A maximum entropy-least squares estimator for elastic origin-destination trip matrix estimation

Chi Xie[a, *], Kara M. Kockelman[a] and S. Travis Waller[a]

[a]*Center for Transportation Research, Department of Civil, Architectural and Environmental Engineering*
The University of Texas at Austin, Austin, TX 78712, United States

Abstract

It is well known that the origin-destination (O-D) flow table of a subnetwork is not only determined by trip generation and distribution, but also by traffic routing and diversion, due to the existence of internal-external, external-internal and external-external flows. This result indicates the variable nature of subnetwork O-D flows. This paper discusses an elastic O-D flow table estimation problem for subnetwork analysis. The underlying assumption is that each cell of the subnetwork O-D flow table contains an elastic demand function rather than a fixed demand rate and the demand function can capture all traffic diversion effect under various network changes. We propose a combined maximum entropy-least squares (ME-LS) estimator, by which O-D flows are distributed over the subnetwork so as to maximize the trip distribution entropy, while demand function parameters are estimated for achieving the least sum of squared estimation errors. While the estimator is powered by the classic convex combination algorithm, computational difficulties emerge within the algorithm implementation until we incorporate partial optimality conditions and a column generation procedure into the algorithmic framework. Comparison results from applying the combined estimator to a couple of numerical examples show that an elastic O-D flow table reflects network flow changes on a significantly improved level than its fixed counterpart when used as input for subnetwork flow evaluations.

Keywords: Origin-destination trip table; elastic demand; maximum entropy; least squares; subnetwork analysis; convex combination; unconstrained optimization; column generation

1. Introduction

In the transportation field, one almost always evaluates and analyzes network flows at an abstract or subnetwork level. For example, a state highway network is a subnetwork of the national highway network, an urban transportation network is a subnetwork of its surrounding regional network, and an urban transit network is a subnetwork of the whole urban transportation system. Here, by subnetwork, we mean a local, closely connected part of the global network, which shares some common spatial or physical features. In practice, a subnetwork may be identified by using some geographical boundary and/or a hierarchy of the existing transportation system, depending on the required modeling extent and level of realism. Extraction of a subnetwork from a full network enables us to confine the network analysis to the local area of interest without dealing with the marginal effect on the global scale.

While it is straightforward to identify and form a subnetwork (see, for example, Chan, 1976, 1989; Eash, 1983; Kaplan, 1984; Taylor, 1988; Rogus, 1996), extracting the corresponding demand pattern or distribution from its full-

---

[*] Corresponding author. Tel.: +1 (512) 471-4622; fax: +1 (512) 475-8744.
E-mail address: chi.xie@mail.utexas.edu.

size counterpart and ensuring that the result of a network flow evaluation conducted in the subnetwork will be consistent with the full-size result are far more difficult tasks (Dowling and May, 1985; Zhou et al., 2006). This is especially apparent and critical when the evaluation outcome from the subnetwork will be reviewed and further analyzed at the global level. Such demand extraction requirement poses a subnetwork origin-destination (O-D) trip matrix estimation problem. If the subnetwork evaluation outcomes resulting from the estimated O-D trip matrix are reliable, a cost-effective network evaluation strategy may be achieved by decomposing a large-size network into a number of smaller subnetworks, working with subnetworks, and then aggregating their results to approximate the full network's state.

For a user-equilibrium network, the simplistic way to estimate a subnetwork O-D trip matrix may be a direct combination of path flows generated from the full network, a procedure depicted by Haghani and Daskin (1984, 1986) and Hearn (1984). This approach results in a subnetwork trip matrix that can induce a traffic flow pattern in the subnetwork exactly the same as in the full network. However, the subnetwork trip matrix derived by this simple aggregation method depends on the traffic assignment result from the full network. It is well known that in general a unique path flow pattern in a user-equilibrium traffic network does not exist. In this sense, this approach does not result in a unique subnetwork trip matrix.

This non-uniqueness issue may be eliminated by two complementary approaches, both exploiting the maximum entropy (ME) principle (see Wilson, 1970). The first approach aims at estimating a most likely, unique path flow pattern in the full network by an entropy-maximizing user-equilibrium traffic assignment algorithm (Rossi, 1989; Janson, 1993; Larsson et al., 2002; Bar-Gera, 2006) and aggregating corresponding path flows to form a subnetwork trip matrix. The second one only requires the set of arc flow rates generated from the full network, which is used as input for an ME method to directly estimate a most likely, unique O-D flow pattern for the subnetwork (Xie et al., 2010).

