



PERGAMON

Transportation Research Part B 37 (2003) 405–422

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

Origin-based algorithms for combined travel forecasting models

Hillel Bar-Gera ^{a,*}, David Boyce ^b

^a *Department of Industrial Engineering and Management, Ben-Gurion University, POB 653, Be'er Sheva 84105, Israel*

^b *Department of Civil and Materials Engineering, University of Illinois at Chicago, IL 60607, USA*

Received 9 April 2001; received in revised form 3 March 2002; accepted 19 March 2002

Abstract

Consistent transportation forecasting models that combine travel demand and network assignment are receiving more attention in recent years. A fixed point formulation for the general combined model is presented. Measures for solution accuracy are discussed. An origin-based algorithm for solving combined models is proposed. Experimental results demonstrate the efficiency of the algorithm in comparison with prevailing alternatives.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Travel forecasting; Combined models; Origin-based assignment; Fixed point

1. Introduction

When planning improvements to transportation systems, various alternatives are considered. Careful evaluation requires forecasts of travel patterns for each alternative. Travel patterns are the result of many choices. Traditional modeling practice considers these choices as a sequential process with unique order: activity location choice (trip generation), joint choice of origin and destination (trip distribution), mode choice and finally route choice (assignment). Despite its intuitive appeal, justification for this order is not as trivial as it may seem.

Travelers usually do not think about modes and routes until they have chosen a destination. In many cases this is simply because they have a fairly good idea about the route of choice and its properties, and even more so about the mode of choice and its properties, for most origins and

* Corresponding author. Tel.: +972-8-6461398; fax: +972-8-6472958.

E-mail address: bargera@bgumail.bgu.ac.il (H. Bar-Gera).

destinations under consideration, prior to choosing their activities. To a certain extent this is true even for choices of work place or residential location, whether made simultaneously or sequentially in one order or another. In view of these observations it seems odd and perhaps even inappropriate to ask which choice comes first.

If all of the conditions that could affect travel choices are known in advance, the order of modeling the different choices should not matter. However, a basic assumption in most forecasting models is that travelers' choices are affected by the level of service of the transportation system. On the other hand, this level of service, and particularly travel times on the roadway network, depend upon the prevailing travel pattern and the associated congestion. The fact that the travel pattern depends on the level of service, which in turn depends on the travel pattern, is one of the main challenges of transportation modeling.

The need to consider congestion effects on route choice became apparent fairly early in the development of travel forecasting models. Early attempts included various computational procedures like quantal loading, origin by origin loading, etc. In recent years user-equilibrium models have gradually replaced previous computational procedures. In these models behavioral assumptions are translated into mathematical conditions that need to be satisfied by the model solution. These well-defined conditions allow one to evaluate approximate solutions, and to examine the convergence of various algorithms.

In contrast to the development and penetration of user-equilibrium route choice models, travel forecasts are still based by and large on sequential procedures. Sequential procedures, even if they are based on user-equilibrium route choice model, still suffer from inconsistent consideration of travel times and congestion effects in the various steps of the procedure. The inconsistent consideration of congestion is a well known and often debated flaw of traditional sequential computational procedures. This flaw was a key issue in the San Francisco Bay Area lawsuit (Garrett and Wachs, 1996). A common remedy for this flaw is to introduce a "feedback" mechanism into the computational procedure, much like quantal loading in its different forms provides a "feedback" mechanism in computational procedures for network assignment. An alternative approach is to state the behavioral assumptions, translate them into mathematical conditions, and seek solutions that satisfy these conditions. Such models are referred to as *combined* or *integrated* models.

The authors believe that whenever possible models must be formulated mathematically. The first goal of this paper is to demonstrate that for most models used in practice a mathematical formulation requires less effort than generally believed, and that there are important benefits to mathematical formulations which are not always appreciated.

Models that combine several travel choices together are far from new. The first mathematical formulation of user-equilibrium assignment by Beckmann et al. (1956) assumed in addition that the flow of travelers between every pair of origin and destination (O–D) is a function of the level of service for that O–D pair. Their convex optimization formulation was later extended to take substitution effects into account by the introduction of origin and/or destination constraints (Evans, 1976). Evans was also the first to present an efficient convergent algorithm for solving this model. Other convex optimization models include the multimode model of Boyce et al. (1983); the multimode, multiclass model of Lam and Huang (1992); and the multimode, multiclass model of Boyce and Bar-Gera (2001). Convex optimization formulations have the advantage of unique solutions and algorithms that are proven to converge. More general combined models were formulated as variational inequalities (VI) by Dafermos (1982), Florian et al. (2002) and others.

Algorithms for combined models are mostly link-based, similar to Evans (1976), with the exception of the route-based algorithm of Lundgren and Patriksson (1998).

In this paper we present a fixed point formulation for combined models, which is relatively simple yet quite general. A general algorithm for combined models is presented, which is based on the origin-based assignment algorithm of Bar-Gera (2002). Special attention is given to models that can be formulated as convex optimization problems, as they have unique solutions. Proof of convergence of the proposed algorithm for convex models is provided, and several measures for solution accuracy are discussed. Computational examples are used to demonstrate the effectiveness of the proposed algorithm in achieving accurate solutions, as compared with the algorithm of Evans and with a “feedback” procedure.

The remainder of the paper is organized as follows. Section 2 presents the general fixed point formulation for combined models. Convex models are discussed in Section 3, including the description of the example solved in Section 7. A brief review of the origin-based assignment algorithm is given in Section 4. The proposed algorithm for combined models is presented in Section 5. Measures for solution accuracy are discussed in Section 6. Computational results are presented in Section 7. Conclusions and suggestions for future research are presented in Section 8.

2. Fixed point formulations of combined models

This section presents a mathematical formulation for the general combined model. Mathematical formulations are important tools for describing the goal of a computational process. Setting the goal is a crucial step that must come prior to any consideration of computational procedures, such as the popular “feedback” mechanism. Only with a clearly stated goal can anyone judge whether a certain procedure performs well or not. Fortunately, combined models can be formulated as fixed point problems in a way that is relatively intuitive and minimal in notation.

Consider a *study area* which is divided into *zones*, during a certain time period of the day in a given year. Let Z denote the set of all zones. For every pair of origin $p \in Z$ and destination $q \in Z$ let d_{pq} denote the O–D flow (persons/h) from p to q . \mathbf{d} is the array of O–D flows. Flows are averaged over the entire modeling period (e.g., the morning peak) and over all work days during a specified year. The time period should not be too long, so that flows within it are fairly steady and reasonably represented by their average. Flows can be estimates for past years, or expected values for future years. In any case, it is important to note that as expected/average values, flows can be fractional and do not need to be integers.

The set of available routes from origin p to destination q is denoted by R_{pq} , and the set of these sets is $\mathbf{R} = \{R_{pq}\}_{p,q \in Z}$. The distribution of travelers from p to q among the routes in R_{pq} is described by a vector of non-negative route proportions (conditional probabilities) $\gamma_{pq} = \{\gamma_{pqr}\}_{r \in R_{pq}}$. γ is the array of route proportion vectors. Route proportions must add up to one for each O–D pair, hence the set of all feasible route proportion arrays is

$$\Gamma(\mathbf{R}) = \left\{ \gamma \in [0, 1]^{|\mathbf{R}|} : \sum_{r \in R_{pq}} \gamma_{pqr} = 1 \quad \forall p, q \in Z \right\} \quad (1)$$

Given γ_{pq} , the implied vector of route flows is $\mathbf{h}_{pq} = \{h_{pqr}\}_{r \in R_{pq}} = d_{pq} \cdot \gamma_{pq}$. The array of route flow vectors is denoted by \mathbf{h} .

