ESTIMATION OF AN ORIGIN-DESTINATION TRIP MATRIX FROM LINK TRAFFIC COUNTS FOR LARGE NETWORKS

William Zhu
David A. Hensher
Prepared for the 7th World Conference on Transportation Research (WCTR) to be held in Sydney, 16 July - 20 July 1995
ESTIMATION OF AN ORIGIN-DESTINATION TRIP MATRIX FROM LINK TRAFFIC COUNTS FOR LARGE NETWORKS

Existing approaches to estimating origin-destination trip matrices (O-D) from traffic counts are often limited by the size of a network. These approaches concentrate on achieving a target O-D matrix obtained elsewhere (e.g., from previous data or a sample survey). With the inconsistency in flow sampling and bias target matrix estimation, incorrect results may be obtained. This paper recognises that specifying a target matrix in the objective function of some existing models may not resolve the problem of uniqueness for congested networks. As an alternative approach to overcoming the bias estimation often associated with the specification of a target trip matrix, the paper formulates a non-linear programming model which incorporates production and attraction information and inconsistency in the traffic counts, and a network model for extending sampled traffic counts to network population flows. Heuristics for solving the formulated models are presented, together with a case study.

William Zhu, ARC Research Fellow
David A. Hensher, Professor of Management

The authors acknowledge the financial support of the Australian Research Council (Grant no. A79131696). We thank the Roads and Traffic Authority of NSW for permission to use the Future Directions Study network.

Institute of Transport Studies
Graduate School of Business
The University of Sydney NSW 2006
Australia
ESTIMATION OF AN ORIGIN-DESTINATION TRIP MATRIX FROM LINK TRAFFIC COUNTS FOR LARGE NETWORKS

WILLIAM ZHU AND DAVID A. HENSHER

Institute of Transport Studies
The University of Sydney NSW 2006
Australia

Abstract

Existing estimation approaches for estimating origin-destination trip matrices (O-D) from traffic counts are limited by large networks. These approaches concentrate on achieving a target O-D matrix obtained elsewhere (eg from previous data or sample data survey). With the inconsistency in flow sampling and bias target matrix estimation, they may lead to incorrect results. This paper realises that specifying a target matrix in the objective function of some existing models may not resolve the problem of uniqueness for congested networks. As an alternative approach to overcome the bias estimation probably caused by specifying a target trip matrix, the paper formulates a non-linear programming model which incorporates production and attraction information and inconsistency in the traffic counts, and a network model for extending sampled traffic counts to network population flows. Heuristics for solving the formulated models are presented.
1. INTRODUCTION

Over the last two decades, research on the estimation of origin-destination trip matrices based on link traffic counts has been increasing. This can in part be attributed to the increasing feasibility and convenience of obtaining link traffic counts. With the installation of counting devices in many cities, traffic counts can be obtained less expensively and quicker than via the traditional O-D data survey technique. Although an O-D sample survey will be executed from time to time to obtain behavioural data on the units being surveyed (e.g., passengers and commodities), the sample sizes are often inadequate for revealing O-D trip profiles. Traffic counts are a less expensive and more easily updated data source. Trip matrix estimation from link traffic counts is evolving as a promising approach to trip matrix estimation.

Approaches to the estimation of O-D trip matrices from link traffic counts on a network have been classified by a number of authors (e.g., Sherali et al., 1994) into three broad categories: statistical estimation methods, models based on maximum entropy/minimum-information theory, and network equilibrium based techniques. These models however can be generalised as two mathematical forms. The first is formulated as follows (Yang et al., 1992)

$$\min D(t, \tilde{t}) \quad \text{subject to} \quad M(t) = \tilde{v}$$

where $t$ is the O-D trip matrix to be estimated, represented in a column vector, $\tilde{t}$ is a column vector representing a target O-D trip matrix and $v$ is a column vector representing observed traffic volumes on a subset of links of a network.

$D(t, \tilde{t})$ is a function of the generalised distance between the estimated O-D trips $t$ and the target O-D trip matrix $\tilde{t}$. Typical forms of the function are entropy (Van Zuylen and Willumsen, 1980) and $L_p$ norm (Cascetta, 1984).

$M(t)$ is referred to as the assignment map (or operator) from the O-D trip matrix $t$ to the observed link flows $\tilde{v}$ (Cascetta and Nguyen, 1988). In general, the inverse of $M(t)$ may not be unique. Therefore, a target O-D trip matrix $\tilde{t}$ is proposed to minimise the generalised distance between the estimated and target O-D trip matrices.

The estimation of the O-D trip matrix in this type of model is constrained by the observed link counts, referred to herein as a counts-constrained (CC) model.

Earlier research has concentrated on uncongested networks, with a linear map of primary proposition (Turnquist and Gur, 1979; Van Zuylen and Willumsen, 1980; Bell, 1984; Cascetta, 1984; Brenninger-Gothe et al., 1989): $M(t) = Pt$, where $P$ is a proportion matrix whose element is a proportion of trips of an O-D pair using an observed link.
Due to the fact that the linear mapping is not appropriate for congested networks where the trip matrix is generally assigned to the network with user equilibrium, $M(t)$ is specified as a user equilibrium assignment operator consistent with Wardrop’s first principle (Wardrop, 1952). The user equilibrium assignment based approaches include Nguyen’s (1977) model in which the trip matrix is estimated to reproduce the observed O-D travel times, Fisk’s (1988) entropy maximisation model with user-equilibrium constraints, and distribution/assignment calibration model of Fisk and Boyce (1983). Further discussions and applications of these models can be found in Nguyen (1984), LeBlanc and Farhangian (1982), Sheffi (1985), Sheffi and Barnhart (1982).