However, all of these methods mentioned above merely produce a fixed O-D trip matrix, which may not sufficiently reflect the variable nature of subnetwork O-D flows. Without loss of generality, we may simply categorize O-D trips in a subnetwork into four groups in terms of their departure and arrival locations: 1) internal-internal trips; 2) internal-external trips; 3) external-internal trips; and 4) external-external trips. It is obvious that subnetwork O-D trips, especially those from the last three categories, are largely determined by the traffic routing result over the full network. Any network change, either inside or outside the subnetwork, may change the traffic routing pattern and hence the subnetwork O-D trip distribution. A fixed O-D trip matrix cannot capture this global-local routing connection and inevitably lose a certain level of modeling accuracy when a subnetwork analysis is conducted. An exclusive case is due to Boyles (2010), who recently suggested adding dummy arcs to mimic this global-local connection. The focus of his work is on the estimation of cost function parameters of these added dummy arcs.

In contrast to Boyles' attempt at imitating variable subnetwork O-D flows, which is mainly on the supply side, we suggest a simple variable demand approach. Our approach does not involve any network topology addition, which is often very arbitrary to modelers. Specifically, in this paper, we intend to estimate a subnetwork trip matrix such that each cell of the matrix is an elastic demand function instead of a fixed demand rate. The focus of this paper is on a convex optimization model and its solution method for estimating the demand function parameters.

The remainder of this paper is structured as follows. We first review a variety of O-D trip matrix estimation methods in Section 2, with a concentration on those methods relying on traffic count data. In Section 3, the optimization model for parameter estimation is described in detail, as well as its solution properties analyzed. A convex combination algorithm for problem solutions is then sketched in Section 4, with an elaboration on the problem conversion and decomposition and the design of a column generation procedure. As we will see, the utilization of these algorithmic devices greatly improves the solution efficiency. Finally, we present the results from applying the method to a couple of numerical examples in Section 5 and conclude the paper with some research extensions in Section 6.

2. Existing O-D flow estimation methods

Various O-D trip matrix estimation problems with different modeling rationales and for different application contexts appeared in the literature. The model formulations and solution methods for these problems differ substantively from one another in terms of the theoretical basis (e.g., law of gravity, ME, and least squares (LS)),

traffic assignment principle (e.g., all-or-nothing assignment, user-equilibrium assignment, and proportional assignment), and input data availability (e.g., trip productions/attractions, traffic counts, travel times, target trip matrix, and combination of them).

The most frequently studied trip matrix estimation methods arguably are those relying on traffic counts, either exclusively or partially. Our discussion below is focused on this type of methods. The reason for its popularity is largely due to the wide availability and relatively high quality of traffic count data. For infrastructure management and transportation planning purposes, almost every state department of transportation collects and assembles traffic count data on a regular basis (e.g., nationwide archived traffic count data can be retrieved from the U.S. Highway Performance Monitoring System (HPMS)). Continuous monitoring and collection of traffic counts has become a routine task at many traffic management centers so that the evolution of trip rates and flow conditions can be readily tracked.

It is well known that trip matrix estimation based on arc traffic counts may be regarded as a reverse process to traffic assignment, which estimates arc flow rates based on a trip matrix. Unlike the dominant use and wide acceptance of the user-equilibrium principle (or its variants) used in traffic assignment, trip matrix estimation problems have been formulated by a few different O-D flow distribution rationales, such as ME, LS, maximum likelihood (ML), and so on.

The ME theory for trip matrix estimation based on traffic counts was first introduced by Van Zuylen and Willumsen (1980) and Willumsen (1981). By assuming that the input traffic flow pattern follows a known proportional routing scheme, these models resort to a simple iterative balancing method for solutions. In the same modeling framework, Nguyen (1981) formulated an ME problem which synthesizes both traffic count data and trip production/attraction data. This model seeks to maximally use all available information. However, due to the potential inconsistency between different data sources, this model may not necessarily have a feasible solution. Fisk (1988) imposed the user-equilibrium routing principle on a similar trip matrix estimation problem, resulting in an ME model subject to a variational inequality constraint. The value of her model is more theoretical than practical, however. Due to the nonconvex feasible region, in general, this problem is difficult to solve for its global optimality.

