A FUZZY SET APPROACH TO ESTIMATING OD MATRICES IN CONGESTED BRAZILIAN TRAFFIC NETWORKS

Leslie Richard Foulds
Hugo A. D. do Nascimento*
Iacer C. A. C. Calixto†
Bryon Richard Hall [1]
Humberto Longo

Instituto de Informática – INF
Instituto de Matemática e Estatística – IME [1]
Universidade Federal de Goiás – UFG
Goiânia–GO, Brasil

lesfoulds@inf.ufg.br
hadn@inf.ufg.br
iacercalixto@gmail.com
bryon@mat.ufg.br
longo@inf.ufg.br

Resumo

Apresentamos um método de estimação de matrizes origem–destino (OD) para redes viárias urbanas congestionadas. Assume-se que os dados disponíveis incluem estimações incompletas e imprecisas de: contagens de tráfego, entradas OD, partidas em origens e chegadas em destinos. O método consiste de uma sequência de programas lineares *fuzzy* e foi projetado especialmente para atender à realidade de cidades brasileiras de médio a grande porte. Quando não há uma alocação de tráfego de equilíbrio que corresponda aos dados de entrada disponíveis, o método produz um conjunto de alocações de tráfego – e matrizes OD correspondentes – dentro de um espectro que abrange (i) das soluções que satisfaçam as estimações de entrada até (ii) as soluções que satisfaçam uma alocação de equilíbrio. Testou-se o método com dois exemplos numéricos, um deles proposto pelos autores e outro um teste clássico disponível na literatura.
Palavras-Chave: Matriz OD, Rede de tráfego, *Fuzzy sets*.
Área principal: L&T – Logística e Transportes.

Abstract

In this paper we describe a new method for estimating origin–destination (OD) matrices for congested urban traffic networks. It is assumed that the input data includes incomplete, imprecise estimates of: link counts, trip table entries, numbers of departures from origins and numbers of arrivals at destinations. The method is based on a sequence of fuzzy linear programs and is designed especially for the particular characteristics of medium-to-large Brazilian cities. When there doesn't exist a user–equilibrium traffic assignment that corresponds to the input data, the method provides a range of traffic assignments and their related OD matrices, within the spectrum of (i) satisfaction of the inputted estimates and (ii) a user–equilibrium assignment. The method has been tested on two numerical examples, one proposed by the authors and the other a classic one from the literature.
Keywords: OD matrix, Traffic network, Fuzzy sets.
Main area: L&T – Logistics and Transportation.

*Hugo do Nascimento is partially sponsored by a Scholarship of Research Productivity from CNPq (309463/2009-2).
†Iacer Calixto receives a master scholarship from CAPES.

1 Introduction

As urban populations expand and streets become increasingly congested, city planners need comprehensive transport plans for appropriate user service while supporting economic development and reducing vehicle pollution. Reliable estimation of origin–destination (OD) travel demands is often a crucial first step in achieving these objectives. Over the last 35 years, various methods based on relatively inexpensive link (network arc) counts have been developed for estimating OD matrices. A key issue in the estimation of a trip matrix based on link counts is the identification of the origin–destination pairs whose trips use a particular network arc. The estimation of OD matrices and link counts are connected by the assignment of traffic to the various routes between each OD pair. Two types of assignments have been traditionally proposed to model user behaviour, proportional and equilibrium assignment.

With proportional assignment, the proportion of travel demand between an OD–pair does not depend on the OD matrix. Over the years, this line of research has been extensively examined and improved. See for example, Van Zuylen and Willumsen (1980), Nguyen (1984) and Brenninger-Göthe et al. (1989).

User–equilibrium assignment was first enunciated by Wardrop (1952). With equilibrium assignment based on Wardrop's Principles, the travel demand is distributed such that each route cost (of all used routes) between any OD pair is equal and no greater than that of any unused route. User–equilibrium assignment is assumed throughout this article because it more accurately models congestion effects than proportional assignment. Gur et al. (1980) developed an iterative technique based on equilibrium assignment that employs the Frank–Wolfe algorithm as a subroutine with their respected LINKOD system. Also, Sherali et al. (2003) have presented a user–equilibrium based heuristic that employs linear programming (LP) iteratively. Further, these authors have presented evidence that their heuristic outperforms the maximum entropy approach of Van Zuylen and Willumsen (1980), as implemented by Bromage (1991). Finally, the various main types of OD matrix estimation methods have been surveyed by Viti (2008).

There have been numerous reports of OD estimation methods based on relatively inexpensive traffic link counts. However, many of the reported methods are designed for Northern Hemisphere situations that differ significantly from most medium-to-large Brazilian urban traffic networks. Furthermore, many of the methods rely on comprehensive, exact input data and assume that users have perfect knowledge of their traffic environment. We describe a fuzzy linear programming-based OD matrix estimation method that deals with input data that is both imprecise and incomplete. The method has been developed to overcome the deficiencies mentioned above and is designed especially for congested Brazilian urban traffic networks.

We are concerned with simulating a given congested urban traffic network, rather than organizing its daily operations. We therefore focus on networks that are static, that is, they are studied for one time period only. An origin–destination (OD) matrix has rows and columns that represent the origins and destinations of users of the traffic network and each entry represents the corresponding OD travel demand. We wish to estimate the entries in this matrix from input data obtained from observations of the actual traffic flows.

2 Developing an OD Estimation Model

In order to develop and compare network models that attempt to predict the OD matrix, we adopt the notation introduced by Sherali et al. (2003). Let $G = (N, A)$ be a given traffic network with node set N of n nodes and arc set A. The nodes in N may represent either actual intersections of roads/streets or more general zones. In either case, they are the origins and/or the destinations of trips by the users of G. Let O be the set of given origin nodes in N and D be the set of given destination nodes in N. Usually sets O and D have many nodes in common. Let OD be the set of origin–destination pairs for all possible OD paths that could be reasonably used in any rational, final traffic assignment. Let T_{ij} be the i–j entry in the target origin–destination matrix. That is, the number of users with origin node i and destination node j. We wish to estimate the matrix $T = (T_{ij})_{n\times n}$. Let Q_{ij} be a working estimate of T_{ij}. The values Q_{ij} will vary during the use of the method as it progressively assigns users to OD paths, and the final value of Q_{ij} will be our estimate of T_{ij}.