No matter what the form of the assignment operator $M(t)$, the CC model has an inherent limitation for inconsistent traffic counts. In reality, inconsistency in traffic counts is not unusual because traffic counts cannot be guaranteed error free. Even though the traffic counts are collected error free, the actual flow may not exactly conform to a user equilibrium solution. A number of methods for eliminating and preprocessing inconsistencies have been proposed (Van Zuylen and Branston, 1982; Bell, 1983; Van Zuylen and Willumsen, 1984; and Carey and Revelli, 1986; Jornsten and Wallace, 1993). However, to date there has been no method which can efficiently and systematically resolve the inconsistency problem, especially for a congested network. It is arguable that even with consistent traffic counts, there might not exist a trip matrix that exactly reproduces the link flows because the methods for obtaining the assignment operator $M(t)$ are mostly heuristic based.

Many researchers have realised the inconsistency problem in the CC model and have proposed alternative models. These models are classified herein into the second mathematical form:

$$\min_{t,v} D(t, \overline{t}) + E(v, \overline{v}) \quad \text{subject to} \quad M(t) = v$$

(2)

where, $E(v, \overline{v})$ is a function of the generalised distance between estimated flow $v$ and observed traffic counts $\overline{v}$ on the set of observed links. Typical forms include generalised least squares (Bell, 1984), total expected penalty (Jornsten and Wallace, 1993) and $L_p$ norm.

This model is constrained by estimated link flows instead of observed link counts, and is referred to as a flows-constrained (FC) model. The FC model will produce a trip matrix identical to the one obtained by the CC model if and only if the estimated link flows $v$ in the FC model is such that $E(v, \overline{v}) = 0$.

With the FC model, the generalised least squares approach (Bell, 1984; Cascetta, 1984; Bell, 1991) and the stochastic programming approach (Jornsten and Wallace, 1993) are not applicable for congested networks because they use the proportional assignment of flows for the estimated trip matrix. Although the bilevel programming approach (Yang, 1992) decomposes $M(t)$ from the FC model into a lower level
decision making which evolves from solving a traffic assignment problem, it is computationally limiting for large networks.

Both CC and FC models specify a target O-D trip matrix for the estimated trip matrix. This specification limits the ability to apply statistical procedures to the application of trip estimation models. Specifying a target trip matrix will require extensive study of travel data which is not always readily available. It will be shown below that biased target information will lead to biased estimate. In certain circumstances this implies that specifying a target O-D trip matrix is a much more important task than the task of estimating an O-D trip matrix from traffic counts. Estimating an O-D trip matrix is precisely what we hope to produce (Barbour and Fricker, 1994).

Given the difficulty in specifying a target trip matrix, Barbour and Fricker (1994) proposed the SHAPE-2 algorithm, a heuristic method based on shortest augmenting paths. Sherali et al. (1994) proposed a preliminary linear programming model which can be modified to optionally incorporate the inconsistent traffic counts and uniqueness by adding the target matrix constraint. The model requires traffic counts on all links of a network and has computational limitations for large networks. No systematic method has been proposed, to our knowledge, for extending the sample traffic counts to entire network flows.

This paper proposes an alternative approach to the trip estimation problem using traffic counts by eliminating the specification of a target matrix. The objective of the paper is to identify the problem of inconsistencies in traffic counts, the possible biased estimation in the presence of a pre-specified target trip matrix, and the computational limitations associated with large networks. The paper is organised as follows. Section 2 introduces all notation and discusses some key factors in O-D trip estimation using traffic counts. Section 3 establishes the O-D trip estimation model and proposes an heuristic method for the model. Section 4 develops a systematic approach to extend sample traffic counts into entire network flows and discusses optimal sampling strategies. The remaining sections present a case study and the major conclusions.

2. NOTATION

Define a directed and connected network by \((N, \Omega)\), where \(N\) is the set of nodes and \(\Omega\) is the set of links. The set of links with observed traffic counts are denoted as \(A\) and the set of links without observed traffic counts are denoted as \(B\). A link from node \(i\) to node \(j\) is given as \((i, j)\). A path from node \(i\) to node \(j\) is defined as \(P_{ij}\).

Demand zones are classified according to geographic and socio-economic criteria, as an overlay for the network. Each zone contains a set of nodes of the network. These zones are mutually exclusive and in aggregate they form all the nodes of the network. Using set notation, if the urban area is divided into \(m\) zones \(Z_1, Z_2, \ldots, Z_m\) and each zone \(Z_i\) contains a subset of nodes of the network \((i=1,2,\ldots,m)\), \(Z_i\) and \(Z_j\) must be
mutually exclusive: $Z_i \cap Z_j = \emptyset (i \neq j, 1, 2, \ldots, m)$ and the total union of these zones contains all the nodes of the network: $\bigcup_{i=1}^{m} Z_i = N$.

A shortest free flow travel time (SFT) path between two nodes is defined as a shortest travel time path with free flows on each link of the path. Speed limits on each link differentiate the shortest distance path based on the link lengths of the path. Drivers prefer using the shortest time routes rather than the shortest distance routes. For a congested network, the total travel time passing the SFT path may not represent the actual shortest travel time path between the two nodes.