An alternative trip matrix estimation approach incorporating equilibrium traffic flows was pursued by Nguyen (1977, 1984), Turnquist and Gur (1979), LeBlanc and Farhangian (1982) and Yang et al. (1994). This approach uses O-D travel costs, the minimum of competing path travel costs, as input. Given the arc cost functions, arc travel costs and path travel costs can be readily calculated from arc flow rates. The approach makes no assumption on trip distribution; however, the main drawback of this approach is that alternative optimal solutions may exist. To make the problem converge to a unique trip matrix, some extra information, for example, a target trip matrix or the ME assumption is needed.

When a (complete or incomplete) target O-D trip matrix or reference O-D cost matrix is also available, LS is often employed as a standard criterion for trip matrix estimation. Carey et al. (1981) proposed an ordinary LS estimator and McNeil and Hendrickson (1985) extended this work by a generalized least squares (GLS) method. By assuming that the observed traffic flow pattern follows a prespecified proportional assignment, Cascetta (1984), Hendrickson and McNeil (1984) and Bell (1991) also formulated their trip matrix estimation models based on GLS. Yang et al. (1992) developed a GLS estimator that explicitly accommodates a user-equilibrium traffic assignment model, which results in a bi-level, nonconvex optimization problem. In the same framework of GLS, Nie et al. (2005) and Nie and Zhang (2010) further proposed a decoupling method and a relaxation method to convert this bi-level problem into a single-level formulation. Other model estimation or fitting techniques used for trip matrix estimation include the least absolute norm approach (Yang et al., 1991; Sherali et al., 1994; Nie and Lee, 2002) and the integrated square error approach (Gajewski et al., 2002).

The joint use of (full or partial) traffic count and target matrix information has also resulted in other types of trip matrix estimation methods, such as Bayesian inference approach (Maher, 1983; Lo et al., 1996) and ML approach (Bell, 1983, 1985; Spiess, 1987; Lo et al., 1996). Despite various parameter assumptions and optimization principles implied in these methods, their common spirit is to determine a trip matrix that represents some form of trade-offs between a target trip matrix and observed traffic counts. These trade-offs appear in either model constraints or objective functions. Due to the requirement of a target trip table, none of them could be applied to the subnetwork trip matrix estimation in our case.

One way to accommodate variable O-D trip rates for subnetwork analysis is to construct an O-D trip matrix with elastic demands. Few of the studies listed above are concerned with elastic O-D trip matrix estimation except Carey et al. (1981) and McNeil and Hendrickson (1985). These authors assume that O-D trip rates are governed by a direct demand function; instead of estimating trip rates directly, their models aim at estimating parameters of the demand function. Here a similar idea is adopted for capturing the global traffic routing effect on the subnetwork O-D trip matrix and is combined with the ME concept for the construction of a more sophisticated model.

3. A combined ME-LS model

For discussion convenience, let us first define a strongly connected subnetwork $G = (N, A)$, which is part of a larger or full network, where N and A are the node and arc sets of the subnetwork, and define a set of traffic flow patterns on it, corresponding to a set of network scenarios I. Two special subsets of N are an origin node set R and a destination node set S, i.e., $R \subseteq N$ and $S \subseteq N$. Here, by a traffic flow pattern, we mean a complete set of observed (or estimated) scenario-specific arc flow rates, $\mathbf{v}^i = [\hat{v}_a^i]$, $\forall i \in I$, on the given subnetwork. Such a traffic flow pattern represents an observed network flow sample.

The purpose of this research is to develop an optimization-based method for the construction of a function-embedded O-D flow table. Different from the structure of traditional O-D flow tables, each cell of our table contains a demand function instead of a fixed demand rate. This function-based table setting allows O-D flow rates to be adjusted in terms of network changes and resulting congestion levels, when the network flow pattern is evaluated. Needless to say, this adoption of O-D demand functions implicitly suggests that an elastic-demand traffic assignment procedure should be followed for network flow evaluations. This assumption is formally stated as follows.

