A COMPARISON OF METHODS FOR DYNAMIC ORIGIN-DESTINATION MATRIX ESTIMATION

Nanne van der Zijpp

Faculty of Civil Engineering, Delft University of Technology
P.O.Box 5048 2600 GA Delft NL
*Email zijpp@ct.tudelft.n*l

Abstract: An overview of existing dynamic OD-estimation methods is given, and a statistical technique to estimate parameters in dynamic travel demand models based on Bayesian inference is proposed. Using both synthesized and empirical data the new technique is compared to traditional ones such as the Kalman filter and least squares. This shows the new technique results in more accurate estimates of OD-split proportions. Surprisingly however, this does not result by default in more accurate predictions of link flow volumes.

Key-words: Dynamic OD-matrix estimation; Kalman filter; Bayesian Inference

1. INTRODUCTION

On-line prediction and control of freeway traffic usually requires estimates of time varying travel demand. Time varying travel demand is summarized in dynamic origin-destination (OD) matrices, and the estimation of these has been an active area of research over the last two decades. The estimation of OD-matrices from traffic counts may be considered as the reversal of another well-known problem, that of traffic assignment. In general, the unique reversal of traffic assignment is not possible, as many OD-matrices match a given set of traffic counts. However, there are various possibilities to define a measure of plausibility over the space of OD-matrices. By doing so, a 'best' dynamic OD-matrix is implied.

For dynamic matrices it is plausible to base such a measure on the similarity between the matrices corresponding to consecutive periods, expressing that traffic emerges as a result of slowly evolving travel demand. The methods that are central to the present paper are based on a distance measure that applies to the structure of the OD-matrix, as proposed in e.g. (Cremer and Keller, 1981; Nihan and Davis, 1987).

OD-split-matrix

This structure is expressed by the OD-split-matrix which is obtained from an OD-matrix by dividing each OD-matrix cell-value by its corresponding row-total. The resulting cell-values are referred to as split proportions and express for each network entry the fraction of traffic that is destined for each network exit. As a result split proportions are non-negative and smaller than unity. Moreover, the split proportions within one row of the OD-split matrix, i.e. those corresponding to one network entry, add up to unity.

The various approaches that have been presented over the years may be categorized depending on the traffic characteristics that are taken into account, the data-sources that are used and the estimation procedures that are employed.

With respect to traffic characteristics, the problem may include the estimation of route choice proportions, travel time (see e.g. Chang and Wu, 1994), and travel time dispersion (see e.g. Bell, 1991). With respect to traffic data, these may consist of time-series of traffic counts only (see Cremer and Keller, 1991, 1987; Nihan and Davis, 1987; VanderZijpp and Hamerslag, 1994), or may be supplemented with historical data or with observations from probe-vehicles or license-plate readers (see VanderZijpp, 1996, 1997). Thirdly, the estimation procedure that is employed may be based on neural networks, least squares, maximum likelihood, Kalman filter techniques, or Bayesian inference.

The present paper attempts to isolate the influence of the estimation procedure that is used, and presents a comparison between least squares, constrained least squares, the Kalman filter and a Bayesian updating algorithm that was presented in (VanderZijpp, 1996).

The latter Bayesian estimator was developed especially to deal with the non-negativity constraints that apply to the split proportions. Due to these constraints the probability distribution, of which the estimated split proportions should reflect the mean, have a typical asymmetric shape. Therefore, it is hypothesised that a significant improvement over the traditional methods can be obtained if instead of the point that maximizes the probability, the centre of the probability mass can be computed.

The Bayesian method is compared to other methods, such as least squares, constrained least squares and the Kalman filter, on the basis of both synthesized and empirical data. The latter data were extracted from the traffic monitoring system on the Amsterdam beltway, but unfortunately are not accompanied by a directly observed dynamic OD-matrix. The evaluation on the basis of empirical data gives rise to the interesting methodological issue of how to compare different methods in absence of a directly observed OD-table.