When the SFT path is saturated, i.e. at least one link’s traffic volume has reached its capacity, the paths with the total free flow travel time less than or equal to the saturated travel time are defined as feasible paths. The SFT path is one of the feasible paths. Travellers will still continue to travel on feasible paths even when the SFT path is saturated. Unless the SFT path is over-saturated, the routes travelled by trip makers according to the Wardrop’s first equilibrium principle (Wardrop, 1952) must be feasible routes. If we assume that only feasible paths are chosen by trip makers, we may drastically reduce the route choice set for the trip makers for a large network. Let us denote the set of feasible paths between nodes $i$ and $j$ as $K_{ij}$ and the set of infeasible paths as $\overline{K}_{ij}$. $K_{ij} \cup \overline{K}_{ij}$ forms all paths connecting nodes $i$ and $j$, denoted as $K$.

The O-D trip matrix estimation task using traffic counts is to estimate the number of trips from zone $Z_i$ to $Z_j (i, j = 1, 2, \ldots, m)$ to reproduce the traffic counts on links in $A$. Without additional information, we may not be able to derive a unique solution, even though the traffic counts are obtained in a user equilibrium pattern for the entire network. Consider the following simple network with only three nodes and three links. Each node also represents a zone.

![Figure 1. Example network with three nodes and three links](image-url)
With full traffic counts on the network, we can verify that the O-D trip matrix \( \{ t_{ij} \} \) satisfying the following equations will be able to reproduce the observed traffic counts.

\[
\begin{align*}
    t_{12} + t_{13} + t_{32} &= \bar{v}_{12} \\
    t_{23} + t_{21} + t_{13} &= \bar{v}_{23} \\
    t_{31} + t_{32} + t_{21} &= \bar{v}_{31}
\end{align*}
\]

(3)

The O-D trip matrices satisfying Equations (3) are not unique. To overcome the problem of uniqueness, most existing approaches specify a target O-D matrix in the objective function. However, specifying a target trip matrix does not solve the problem for the following reasons.

(a) The above example shows that with a target trip matrix, intra-zonal trips will most likely match exactly the specified target intra-zonal trips.

(b) Some inter-zonal trips will match exactly the specified target inter-zonal trips if traffic counts are observed only on a subset of network links. In the above example, if there was not a traffic count on link (3,1), the number of trips from node 3 to node 1 would exactly match the specified target.

(c) As discussed in the Appendix, with a general user-equilibrium assignment for congested networks, both counts and flows constrained models may not produce a unique O-D trip matrix even if a target matrix is specified.

The above point suggests that specifying a target trip matrix may lead to biased estimation of a correct trip matrix. Thus in practical applications more effort may be devoted to examining a target matrix rather than the trip estimation itself in order to avoid bias estimation. If justified, the significance of estimating a trip matrix from traffic counts might be neglected. In view of this, we formulate the estimation problem in a similar model to the FC model but with (i) omission of a target matrix and (ii) the addition of production and attraction constraints:

\[
\begin{align*}
\min_{t,v} & E(v, \bar{v}) \\
\text{subject to} & \\
M(t) &= v \\
\sum_{j \in D} t_{ij} &= o_i & i \in O \\
\sum_{i \in O} t_{ij} &= d_j & j \in D
\end{align*}
\]

(4)

where \( O \) is the set of origin zones and \( D \) the set of destination zones; \( o_i \) is the total number of trips from the origin zone \( i \), or trip production, and \( d_j \) is the total number of trips to the destination zone \( j \), or trip attraction. The production and attraction constraints are also called row and column constraints (Carey, 1981).

Institute of Transport Studies
Whether (4) is a better model system than the FC model needs confirmation. From an application perspective, it is easy to implement because trip production and attraction are easier to obtain. The estimation of trip production and attraction requires a major investigation on population and employment densities. A realistic application requires that an O-D trip matrix satisfy the trip production and attraction constraints.

Substituting the flow constraint \( v = M(t) \) into the objective function, we obtain the following transportation problem with a general objective function where traffic assignment is involved.

\[
\min_{t, v} E(M(t), \bar{v})
\]

subject to

\[
\sum_{j \in D} t_{ij} = o_i \quad i \in O \tag{5}
\]

\[
\sum_{i \in O} t_{ij} = d_j \quad j \in D
\]

With transportation constraints, the model will always have feasible solutions and thus eliminate the inconsistencies. For an uncongested network, \( M(t) \) can take the linear form \( M(t) = Pt \). If we assume the objective function is quadratic, then it can be written as \( E(M(t), \bar{v}) = (Pt - \bar{v})'Q(Pt - \bar{v}) \), where \( Q \) is a positive finite weighting matrix. \( Q \) can also be interpreted as a dispersion matrix (Maher, 1983) or the variance-covariance matrix (Brenninger and Jornsten, 1989) of the random error term \( Pt - \bar{v} \) where the objective is interpreted as generalised least squares regression model. For the quadratic objective function and transportation constraints, a number of algorithms can be applied to derive an optimal global solution (Carey 1981).

Although many researchers have proposed methods for solving the transportation problem (see Carey (1981) for a review), without exception they all assume the convexity of the objective function. When the general user equilibrium assignment is involved, the objective function of model (5) may not be convex. Furthermore, it is unknown whether \( M(t) \) is a continuously differentiable function because there is no explicit form for the user equilibrium assignment operator \( M(t) \). With this type of general objective function, the problem becomes very complex. Therefore, heuristics are utilised.