In working with these arrays of vectors it is convenient to consider two types of products. The dot product is interpreted as the sum of the product of the elements, similar to a vector dot product, that is $\mathbf{x} \cdot \mathbf{y} = \sum_{pqr} x_{pqr} \cdot y_{pqr}$. The cross product is interpreted as a dot product of array elements, one by one. That is $\mathbf{z} = \mathbf{x} \times \mathbf{y}$ means $z_{pq} = x_{pq} \cdot y_{pq}$. As the algebraic product of matrices is not used in this paper, there should not be any confusion with this notation. Using these conventions, the relationship between O–D flows, route proportions and route flows can be written in short form as $\mathbf{h}(\mathbf{d}, \gamma) = \mathbf{d} \times \gamma$.

According to the user-equilibrium principle of Wardrop (1952), each traveler seeks to minimize the cost associated with their chosen route; therefore, at equilibrium the cost of every used route cannot be greater than the cost of any alternative route. The term cost is interpreted as a general measure of dis-utility, which incorporates travel time. Let $\mathbf{c} = \{\{c_{pqr}\}_{r \in R_{pq}}\}_{p,q \in Z}$ be the array of route cost vectors, which is a continuous function of the travel pattern, $\mathbf{c} = \mathbf{C}(\mathbf{h})$. The set of routes of minimum cost for a given O–D pair p, q is denoted by $R_{pq}^*(\mathbf{c}) = \operatorname{argmin}\{c_{pqr} : r \in R_{pq}\}$. The array of such sets is denoted by $\mathbf{R}^*(\mathbf{c})$. For any non-empty subset of routes \mathbf{R}' ; $\emptyset \subsetneq R'_{pq} \subseteq R_{pq}$, define the set of feasible route proportion arrays that are limited to \mathbf{R}' as

$$\Gamma(\mathbf{R}') = \{\gamma \in \Gamma(\mathbf{R}) : \gamma_{pqr} = 0 \quad \forall r \notin R'_{pq} \quad \forall p, q \in Z\} \tag{2}$$

In particular the set of minimum cost assignments is $\Gamma(\mathbf{R}^*(\mathbf{c}))$. It is obvious without any derivation that the travel pattern $\{\mathbf{d}, \gamma\}$ satisfies the user-equilibrium requirements iff

$$\gamma \in F_1(\mathbf{d}, \gamma) = \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma)))) \tag{3}$$

In other words, user-equilibrium route proportions must belong to the set of feasible route proportions that are limited to the set of minimum cost routes, where route costs correspond to route flows that result from the chosen route proportions.

We assume that the array of O–D flows is a continuous upper-bounded function of O–D costs, $\mathbf{d} = \Phi(\mathbf{u})$, where $\mathbf{u} = \{u_{pq}\}_{p,q \in Z}$ is the array of O–D costs. O–D costs equal average route costs, weighted by flow, $\bar{U}_{pq}(\mathbf{c}, \gamma) = \gamma_{pq} \cdot \mathbf{c}_{pq}$, or $\bar{\mathbf{U}}(\mathbf{c}, \gamma) = \gamma \times \mathbf{c}$. The fixed point formulation of the combined model is

$$\{\mathbf{d}, \gamma\} \in F_2(\mathbf{d}, \gamma) = \{\Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma))), \gamma)\} \times \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma)))) \tag{4}$$

or equivalently

$$\mathbf{d} = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma))), \gamma) \tag{5}$$

$$\gamma \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h}(\mathbf{d}, \gamma)))) \tag{6}$$

These equations state that at equilibrium O–D flows must correspond to prevailing O–D costs, and at the same time the user-equilibrium conditions must be satisfied.

For user-equilibrium solutions $\bar{U}_{pq}(\mathbf{c}, \gamma) = U_{pq}^*(\mathbf{c}_{pq}) \equiv \min\{c_{pqr} : r \in R_{pq}\}$. Therefore, in the above formulation, we can replace $\bar{\mathbf{U}}$ with \mathbf{U}^* and obtain an equivalent formulation. The importance of using average O–D costs is discussed in Section 5.

This formulation can be easily extended to multimode and multiclass models, by adding a mode subscript m and a class superscript l to all variables. In other words if we let $\mathbf{d} = \{d_{mpq}^l\}$; $\mathbf{R} = \{R_{mpq}^l\}$; $\gamma = \{\gamma_{mpqr}^l\}$; $\mathbf{h} = \{h_{mpqr}^l\}$; $\mathbf{c} = \{c_{mpqr}^l\}$, and adapt the interpretation of \mathbf{R}^* , Γ , Φ , \mathbf{C} , and

$\bar{\mathbf{U}}$ accordingly, then Eq. (4) is a mathematical formulation of a generic multimode, multiclass model.

Solution existence is demonstrated by Kakutani's extension to Brouwer's fixed point theorem (Kakutani, 1941; Nikaido, 1968, Theorem 4.4, p. 67). Nikaido defines a set-valued mapping $f: X \rightarrow 2^Y$, where 2^Y represents the set of subsets of Y , to be closed if $x^k \rightarrow x; y^k \rightarrow y; y^k \in f(x^k)$ implies $y \in f(x)$. Every continuous function is closed; hence $\Phi, \bar{\mathbf{U}}, \mathbf{C}$ are closed. \mathbf{R}^* is closed, with discrete topology on \mathbf{R} , and Γ is also closed. Therefore, F_2 is closed. $\Gamma(\mathbf{R}')$ is convex for every set of routes \mathbf{R}' , hence $F_2(\mathbf{d}, \gamma)$ is convex for every (\mathbf{d}, γ) . Due to the upper bound on O–D flows, M , the set of feasible solutions, $[0, M]^{|Z| \times |Z|} \times [0, 1]^{|R|}$ is non-empty, compact and convex. Under these conditions, Kakutani's extension to Brouwer's fixed point theorem guarantees that the map F_2 has a fixed point. In other words, there is at least one solution for the combined model in Eq. (4). In the most general case there may be more than one solution, a problem that is discussed in the next section.

3. Convex formulations of combined models

In this section we consider a limited family of combined models that can be formulated as convex optimization problems, with the advantages of unique solutions and algorithms that are proved to converge. In principle, it is legitimate to use a computational procedure that is not proven to converge, as long as the resulting solutions are sufficiently accurate. However, in practice, it may be difficult and expensive to change the computational procedure. The risk of having to do so should be considered with caution.

Simple mathematical problems, like quadratic equations, can have multiple solutions. In some situations any solution is good enough, but this is not the case in transportation modeling. Imagine two scenarios A and B , where each of the resulting models have several solutions, $A_i; i = 1, \dots, n; B_j; j = 1, \dots, m$. Can we find all solutions? If we use one solution for each scenario, can we tell whether the differences we find are truly because of the differences in the scenarios, or perhaps are they specific to the chosen solutions A_i and B_j ? We believe that in many cases it is better to accept the limitations of convex optimization models in representing certain features of the transportation system, in order to avoid the risks of ill behaved problems and/or algorithms.