2. PROBLEM DEFINITION

The problem is to determine a dynamic OD matrix for a transport network. It is assumed that time series of observations of all entering volumes and a subset of the internal link volumes are available. The following symbols will be used, sizes of vectors and matrices are implicit in their definitions:

m, n, h	Number of entries, number of exits, and number of link volume observations (excluding entry volumes).
i, j, k	Indices corresponding to entries, exits, and link volume observations
$q(t)$	Vector of entry volume observations in period t, t=1,2,…
$f_{ij}(t)$	Number of trips for entry-exit (EE) pair i-j with departure period t, i=1,…m, j=1,…n.
$y(t)$	Vector of link volume observations in period t. Element k of this vector contains the number of trips with departure period t that traverse location k.
τ	Path-link observation incidence map. τ_{ijk}=1 if route i-j uses link k and zero otherwise, k=1,2,…h

3. CLASSIFICATION OF DYNAMIC OD-ESTIMATION METHODS

Methods described in literature for estimating dynamic OD-matrices can be arranged in a number of ways, depending on their underlying model assumptions, the estimation techniques used, data used and details of their implementation.

3.1 Partitioning of the time-axis.

Data are usually available in time series of observations. They may be aggregated in time, e.g. to one or five minutes. Often the time-span to which a data record refers must be inferred from a time-stamp that accompanies each record.

Variables used in models corresponding to these observations need not necessarily use the same partitioning of the time axis. For practical reasons one may wish to convert observed data to a new time coordinate system. Usually this is done by using a regular grid that applies to all locations in the network. However, there are some practical advantages attached to using a *moving time coordinate system* (MTCS), in which the partitioning of the time-axis depends on the location. In a MTCS the boundaries between periods are given by time-space trajectories (*Van Der Zijpp, 1996*). This means that vehicles largely travel within one time-zone, and consequently only contribute to observations assigned to that period.

3.2 Dealing with travel time

The fact that vehicles need some time to travel through a network can be dealt with in a number of ways. A first possibility is to ignore travel times. This limits the applicability of models to small networks, as the ratio travel time / period length must not exceed a certain number. A second possibility is to take travel times into account, but to assume that they are known, for example from a traffic monitoring system. A third, yet unexplored, possibility would be to consider the link travel times as an extra set of unknowns to the problem. In the present paper we use the second approach, i.e. we assume that travel times are known and do not attempt to estimate them.

3.3 Dealing with travel time dispersion

Further refinements are obtained by taking travel-time *dispersion* into account. This can be done either by assuming a certain distribution of travel times, see e.g. (*Bell*, *1991*b) or by introducing extra unknowns and (*Chang and Wu*, *1994*). Such approaches require the estimation of extra parameters and hence do not lead to a reduced error of estimation by default. We do not consider them in the present paper.

3.4 Dealing with route choice

Also the estimation of route choice proportions introduces extra unknown parameters and may be avoided, for example by using a route choice model based on (estimated) travel times. In the present paper we avoid this issue by confining ourselves to corridors rather than networks.

3.5 Data requirements

Yet another way of subdividing dynamic OD-estimation methods is based on the input data that are used. These may be prior OD-matrices, time series of traffic counts or trajectory counts based on probe vehicles. Again we confine ourselves to a simple case where only time series of traffic counts are used.

3.6 Model assumptions

As discussed in the introduction, the problem of estimating dynamic OD-matrices from time-series of traffic counts is under-specified and extra assumptions are needed to define a unique solution. Imposing a model of travel demand is one option, but requires a certain level of temporal and spatial aggregation. This conflicts with the requirement of dynamic output. The remaining options are tracking OD-cell values and tracking OD-split proportions. From a methodological point of view the latter option is preferred: Due to the variation of entry volumes consecutive observations represent independent linear combinations of split proportions. This is not the case when OD-cell values are considered, as was done e.g. by (Ashok and Ben-Akiva, 1993). As stated in the introduction, the present paper deals with estimating split-proportions. Summarising, the model that is used in the paper is defined by the following equations:

$$E[f_{ij}(t)] = q_i(t)\, b_{ij}(t) \qquad (1)$$

$$b_{ij}(t) = b_{ij}(t) + v_{ij}(t) \qquad (2)$$

Where $b(t)$ denotes the vector of split proportions and $v(t)$ denotes a vector of small (zero mean) increments to these proportions.