Suppose that initial travel demands are given for some of the origin–destination pairs in OD. These values are estimates of some of the entries of the final, target OD matrix T. Note that this is the matrix that we are trying to find, not a seed matrix that may represent inaccurate or outdated data. It is assumed that the Q_{ij} values are not known precisely but are "triangulated" by estimates Q'_{ij} (> 0) and by deviations b^L_{ij} and b^U_{ij}, that are all given constants. For some pairs $(i, j) \in OD$ we are given Q'_{ij} as the best known estimate of T_{ij} and it is desirable to find a final value of Q_{ij} that is as close as possible to Q'_{ij}. So,

$$Q'_{ij} - b^L_{ij} \leq Q_{ij} \leq Q'_{ij} + b^U_{ij}, \quad \text{for } (i,j) \in OD, \text{ where } Q'_{ij} > 0. \tag{2.1}$$

Of course, in each case, $Q'_{ij} - b^L_{ij} \geq 0$. It may well be that no such estimates are available for certain i–j pairs in OD. In these cases the corresponding travel demands are termed "unspecified" and the Q_{ij} values are not constrained as in (2.1) above.

It is our experience in modelling Brazilian urban traffic networks that it is often difficult to obtain reliable estimates of the origin–destination travel demands Q'_{ij} for a significant number of i–j pairs. We have found it far easier to obtain reliable estimates of both the total number of users that depart from various origins and the total number of users that arrive at various destinations. Along with the Q'_{ij} estimates that can be identified, the origin departure and destination totals are also useful in estimating the matrix T.

To incorporate these latter estimates, let Q_V be the set of OD paths $(i, j) \in OD'$ where Q'_{ij} is given; O_V be the set of nodes $i \in O$, where O'_i is given; D_V be the set of nodes $j \in D$ where D'_j is given; O_i be the total number of users that depart from origin i, where $i \in O_V$; and D_j be the total number of users that arrive at destination j, where $j \in D_V$.

Similar to the discussion above for the Q'_{ij} estimates, suppose that initial values of the total number of departures (arrivals) are given for some nodes $i \in O (j \in D)$. It is assumed that the O_i (D_j) values are not known precisely but are "triangulated" by positive best estimates O'_i (D'_j), respectively, and by deviations d^L_i and d^U_i (e^L_j and e^U_j), respectively, that are all given constants. Thus,

$$O'_i - d^L_i \leq O_i \leq O'_i + d^U_i \quad \text{for } i \in O_V, \quad \text{and} \tag{2.2}$$

$$D'_j - e^L_j \leq D_j \leq D'_j + e^U_j \quad \text{for } j \in D_V. \tag{2.3}$$

Of course, in each case, $O'_i - d^L_i \geq 0$. It may well be that no such estimates are available for certain nodes $i \in O$ ($j \in D$). In these cases the corresponding departures (arrivals) are termed "unspecified" and the O_i (D_j) values are not constrained as in (2.2) and (2.3) above.

Let A be the directed arcs of G that represent the roads and streets of interest. That is, the arcs (roads/streets) in A directly connect certain adjacent pairs of nodes (intersections/zones) in N. Let A_V be the subset of A for which link counts are known and A_M be the subset of A for which link counts are unknown. Thus $A = A_V \cup A_M$. Let f_α be a working estimate of the arc flow in arc α, for all arcs $\alpha \in A$. The arcs in A_V will have flows reflecting the given link counts. And it may well be that many of the arcs in A_M will also have positive flows. The f_α values corresponding to all the arcs in A do not have a precise value, but an average value. The arcs in A_V can be "triangulated"' by estimates $f'_\alpha (> 0)$ and by deviations a^L_α and a^U_α, that are given constants. For each arc $\alpha \in A_V$ we are given f'_α as the best known estimate of f_α and it is desirable to find a final value of f_α that is as close as possible to f'_α. So,

$$f'_\alpha - a^L_\alpha \le f_\alpha \le f'_\alpha + a^U_\alpha, \quad \text{for all arcs } \alpha \in A_V. \tag{2.4}$$

As before, in each case, $f'_\alpha - a^L_\alpha \ge 0$. No link count estimates are available for the arcs in A_M. In these cases the flows are termed "unspecified" and the corresponding f_α values are not constrained as in (2.4).

We now begin constructing models with the aim of estimating the matrix T. The following notation is defined for each origin–destination pair i–$j \in OD$. Let: n_{ij} be the number of distinct possible paths from origin node i to destination node j that are considered of potential use by users; p^k_{ij} denote the k^{th} least–costly path from origin node i to destination node j; $(p^k_{ij})_\alpha = 1$, if the k^{th} path from origin node i to destination node j contains arc α, and $= 0$ otherwise; and x^k_{ij} be the number of users using the k^{th} path from origin node i to destination node j. In order to understand how the users behave we must find values for all of the x^k_{ij} decision variables. They are constrained as follows:

$$\sum_k x^k_{ij} \ \ge \ Q'_{ij} - b^L_{ij}, \quad \forall\,(i,j) \in Q_V, \tag{2.5}$$

$$\sum_k x^k_{ij} \ \le \ Q'_{ij} + b^U_{ij}, \quad \forall\,(i,j) \in Q_V, \tag{2.6}$$

$$\sum_{j \in D} \sum_k x^k_{ij} \ \ge \ O'_i - d^L_i, \quad \forall\, i \in O_V, \tag{2.7}$$

$$\sum_{j \in D} \sum_k x^k_{ij} \ \le \ O'_i + d^L_i, \quad \forall\, i \in O_V, \tag{2.8}$$

$$\sum_{i \in O} \sum_k x^k_{ij} \ \ge \ D'_j - e^L_j, \quad \forall\, j \in D_V, \tag{2.9}$$

$$\sum_{i \in O} \sum_k x^k_{ij} \ \le \ D'_j + e^U_j, \quad \forall\, j \in D_V, \tag{2.10}$$

$$\sum_{(i,j) \in OD} \sum_k (p^k_{ij})_\alpha . x^k_{ij} \ \ge \ f'_\alpha - a^L_\alpha, \quad \forall\, \alpha \in A_V, \tag{2.11}$$

$$\sum_{(i,j) \in OD} \sum_k (p^k_{ij})_\alpha . x^k_{ij} \ \le \ f'_\alpha + a^U_\alpha, \quad \forall\, \alpha \in A_V, \text{ and} \tag{2.12}$$

$$x^k_{ij} \ \ge \ 0, \quad \forall (i,j) \in OD \text{ and } k = 1, 2, \ldots, n_{ij}. \tag{2.13}$$