Uniqueness and convergence have been demonstrated for a wider family of models, formulated as monotonic VI (e.g. Dafermos, 1982). Wynter (2002) demonstrated uniqueness and convergence under weaker conditions of *nested* monotonicity. The additional theoretical flexibility of (nested) monotonic VI models may be used to relax the assumptions needed by convex optimization formulations, and thus make the model more realistic. However, to the best of our knowledge, examples of more realistic assumptions of practical significance that make use of this additional flexibility have not been presented thus far.

For the discussion of convex optimization models we need to consider the network structure of the transportation system. Let N be the set of nodes and A the set of directed links (roadway segments). A (simple) route segment is a sequence of (distinct) nodes $[v_1, \dots, v_k]$ such that $[v_l, v_{l+1}] \in A \forall 1 \leq l \leq k - 1$. In particular, the route segment $[i, j]$ is the link from node i to node j . (We assume that there is only one link, if any, between every pair of nodes, and that there are no links from a node to itself.) The first node of route segment r is considered its *tail* and denoted by

r_t , and the last node is considered the route's *head* denoted by r_h . In particular by definition $a \equiv [a_t, a_h]$ for every link $a \in A$. The statement $a \subseteq r$ means link a is part of route r . Aggregating route flows through a link over all destinations results in origin-based link flows

$$f_{ap}(\mathbf{h}) = \sum_{q \in Z} \sum_{r \in R_{pq}; a \subseteq r} h_{pqr} \tag{7}$$

Further aggregating those over all origins result in total link flows

$$f_{a\bullet}(\mathbf{h}) = \sum_{p \in Z} f_{ap} = \sum_{r \in R; a \subseteq r} h_{pqr} \tag{8}$$

It is common to assume that route costs are additive over the links, i.e. $c_{pqr} = \sum_{a \in r} t_a$, where $\mathbf{t} = \{t_a\}_{a \in A}$ is the link cost vector. It is also common to assume that link costs depend only on total link flows $\mathbf{t} = \mathbf{t}(\mathbf{f}\bullet)$.

In order to formulate the traffic assignment problem as an optimization problem the cost function must be symmetric, i.e., $\partial C_{pqr} / \partial h_{p'q'r'} = \partial C_{p'q'r'} / \partial h_{pqr} \forall r, r' \in R, \forall p, p', q, q' \in Z$. We make a stronger assumption, that link costs are separable, that is $t_a(\mathbf{f}\bullet) = t_a(f_{a\bullet})$. To ensure uniqueness of the equilibrium link flows we assume that link cost functions are monotonically increasing. These assumptions are satisfied by the most commonly applied BPR-type functions with

$$t_a(f_a) = t_a^0 \left(1 + 0.15 \left(\frac{f_a}{k_a} \right)^4 \right) \tag{9}$$

where k_a is the link capacity, as well as by many variants of the BPR function.

Under these conditions it is well known (Patriksson, 1994) that the route flow array \mathbf{h} satisfies the user-equilibrium conditions iff it minimizes the following objective function:

$$T(\mathbf{h}) = \int_0^{\mathbf{h}} \mathbf{C}(\mathbf{h}') d\mathbf{h}' = \sum_{a \in A} \int_0^{f_{a\bullet}(\mathbf{h})} t_a(x) dx \tag{10}$$

Suppose that the O–D flow model can be represented by a convex optimization problem of the following form:

$$\Phi(\mathbf{u}) = \operatorname{argmin}\{\mathbf{u} \cdot \mathbf{d} + E(\mathbf{d}) : \mathbf{d} \in D\} \tag{11}$$

where E is a continuously differentiable strictly convex function and D is a convex set of feasible O–D flows. The doubly constrained gravity model with negative exponential deterrence used by Evans (1976) belongs to this family when we choose

$$E(\mathbf{d}) = \frac{1}{\mu} \sum_{p,q \in Z} d_{pq} \cdot (\ln(d_{pq}) - 1) \tag{12}$$

$$D = \left\{ \mathbf{d} \in \mathbb{R}_+^{Z \times Z} : \sum_{q \in Z} d_{pq} = \bar{d}_{p\bullet}; \sum_{p \in Z} d_{pq} = \bar{d}_{\bullet q} \right\} \tag{13}$$

where $\bar{d}_{p\bullet}, \bar{d}_{\bullet q}$ are the given totals for origins and destinations respectively. The resulting O–D flows have the LOGIT form

$$d_{pq} = A_p \cdot B_q \cdot \exp(-\mu u_{pq}) \tag{14}$$

where A_p and B_q are balancing factors that correspond to the Lagrange multipliers of the constraints on total flow from an origin and on total flow to a destination.

The combined model objective function is $G(\mathbf{h}) = T(\mathbf{h}) + E(\mathbf{d}(\mathbf{h}))$ and the convex formulation is

$$\{\mathbf{d}, \boldsymbol{\gamma}\} \in \operatorname{argmin}\{G(\mathbf{d} \times \boldsymbol{\gamma}) : \mathbf{d} \in D; \boldsymbol{\gamma} \in \Gamma(\mathbf{R})\} \quad (15)$$

Computational examples in this paper use a more realistic multimodal model, which is similar to the model presented in Boyce and Daskin (1997). The main inputs to this model are: the flow of person trips per hour from each origin $\bar{d}_{p\bullet}$; the flow of person trips per hour to each destination $\bar{d}_{\bullet q}$; free flow travel times tt_a^0 , capacities k_a , and lengths l_a for each link on the roadway network; parking costs pc_z and walking times to or from the parking place wt_z for each zone; in vehicle travel times c_{tpq}^{ivtt} , fares c_{tpq}^{fare} , and out of vehicle times c_{tpq}^{ovt} when traveling by transit from origin p to destination q (these are fixed regardless of flows); and truck flows d_{pq}^{truck} by O–D in passenger cars equivalents.

Link travel time functions are of the BPR form

$$tt_a(f_a) = tt_a^0(1 + 0.15(f_a/k_a)^4) \quad (16)$$

Auto operating costs, including gasoline consumption, are a linear function of link length and travel time,

$$oc_a = \eta_1 tt_a(f_a) + \eta_2 l_a \quad (17)$$

Link generalized costs are

$$t_a(f_a) = \beta_{\text{time}}^{\text{au}} tt_a(f_a) + \beta_{\text{cost}}^{\text{au}} oc_a(f_a) \quad (18)$$

where the β 's are calibration parameters. Parking costs and walking times are components of the additional auto costs, defined as

$$ac_{apq} = \beta_{\text{park}}^{\text{au}} \frac{pc_p + pc_q}{2} + \beta_{\text{walk}}^{\text{au}} (wt_p + wt_q) \quad (19)$$

Route generalized costs by auto are $c_{apqr} = ac_{apq} + \sum_{a \in r} t_a$. O–D generalized costs by auto are $\bar{U}_{apq}(\mathbf{c}, \boldsymbol{\gamma}) = \gamma_{apq} \cdot \mathbf{c}_{apq}$, as before. O–D generalized costs by transit are

$$u_{tpq} = \beta_{\text{bias}}^{\text{tr}} + \beta_{\text{ivtt}}^{\text{tr}} \cdot c_{tpq}^{\text{ivtt}} + \beta_{\text{fare}}^{\text{tr}} \cdot c_{tpq}^{\text{fare}} + \beta_{\text{ovt}}^{\text{tr}} \cdot c_{tpq}^{\text{ovt}} \quad (20)$$

O–D flows are of the LOGIT form

$$d_{apq} = A_p \cdot B_q \cdot \exp(-\mu u_{apq}) \quad (21)$$

$$d_{tpq} = A_p \cdot B_q \cdot \exp(-\mu u_{tpq}) \quad (22)$$

Flows of person trips by auto are converted to vehicle flows by a predetermined auto occupancy factor, aof. The same route proportions are used for auto flows and for truck flows, hence

$$\mathbf{h}_{pq}(\mathbf{d}, \boldsymbol{\gamma}) = (d_{apq}/\text{aof} + d_{pq}^{\text{truck}})\boldsymbol{\gamma}_{pq} \quad (23)$$

The fixed point formulation in Eq. (4) applies to this model almost directly, except that the definition of $\mathbf{h}_{pq}(\mathbf{d}, \boldsymbol{\gamma})$ mentioned above is slightly different than before.