In searching for a heuristic method for the trip estimation model (5), the strategy is to combine a method which can solve the transportation problem and be a user equilibrium algorithm. The travelling costs from zones to zones for the transportation problem are the travelling times that are the weighted average of travel time using the current estimated traffic flow \( \nu \) and the travelling time used for the transportation in the last iteration. That is, if we denote \( \tilde{u}_{ij} \) as the travel time from zone i to zone j using the estimated traffic flows and \( u_{ij} \) as the travelling time used for the transportation in the last iteration, then the travelling time used for the current transportaion problem is updated as \( u_{ij} \leftarrow u_{ij} + \lambda(\tilde{u}_{ij} - u_{ij}) \), where \( 0 \leq \lambda \leq 1 \) is a constant such that the updated travelling time \( u_{ij} \) is positive for all zone pairs (i, j). The algorithm is described below.
STEP 0  For the observed traffic flows \( \tilde{v} \), determine the O-D travel time \( \tilde{u}_{ij} \) using a shortest path algorithm for each zone pair \((i, j)\). Use the determined O-D travel time as the initial travel time for the transportation problem, i.e. \( u_{ij} = \tilde{u}_{ij} \).

STEP 1  Estimate the O-D trips \( t_{ij} \) using a method for the transportation problem using the travel time \( u_{ij} \).

STEP 2  For the estimated O-D trips \( \{ t_{ij} \} \), calculate traffic flow \( v_{\xi\eta} \) for each link \((\xi, \eta) \in A\) and zonal travel time \( \tilde{u}_{ij} \) for each zone pair \((i, j)\) using a user equilibrium algorithm.

STEP 3  If one of the following stopping criteria is satisfied, then stop. Otherwise, update the generalised impedance \( u_{ij} = u_{ij} + \lambda (\tilde{u}_{ij} - u_{ij}) \) and go to STEP 1.

**Stopping Criteria**

1. Number of iterations exceeds a specified number.

2. The difference of values of the objective function \( E(v, \tilde{v}) \) between the current run and previous run is within a specified number.

3. The maximum difference of O-D trips between the current run and previous run is within a specified number.

**Remark 1.** The proposed heuristic can efficiently solve for large networks because no additional variables have to be introduced. In the execution of the above method, the travel time \( u_{ij} \) used for the transportation problem is constantly updated during each iteration. It is updated from the previous iteration in the way that if the estimated flow \( v_{\xi\eta} \) on the path from zone \( i \) to zone \( j \) is less than the observed flow \( \tilde{v}_{\xi\eta} \), more trips will be assigned between zone \( i \) and zone \( j \) with the reduced impedance \( u_{ij} \), it is increased if estimated flow \( v_{\xi\eta} \) is greater than the observed flow \( \tilde{v}_{\xi\eta} \).

**Remark 2.** In step 1 of the above method, it involves selecting a solution procedure for solving the transportation problem while in step 2 it involves using a method for traffic assignment. The traffic assignment method selected for the case study presented in Section 4 is Frank and Wolfe’s linear approximation method (Frank and Wolfe, 1956), implemented in the EMME/2 package (INRO, 1994). The method selected for solving the transportation problem is the double factor method (Kruithof, 1937) for the transportation model with an entropy objective function. The advantage of the double factor method is that it is efficient and the convergence is guaranteed (Gorman, 1963; Evan and Kirby, 1974; Robillard and Stewart, 1974). In applying the double factor method, we substitute the target trips with travel times as the balancing parameters:
Modified double factor method

**Initialisation:** \( \beta_j^0 = 1 \) for all \( j \), \( t_{ij}^0 = 0 \) for all \( (i, j) \), and set \( n=0 \).

1\textsuperscript{st} Perform row balancing

\[
\alpha_i^n = \frac{1}{\sum_j \beta_j^{n-1} d_j i_u_j} \quad \text{for all } i
\]

and column balancing

\[
\beta_j^n = \frac{1}{\sum_i \alpha_i^n o_j i_u_j} \quad \text{for all } j
\]

2\textsuperscript{nd} Estimate the intermediate O-D trips

\[
t_{ij}^n = \frac{\alpha_i^n \beta_j^n o_j d_i}{u_j} \quad \text{for all } (i, j)
\]

If the \( \max_{i,j} |t_{ij}^n - t_{ij}^{n-1}| < \varepsilon \quad (\varepsilon > 0) \), then stop. Otherwise increase \( n \) by 1 and repeat steps 1\textsuperscript{st} and 2\textsuperscript{nd}.

3. EXTENSION OF TRAFFIC COUNTS TO THE ENTIRE NETWORK FLOW

In reality, traffic counts are obtained for a subset of links of a network, given resource constraints. Direct application of the estimation models using the subset of traffic counts may lead to unexpected results due to the fact that the collected traffic counts cannot be guaranteed error free or in an equilibrium pattern. Preprocessing of the observed traffic counts reveals a necessary step. One of the methods to calibrate the traffic counts is to extend the sample traffic counts into the entire network in a user-equilibrium pattern. During the process of extension, validation and calibration can be undertaken consistent with user-equilibration. With the extended link flows, real traffic counts on unobserved links may also be collected to conduct validation and calibration activities.

The model for extending the observed traffic counts in \( A \) to the whole network traffic flows (ie. estimate the flows on links in \( B \)) is formulated as follows:
Minimize \[
\sum_{i,j \in N} \sum_{P_{ij}, P_{ij}' \in K_{ij}} D_P(P_{ij}, P_{ij}')
\]

subject to
\[
\sum_{j \in N} v_{ij} = \sum_{j \in N} v_{ji} \quad \text{for all } i \in N
\]
\[
0 \leq v_{ij} \leq C_{ij} \quad \text{for all } (i, j) \in \Omega
\] (6)

where,

\(P_{ij}\) and \(P_{ij}'\) are feasible paths from nodes \(i\) to \(j\) \((i, j \in N)\), i.e. \(P_{ij}, P_{ij}' \in K_{ij}\).