We can regard the set of constraints (2.5)–(2.13) as a set of linear equations with unknowns $X = \{x^k_{ij} \mid (i,j) \in OD \text{ and } k = 1, 2, \ldots, n_{ij}\}$. X is called a traffic assignment for G. The number of solutions in X may be none, exactly one, many or infinite. For most practical situations, in order to find the most appropriate assignment X, we usually need some additional information about the way in which users choose their individual OD paths. As is well known, users often choose their OD paths according to their perceived cost of the various paths that are available. Usually, the cost of each path can be obtained as the sum the costs of its constituent arcs. And it is commonly assumed that each individual arc α, has a unit traversal cost $c_\alpha(f_\alpha)$, that depends upon the number of users f_α, of the arc. The f_α values may be calculated using software that produces traffic assignments or which simply attributes a certain cost to each arc as a function of its flows and the characteristics of the arc. As an example, PET-Gyn (Jradi et al. (2009)) is a software that models traffic in Brazil, taking into account the differences in the common urban traffic structure existing in Brazil and many other developing countries. The cost function for each arc depends on the flow in the arc itself, on its traffic lights and also on flow in other arcs that dominate (preferential flow) the arc, such as traffic that is opposed to flow in the arc, for example, two-way streets or intersections. In any case we have $c_\alpha(f)$ as a cost (time) function of flow in the arc α. A function that is in common use was devised by the Bureau of Public Roads

(USA) in 1964:

$$c_\alpha(f_\alpha) = c_\alpha^F[1 + 0.15(f_\alpha/u_\alpha)^4], \tag{2.14}$$

where c_α^F is the congestion–free travel cost in arc α and u_α is the effective capacity of arc α.

If a reliable traffic assignment X has been found by some process, such as the traffic assignment program PETGyn (Jradi et al. (2009)), its x_{ij}^k values can be used to find estimates of the arc flows f_α, as,

$$f_\alpha = \sum_{(i,j)\in OD} \sum_k (p_{ij}^k)_\alpha.(x_{ij}^k), \quad \alpha \in A. \tag{2.15}$$

However, in the present situation of OD matrix estimation it is unlikely that suitable x_{ij}^k values will be available to enable expression (2.15) to be used. Instead, what is assumed throughout the rest of this article is that link counts are pre-specified for the subset A_V and are unavailable for A_M, the rest of the arcs in A. When all possible estimated arc flow values f_α have been identified, estimates of the relevant average OD path traversal costs must be calculated as exact values. Let C_{ij}^k be the unit traversal cost of the k^{th} path from any origin node i to any destination node j for each user, under current conditions for the present time period. Then

$$C_{ij}^k = \sum_{\alpha\in A} (p_{ij}^k)_\alpha.c_\alpha(f_\alpha), \quad \forall\, (i,j) \in OD \text{ and } k = 1, 2, \ldots, n_{ij}. \tag{2.16}$$

In the execution of the method discussed later in this article, it is productive to improve periodically the estimates of the C_{ij}^k's to be used. This is because, for each combination of $(i,j) \in OD$ and $k = 1, 2, \ldots, n_{ij}$, it is essential to identify an estimate of C_{ij}^k that reflects the expected number of users of path p_{ij}^k in the final assignment of flow to be constructed. This necessitates the use of a "bi–level" process that iterates between a method that identifies Q_{ij} values and some simulation process (such as PETGyn). Such accurate C_{ij}^k estimates may enhance the accuracy of the assignment X and ultimately the final OD matrix T.

We now make some assumptions. All of the **least–cost** paths belonging to the set $\{p_{ij}^k \mid k = 1, 2, \ldots, n_{ij}\}$ are known and the cost of each of them is known as a non–negative number C, say. Consider the costs of some of the other (**greater–cost**) paths belonging to the set $\{p_{ij}^k \mid k = 1, 2, \ldots, n_{ij}\}$, that may possibly used. The actual costs of these paths need not be known. Their costs are known only to be greater than C.

The method to be presented in this article does not require any information about arc costs. If we have only the arc costs and not the OD path costs given as input data, we must pre–process by summing the arc costs to find the necessary OD path costs as exact numbers.

We now construct models that aim to produce final values of the x_{ij}^k variables, that is, the final traffic assignment, X. Incidentally, once values for these variables have been identified, the estimates of T can be calculated as:

$$T_{ij} = \sum_k x_{ij}^k, \quad \forall\, (i,j) \in OD. \tag{2.17}$$

We begin by modifying the OD path costs in order to attempt to assign each user to one of his least–cost OD paths. If this can be achieved, the result will be a user–equilibrium assignment. Without loss of generality, we assume from now on that, for each $(i,j) \in OD$, the costs C_{ij}^k, $k = 1, 2, \ldots, n_{ij}$, are ordered in non–decreasing order, that is: $C_{ij}^1 \le C_{ij}^2 \le \cdots \le C_{ij}^n$. Let: $C_{ij}^* = \min\{C_{ij}^k \mid k = 1, 2, \ldots, n_{ij}\}$; $K_{ij} = \{k \mid k = 1, 2, \ldots, n_{ij};\ C_{ij}^k = C_{ij}^*\}$; and $K'_{ij} = \{1, 2, \ldots, n_{ij}\} \backslash K_{ij}$.

We now modify the C_{ij}^k's as follows. Let

$$\begin{aligned} C_{ij}^k &= C_{ij}^*, \quad \text{for } k \in K_{ij}, \\ &= (k-1).M.C_{ij}^*, \quad \text{for } k \in K'_{ij}, \end{aligned} \tag{2.18}$$

where M is a suitably large real number.