A convex formulation requires some additional modifications. The cost function is symmetric, hence $T(\mathbf{h})$ can be defined similar to Eq. (10), with the addition of a constant term for the additional costs ac_{apq} defined in Eq. (19).

$$T(\mathbf{h}) = \int_0^{\mathbf{h}} \mathbf{C}(\mathbf{h}') d\mathbf{h}' = \sum_{a \in A} \int_0^{f_{a \bullet}(\mathbf{h})} t_a(x) dx + \sum_{p, q \in Z} d_{apq} ac_{apq} \quad (24)$$

The set of feasible O–D flows is

$$D = \left\{ \mathbf{d} \in \mathbb{R}_+^{Z \times Z} : \sum_{q \in Z} (d_{apq} + d_{tpq}) = \bar{d}_{p \bullet}; \sum_{p \in Z} (d_{apq} + d_{tpq}) = \bar{d}_{\bullet q} \right\} \quad (25)$$

O–D flows are obtained by Eq. (11) when $E(\mathbf{d}) = (1/\mu) \sum_{p, q \in Z, m=a, t} d_{mpq} (\ln(d_{mpq}) - 1)$. The different definition of $\mathbf{h}_{pq}(\mathbf{d}, \gamma)$ requires the introduction of an auto occupancy factor into the combined model objective function, $G(\mathbf{h}) = \text{aof} \cdot T(\mathbf{h}) + E(\mathbf{d}(\mathbf{h}))$. Further details of the derivation are given in Boyce and Daskin (1997).

4. The origin-based assignment algorithm

This section presents a brief overview of the origin-based assignment algorithm of Bar-Gera (2002). Derivations in this section as well as the following one are necessarily more opaque, given space limits. The main solution variables in this algorithm are *origin-based approach proportions* $\alpha = \{\alpha_{ap}\}_{a \in A, p \in Z}$; $0 \leq \alpha_{ap} \leq 1$; $\sum_{a: a_h=j} \alpha_{ap} = 1 \forall j \in N, \forall p \in Z$. For every origin an a-cyclic restricting subnetwork is chosen, $A_p \subseteq A$; $a \notin A_p \Rightarrow \alpha_{ap} = 0$. Initial restricting subnetworks are trees of minimum cost routes. To update the restricting subnetwork, unused links are removed, v_i —the maximum cost to node i within the restricting subnetwork is computed, and all links $[i, j]$ such that $v_i < v_j$ are added to the restricting subnetwork. Approach proportions for origin p are updated by shifting flows within the restricting subnetwork A_p according to a boundary (piece-wise linear) search in a direction determined by an approximate second order method.

Route proportions are determined by $\gamma_{pqr} = \prod_{a \in r} \alpha_{ap}$. It can be shown that $f_{ap} = \alpha_{ap} \cdot g_{jp}$, where $g_{jp} = \sum_{a: a_h=j} f_{ap}$ is the origin-based node flow, demonstrating that α_{ap} is indeed the proportion of flow on approach a to node a_h for origin p . The availability of route proportions allows one to compute average O–D costs, as well as the assignment of new O–D flows by current route proportions. Due to the restriction to a-cyclic subnetworks, these computations can be done efficiently without route enumeration, in a time that is a linear function of the number of links times the number of origins.

Given the demand \mathbf{d} and the current solution α , the set of restricting subnetworks for the next iteration is defined by the function $\mathbf{A} = \mathcal{A}(\mathbf{d}, \alpha)$, and the next iteration solution is defined by the map $\alpha' \in \Theta^z(\mathbf{d}, \alpha)$. The following properties of \mathcal{A} and Θ^z are used to prove convergence.

Lemma 1. *If $\alpha' \in \Theta^z(\mathbf{d}, \alpha)$ then $T(\mathbf{h}(\mathbf{d}, \alpha')) \leq T(\mathbf{h}(\mathbf{d}, \alpha))$. Equality may hold but only if $\mathbf{h}(\mathbf{d}, \alpha') = \mathbf{h}(\mathbf{d}, \alpha)$ (Bar-Gera, 2002; Lemma 8).*

Lemma 2. For a sequence with fixed restricting subnetworks, the map Θ^α is closed. That is if $\alpha^k \rightarrow \alpha^*$; $\hat{\alpha}^k \rightarrow \hat{\alpha}^*$; $\mathcal{A}(\mathbf{d}, \alpha^k) = \mathbf{A}$; $\hat{\alpha}^k \in \Theta^\alpha(\mathbf{d}, \alpha^k)$; then $\hat{\alpha}^* \in \Theta^\alpha(\mathbf{d}, \alpha^*)$ (Bar-Gera, 2002; Lemma 7,8).

Lemma 3. Given $\alpha^{k+1} \in \Theta^\alpha(\mathbf{d}, \alpha^k)$ and a subsequence K such that $\mathcal{A}(\mathbf{d}, \alpha^{k+l}) = \mathbf{A}^{*l} \forall k \in K, \forall 1 \leq l \leq |N|$; $\alpha^{k+l} \rightarrow \alpha^{*l} \forall 1 \leq l \leq |N|$; $\mathbf{h}(\mathbf{d}, \alpha^{*l}) = \mathbf{h}^* \forall 1 \leq l \leq |N|$, then every limit point of $\{\alpha^k\}$ satisfies the user-equilibrium conditions (Bar-Gera, 2002; part of theorem 2).

In Lemmas 2 and 3 the fixed demand can be replaced with a converging sequence $\mathbf{d}^k \rightarrow \mathbf{d}^*$, and the same proofs hold. For further details about the algorithm, see Bar-Gera (1999).