Assumption 1. (Existence of elastic demand functions) We assume that an elastic demand function exists for each O-D pair of the subnetwork and such a demand function is a function of the travel cost of the associated O-D pair. Although the traffic flow rate of any O-D pair of the subnetwork is a result of the highly complex, network-wide flow routing process in the full network, our simple expectation here is that the O-D travel cost variation can well reflect network changes. By definition, the elastic demand function is in general a decreasing function and may be written in a variety of forms, such as the linear, quadratic, multiplicative, or exponential forms, to name a few. For simplicity, we arbitrarily choose a linear demand function such as: $x_{rs} = b_{rs} + e_{rs}\hat{t}_{rs}$, $\forall r \in R$, $s \in S$, where x_{rs} is the to-be-estimated O-D flow rate, $\hat{t}_{rs}$ is the observed (or estimated) O-D travel cost, b_{rs} is the base O-D flow rate (or the potential maximum O-D flow rate), and e_{rs} is the O-D demand elasticity rate. In most traffic environments, we should expect $b_{rs} > 0$ and $e_{rs} < 0$, given that the O-D flow rate x_{rs} is positive and the elastic demand function has a decreasing cost-flow relationship. ∎

Traffic counts may contain errors or noises due to a variety of measurement or disturbance factors. It potentially brings conflicts with the flow conservation constraint and raises the data inconsistency concern (Van Zuylen and Willumsen, 1980). However, given the prevalent existence of origin and destination nodes in a network, the data inconsistency issue is automatically eliminated.

Assumption 2. (Prevalence of origin and destination nodes) We assume that every node in the subnetwork is potentially an origin and destination node, i.e., $R = N$ and $S = N$. This assumption relaxes the potential flow inconsistency issue within the input set of arc flow rates that share the same head or tail node, and accordingly reduces the modeling and solution complexity. ∎

However, we will suggest an alternative model later in this paper, which can accommodate the more general situation of non-origin and non-destination nodes existing in the subnetwork with input data errors or noises. In the case that any node r cannot be an origin, we just need to set $x_{rs} = 0$, $\forall s \in S$; similarly, if any node s cannot be a destination, we simply set $x_{rs} = 0$, $\forall r \in R$. These following modeling assumptions are used throughout this text unless stated otherwise.

3.1. Model formulation

Given the modeling settings and assumptions listed above, we consider the subnetwork O-D flow estimation problem in terms of two optimization principles simultaneously. The first optimization principle assumes that underlying each given network flow pattern, O-D flows are distributed according to the ME theory. The second principle adopted here is to estimate the parameters of the given elastic demand function of each O-D pair. LS, as a

standard optimization criterion for unbiased model estimation, especially linear regression models, is chosen here as a parameter estimation technique to identify the optimal set of parameter values of the elastic demand function that provides the closest match between the O-D flow rates estimated from the elastic demand function and from the given network flow samples.

The combination of ME and LS results in the following optimization model:

$$\min \ z(\mathbf{f}, \mathbf{b}, \mathbf{e}) = \sum_i \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) + w \sum_i \sum_{rs} \left(d_{rs}(\hat{t}_{rs}^i | b_{rs}, e_{rs}) - x_{rs}^i \right)^2 \tag{1.1}$$

$$\text{subject to} \ \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} = \hat{v}_a^i \qquad\qquad \forall a \in A, i \in I \tag{1.2}$$

$$f_{k,rs}^i \geq 0 \qquad\qquad \forall k \in K_{rs}, r \in R, s \in S, i \in I \tag{1.3}$$

where the O-D flow rate x_{rs}^i is defined as

$$x_{rs}^i = \sum_k f_{k,rs}^i \qquad\qquad \forall r \in R, s \in S, i \in I \tag{1.4}$$

where x_{rs}^i is the flow rate between O-D pair r-s under network scenario i, $f_{k,rs}^i$ is the flow rate on path k connecting O-D pair r-s under network scenario i, $\delta_{a,k}^{rs}$ is the arc-path incidence indicator (i.e., $\delta_{a,k}^{rs} = 1$, if arc a is part of path k connecting O-D pair r-s; $\delta_{a,k}^{rs} = 0$, otherwise), w is a weighting coefficient, and other variables and parameters can be referred to in the previous text of this section.

In the optimization model presented above, the elastic demand function $d_{rs}(\cdot)$ is defined as, following Assumption 1,

$$d_{rs}(\hat{t}_{rs}^i | b_{rs}, e_{rs}) = b_{rs} + e_{rs}\hat{t}_{rs}^i \qquad\qquad \forall r \in R, s \in S \tag{2}$$

where the demand function parameters b_{rs} and e_{rs} are both decision variables of the model.