In the numerical examples solved by the method developed in the article, M is set to 10. Note also, because of the order of the C_{ij}^k's, that $\forall k \in K'_{ij}$, $k > 1$. We now construct models that attempt, where it exists, to identify a user–equilibrium assignment that is subject to the constraints on the OD travel demands, total departures, total arrivals and on the link counts. It can be seen from (2.5) to (2.12) that the variables Q_{ij}, O_i, D_j and f_α are not known precisely but are triangulated. At this point we could construct solution models based on standard LP, such as the model $\boldsymbol{M_0}$ given below:

$$\text{Minimize} \quad \sum_{(i,j)\in OD} \sum_k C_{ij}^k . x_{ij}^k,$$
subject to the relationships: (2.5), (2.6),..., (2.13).

Suppose that the original OD path costs, and not the revised costs defined in (2.18), are used in M_0. Then M_0 will identify a system optimal assignment. Although not realistic for modelling practical, congested traffic networks, a system optimal assignment is useful in the attempt to establish bounds on user costs and on exhaust gas emissions. Solution approaches similar to model M_0 accept that imprecision is equivalent to randomness. Here, randomness in concerned with the uncertainty of the membership of an exact set. That is, it is equally likely and acceptable that any variable can be set to any value within its range.

3 OD Matrix Estimation in a Fuzzy Environment

Recall that it is desirable to find values of Q_{ij}, O_i, D_j and f_α that are as close as possible to Q'_{ij}, O'_i, D'_j and f'_α, respectively. Because of this, we reject the uniform distribution approach (randomness) of standard LP. Instead, we choose a fuzzy set approach to OD matrix estimation. The basic elements of the theory of fuzzy sets were introduced by Zadeh (1965) and their application to linear programming in a fuzzy environment has been popularized by Zimmermann (1983).

We shall create linear programming models that have fuzzy variables as input data. The authors are aware of only two relevant articles on the application of fuzzy sets to OD matrix estimation. Liao and Wang (1996) have addressed the fuzzy resolution of the infeasibility of user equilibrium in traffic assignment problems. These authors have developed some mathematical results for the case where the relative values of the link counts and the structure of the network mean that a feasible user–equilibrium assignment does not exist. They use fuzzy set theory to identify which link counts must be relaxed to enable user equilibrium to be achieved within the relaxed link count limits. And Biletska et al. (2009) developed a dynamic two–step method for short–time OD matrix estimation at a complex signalized junction for one traffic light cycle using fuzzy–timed high–level Petri nets. Since the data used to estimate the matrix are imprecise, the authors represent them as fuzzy numbers.

To indicate the fuzzy nature of the Q_{ij}'s, O_i's, D_j's and f_α's values we substitute fuzzy versions of them in the relationships (2.1)–(2.4). The fuzzy versions are denoted by: $\tilde{Q}_{ij}$, $\tilde{O}_i$, $\tilde{D}_j$ and $\tilde{f}_\alpha$. We then have the following fuzzy linear model $\boldsymbol{M_1}$:

$$\text{Minimize} \quad \sum_{(i,j)\in OD} \sum_k C_{ij}^k . x_{ij}^k \quad (= z) \tag{3.1}$$

subject to:

$$\sum_k x_{ij}^k = \tilde{Q}_{ij}, \forall\, (i,j) \in Q_V, \tag{3.2}$$

$$\sum_{j\in D} \sum_k x_{ij}^k = \tilde{O}_i, \ \ \forall\, i \in O_V, \tag{3.3}$$

$$\sum_{i\in O}\sum_{k} x_{ij}^{k} = \tilde{D}_j, \forall\, j \in D_V, \tag{3.4}$$

$$\sum_{(i,j)\in OD}\ \sum_{k} (p_{ij}^{k})_{\alpha}.x_{ij}^{k} = \tilde{f}_{\alpha}, \forall\, \alpha \in A_V, \tag{3.5}$$

$$x_{ij}^{k} \geq 0,\ \ \forall\,(i,j) \in OD \text{ and } k = 1,2,\ldots,n_{ij}. \tag{3.6}$$

It is proposed to solve model M_1 by following the fuzzy linear programming approach of Zimmermann (1983). First, we calculate upper and lower bounds on the set of all possible optimal solution values that z can take on while the variables $\tilde{Q}_{ij}$, $\tilde{O}_i$, $\tilde{D}_j$ and $\tilde{f}_{\alpha}$ vary between their upper and lower limits. The bounds are obtained by solving the crisp linear programming models M_0 stated before and M_2 that is given below. We can calculate a lower bound z_L say, on z as the optimal solution to model M_0. We can calculate an upper bound z_U say, on z by solving the following LP (model $\boldsymbol{M_2}$):

$$\begin{array}{ll} \text{Minimize} & \sum_{(i,j)\in OD}\sum_k C_{ij}^k.x_{ij}^k \quad (= z_U) \\ \text{subject to:} & \end{array}$$

$$\begin{array}{rcll} \sum_{k} x_{ij}^{k} & \geq & Q'_{ij}, & \forall\,(i,j) \in Q_V, \\ \sum_{j\in D}\sum_{k} x_{ij}^{k} & \geq & O'_{i}, & \forall\, i \in O_V, \\ \sum_{i\in O}\sum_{k} x_{ij}^{k} & \geq & D'_{j}, & \forall\, j \in D_V, \\ \sum_{(i,j)\in OD}\sum_{k} (p_{ij}^{k})_{\alpha}.x_{ij}^{k} & \geq & f'_{\alpha}, & \forall\, \alpha \in A_V, \quad \text{and condition (3.6).} \end{array}$$

It is assumed that z_L and z_U are both finite. According to the theory of fuzzy sets as applied to linear programming, the following problem is equivalent to model M_1:

$$\begin{array}{rcl} \text{Find } X & = & \{x_{ij}^k \mid (i,j) \in OD,\ k = 1,2,\ldots,n_{ij}\},\ \text{such that:} \\ & & \sum_{(i,j)\in OD}\ \sum_{k} C_{ij}^{k} x_{ij}^{k} \leq z_U \ \text{and the constraints (3.2)–(3.6) hold.} \end{array} \tag{3.7}$$

A fuzzy set is a function that maps a potential member of the set to a number between zero and one, indicating its actual degree of membership. For an optimization problem with fuzzy features Bellman and Zadeh (1970) proposed optimizing its objective function and all its constraints simultaneously. In order to determine the optimal solution point, both the objective function and all the constraints must be characterized by membership functions and they must be linked by the linguistic conjunctions "or" (for maximization) and "and" (for minimization). The aim is to satisfy a fuzzy objective function and fuzzy constraints that all receive the same treatment. When there is no such difference, the fuzzy optimization problem is termed **symmetric**.