5. An origin-based algorithm for combined models

In this section we present the proposed origin-based algorithm for combined models. The discussion focuses on convex optimization models. A generalization for fixed point models is also discussed. The general scheme of the algorithm is presented in Fig. 1. Stopping conditions for the algorithm are discussed in Section 6. As shown in the figure, the main addition to the assignment algorithm is a procedure to update O–D flows, while keeping route proportions fixed. Given a current solution, $\{\mathbf{d}^k, \gamma^k\}$, subproblem O–D flows are determined according to average O–D costs $\hat{\mathbf{d}}^k = \Phi(\gamma^k \times \mathbf{C}(\mathbf{d}^k \times \gamma^k))$. New O–D flows are obtained by a weighted average $\mathbf{d}_\lambda^{k+1} = (1 - \lambda) \cdot \mathbf{d}^k + \lambda \cdot \hat{\mathbf{d}}^k$, where $0 \leq \lambda \leq 1$ is a chosen step size. For brevity let $\mathbf{h}^k = \mathbf{d}^k \times \gamma^k$; $\hat{\mathbf{h}}^k = \hat{\mathbf{d}}^k \times \gamma^k$; $\mathbf{h}_\lambda^{k+1} = \mathbf{d}_\lambda^{k+1} \times \gamma^k = (1 - \lambda) \cdot \mathbf{h}^k + \lambda \cdot \hat{\mathbf{h}}^k$; $\mathbf{c}^k = \mathbf{C}(\mathbf{h}^k)$; $\mathbf{u}^k = \gamma^k \times \mathbf{c}^k$.

Initialization:

```

Let  $\mathbf{u} = \mathbf{U}^*(\mathbf{C}(0))$ 
Let  $\mathbf{d}^0 = \Phi(\mathbf{u})$ 
for  $p$  in  $Z$  do
   $A_p$  = tree of minimum cost routes from  $p$ 
   $\mathbf{f}_p$  = all or nothing assignment using  $A_p$ 
end for

```

Main loop:

```

for n=1 to number of main iterations
  Update O-D flows, retain route proportions
  for  $p$  in  $Z$  do
    update restricting subnetwork  $A_p$ 
    update origin-based approach proportions  $\alpha_p$ 
  end for
  for m=1 to number of inner iterations
    for  $p$  in  $Z$  do
      update origin-based approach proportions  $\alpha_p$ 
    end for
  end for
end for

```

Fig. 1. An origin-based algorithm for combined models.

Lemma 4. $(\hat{\mathbf{h}}^k - \mathbf{h}^k)$ is a descent direction of G , equality holds only if $\hat{\mathbf{d}}^k = \mathbf{d}^k$.

Proof.

$$\nabla_{\mathbf{h}}G(\mathbf{h}^k)(\hat{\mathbf{h}}^k - \mathbf{h}^k) = \mathbf{c}^k(\hat{\mathbf{h}}^k - \mathbf{h}^k) + \nabla_{\mathbf{d}}E(\mathbf{d}^k)(\hat{\mathbf{d}}^k - \mathbf{d}^k) = \mathbf{u}^k(\hat{\mathbf{d}}^k - \mathbf{d}^k) + \nabla_{\mathbf{d}}E(\mathbf{d}^k)(\hat{\mathbf{d}}^k - \mathbf{d}^k) \quad (26)$$

The convexity of E implies that

$$\nabla_{\mathbf{d}}E(\mathbf{d}^k)(\hat{\mathbf{d}}^k - \mathbf{d}^k) \leq E(\hat{\mathbf{d}}^k) - E(\mathbf{d}^k) \quad (27)$$

The choice of $\hat{\mathbf{d}}^k$ and the convex formulation of the O–D flow model in Eq. (11) implies that

$$\mathbf{u}^k\hat{\mathbf{d}}^k + E(\hat{\mathbf{d}}^k) \leq \mathbf{u}^k\mathbf{d}^k + E(\mathbf{d}^k) \quad (28)$$

Substituting Eqs. (27) and (28) into Eq. (26) shows that indeed

$$\nabla_{\mathbf{h}}G(\mathbf{h}^k)(\hat{\mathbf{h}}^k - \mathbf{h}^k) \leq 0 \quad (29)$$

Furthermore, by the strict convexity of E equality holds only if $\hat{\mathbf{d}}^k = \mathbf{d}^k$. \square

Notice that if minimum O–D costs were used instead of average O–D costs, descent could not be guaranteed. This is demonstrated by the following simple example. Consider a system with one O–D pair, two independent routes with linear cost functions $C_i(h_i) = 10 + h_i/100$; $i = 1, 2$ and a linear O–D flow function $\phi(u) = 2500 - 100u$; $E(d) = d^2/200 - 25d$. It is easy to see that the equilibrium solution is $d = 1000$; $h_1 = h_2 = 500$; $c_1 = c_2 = 15$. Suppose the current solution is $h_1 = 1000$; $h_2 = 0$, or equivalently $d = 1000$; $\gamma_1 = 1$; $\gamma_2 = 0$. Route costs are $c_1 = 20$; $c_2 = 10$, minimum cost is 10 while the average cost weighted by the flow is 20. As a result $\hat{d}^{\text{MIN}} = \phi(10) = 1500$; $\hat{d}^{\text{AVE}} = \phi(20) = 500$. Since we keep the current route proportions, the corresponding route flows are: $\hat{h}_1^{\text{MIN}} = 1500$; $\hat{h}_1^{\text{AVE}} = 500$; $\hat{h}_2^{\text{MIN}} = \hat{h}_2^{\text{AVE}} = 0$. The directional derivative of G in the average cost case is $-2500 < 0$, while in the minimum cost case it is $2500 > 0$.

Lemma 4 and the convexity of G ensures that $G(\mathbf{h}_\lambda^{k+1}) \leq G(\mathbf{h}^k)$ for every step size λ such that

$$\nabla_{\mathbf{h}}G(\mathbf{h}_\lambda^{k+1})(\hat{\mathbf{h}}^k - \mathbf{h}^k) \leq 0 \quad (30)$$

The algorithm chooses the maximal step size $\lambda = 2^{-k}$; $k = 0, 1, 2, \dots$ that satisfies the condition in Eq. (30). In order to obtain a closed algorithmic map, a more flexible step size choice is permitted,

$$\Lambda(\mathbf{h}^k, \hat{\mathbf{h}}^k) = \left\{ \lambda \in [0, 1] : \begin{array}{l} \nabla_{\mathbf{h}}G(\mathbf{h}_\lambda^{k+1}) \cdot (\hat{\mathbf{h}}^k - \mathbf{h}^k) \leq 0 \\ \nabla_{\mathbf{h}}G(\mathbf{h}_{2\lambda}^{k+1}) \cdot (\hat{\mathbf{h}}^k - \mathbf{h}^k) \geq 0 \text{ or } 2\lambda \geq 1 \end{array} \right\} \quad (31)$$

The map for O–D flow update, $\Theta^{\mathbf{d}}$, is defined as follows:

$$\Theta^{\mathbf{d}}(\mathbf{d}^k, \boldsymbol{\gamma}^k) = \left\{ \mathbf{d} = (1 - \lambda)\mathbf{d}^k + \lambda\hat{\mathbf{d}}^k : \begin{array}{l} \hat{\mathbf{d}}^k = \Phi(\boldsymbol{\gamma}^k \times \mathbf{c}(\mathbf{d}^k \times \boldsymbol{\gamma}^k)) \\ \lambda \in \Lambda(\mathbf{d}^k \times \boldsymbol{\gamma}^k, \hat{\mathbf{d}}^k \times \boldsymbol{\gamma}^k) \end{array} \right\} \quad (32)$$

Lemma 5. $\Theta^{\mathbf{d}}$ is a closed map.