Here, the O-D travel cost $\hat{t}_{rs}^i$ is assumed to equal the minimum travel cost of all paths between O-D pair r-s,[*] which can be calculated in advance as follows,

$$\hat{t}_{rs}^i = \min_{k \in K_{rs}} \left\{ \sum_a \hat{t}_a^i \delta_{a,k}^{rs} \right\} \qquad\qquad \forall r \in R, s \in S, i \in I \tag{3}$$

where $\hat{t}_a^i$ is the arc travel cost of arc a, whose value is determined by a continuously increasing, convex arc cost function as follows, given the free-flow arc travel cost t_a^0, the arc capacity c_a, and the function parameters α_a and β_a,

$$\hat{t}_a^i = t_a^0 \left(1 + \alpha_a \left(\frac{\hat{v}_a^i}{c_a} \right)^{\beta_a} \right) \qquad\qquad \forall a \in A, i \in I \tag{4}$$

As a summary, the result of solving this optimization problem includes two parts: the optimal O-D flow pattern $\mathbf{x}^{i*} = \{x_{rs}^{i*}\}$ for each network scenario i and the optimal demand function parameters b_{rs}^* and e_{rs}^* for each O-D pair r-s. The model is designed to simultaneously maximize the diversity or entropy associated with each O-D flow pattern (i.e., the first term of the objective function) and minimize the difference of the O-D flow rates estimated from the elastic demand function and from the entropy-maximizing result (i.e., the second term of the objective

---

[*] Use of the minimum path cost as the O-D cost is valid only when the given traffic flow pattern is in the user-equilibrium status, though the user-equilibrium assumption about the given flow pattern is not required by the present model as well as the solution method described in the subsequent section. Other traffic flow patterns could be assumed too. But the O-D cost should be accordingly reevaluated. For example, if the given traffic flow pattern follows a stochastic user-equilibrium principle, then the O-D cost is the expected minimum path cost.

function). The weighting coefficient w plays a role in balancing the two optimization components. Specifically, if $w \to 0$, the optimization model reduces to, for each network scenario i,

$$\min \quad \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) \tag{5.1}$$

$$\text{subject to} \quad \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} = \hat{v}_a^i \qquad\qquad \forall a \in A \tag{5.2}$$

$$f_{k,rs}^i \geq 0 \qquad\qquad \forall k \in K_{rs}, r \in R, s \in S \tag{5.3}$$

where, again, the O-D flow rate x_{rs}^i is,

$$x_{rs}^i = \sum_k f_{k,rs}^i \qquad\qquad \forall r \in R, s \in S \tag{5.4}$$

which is simply a standard ME model for network scenario i. This reduced model for individual network scenarios can be found in Xie et al. (2010). An important feature of this pure ME model is that its O-D flow estimates can be used to produce the exactly same network flow pattern as the given input pattern, if the network is not altered and if the subsequent traffic assignment procedure follows the same routing principle as the one implied by the given input flow pattern (see Property 1 in Xie et al. (2010)). This property, however, is not held by the optimization model in this paper, because the elastic demand function would produce an O-D flow matrix that is more or less different from the one that is estimated directly from the input flow pattern. This O-D flow rate difference is $b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i$, for O-D pair r-s under network scenario i. On the other hand, if $w \to +\infty$, the optimization model becomes,

$$\min \quad \sum_i \sum_{rs} (b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i)^2 \tag{6}$$

subject to   (1.2)-(1.4)

which produces a set of O-D flow patterns simultaneously in terms of the LS criterion only. However, this set of O-D flow patterns does not imply any behavior rationale (because the ME term is removed), so its usefulness is of less interest here.

3.2. Solution properties

Property 1. (Existence of feasible solutions) Given Assumption 2, i.e., every node in the subnetwork is potentially an origin and destination node, there always exists a feasible solution to the original optimization problem given an arbitrary set of positive arc flow rates, $[x_{rs}^i]$, where $x_{rs}^i > 0$, $\forall r \in R, s \in S$, under each network scenario i.

Proof. This property can be proved by a two-stage solution process.

First, we show that there always exists a feasible O-D flow pattern under each network scenario i. Note that a feasible O-D flow pattern is defined by a system of linear equations:
$\{\mathbf{x}^i = [x_{rs}^i]: \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} = \hat{v}_a^i, \forall a, i, f_{k,rs}^i \geq 0, \forall k, r, s, i, x_{rs}^i = \sum_k f_k^{rs}, \forall r, s, i\}$. Its feasibility is equivalent to the feasibility of the reduced linear system of $f_{k,rs}^i$: $\{\mathbf{f}^i = [f_{k,rs}^i]: \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} = \hat{v}_a^i, \forall a, i, f_{k,rs}^i \geq 0, \forall k, r, s, i, \}$. Given that every $f_{k,rs}^i \geq 0$, $\forall k \in K_{rs}, r \in R, s \in S$, where $R = S = N$, exists, following Carey and Revelli (1986), the optimal solution of a quadratic program for each network scenario i,