4. OVERVIEW OF ESTIMATION METHODS

4.1 Least squares (DLS)

The simplest way to estimate OD matrices is by using (discounted) least squares:

$$\hat{\boldsymbol{b}}(t) \equiv \underset{\boldsymbol{b}}{\text{argmin}}\; J(\boldsymbol{b}, t) \qquad (3)$$

where the target J is defined by:

$$J(b, t) = \sum_{k=1}^{t} \lambda^{t-k} \| y(k) - H(k)b \|^2 \qquad (4)$$

and the non-zero elements of measurement matrix H are given by:

$$H_{k,\,(i-1)n+j} = \tau_{ijk} \qquad \forall ijk \qquad (5)$$

The parameter λ, $0 \le \lambda \le 1$, determines the weight that is put on older observations. The advantage of the DLS method over other methods that are presented in the paper is that it also can be applied if not exit volumes are observed and its ease of implementation.

4.2 Discounted constrained least squares (DCLS)

The DCLS estimate for the split probabilities (*Nihan and Davis*, *1987*; *Cremer and Keller, 1987*) is given by:

$$\hat{\boldsymbol{b}}(t) \equiv \underset{\boldsymbol{b}}{\text{argmin}}\; J(\boldsymbol{b}, t) \qquad \mathbf{0} \le \boldsymbol{b} \le \mathbf{1} \qquad \pi'\boldsymbol{b} = \mathbf{1} \qquad (6)$$

where the non-zero elements of matrix π are given by:

$$\pi_{i,\,(i-1)n+j} = 1 \qquad \forall ij \qquad (7)$$

The advantage of the DCLS method is its superior performance relative to the DLS method.

4.3 Kalman filter (KF)

The Kalman filter is a recursive procedure that under certain assumptions produces unbiased minimum variance estimates. Although in general these assumptions will not be completely satisfied, the procedure has many advantages such as its flexibility and its ease of implementation.

To use the Kalman filter two additional matrices need to be specified: a first one (S_t) corresponding to the covariance matrix of the component $v(t)$ in the random walk equation $b(t+1) = b(t) + v(t)$, and a second one (R_t) corresponding to the error term $w(t)$ in the measurement equation $y(t) = H(t)\, b(t) + w(t)$. Guidelines how to choose S_t and R_t are given in (Vander Zijpp,1996). The KF update equations are given by:

$$\begin{aligned} \hat{\boldsymbol{b}}(t) &= \hat{\boldsymbol{b}}(t\text{-}1) + K_t[y(t)\text{-}H(t)\hat{\boldsymbol{b}}(t\text{-}1)] \\ K_t &= \Sigma_b(t)H'(t)[H(t)\Sigma_b(t)H'(t)+R_t]^{-1} \\ \Sigma_b(t+1) &= \Sigma_b(t)+S_t-\Sigma_b(t)H'(t)\ldots \\ & [H(t)\Sigma_b(t)H'(t)+R_t]^{-1}H(t)\Sigma_b(t) \end{aligned} \qquad (8)$$

4.4 Bayesian updating (BU)

A traditional interpretation of the Kalman filter is the Bayesian one. Here it is assumed that a multivariate normal (MVN) prior distribution of $b(t)$ is given and is updated with the information contained in the MVN distributed measurements $y(t)$. In this interpretation the vector $\hat{\boldsymbol{b}}(t)$ and matrix $\Sigma_b(t)$ are the mean and covariance of the MVN distribution of $b(t)$.