We seek to transform a symmetric fuzzy model into a crisp, deterministic model by defining appropriate membership functions. Fuzzy goals and fuzzy constraints can be defined precisely as fuzzy sets in the space of alternatives. A fuzzy decision then may be viewed as an intersection of the given goals and constraints. We now define membership functions for (3.1) to (3.5), where $z(x)$ is the optimal solution value to Model M_1. We assume that the objective function $z(x)$ must be essentially less than or equal to a given aspiration level z_L say, a given finite real number. However $z(x)$ is fuzzy in the sense that the values of $z(x)$ can be as high as z_U say, another given finite real number such that $z_L < z_U$. As shown by Klir and Yuan (1995), in this case the membership function of the fuzzy set of optimal solutions for this objective is:

$$\mu_z(x) = \begin{cases} 1, & \text{if } z(x) < z_L, \\ (z_U - z(x))/(z_U - z_L), & \text{if } z_L \leq z(x) \leq z_U, \\ 0, & \text{if } z_U < z(x); \end{cases} \tag{3.8}$$

$\forall\,(i,j) \in OD$ where $Q'_{ij} > 0$ and $k = 1,2,\ldots,n_{ij}$, let,

$$\mu_{ij}(x) = \begin{cases} (\sum_k x_{ij}^k - (Q'_{ij} - b_{ij}^L))/b_{ij}^L, & \text{if } Q'_{ij} - b_{ij}^L \leq \sum_k x_{ij}^k \leq Q'_{ij}, \\ (Q'_{ij} + b_{ij}^U - \sum_k x_{ij}^k)/b_{ij}^U, & \text{if } Q'_{ij} \leq \sum_k x_{ij}^k \leq Q'_{ij} + b_{ij}^U, \\ 0, & \text{otherwise;} \end{cases} \tag{3.9}$$

$\forall\, i \in O$ where $O'_i > 0$, let,

$$\mu_i^O(x) = \begin{cases} (\sum_{j\in D}\sum_k x_{ij}^k - (O'_i - d_i^L))/d_i^L, & \text{if } O'_i - d_i^L \leq \sum_{j\in D}\sum_k x_{ij}^k \leq O'_i, \\ (O'_i + d_i^U - \sum_{j\in D}\sum_k x_{ij}^k)/d_i^U, & \text{if } O'_i \leq \sum_{j\in D}\sum_k x_{ij}^k \leq O'_i + d_i^U, \\ 0, & \text{otherwise;} \end{cases} \tag{3.10}$$

$\forall\, j \in D$ where $D'_j > 0$, let,

$$\mu_j^D(x) = \begin{cases} (\sum_{i\in O}\sum_k x_{ij}^k - (D'_j - e_j^L))/e_j^L, & \text{if } D'_j - e_j^L \leq \sum_{i\in O}\sum_k x_{ij}^k \leq D'_j, \\ (D'_i + e_j^U - \sum_{i\in O}\sum_k x_{ij}^k)/e_j^U, & \text{if } D'_j \leq \sum_{i\in O}\sum_k x_{ij}^k \leq D'_j + e_j^U, \\ 0, & \text{otherwise;} \end{cases} \tag{3.11}$$

$\forall\, \alpha \in A_V$ where $f'_\alpha > 0$, let,

$$\mu_\alpha(x) = \begin{cases} (\sum_{(i,j)\in OD}\sum_k (p_{ij}^k)_\alpha - (f'_\alpha - a_\alpha))/a_\alpha, & \text{if } f'_\alpha - a_\alpha \leq \sum_{(i,j)\in OD}\sum_k (p_{ij}^k)_\alpha . x_{ij}^k \leq f'_\alpha, \\ (f'_\alpha + a_\alpha - \sum_{(i,j)\in OD}\sum_k (p_{ij}^k)_\alpha . x_{ij}^k)/a_\alpha, & \text{if } f'_\alpha \leq \sum_{(i,j)\in OD}\sum_k (p_{ij}^k)_\alpha . x_{ij}^k \leq f'_\alpha + a_\alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

We now use these membership functions to define a fuzzy optimization problem that is equivalent to model M_1. Bellman and Zadeh (1970) highlighted the main pillar of decision making in a fuzzy environment as $S = Z \cap C$, where Z is the fuzzy objective function, C is the set of fuzzy constraints and S = the set of decisions that can be taken in the fuzzy environment represented by Z and C combined. From the theory of fuzzy sets, the membership grade of any element in the set comprising the intersection of any two fuzzy sets A and B can be defined as the minimum of the membership functions of A and B. Let S be the set of decisions that can be taken in the fuzzy environment represented by (3.8)–(3.12). Because Z and C must be satisfied simultaneously, then, as demonstrated by Lai and Hwang (1992), we can define the membership function of S as: $\mu_S(x) = \min\{\mu_z(x), \mu_C(x)\}$. Then we can construct the following decision model for M_1:

$$\mu_S(x) = \min\{\mu_z(x), \mu_{ij}(x), \mu_i(x), \mu_j(x), \mu_\alpha(x) \mid (i,j) \in OD,\ i \in O_V,\ j \in D_V,\ \alpha \in A_V\}. \tag{3.13}$$

We now create an equivalent model for $\mu_S(x)$ by defining its maximizing decision as:

$$x^* = \max_{x\in X} \mu_S(x) = \max_{x\in X}[\min\{\mu_z(x), \mu_C(x)\}],$$

where x^* is the optimal solution to the fuzzy model in the original scale. In order to find x^*, we introduce an auxiliary variable λ, thus allowing the model to be aggregated into an equivalent, crisp LP by using the max–min operator as (model $\boldsymbol{M_3}$):

$$x^* = \max_{x\in X}[\min\{\mu_z(x), \mu_{ij}(x), \mu_i(x), \mu_j(x), \mu_\alpha(x) \mid (i,j) \in OD,\ i \in O_V,\ j \in D_V,\ \alpha \in A_V\}]. \tag{3.14}$$