\(D_P(P_{ij}, P_{ij}')\) is the function of general distance between the travel times \(T(P_{ij})\) and \(T(P_{ij}')\) parsing paths \(P_{ij}\) and \(P_{ij}'\) respectively. For example, \(D_P(P_{ij}, P_{ij}')\) can be expressed as \(D_P(P_{ij}, P_{ij}') = (T(P_{ij}) - T(P_{ij}'))^2\).

\(v_{ij}\) is the traffic flow on link \((i, j)\) \(B\). \(v_{ij}\) is known as a decision variable in the model. Otherwise it is known as a constant in the model which represents an observed traffic count.

If a path \(P_{ij}\) from node \(i\) to node \(j\) in the network is represented by a set of links,
\[
P_{ij} = \{(i, n_1), (n_2, n_3), \ldots, (n_r, j)\}
\] (7)

with \(n_0 = i\) and \(n_{r+1} = j\), the travel time of traversing path \(P_{ij}\) can be represented by
\[
T(P_{ij}) = \sum_{k=0}^{r} T(v_{n_kn_{k+1}})
\] (8)

where \(v_{ij}\) is the traffic flow on link \((i, j)\).

\(T(v)\) is the volume delay function of the link traffic volume \(v\). There are a number of specifications in the literature for the volume delay function, a well known form is that introduced by the Bureau of Public Roads (BPR, 1964):
\[
T(v) = T_0(1 + \alpha(v/C)^\beta)
\] (9)

where \(T_0\) is the free-flow travel time, \(C\) is the capacity of the link, \(\alpha\) and \(\beta\) are parameters.

Davidson (1966, 1978) proposed a volume-delay function which is widely used in Australia:
where \( J \) is a delay parameter reflecting the road type (e.g., level of internal friction within the traffic stream) and \( \rho \) is a predetermined constant, usually in the range of \([0.85, 0.95]\) (Taylor 1992).

Akcelik (1991) proposed a statistically based time-independent volume delay function based on Davidson’s function:

\[
T(v) = T_0 \left( 1 + 0.25r \left( v - 1 \right) + \sqrt{(v - 1)^2 + \frac{8J}{r}v} \right)
\]

(11)

where, \( r \) is the ratio of flow period to free flow travel time and \( J \) is the delay parameter as defined in the Davidson function.

The flow conservation constraints of the model ensure that the flows are conserved at intermediate nodes of the network. If a node is connected to a centroid of a zone, then the flows from and to the zone should also be included in the conservation constraints. However, the centroids are not present in the conservation constraints. The results of the models will also contain the flows from and to zones and therefore production and attraction of each zone can be obtained. These estimated productions and attractions can be compared with the real productions and attractions to gain the confidence about the quality of the production and attraction data.

The model may not produce feasible solutions for inconsistent traffic counts. Especially when \( A = 0 \), the traffic counts collected for the entire network cannot be guaranteed to satisfy the conservation equations. Therefore, the model needs to be modified to accommodate the inconsistencies in traffic counts:

\[
\begin{align*}
\text{Minimize} & \quad \sum_{i,j \in N} \sum_{P_{ij}, P_{ij} \in K_{ij}} D_{ij}(P_{ij}, P_{ij}) + \gamma \sum_{i \in N} \delta_i^+ + \delta_i^- \\
\text{subject to} & \quad \sum_{j \in N} v_{ij} - \sum_{j \in N} v_{ji} + \delta_i^+ - \delta_i^- = 0 \quad \text{for all } i \in N \\
& \quad 0 \leq v_{ij} \leq C_{ij} \quad \text{for all } (i, j) \in \Omega \\
& \quad \delta_i^+, \delta_i^- \geq 0 \quad \text{for all } i \in N
\end{align*}
\]

(12)

where \( \delta_i^+ \) and \( \delta_i^- \) are non-negative artificial decision variables to permit positive or negative deviations in the conservation constraints, borrowing the idea from goal programming. The parameter \( \gamma \) in the objective function represents a scaling constant which shows the importance of the total error term \( \sum_{i \in N} \delta_i^+ + \delta_i^- \) in the objective.
function. The value of $\gamma$ should be in the range of $0 \leq \gamma \leq \sum_{(i,j) \in \Omega} T_{ij}(C_{ij})$, where $T_{ij}(C_{ij})$ is the volume delay function for link $(i, j)$ with traffic flow $v_{ij} = C_{ij}$, the capacity of link $(i, j)$ (i.e. the link is saturated).

The model will produce a feasible solution $v_{ij} = 0$ for all links $(i, j) \in \Omega$ when $A = \emptyset$. However, this may not be unique because flows may exist satisfying the conservation constraints which may also achieve the minimisation of the objective function. Figure 2 gives an illustration.

![Figure 2. Simple network showing the multiple solutions to the flow extension](image)

In the above network (Figure 2), we assume nodes 2, 3, 4, 5 are centroids. Any flow solutions that satisfy the conservation constraint at node 1 will achieve the minimisation of the objective function because there is no second feasible path from nodes to nodes. A unique solution can be obtained only if traffic counts are observed for three of the four links.

This begs the question: what is the minimum number of links in the observed link set $A$ for a general network so that the model will produce a unique solution, and what type of the structure should these links be?

The answer depends on the structure of the network. In the above example, we can conclude that there should be $n-2$ links for the star type network of $n$ nodes. With the line network as depicted in Figure 3, only one link is sufficient.
For other types of networks, further research is needed. In this paper, we recommend using a minimum spanning tree algorithm to choose $n-1$ links for sample traffic counts. The search of the minimum spanning tree is guided by the weights associated with each link. The weights may be associated with the cost of obtaining traffic counts or the importance of links in the entire network.