However, another Bayesian interpretation is also possible. Assume that the prior distribution is not MVN but *truncated* MVN, i.e. has a shape identical to MVN, but is confined to the interval $b(t) \in [0,1]$, see Fig. 1. For this case it can also be shown that performing a Bayesian update with an MVN distributed observation $y(t)$ results in a TMVN posterior distribution (see VanderZijpp and Hamerslag, 1994). Moreover, the resulting distribution is still characterized by the parameters given by the recursion (8).

In this interpretation the vector $\hat{\boldsymbol{b}}(t)$ is no longer the mean and therefore not the best possible point estimate for $b(t)$. In fact there is no tractable expression to compute the mean as this would require evaluating a high dimensional integral. Instead a practical approach is to compute the average of a large set of numbers sampled from the truncated MVN distribution.

5. EXPERIMENTS

To give an impression of the performance of the methods described in this paper a number of experiments were done. A first series of experiments was done using synthesized data. In a second series of experiments empirical data from the Amsterdam beltway were used.

5.1 Experiments with synthesized data

Test-data. In a first series of experiments test-data

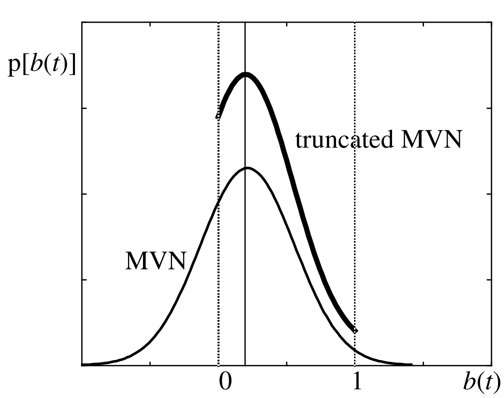

Fig. 1. A normal distribution (below) and its corresponding truncated distribution.

are generated according to six preset specifications. The test-data comprise entry-exit (EE) flows as well as synthesized traffic counts. The generation of the test-data involves the following steps:

1. *Generation of the split probabilities*. The split probabilities are generated by the random walk model $b(t+1) = b(t) + v(t)$ and initialized with a randomly generated vector $b(0)$. The following specification for the covariance matrix of $v(t)$ (S_t) is used:

$$S_t = \sigma_{\boldsymbol{b}}^2 I \tag{9}$$

2. *Generation of the entry flow volumes*. The entry volumes are sampled from a normal distribution of which mean and variance equal a parameter $\bar{\boldsymbol{q}}$.
3. *Generation and assignment of the EE-flows*. The EE-flows are sampled from the multinomial distribution, using the probabilities generated under step -1-.
4. *Generation of the entry- and link volume observation errors*. To the generated entry-flows and link-flows an zero mean error is added with covariance matrices given respectively:

$$\Phi = \sigma_{\boldsymbol{q}}^2 I \tag{10}$$

$$\Theta = \sigma_{\boldsymbol{y}}^2 I \tag{11}$$

In order to reduce random effects, for every set of parameters described in table 1, ten independent datasets are generated. Each estimation method will be applied to all of these sets, after which the errors of estimation will be averaged. Each set consist of 48 periods and was generated bearing in mind a period length of ten minutes. Network 1 represents the default configuration. Each of the specifications 2-6 differs only in one parameter from this configuration.

Table 1: Network specifications
-values of parameters-

	Network					
	1	2	3	4	5	6
m	4	4	4	4	4	*6*
n	4	4	4	4	4	*6*
$\sigma_{\boldsymbol{b}}^2$	10^{-4}	*10^{-2}*	*0*	10^{-4}	10^{-4}	10^{-4}
$\bar{\boldsymbol{q}}$	100	100	100	100	100	100
$\sigma_{\boldsymbol{q}}^2$	100	100	100	*10*	100	100
$\sigma_{\boldsymbol{y}}^2$	100	100	100	100	*10*	100

Estimation methods. The methods that are compared are divided in the following categories:

1. The Least squares method (LS) see equation (3). After some experimenting, the discounting param-

eter λ was set to 1- σ_b^2.