In order to solve model M_3 we introduce an auxiliary variable λ. This allows the model to be aggregated into the following, equivalent crisp LP (model $\boldsymbol{M_4}$):

$$\begin{array}{llll} \text{Maximize} & \lambda, & & \\ \text{subject to:} & \lambda \leq & \mu_z(x), & \\ & \lambda \leq & \mu_{ij}(x), & \forall\,(i,j) \in Q_V, \\ & \lambda \leq & \mu_i^O(x), & \forall\, i \in O_V, \\ & \lambda \leq & \mu_j^D(x), & \forall\, j \in D_V, \\ & \lambda \leq & \mu_\alpha(x), & \forall\, \alpha \in A_V, \\ & x \geq & 0 \text{ and} & \\ & 0 \leq & \lambda \leq 1. & \end{array} \tag{3.15}$$

Substituting the relationships (3.8)–(3.12) we have the following crisp model (M_5 – **FLIPSOD**[1])

$$
\begin{array}{rcll}
\text{Maximize } \lambda & & & \\
\text{subject to:} & & & \\
\sum_{(i,j)\in OD} \sum_k C_{ij}^k x_{ij}^k + (z_U - z_L)\lambda & \leq & z_U, & \\
\sum_k x_{ij}^k - b_{ij}^L \lambda & \geq & Q'_{ij} - b_{ij}^L, & \forall\,(i,j) \in Q_V, \\
\sum_k x_{ij}^k + b_{ij}^U \lambda & \leq & Q'_{ij} + b_{ij}^U, & \forall\,(i,j) \in Q_V \\
\sum_{j\in D} \sum_k x_{ij}^k - d_i^L \lambda & \geq & O'_i - d_i^L, & \forall\, i \in O_V, \\
\sum_{j\in D} \sum_k x_{ij}^k + d_i^U \lambda & \leq & O'_i + d_i^U, & \forall\, i \in O_V \\
\sum_{i\in O} \sum_k x_{ij}^k - e_j^L \lambda & \geq & D'_j - e_j^L, & \forall\, j \in D_V, \\
\sum_{i\in O} \sum_k x_{ij}^k + e_j^U \lambda & \leq & D'_j + e_j^U, & \forall\, j \in D_V, \\
\sum_{(i,j)\in OD} \sum_k (p_{ij}^k)_\alpha . x_{ij}^k - a_\alpha^L \lambda & \geq & f'_\alpha - a_\alpha^L, & \forall\, \alpha \in A_V, \\
\sum_{(i,j)\in OD} \sum_k (p_{ij}^k)_\alpha . x_{ij}^k + a_\alpha^U \lambda & \leq & f'_\alpha + a_\alpha^U, & \forall\, \alpha \in A_V, \\
x_{ij}^k & \geq & 0 & \forall\,(i,j) \in OD,\ k = 1, 2, \ldots, n_{ij} \text{ and} \\
0 \;\leq\; \lambda & \leq & 1. &
\end{array}
\tag{3.16}
$$

Model M_5 is to be used iteratively to gradually improve the estimates of the C_{ij}^k's that are used. This is achieved by progressively reducing the value of z_U in each cycle and thus forcing the assignment to approach user–equilibrium. For each path p_{ij}^k, it is essential to identify an estimate of C_{ij}^k that reflects the expected number of users of the path in the final assignment of OD travel demand to be constructed. This is achieved by inputting the current iteration's estimated link counts f_α — found as a by–product of the OD matrix estimation — into a relation of the form of (2.14) to update the next iteration's arc costs and OD path costs. This implies that initial estimates of all the needed C_{ij}^k's are available. Estimates of all necessary OD path costs must be provided. For the initial iteration of the method, costs for arcs in A_V are calculated by using the given link counts in expression (2.14) and for arcs in A_M by initially setting the flows to zero. The new C_{ij}^k values are then used iteratively in Model M_5. Recall that once values for an assignment $X = \{x_{ij}^k \mid (i,j) \in OD,\ k = 1, 2, \ldots, n_{ij}\}$ have been identified, the estimates of T can be calculated as: $T_{ij} = \sum_k x_{ij}^k,\ \forall\,(i,j) \in OD$. The process is repeated until there is no significant change in the entries of T.

It is important to state that the chosen solution at the end of a given cycle used as reference to update the arc costs in the next cycle is always an assignment that is either a user–equilibrium assignment or one that is as close as possible to one. If there are multiple assignments, the one with the lowest total user cost amongst those will be chosen.

4 Computational experience with FLIPSOD

We now illustrate FLIPSOD by using it to solve a simple numerical problem that does not possess congestion affects.The problem has 8 nodes, $N = \{A, B, C, D, E, F, X, Y\}$ and 13 arcs, $A = \{AX_1(10), BX_2(12), XC_3(15), XD_4(11), YE_5(10), YF_6(16), DC_7(9), CE_8(19), XY_9(25), CY_{10}(12), DY_{11}(13)\}$. The (constant) arc costs are given in parentheses and their labelling numbers are given as subscripts. The remaining input data is: $AV = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $OD = \{AC, AD, AE, AF, BC, BD, BE, BF, CE, CF, DC,\ DE, DF\}$, $(Q'_{AC},\ Q'_{AD},\ Q'_{AE},\ Q'_{AF},\ Q'_{BC},\ Q'_{BD},\ Q'_{BE},\ Q'_{BF},\ Q'_{CE},\ Q'_{CF},\ Q'_{DE},\ Q'_{DF},\ Q'_{DC})$ = (53, 47, 30, 30, 63, 48, 27, 32, 129, 24, 134, 24, U) where 'U' denotes 'unspecified', $(b_{AC},\ b_{AD},\ b_{AE},\ b_{AF},\ b_{BC},\ b_{BD},\ b_{BE},\ b_{BF},\ b_{CE},\ b_{CF},\ b_{DE},\ b_{DF},\ b_{DC})$ = (11, 9, 6, 6, 13, 10, 5, 6, 26, 5, 27, 5, U), $(O'_A,\ O'_B,\ O'_C,\ O'_D)$ = (160, 170, 153, U), $(d_A,\ d_B,\ d_C,\ d_D)$ = (32, 34, 31, U), $(D'_C,\ D'_D,\ D'_E,\ D'_F)$ = (U, 95, 320, 110), $(e_C,\ e_D,\ e_E,\ e_F)$ = (U, 19, 64, 22), $(f'_1,\ f'_2,\ f'_3,\ f'_4,$

[1]**F**uzzy **L**inear **I**nteger **P**rogramming for **S**tatic **O**rigin-**D**estination matrix estimation.

f'_5, f'_6, f'_7, f'_8, f'_9, f'_{10}, f'_{11}) = (160, 170, 170, 144, 178, 110, 130, 141, U, U, U), (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 , a_9, a_{10}, a_{11})= (16, 17, 17, 14, 18, 11, 13, 14, U, U, U).