Given a minimum spanning tree with traffic counts, any unobserved link $(i, j)$ of the network will form a loop with links in the spanning tree. Thus an alternative path from node $i$ to node $j$ exists and the minimisation of the generalised distance between the travel times of paths can be carried out given that the link $(i, j)$ and the alternative path are feasible paths. In many cases, the conservation constraints will ensure the model produces a unique solution even if the alternative path is not feasible. A loop network with a large number of nodes is a good example (Figure 4).
Solving the flow extension model (9) involves enumerating feasible paths and deriving travel times for each feasible path. With a large network, it is almost impossible to use non-linear optimisation techniques. A sequential solution procedure is proposed which involves three steps:

We assume that the links with traffic counts are in the form of a minimum spanning tree. Then for each unobserved link, there must be an alternative path with observed counts connecting the end points of the link.

Denote $A'$ as the set of links with estimated link flows and $B'$ as the set of unestimated links. Initially set $A' = A$ and $B' = B$.

**STEP I**  
*Node Conservation.* Find a node $i \in N$ with only one link $(i, j)$ (or $(j, i)$) of $B'$ from (or to) the node and solve the conservation equation for the link at node $i$ with $\delta_{ij} = \delta_{ji} = 0$. Remove the link from $B'$: $B' = B' - \{(i, j)\}$ and $\Omega' = \Omega' - \{(i, j)\}$ (or $B' = B' - \{(j, i)\}$ and $\Omega' = \Omega' - \{(j, i)\}$). Repeat the process until no such node is found.

**STEP II**  
*Objective minimisation.*

1. Find a link $(i, j) \in B'$ with the minimum difference between the link length and the length of the shortest path $P \in K_{ij}$ from $i$ to $j$ in the network $(N, \Omega')$.

2. Solve the following optimisation problem with one variable $v_{ij}$ for link $(i, j)$:
Minimize \((T_y(v_{ij}) - T(P))^2\)

subject to
\[0 \leq v_{ij} \leq C_{ij}\]

where, \(T_y(v_{ij})\) is the volume delay function for link \((i, j)\) with capacity \(C_{ij}\).

(3) Remove \((i, j)\) from \(B'\): \(B' = B' - \{ (i, j) \}\) and \(\Omega' = \Omega' = \{ (i, j) \}\), and go back to (1) until \(B' = \emptyset\).

Because of the minimum spanning tree of the observed links, each unobserved link will eventually be assigned an estimated flow by the non-linear programming model in (2). However, the obtained flows may not satisfy the conservation constraint on each node. The next step is to balance the conservation for flows at each node.

**STEP III  Node conservation balance**

(1) Find a node with the maximum difference in out-flows and in-flows, e.g. to find a node \(i\) such that
\[
\left| \sum_{j \in N} v_{ij} - \sum_{j \in N} v_{ji} \right| = \max_{i' \in N} \left| \sum_{j \in N} v_{i'j} - \sum_{j \in N} v_{ji} \right|.
\]

(2) Two cases:

**CASE 1**

If \(\sum_{j \in N} v_{ij} > \sum_{j \in N} v_{ji}\), then we try to decrease \(v_{ij}\) and increase \(v_{ji}\) so as to minimise \(\sum_{j \in N} v_{ij} - \sum_{j \in N} v_{ji}\).

(i) Sort \(V_i = \{ (i, j) \}\) according to the descending order of \(\{ v_{ij} \}\) and sort \(U_i = \{ (j, i) \}\) according to the descending order of \(\{ C_{ji} - v_{ji} \}\). The resulting link lists are still denoted as \(V_i\) and \(U_i\) respectively.

(ii) If \(\sum_{j \in N} (C_{ij} - v_{ij}) < \sum_{j \in N} (C_{ji} - v_{ji})\), then for the link \((j, i)\) on the top of \(U_i\), calculate the flow increment for link \((j, i)\):
\[ \Delta_{ji} = \min\left\{ C_{ji} - v_{ji}, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'}, \min\left\{ 0, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'} \right\} \right\} \]

and set \( v_{ji} = v_{ji} + \Delta_{ji} \) and \( U_i = U_i - \{ (j, i) \} \).

Otherwise, for the link \((i, j)\) on the top of \(V_i\) calculate the flow decrement for link \((i, j)\):

\[ \delta_{ij} = \min\left\{ v_{ij}, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'}, \min\left\{ 0, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'} \right\} \right\} \]

and set \( v_{ij} = v_{ij} - \delta_{ij} \) and \( V_i = V_i - \{ (i, j) \} \).

(iii) Repeat (ii) until the flows are conserved at node \(i\): \( \sum_{j \in N} v_{ij} = \sum_{j \in N} v_{ji} \), or \( U_i \) and \( V_i \) are empty.

**CASE 2**

if \( \sum_{j \in N} v_{ij} < \sum_{j \in N} v_{ji} \), increase \( v_{ij} \) and decrease \( v_{ji} \) so as to minimise

\[ \sum_{j \in N} v_{ij} - \sum_{j \in N} v_{ji} . \]

(i) Sort \( V_i = \{ (i, j) \} \) according to the descending order of \( \{ C_{ij} - v_{ij} \} \) and sort \( U_i = \{ (j, i) \} \) according to the descending order of \( \{ v_{ji} \} \). The resulting link lists are still denoted as \( V_i \) and \( U_i \) respectively.