Proof. Φ is continuous and Λ is closed hence $\Theta^{\mathbf{d}}$ is closed. \square

Theorem 1. *The algorithm in Fig. 1 converges to the optimal solution of the combined convex formulation model in Eq. (15).*

Proof. There is a subsequence K such that $\mathcal{A}(\mathbf{d}^{k+l}, \boldsymbol{\alpha}^{k+l}) = \mathbf{A}^{*l} \forall k \in K, \forall 1 \leq l \leq |N|$; $\mathbf{d}^{k+l} \rightarrow \mathbf{d}^{*l} \forall 1 \leq l \leq |N|$; $\boldsymbol{\alpha}^{k+l} \rightarrow \boldsymbol{\alpha}^{*l} \forall 1 \leq l \leq |N|$. By Lemma 5 $\mathbf{d}^{*l+1} \in \Theta^{\mathbf{d}}(\mathbf{d}^{*l}, \boldsymbol{\alpha}^{*l})$ and by Lemma 2 $\boldsymbol{\alpha}^{*l+1} \in \Theta^{\boldsymbol{\alpha}}(\mathbf{d}^{*l+1}, \boldsymbol{\alpha}^{*l})$. By Lemmas 1 and 4 the objective function is a monotonically non-increasing bounded sequence, hence $T(\mathbf{d}^{*l}, \boldsymbol{\alpha}^{*l}) = T^*$; furthermore, $\mathbf{d}^{*l+1} = \mathbf{d}^{*l} = \mathbf{d}^*$ and $\mathbf{h}(\mathbf{d}^{*l+1}, \boldsymbol{\alpha}^{*l+1}) = \mathbf{h}(\mathbf{d}^{*l}, \boldsymbol{\alpha}^{*l}) = \mathbf{h}^*$. By Lemma 3, every limit point of $\{\boldsymbol{\alpha}^k\}$ satisfies the user-equilibrium conditions. Therefore $\gamma(\boldsymbol{\alpha}^{*l}) \times \mathbf{C}(\mathbf{h}^*) = \mathbf{u}^*$, and $\mathbf{d}^* = \Phi(\mathbf{u}^*)$. \square

The algorithm presented above can be easily extended to any general fixed point combined model. As discussed in Bar-Gera (2002), the computations in the assignment algorithm require nothing more than the ability to compute link costs and the ability to estimate link cost derivatives. Therefore, the assignment algorithm can be used as a computational procedure for problems with general cost structure, even without convex formulation, only that in such cases convergence cannot be proven. In order to develop a computational procedure for a general fixed point combined model, the only change needed in the algorithm presented above is the choice of a step size in the procedure for updating O–D flows. A typical choice in the case where a convex optimization formulation is lacking is $\lambda = 1/k$, where k is the main iteration index. The questions of convergence and performance of the resulting algorithm remain a subject for future research.

6. Accuracy measures

One of the main advantages of mathematically formulated models is the ability to evaluate solutions by well defined accuracy measures, and hence to determine whether a solution is sufficiently accurate for the specific analysis under consideration.

A common measure for solution accuracy in convex optimization problems is the gap between objective function value and its lower bound. As shown in Evans (1976), with any feasible solution (\mathbf{d}, γ) ; $\mathbf{h} = \mathbf{d} \times \gamma$, there is an associated lower bound,

$$\text{LB} = E(\mathbf{d}') + T(\mathbf{h}) + \nabla T(\mathbf{h}) \cdot (\mathbf{h}' - \mathbf{h}) \tag{33}$$

where $\mathbf{d}' = \Phi(\mathbf{U}^*(\mathbf{C}(\mathbf{h})))$; $\gamma' \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h})))$; and $\mathbf{h}' = \mathbf{d}' \times \gamma'$. In iterative methods a tighter estimate of the gap can be obtained by comparing the lowest objective function value (LOF), which is always from the last iteration, with the highest lower bound (HLB), which may be from previous iterations. In the following we consider the relative gap, defined as $\text{RG} = (\text{LOF} - \text{HLB})/|\text{HLB}|$.

Gap and relative gap are useful as global measures of accuracy; however, they are not easy to interpret intuitively. Therefore it is not easy to choose a gap level that guarantees sufficient accuracy. As an alternative we propose to consider separate measures of accuracy for O–D flows and for assignment.

In the case of O–D flows it is natural to compare O–D flows in the current solution \mathbf{d} , with the O–D flows that result from the costs of travel under current conditions. For the latter we can choose either minimum O–D costs $\mathbf{d}' = \Phi(\mathbf{U}^*(\mathbf{C}(\mathbf{h})))$ or average O–D costs $\mathbf{d}'' = \Phi(\bar{\mathbf{U}}(\mathbf{C}(\mathbf{h})), \gamma)$. Both comparisons lead to similar results. We shall use \mathbf{d}' simply because average O–D costs are

not available for some algorithms. Possible aggregate measures of accuracy are the maximum positive difference, $\max\{d'_{pq} - d_{pq} : d'_{pq} \geq d_{pq}\}$, the maximum negative difference, $\max\{d_{pq} - d'_{pq} : d'_{pq} \leq d_{pq}\}$, and the total misplaced O–D flow, $\sum_{p,q \in Z} |d'_{pq} - d_{pq}|$. All units are person trips/hour.

The intuitive interpretation of these measures can be very helpful in setting conditions for sufficiently accurate solutions. For example, consider a study that examines the impact of a new commercial facility, which is expected to attract 1000 trips/h during the morning peak. It would be reassuring to know that the total misplaced O–D flow in the solution is less than 100 trips/h. A total misplaced O–D flow of 1000 trips/h may still be acceptable, assuming that it is spread over a wide region. But, a total misplaced O–D flow of 10,000 trips/h is probably not acceptable, as it is quite likely to have significant influence on the results of the study.

Assignment accuracy measures can be based on the distribution of *excess cost*, $ec_r = c_{pqr} - u_{pq}^*(\mathbf{c})$, among used routes. In particular we shall use the *average excess cost*, $AEC_a = (1/d_{\bullet\bullet}) \sum_{r \in R} h_r ec_r$, where $d_{\bullet\bullet} = \sum_{p,q \in Z} d_{pq}$ is the total O–D flow (on the road network). This is equivalent to a normalized gap of the fixed demand problem.

Setting requirements for assignment accuracy is more challenging, since typically the goal is to make sure that link flows are sufficiently close to the true equilibrium solution. A case study (Boyce and Bar-Gera, 2002) examined the impact of adding a pair of freeway ramps in the Delaware Valley Region. The goal of the study was to estimate flow differences on links in the vicinity of the proposed improvement between the build and no-build scenarios. It was found that solutions should have average excess cost less than 0.001 vehicle-min, so that estimates for freeway links shall be within 3% from the true equilibrium solution, and estimates for arterials shall be within 10% from the true equilibrium solution. As additional case studies are conducted on different networks and for various levels of congestion, more definite recommendations will be available for practitioners.

A solution is considered to be sufficiently accurate only if it satisfies both conditions, that is if it has average excess cost less than, say, 0.001 vehicle-min, and total misplaced O–D flow less than, say, 1000 trips/h. Notice that these measures of accuracy are not limited to convex optimization problems, and can be used in any general combined model.