There are 33 OD paths, the first 13 of which are least–cost paths, one for each OD pair. When Models M_0 and M_2 are used with the OD path costs defined as in (2.18), they produce a lower bound $z_L = 17562$ and an upper bound of $z_U = 34325$. FLIPSOD produces an initial assignment with total user cost $z = 34325$ with $\lambda = 1$. The OD matrix estimates are identical to the original Q'_{ij} estimates above, except that Q'_{DC} was originally unspecified and has been assigned the value $T_{DC} = 130$. The link counts are also identical to the original f'_α estimates above except that f'_9, f'_{10} and f'_{11} were originally unspecified and have been assigned the values 16, 66 and 207, respectively. However, the corresponding assignment is not a user–equilibrium assignment in the sense that some users are assigned paths that exceed their individual least cost. Due to the structure of the network and the relative values of the input data, there does not exist a user–equilibrium assignment that corresponds to the given best estimates.

As the upper bound z_U is progressively reduced from the value 34325, FLIPSOD produces a series of assignments. The corresponding Q_{ij}, O_i, D_j and f_α values move progressively away from the given best estimates and towards either the given upper or lower limits of the best estimates. At $z_U = z_L =$ 17562, FLIPSOD produces a user–equilibrium assignment with total user cost $z = 17562$ and $\lambda = 0$. The corresponding OD matrix estimates are $(T_{AC}, T_{AD}, T_{AE}, T_{AF}, T_{BC}, T_{BD}, T_{BE}, T_{BF}, T_{CE}, T_{CF}, T_{DE}, T_{DF}, T_{DC}) = (57, 38, 24, 25, 50, 55, 22, 26, 103, 29, 160, 19, 117)$ and the link counts are (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}) = (144, 153, 153, 144, 160, 99, 117, 149, 0, 29, 230) which are somewhat different from the original ones.

FLIPSOD has also been compared with the well–respected LP–based, iterative OD matrix estimation heuristic of Sherali et al. (2003). Both methods were applied to the well-known "Corridor Problem" of Gur et al. (1980) that does possess congestion affects.The Corridor Problem has 12 nodes, eight arcs, multiple paths between certain OD pairs and multiple user–equilibrium assignments. For each arc, constant capacity and congestion-free travel cost parameters are given. Preliminary link counts are also provided and these enable preliminary arc costs and consequently, OD path costs to be calculated using (2.14). These costs are calculated in subsequent iterations of the methods by inputting resulting arc flow counts into expression (2.14).

Sherali et al. (2003) tested their method, denoted by SA (TT), on 12 input datasets that arise from all combinations of three different sets of available link counts and four different trip tables. In our case, only three of these trip tables are of interest, since one of them (the "no information trip table") contains data that is not within the given, fuzzy ranges of the best available estimates — an assumption of FLIPSOD.

The three trip tables that were used as parts of separate input datasets are: (i) the "correct trip table" (CTT), which produces a user–equilibrium traffic assignment; (ii) the "alternative user–equilibrium trip table" (ATT), which leads to a different user–equilibrium traffic assignment; and (iii) the "small error trip table" (SETT), which was constructed from small variations to the entries of CTT, in other words, a "noisy" version of CTT. The numerical instance of the Corridor Problem has counts available for all of its arcs. For the purposes of testing, following Sherali et al. (2003), we used three different versions of the availability of link counts: (i) link counts in all the 18 arcs of the network, (ii) link counts in 67% of the arcs of the network and (iii) link counts in 50% of the arcs of the network.

For the use of FLIPSOD, it is assumed that the entries in the three given OD target matrices and for the link counts were inexact and could vary by up to ±20% from the given entries as fuzzy numbers. However, the given entries are still considered to be the best available estimates. FLIPSOD is designed to produce for each given OD matrix entry or arc count, an outputted value that is as close as possible to the given central estimate. When applying FLIPSOD to ATT with 50% of the link counts, we obtained 11 different traffic assignments. Their corresponding OD matrix estimates are shown in Figure 1 and they are numbered $1, 2, \ldots, 11$. As we move from output 1 to output 2, . . . , and progressively to output 11, the corresponding values of λ decrease and their assignments progress from reflecting the given link counts to approaching user equilibrium. As the upper bound z_U on the total user cost is progressively reduced from an initial value of 817005.8 to a final value of 675677.2, the value of λ diminishes from 1.0000 to a final value of 0.0009. The latter value corresponds to a user equilibrium assignment with the given link counts and OD matrix inputs matched exactly.

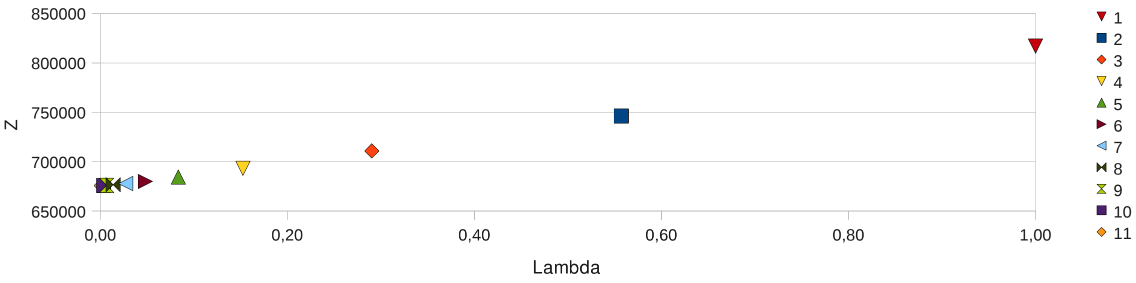

Figure 1: The FLIPSOD solution pool for ATT.