(ii) If \( \sum_{j \in N} (C_{ij} - v_{ij}) > \sum_{j \in N} (C_{ji} - v_{ji}) \), then for the link \((i, j)\) on the top of \(V_i\), calculate the flow increment for link \((i, j)\):

\[ \Delta_{ij} = \min\left\{ C_{ij} - v_{ij}, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'}, \min\left\{ 0, \sum_{j' \in N} v_{j'} - \sum_{j' \in N} v_{j'} \right\} \right\} \]

and set \( v_{ij} = v_{ij} + \Delta_{ij} \) and \( V_i = V_i - \{ (i, j) \} \).

Otherwise, for the link \((j, i)\) on the top of \(U_i\) calculate the flow decrement for link \((j, i)\).
\[ \delta_{ji} = \min \left\{ v_{ji}, \sum_{j \in N} v_{fj} - \sum_{j \in N} v_{ji}', \min \left\{ 0, \sum_{j \in N} v_{j'i} - \sum_{j \in N} v_{fj}' \right\} \right\} \]

and set \( v_{ji} = v_{ji} - \delta_{ji} \) and \( U_i = U_i - \{ (j, i) \} \).

(iii) Repeat (ii) until the flows are conserved at node \( i \):
\[
\sum_{j \in N} v_{ji} = \sum_{j \in N} v_{ji'}, \quad \text{or} \quad U_i \text{ and } V_i \text{ are empty.}
\]

(3) Repeat (1) and (2) until the stopping criteria are met at all nodes:
\[
\left| \sum_{j \in N} v_{ji} - \sum_{j \in N} v_{ji'} \right| \leq \varepsilon \quad (\varepsilon > 0).
\]

5. A CASE STUDY

A case study was conducted to examine the proposed methods. The data was drawn from the Future Directions Study (FDS) Network for Sydney, created by the Roads and Traffic Authority (RTA) of New South Wales, Australia. The proposed methods were applied to the FDS network, with the original 78 zones aggregated into 14 contiguous zones. The links of the FDS network were aggregated into a single link between each zone pair (Hensher et al., 1995). The resulting network is called a zonal network, depicted as follows (Figure 5).
The observed traffic counts were obtained from the equilibrium results produced by the EMME/2 package (INRO, 1994) with the following O-D trip matrix (Table 1).

The O-D trip matrix is obtained by aggregating the detailed zonal trips and modified to produce the traffic flows which are in the time equilibrium pattern by Wardrop’s first Principle. Therefore, the obtained traffic counts are in the equilibrium pattern.

To apply the developed method to estimate the O-D trip matrix based on the equilibrium traffic flows produced by the EMME/2 package, the $L_2$ norm function is used as the generalised distance between the observed and estimated traffic flows in the objective function. We set $\lambda = 0.5$ and use the BPR function (BPR, 1964) with $\alpha = 2.5$ and $\beta = 2.5$ for the application of the proposed method for estimating the O-D trip matrix based on the equilibrium traffic counts. The EMME/2 package was used to produce the equilibrium traffic flows for the estimated O-D trip matrix during the iteration of the algorithm. The stopping criteria set for the termination of the algorithm are trip and flow difference between two successive iterations, set as 1 and 10 respectively. The actual running of the algorithm involved only six iterations and was terminated by the flow difference of 4.24. The algorithm is not sensitive to the value of $\lambda$. The resulting estimated O-D trip matrix is presented in Table 2.

Two statistical measures of ‘closeness’ for both the estimated trips and flows were employed: the root mean squared error (RMSE) and mean absolute error (MAE) (Sherali et al, 1994):
\[ RMSE_t = \sqrt{\sum_{(i,j) \in OD} (t_{ij} - \tilde{t}_{ij})^2 / |OD|} \]

\[ MAE_t = \sum_{(i,j) \in OD} |t_{ij} - \tilde{t}_{ij}| / |OD| \]

for the estimated O-D trips, and

\[ RMSE_v = \sqrt{\sum_{(i,j) \in A} (v_{ij} - \tilde{v}_{ij})^2 / |A|} \]

\[ MAE_v = \sum_{(i,j) \in A} |v_{ij} - \tilde{v}_{ij}| / |A| \]

for the estimated network flows.

\( t_{ij} \) and \( \tilde{v}_{ij} \) are the true trips (real target) and observed flows respectively, \( OD \) and \( A \) are the sets of O-D pairs and observed network links respectively. Relative errors can be obtained by dividing the absolute errors by the average of the true values (Yang 1992), i.e.

\[ RMSE(\%)_t = \frac{RMSE_t}{\sum_{(i,j) \in OD} t_{ij} / |OD|} \]

\[ MAE(\%)_t = \frac{MAE_t}{\sum_{(i,j) \in OD} t_{ij} / |OD|} \]

for the estimated trips, and

\[ RMSE(\%)_v = \frac{RMSE_v}{\sum_{(i,j) \in A} v_{ij} / |A|} \]

\[ MAE(\%)_v = \frac{MAE_v}{\sum_{(i,j) \in A} v_{ij} / |A|} \]

Similar definitions of RMSE and MAE can also applied to the travel times between zones. Table 3 summarises the performance of the proposed method on these measures for the estimated network flows.

Table 3. Statistical measures for the proposed method applied to the Sydney network

<table>
<thead>
<tr>
<th></th>
<th>IterDif</th>
<th>ObsrDif</th>
<th>RMSE</th>
<th>MAE</th>
<th>RMSE(%)</th>
<th>MAE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time</td>
<td>0.39</td>
<td>1646.77</td>
<td>117.63</td>
<td>42.64</td>
<td>0.71</td>
<td>0.26</td>
</tr>
<tr>
<td>Trip</td>
<td>2.40</td>
<td>7882.96</td>
<td>563.07</td>
<td>198.07</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>Flow</td>
<td>4.24</td>
<td>14542.64</td>
<td>1979.00</td>
<td>1040.65</td>
<td>0.10</td>
<td>0.05</td>
</tr>
</tbody>
</table>

In Table 3, \( IterDif \) stands for the Euclidean distance of attributes (travel time, trip, and flows) between the current and previous iterations and \( ObsrDif \) is the Euclidean distance between the estimated and observed attributes. From Table 3, we can see that the both absolute and relative errors are small for the three attributes. This indicates that the method developed in this paper produces satisfactory result for the Sydney network.