7. Experimental results

This section presents computation results comparing the convergence of the proposed origin-based algorithm with the convergence of the algorithm of Evans, and with a feedback procedure. In every iteration of the Evans algorithm, a step size is chosen by Eq. (31) to average route flows of the current solution (\mathbf{d}, γ) ; $\mathbf{h} = \mathbf{d} \times \gamma$, with those of a subproblem solution $\mathbf{d}' = \Phi(\mathbf{U}^*(\mathbf{C}(\mathbf{h})))$; $\gamma' \in \Gamma(\mathbf{R}^*(\mathbf{C}(\mathbf{h})))$; $\mathbf{h}' = \mathbf{d}' \times \gamma'$. In every iteration of the feedback procedure, the same subproblem O–D flows \mathbf{d}' are assigned to the road network by Frank–Wolfe user-equilibrium assignment (LeBlanc et al., 1975) to an average excess cost of 1 min equivalents, or 30 iterations, whichever comes first. Current solution and subproblem solution are averaged by a step size of $1/k$, where k is the feedback iteration counter. All codes are written in C. All three procedures were solved on a Compaq Alpha Unix Server model DS20E, with CPU speed of 666 MHz, and 256 MB RAM.

The algorithms were applied to two test problems; both follow the structure of the multimodal model described in Section 3. One test problem is based on the detailed Chicago Regional



Fig. 2. Chicago Regional Road Network.

Network, with 1790 zones, 12,982 nodes, 39,018 road links, presented in Fig. 2. The model is for 1 h during the morning peak with total O–D flow of about 1.4 million person trips per hour. The second test problem is based on an aggregated version of the previous network, covering a slightly smaller area, which is referred to as the Chicago Sketch Network. In this case there are 387 zones, 933 nodes, 2950 road links, and total O–D flow of about 1.25 million person trips per hour. Both models were calibrated using household travel survey data, and validated using Census data.

Figs. 3–7 show the convergence of the various measures as a function of CPU time for the Chicago Regional Network. All figures demonstrate the superior performance of the origin-based algorithm, and the inferior performance of the feedback procedure. The first solution produced by the feedback procedure, after 25 min of CPU time, has a relative gap of 120%, average excess cost of 117 min equivalents, and misplaced O–D flow of 2.5 million trips/h. In the same amount of

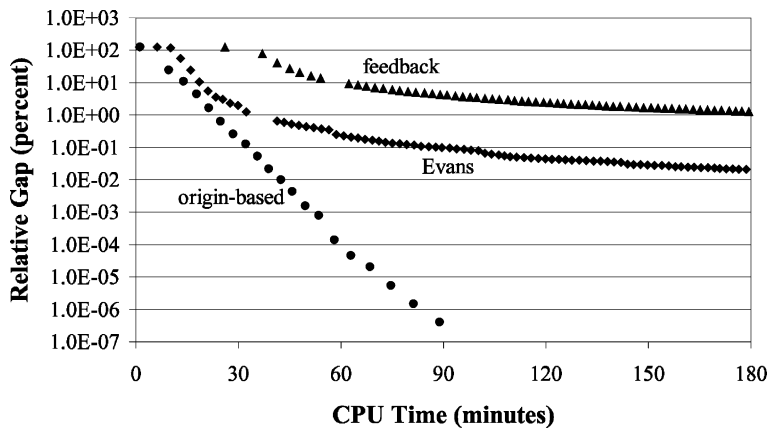


Fig. 3. Overall solution convergence for the Chicago Regional Network.

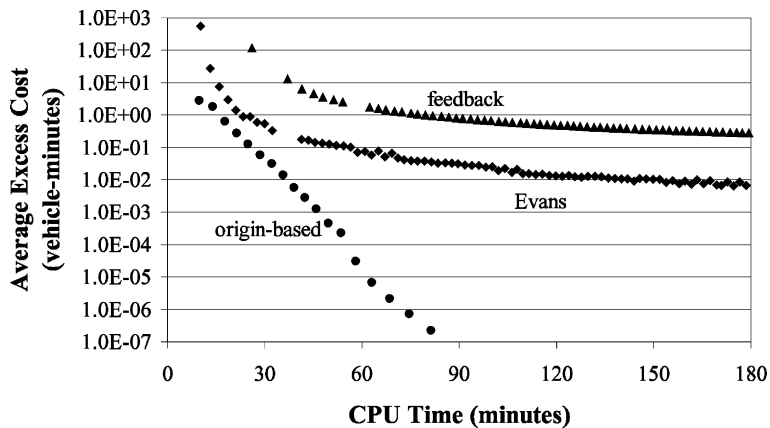


Fig. 4. Assignment convergence for the Chicago Regional Network.

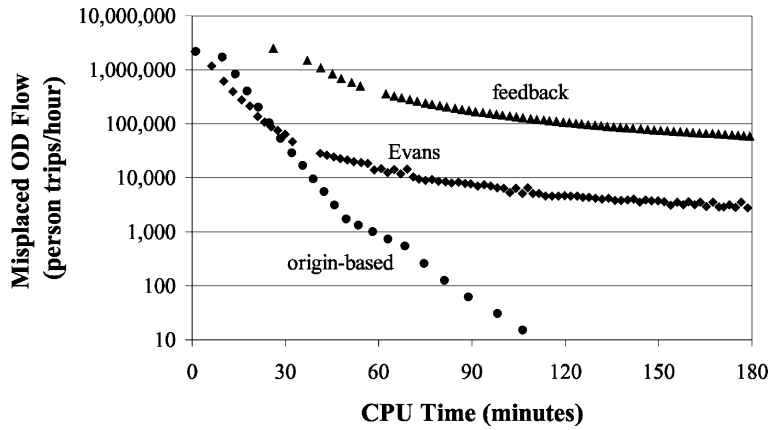


Fig. 5. O–D flow convergence for the Chicago Regional Network.

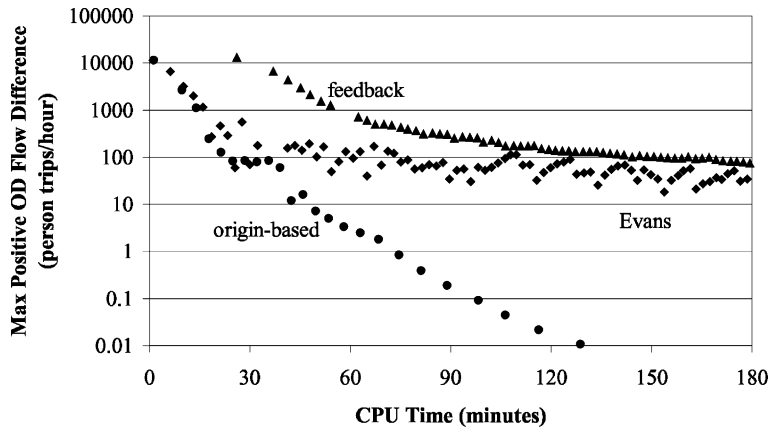


Fig. 6. Maximum positive O–D flow difference for the Chicago Regional Network.

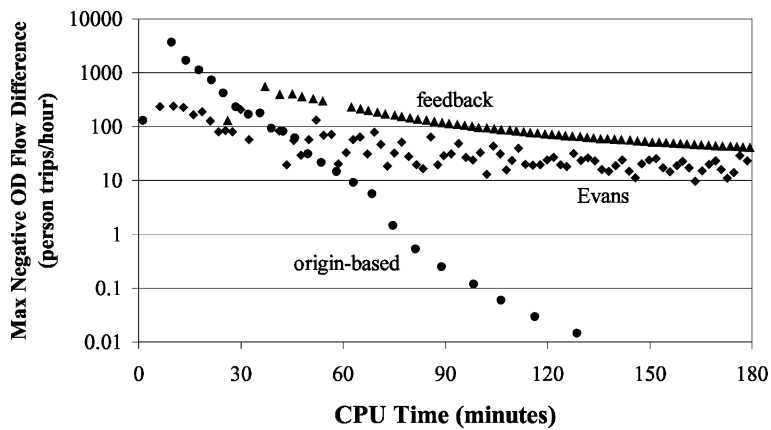


Fig. 7. Maximum negative O–D flow difference for the Chicago Regional Network.