At the end of every cycle, FLIPSOD outputs an OD matrix and an assignment that reflects the travel demand of the matrix. In the course of its operation, FLIPSOD produces a spectrum of matrices and their related assignments. The spectrum has one extreme that is a matrix and an assignment that has flows that are as close as possible to the central estimates of link counts, etc. The other extreme of the spectrum is a matrix and an assignment that is as close as possible to user equilibrium. For all the three trip tables evaluated (CTT, ATT and SETT), FLIPSOD produced matrices within the spectrum just mentioned above. The OD matrix found for CTT and one of the matrices found for SETT are the same as the ones found by the SA (TT) method. Like SA (TT), FLIPSOD found all the matrices in reasonable computational time.

In order to compare the outputs of the two methods, two statistical measures, the percentage root mean square error and the percentage mean absolute error were used. For CTT, FLIPSOD reproduced the same results as those achieved by the SA (TT) method.That is, FLIPSOD matches the given link counts and identifies CTT for all three levels of link count availability. FLIPSOD also found the matrix that corresponds to a user equilibrium assignment for ATT. Outcomes for ATT with SA (TT) were not reported by Sherali et al. (2003). The results of FLIPSOD and SA (TT) for the SETT trip table can be seen in the Table 1.

Table 1: Results obtained by SA (TT) and FLIPSOD with $SETT$.

	SA (TT)				FLIPSOD			
TCA[1]	%RMSE (f_α)	%MAE (f_α)	%RMSE (OD)	%MAE (OD)	%RMSE (f_α)	%MAE (f_α)	%RMSE (OD)	%MAE (OD)
50%	0	0	13.77	8.7	0.54	0.42	13.91	8.29
67%	0	0	13.42	8.18	0.32	0.25	13.82	8.15
100%	0	0	0	0	0.33	0.25	13.81	8.14

[1] Traffic counts availability.

The main difference between the two methods occurs when there are 100% of link counts available. In that situation, SA (TT) finds the matrix that corresponds to a user equilibrium assignment and FLIPSOD produces a matrix with results close to those obtained with the 50% and 67% of link counts available. In the cases of 50% and 67% of link count availability the results found by both methods are very close to each other. However, the results obtained by SA (TT) are slightly better. The difference between the results found by the two methods is at most 0.4% ($\%MAE(OD)$ with 50% of link counts available). For the other statistical measure, the differences are even smaller.

5 Conclusions and Summary

We have reviewed issues concerned with the estimation of OD matrices in congested urban traffic networks when the input data is incomplete and imprecise. We have presented an iterative linear estimation approach, called FLIPSOD, that utilizes the theory of fuzzy sets in order to deal with the imprecision and incompleteness of the given input estimates. Sometimes a user–equilibrium assignment that reflects the given input data does not exist. In this case FLIPSOD has the useful feature that it provides a range of traffic assignments and their corresponding OD matrix estimates, reflecting the spectrum within the range between insistence on the best estimates within fuzzy limits and a user–equilibrium assignment. Computational experience with the model compares favourably with that of a well–respected method. We believe that it will become a useful tool for traffic planners. The authors are currently testing FLIPSOD on large scale Brazilian city networks and comparing the results with those produced by existing methods.

References

Bellman, R. E. and Zadeh, L. A. (1970). Decision-making in a fuzzy environment. *Management Science*, 17(4):B141–B164.

Biletska, K., Masson, M.-H., Midenet, S., and Denaeux, T. (2009). Short-time od matrix estimation for a complex junction using fuzzy-timed high-level petri nets. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th Internat. IEEE Conference on*, pages 1–6.

Brenninger-Göthe, M., Jörnsten, K. O., and Lundgren, J. T. (1989). Estimation of origin-destination matrices from traffic counts using multiobjective programming formulations. *Transportation Research Part B: Methodological*, 23(4):257–269.

Bromage, E. J. (1991). The highway emulator model – release 4.0. Technical report, Central Transportation Planning Staff, Boston, Massachusets.

Gur, Y. J., Turnquist, M., Schneider, M., Leblanc, L., and Kurth, D. (1980). *Estimation of an origin-destination trip table based on observed link volumes and turning movements*. Dept. of Transportation, Federal Highway Administration, [Office of Research], Traffic Systems Division.

Jradi, W. A. R., do Nascimente, H. A. D., Longo, H., and Hall, B. (2009). Uma Arquitetura de Software Interativo para Apoio à Decisão na Modelagem e Análise do Tráfego Urbano. In *Anais do XLI SBPO*, pages 1324–1331.

Klir, G. J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic – theory and applications*. Prentice Hall, New Jersy, USA.

Lai, J. J. and Hwang, C. L. (1992). *Fuzzy mathematical programming. Lecture Notes in Economic and Mathematical Systems*. Springer–Verlag, Berlin, Germany.

Liao, H. L. and Wang, H. F. (1996). Fuzzy resolution on the infeasibility of user equilibrium in traffic assignment problem. In *Proceedings of the 1996 Asian Fuzzy Systems Symposium on Soft Computing in Intelligent Systems and Information Processing*, pages 103–108, Kenting, Taiwan.

Nguyen, S. (1984). *Transportation Planning Models*, chapter Estimation of Origin–destination Matrices from Observed Flows, pages 363–380. Elsevier Science Publishers.

Sherali, H. D., Narayanan, A., and Sivanandan, R. (2003). Estimation of origin-destination trip-tables based on a partial set of traffic link volumes. *Transportation Research Part B: Methodological*, 37(9):815–836.

Van Zuylen, H. J. and Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3):281–293.

Viti, F. (2008). State-of-art of O-D Matrix Estimation Problems based on traffic counts and its inverse Network Location Problem: perspectives for application and future developments. Technical report, Centre of Industrial Engineering, Katholieke Universiteit Leuve, Belgium. Working paper prepared for the BMW project. `http://www.mech.kuleuven.be/cib/verkeer/dwn/pub/P2008F.pdf`.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36):352–362.

Zadeh, L. (1965). Fuzzy sets. *Information Control*, 8:338–353.

Zimmermann, H.-J. (1983). Fuzzy mathematical programming. *Computers & OR*, 10(4):291–298.