akiva and S. Lerman. *Discrete Choice Analysis: Theory and Application to travel demand.* M.I.T. Press, 1985.

[4] E. Cascetta, D. Inaudi, and G. Marquis. Dynamic Estimators of Origin-Destination Matrices using Traffic Counts *Transportation Science*, 27(4):363–373, 1993.

[5] E. Cascetta and S. Nguyen. A unified framework for estimating or updating Origin/Destination matrices from traffic counts. *Transportation Research*, 22B:437–455, 1989.

[6] C.K. Chui and G. Chen. *Kalman Filtering with Real-Time Applications.* Springer-Verlag, 1991.

[16] See for example Chui[6].

[7] M. Cremer and H. Keller. A new class of Dynamic methods for the identification of Origin-Destination Flows. *Transportation Research*, 21B(2):117–132, 1987.

[8] A. Gelb, editor. *Applied Optimal Estimation.* M.I.T. Press, 1974.

[9] H. Keller and G. Ploss. Real-Time Identification of O-D Network Flows from Counts for Urban-Traffic Control. In N.H. Gartner and N.H.M. Wilson, editors, *International Symposium on Transportation and Traffic Theory*, pages 267–284. Elsevier Science Publishing Company Inc., 1987.

[10] N.L. Nihan and G.A. Davis. Recursive estimation of Origin-Destination matrices from input/output counts. *Transportation Research*, 21B(2):149–163, 1987.

[11] I. Okutani. The Kalman Filtering Approach in Some Transportation and Traffic Problems. In N.H. Gartner and N.H.M. Wilson, editors, *International Symposium on Transportation and Traffic Theory*, pages 397–416. Elsevier Science Publishing Company Inc., 1987.

Acknowledgements

Funding for this research was provided in part by Lincoln Laboratory, M.I.T. and the Intelligent Engineering Systems Laboratory. Part of the computations were performed on the MIT Cray XMP EA/464 super-computer. We are thankful to John Judge at the Massachusetts Turnpike Authority and Yong Chang, Tom Lisco and James Parker at the Central Transportation Planning Staff, Boston for helping us obtain turnpike data. We greatly benefited from comments by Michael Bell, Jim Culbert, Tony Hotz, and Megan Khoshyaran. Finally, we received helpful comments from Carlos Daganzo and two anonymous referees.