CPU time, the Evans algorithm reaches relative gap of 3.1%, average excess cost of 0.9 min equivalents, and misplaced O–D flow of 87,000 trips/h; while the origin-based algorithm reaches relative gap of 0.6%, average excess cost of 0.12 min equivalents, and misplaced O–D flow of 103,000 trips/h.

The proposed assignment accuracy condition, average excess cost of 0.001 vehicle-min, is met by the origin-based algorithm after 50 min. After 200 iterations and more than 6 h of CPU time, the Evans algorithm produces a solution with average excess cost of 0.002 vehicle-min, that still does not satisfy the requirement. After 200 iterations and more than 7 h of CPU time, the feedback procedure produces a solution with average excess cost of 0.1 vehicle-min which is far from satisfying the requirement.

The proposed condition for O–D flow accuracy, total misplaced O–D flow of 1000 trips/h or less, is met by the origin-based algorithm after 58 min of CPU time. After 200 iterations and 6 h of CPU time the algorithm of Evans only approaches this accuracy, with total misplaced O–D flow of 1300 trips/h. As for the feedback procedure, even after 7 h of CPU time and 200 iterations the resulting solution still suffers from total misplaced O–D flow of more than 19,000. It seems that such solution is not accurate enough in any type of analysis.

The origin-based algorithm stops after about 2 h of CPU time when it reaches computational accuracy limits, which is due to the accuracy of origins and destinations constraints ($1E-9$) in the demand model.

Similar trends were observed for the Chicago Sketch Network. The origin-based algorithm produces solutions with average excess cost better than 0.001 vehicle-min in less than 2 min, and solutions with total misplaced O–D flow less than 1000 trips/h in 2.7 min. The origin-based algorithm reaches computational accuracy limits after 5 min of CPU time. After 200 iterations and 21 min of CPU time the Evans algorithm produced a solution with average excess cost of 0.0025 vehicle-min, and with total misplaced O–D flow of 1315 person trips per hour. This solution could be accepted for analyses that are relatively insensitive. After 200 iterations and 23 min of CPU time the feedback procedure produced a solution with average excess cost of 0.08 vehicle-min, and with total misplaced O–D flow of 14,208 person trips per hour. Such solution is probably not acceptable in any type of analysis.

8. Conclusions and future research

Traditional travel forecasting methods were based on sequential computational procedures. The need for integrated or combined models is becoming more and more convincing, in view of court decisions and legislative mandates in the US in the last decade. The fixed point formulation presented in this paper seems to be a natural tool to formulate general combined models mathematically, including most models used in practice. The need for intuitive accuracy measures leads to separate consideration of assignment accuracy and the accuracy of O–D flows. Experimental results demonstrate that solutions that may seem well converged by general measures like relative gap, may still exhibit substantial inaccuracy in terms of O–D flows. The results show that algorithm choice can have substantial impact on computation time, especially when accurate solutions are desired. In particular, we found that the proposed origin-based algorithm is more efficient

than the algorithm of Evans, and much more efficient than one commonly applied feedback procedure. Other feedback procedures based on link speeds are even less satisfactory.

The promising results in this paper suggest that similar algorithms are likely to be efficient in solving models with other structures, including models that do not have convex optimization formulation. These remain subjects for future research.

Acknowledgements

The authors wish to thank the Chicago Area Transportation Study, Chicago, IL. for providing the Chicago Network Data. Financial support was kindly provided by the Kreitman Foundation, Ben-Gurion University of the Negev, Beer-Sheva, Israel, and by the National Science Foundation through the National Institute of Statistical Sciences, Research Triangle Park, NC.

References

- Bar-Gera, H., 1999. Origin-based Algorithms for Transportation Network Modeling. Ph.D. thesis, University of Illinois at Chicago, USA.
- Bar-Gera, H., 2002. Origin-based Algorithm for the Traffic Assignment Problem, *Transportation Science*, in press.
- Beckmann, M., McGuire, C.B., Winston, C.B., 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
- Boyce, D., Bar-Gera, H., 2001. Network Equilibrium Models of Travel Choices with Multiple Classes. In: Lahr, M.L., Miller, R.E. (Eds.), *Regional Science in Economic Analysis*. Elsevier Science, Oxford, UK, pp. 85–98, Chapter 6.
- Boyce, D., Bar-Gera, H., 2002. Convergence of Traffic Assignments: How Much Is Enough? The Delaware Valley Region Case Study. Working paper, Department of Civil and Materials Engineering. University of Illinois at Chicago, Chicago, IL, USA.
- Boyce, D., Chon, K.S., Lee, Y.J., Lin, K.T., LeBlanc, L.J., 1983. Implementation and evaluation of combined models of location, destination, mode and route choice. *Environment and Planning A* 15, 1219–1230.
- Boyce, D., Daskin, M., 1997. Urban Transportation. In: ReVelle, C., McGarity, A.E. (Eds.), *Design and Operation of Civil and Environmental Engineering Systems*. John Wiley and Sons, New York, pp. 277–341, Chapter 7.
- Dafermos, S., 1982. The general multimodal network equilibrium problem with elastic demand. *Networks* 12, 57–72.
- Evans, S.P., 1976. Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment. *Transportation Research* 10, 37–57.
- Florian, M., Wu, J.H., He, S., 2002. A Multi-Class Multi-Mode Variable Demand Network Equilibrium Model with Hierarchical Logit Structures. In: Gendreau, M., Marcotte, P. (Eds.), *Transportation and Network Analysis: Current Trends*, Kluwer Academic Publishers, Dordrecht, pp. 119–133, Chapter 8.
- Garrett, M., Wachs, M., 1996. *Transportation Planning on Trial: The Clean Air Act and Travel Forecasting*. Sage Publications, Thousand Oaks, CA, USA.
- Kakutani, S., 1941. A Generalization of Brouwer's Fixed Point Theorem. *Duke Math. J.* 8 (3).
- Lam, W.H.K., Huang, H.J., 1992. A combined trip distribution and assignment model for multiple user classes. *Transportation Research* 26B, 275–287.
- LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9, 309–318.
- Lundgren, J.T., Patriksson, M., 1998. The combined distribution and stochastic assignment problem. *Annals of Operations Research* 82, 309–329.
- Nikaido, H., 1968. *Convex Structures and Economic Theory*. Academic Press, New York.

Patriksson, M., 1994. *The Traffic Assignment Problem—Models and Methods*. VSP, Utrecht, Netherlands.

Wardrop, J.G., 1952. Some Theoretical Aspects of Road Traffic Research. *Proceedings of the Institution of Civil Engineers, Part II* 1, 325–378.

Wynter, L., 2002. A Convergent Algorithm for the Multimodal Traffic Equilibrium Problem. INRIA Research Report 4125, Institut National de Recherche en Informatique, Le Chesnay Cedex, France.