2. The discounted constrained least squares method (DCLS), see equation (6).
3. The Kalman filter(KF) method, see equation (8).In accordance with the way the test-data are generated the matrix S_t is set to $\sigma_b^2 I$. The observation error covariance matrix R_t is set to a diagonal matrix with the average observed values on its diagonal:
4. $R_t = \mathrm{diag}(\bar{y})$
5. The Bayesian Updating method (BU), see again equation (8), and section 4.4.

Evaluation criterion. The following measure is used as an evaluation criterion:

$$\mathrm{RMSE}(t)= \sqrt{\frac{1}{N}\sum_{i,j}\left(q_i(t)\,\bar{b}_{ij}(t)-f_{ij}(t)\right)^2} \qquad (12)$$

where N represent the number of connected EE-pairs. For each period and network specification in table 1 the measure is averaged over the number of datasets that are generated. The averages over the last 40 periods of the measure are summarized in table 2.

Results. The numbers in table 2 relate to the specifications in table 1. It is clear that the best performance is obtained from the Bayesian updating method followed by the Kalman method, discounted least squares method and finally the least squares method.

For the first network specification the error of estimation is plotted as a function of the number of periods in figure 2. The main difference between the performance of the methods is the number of periods needed to reach a certain level of accuracy. In this respect the advantages of the BU methods are evident. This also makes the Bayesian updating the preferred method under less favourable circumstances, where the network may involve a larger number of EE-pairs, the measurements contain larger errors, or the rate of change in the split probabilities might be higher.

Table 2: EE-Flow Errors
- average over 10 computations (trips/period)-

	network					
	1	2	3	4	5	6
LS	21.49	30.09	21.98	22.28	16.98	25.59
FCLS	19.34	25.22	19.88	18.86	16.25	20.96
KF	17.97	20.07	18.85	16.71	14.74	18.70
BU	14.28	14.70	15.11	12.58	13.00	14.09

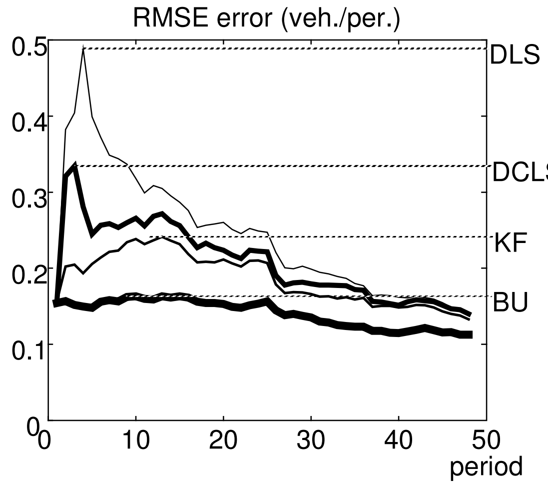

Fig. 2. Error of estimation for respectively the DLS,DCLS, KF and BU methods, when applied to network specification 1 (averaged over 10 simulation runs).

5.1 Experiments with empirical data

A second series of experiments has been done using empirical data collected at the Amsterdam beltway. Nine days are selected which are free from major incidents and disruptions in the data collection system. The selected network covers 11 kilometres of the anti clockwise direction of the Beltway (see Fig. 3). This corridor contains 5 entry ramps and 5 exit ramps (see the emphasised links). Average travel times have been estimated in two independent ways (using speed observations and flow observations) and have been averaged. Using a mean travel speed, all flow observations are converted to an MTCS (see section 3.1), using intervals of 5 and 10 minutes. As an evaluation