6. CONCLUSION
Taking into account the inconsistency in traffic counts and the disadvantage of specifying a target matrix in estimating an origin-destination (O-D) trip matrix from traffic counts, this paper formulates a non-linear programming model for the trip matrix estimation problem from traffic counts without the specification of a target trip matrix. Rather, the model incorporates the trip production and attraction information which is relatively easier to collect than the target trip matrix. The formulated model has the advantages that it incorporates inconsistent traffic counts, it does not require a full set of network counts, and it will always produce feasible solutions. A heuristic method combining the traditional double factor method and user-equilibrium assignment is proposed for the model.

A network model is formulated to extend sampled traffic counts to network population flows in the way that they are specified in a user-equilibrium pattern consistent with Wardrop's first principle. For the extended network flows, validation and calibration can be carried out with the real network flows. A heuristic algorithm is proposed for the network model which considers the balance of travel time between a link and the alternative shortest path associated with the end points of the link, and the node flow conservation balance. A sampling strategy utilising the minimum spanning tree algorithm is outlined for an optimal traffic counting.

The proposed approach is applied to the Sydney network. Satisfactory results are obtained based on indices of root mean square error (RMSE) and the mean absolute error (MAE) for both network flows and O-D trip matrices.

Further research to compare our approach with the target matrix approach for real networks is recommended. In addition, there is a need to (i) investigate the convergence of the proposed heuristic methods, (ii) evaluate alternative heuristics to obtain more accurate estimates, and (iii) to examine the statistical variability of the sample counts.

Acknowledgement - The authors acknowledge the financial support the Australian Research Council (Grant no. A79131696). We thank the Roads and Traffic Authority of NSW for permission to use the Future Directions Study network.

REFERENCES

Akcelik R. (1991) Travel time functions for transport planning purposes: Davidson’s function, its time-dependent form and an alternative travel time function, Australian Road Research, 21(3), 49 - 59.


APPENDIX: SOLUTION UNIQUENESS OF EXISTING MODELS FOR ESTIMATING A TRIP MATRIX FROM LINK TRAFFIC COUNTS

Both CC and FC models will produce unique solutions if the objective functions are convex and the constraints are concave. For congested networks, the non-linear user-equilibrium assignment operator $M(t)$ may not turn out to be concave since there is no explicit form for $M(t)$. Therefore the uniqueness of the solutions to both either CC or FC model is questionable. Further research may be needed to show that there are non-linear user-equilibrium assignment operators for congested networks that are concave.

For uncongested networks, when linear assignment is applied: $M(t) = Pt$ and quadratic objective functions are used if it is known that both CC and FC models will produce unique solutions since the objective functions become convex and the constraints concave. This result can also be proven in matrix form as follows.

**Counts constrained (CC) models:**

\[
\text{Minimize} \quad (t - \bar{t})^T Q(t - \bar{t}) \\
\text{subject to} \quad Pt = \bar{v}
\]

where, $Q$ is a finite positive definite matrix.

The problem may be solved by forming the Lagrangian equation

\[
L(t, \lambda) = (t - \bar{t})^T Q(t - \bar{t}) + \lambda^T (Pt - \bar{v})
\]

where $\lambda$ is the vector of Lagrangian multipliers. Differentiate with respect to $t$ and $\lambda$, the necessary and sufficient conditions for a solution is given by:
\[
\frac{\partial L}{\partial t} = Q(t - \tilde{t}) + P^T \lambda = 0
\]
\[
\frac{\partial L}{\partial \lambda} = Pt - \bar{v} = 0
\]

Drive \( t \) from (13): \( t = \tilde{t} + Q^{-1} P^T \lambda \), and substitute into (14), we have

\[
PQ^{-1} P^T \lambda = \bar{v} - P \tilde{t}
\]

If we assume that the rows of the proportion matrix \( P \) are linearly independent, then the matrix \( PQ^{-1} P^T \) is non-singular. Therefore \( \lambda \), thus \( t \), can be uniquely solved.

**Flows constrained (FC) models:**

Minimize \( (t - \tilde{t})^T Q(t - \tilde{t}) + (\bar{v} - \tilde{v})^T R(\bar{v} - \tilde{v}) \)

subject to \( Pt = \nu \)

where, \( Q \) and \( R \) are finite positive definite matrices.

Substitute the constraints into the objective function and differentiate with respect to \( t \), the necessary and sufficient condition for a solution is given by

\[
Q(t - \tilde{t}) + P^T R(Pt - \bar{v}) = 0
\]

or

\[
(Q + P^T RP)t = Q \tilde{t} + P^T R \bar{v}
\]

Since \( P^T RP \) is non-negative definite, \( Q + P^T RP \) is positive definite. Therefore, \( t \) is uniquely determined.

From the above proofs, it can be seen that the uniqueness of a solution to the CC model requires that the rows of the proportion matrix are linearly independent while the FC model does not require this constraint. Therefore it is clear that the FC model can incorporate the inconsistencies in both flows and constraints. This appendix also derives solutions to both CC and FC models with a quadratic objective function and linear constraints for uncongested networks.