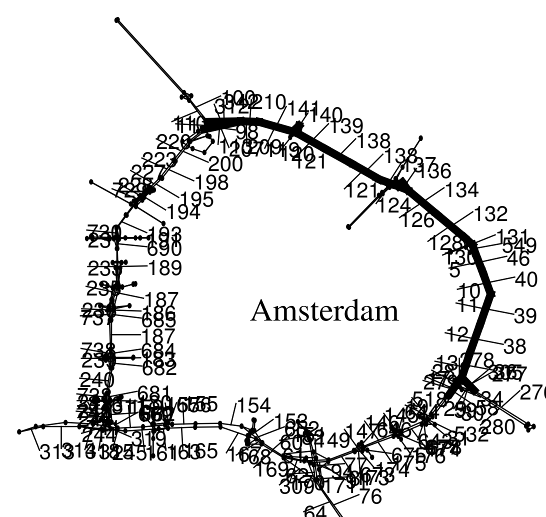

Fig. 3. Representation of the Amsterdam network. The numbers represent the induction loop locations. The emphasized links represent the network that was selected for further tests.

criterion the root mean squared error of the link flow predictions is used, where link flows are predicted based on split proportions that were estimated one period before.

Results. The results are shown in table 3. Surprisingly the accurate estimates of split proportions obtained with the BU method could not be translated into accurate link flow predictions. Detailed analysis has revealed two explanations for this. Firstly the computation of point estimates for the BU method as described in section 4.4 could not always be completed successfully due to low probabilities of sampling values within the interval [0,1]. In such cases the algorithm switches to a suboptimal method in which each split proportion is considered separately and correlations are ignored. Although this does not effect the RMSE defined by (13) to a great extent it might cause a bias in the link flow prediction. A second explanation is that the error measure that is used is quite similar to the objective functions that are minimized by the DCLS and KF methods. A quick fix to these problems would be to rely on the point estimates produced by the KF method in cases where point estimates for the BU method can not be successfully computed. However, further research is needed on this subject.

Table 3: RMSE predicted link-flow volumes, averaged over 9 days (veh./period)

method	interval (min.)	
	5	10
DCLS	13.02	22.19
KF	11.82	19.92
BU	13.48	29.95

6. CONCLUSIONS

The estimation method that is used in dynamic OD estimation has a large effect on the error of estimation. Simulation results indicate that the most accurate estimates of split proportions are obtained by a new method based on Bayesian updating. However, further research is still needed to translate these estimates in accurate link flow predictions.

7. REFERENCES

Ashok, K. and Ben-Akiva, M.E. (1993) Dynamic Origin-Destination Matrix Estimation and Prediction for Real-Time Traffic Management Systems, *Proc. 12th Int. Symp. on Transportation and Traffic Theory*, Berkeley, C.F. Daganzo (Ed)

Bell, M.G.H. (1991) The Real Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion, *Transportation Research-B*, **Vol. 25-B**, pp. 115-125

Chang, G. and Wu, J. (1994) Recursive Estimation of Time-Varying Origin-Destination Flows from Traffic Counts in Freeway Corridors, *Transportation Research-B*, **Vol. 28B**, pp. 141-160

Cremer, M. and Keller, H. (1981) Dynamic Identification of Flows from Traffic Counts at Complex Intersections, **In:** *Proc. 8th Int. Symposium on Transportation and Traffic Theory*, University of Toronto Press, Toronto Canada

Cremer, M. and Keller, H. (1987) A New Class of Dynamic Methods for the Identification of Origin-Destination Flows, *Transportation Research-B*, Vol.**21B**, No.2, pp.117-132

Nihan, N.L. and Davis, G.A. (1987) Recursive Estimation of Origin-Destination Matrices from Input/Output Counts, *Transportation Research B*, Vol. **21B**, No.2, pp. 149-163

Van der Zijpp, N.J. and Hamerslag, R. (1994) An Improved Kalman Filtering Approach to Estimate Origin-Destination Matrices for Freeway Corridors, *Transportation Research Records*, No. 1443, pp. 54-64

Van der Zijpp, N.J. (1996) Dynamic OD-Matrix Estimation on Freeway Networks *PhD Thesis*, Delft University of Technology

Van der Zijpp, N.J. (1997) Dynamic OD-Matrix estimation from Traffic Counts and Automated Vehicle Identification Data, *submitted to the 1997 annual meeting of the Transportation Research Board (TRB)*, Washington, January 1997