$$\min \quad \sum_a (y_a^i)^2$$

$$\text{subject to} \quad \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} + y_a^i = \hat{v}_a^i \qquad\qquad \forall a \in A$$

$$f_{k,rs}^i \geq 0 \qquad\qquad\qquad \forall k \in K_{rs}, r \in R, s \in S$$

$$y_a^i \geq 0 \qquad\qquad\qquad \forall a \in A$$

is $y_a^{i*} = 0, \forall a$, where y_a^i is a slack variable for arc a. Thus, the existence of a feasible O-D flow pattern under network scenario i is guaranteed.

Furthermore, after collecting the feasible O-D flow patterns under all network scenarios as inputs, we then only need to consider the second term of the objective function, which is a pure LS problem. Given that b_{rs} and e_{rs}, $\forall r \in R, s \in S$, do not appear in any constraint, this LS problem is actually an unconstrained optimization problem and its feasible solutions always exist. Thus, we produce a feasible solution via this two-stage process and we know that the original optimization problem at least contains this feasible solution. ∎

Property 2. (Equivalence of the optimal solution) The optimality conditions of the ME-LS problem with respect to path flows under each network scenario can be equivalently described as: all used paths (i.e., paths with a positive flow rate) between an O-D pair have their entropy impedance equal to the O-D entropy impedance plus an O-D flow estimation error term; all unused paths (i.e., paths with a zero flow rate) are associated with a path entropy impedance greater than or equal to the O-D entropy impedance plus the error term.

Proof. The optimality conditions of the problem with respect to path flow $f_{k,rs}^i$ can be analyzed through checking the Lagrangian dual of the model formulation. Given $\mathbf{f} = \{\mathbf{f}^i\} = \{[f_{k,rs}^i]\}$, $\mathbf{b} = [b_{rs}]$ and $\mathbf{e} = [e_{rs}]$, the Lagrangian dual after relaxing the flow conservation constraint (1.2) is,

$$L(\mathbf{f}, \mathbf{b}, \mathbf{e}, \boldsymbol{\lambda}) = \sum_i \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) + w \sum_i \sum_{rs} (b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i)^2$$
$$+ \sum_i \sum_a \lambda_a^i \left(\hat{v}_a^i - \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} \right) \tag{7}$$

where $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}^i\} = \{[\lambda_a^i]\}$ is the set of Lagrangian multipliers. Given the first-order derivative of the Lagrangian dual with respect to $f_{k,rs}^i$, i.e.,

$$\frac{\partial L(\mathbf{f}, \mathbf{b}, \mathbf{e}, \boldsymbol{\lambda})}{\partial f_{k,rs}^i} = \ln x_{rs}^i - 2w(b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i) - \sum_a \lambda_a^i \delta_{a,k}^{rs} \tag{8}$$

the optimality conditions with respect to $f_{k,rs}^i$ can be expressed as follows,

$$f_{k,rs}^i = 0 \Rightarrow \ln x_{rs}^i - 2w(b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i) \geq \sum_a \lambda_a^i \delta_{a,k}^{rs} \tag{9.1}$$

$$\ln x_{rs}^i - 2w(b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i) = \sum_a \lambda_a^i \delta_{a,k}^{rs} \Rightarrow f_{k,rs}^i \geq 0 \tag{9.2}$$

Here, we define $-\ln x_{rs}^i$ the minimum path entropy impedance among all paths connecting O-D pair r-s, $-\lambda_a^i$ the entropy impedance of arc a, and $-\sum_a \lambda_a^i \delta_{a,k}^{rs}$ the entropy impedance of path k connecting O-D pair r-s, all under network scenario i. Note that $b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i$ is the O-D flow estimation error by the demand function for O-D pair r-s under network scenario i. So we simply call $2w(b_{rs} + e_{rs}\hat{t}_{rs}^i - x_{rs}^i)$ the scaled estimation error term (which is scaled by $2w$). Given these definitions, now we can claim that the optimality conditions for path flow rate $f_{k,rs}^i$, expressed by Equation (9), are: the sum of arc entropy impedances along any path k connecting O-D pair r-s is greater than or equal to the minimum path entropy impedance of this O-D pair (or the O-D entropy impedance) plus a scaled O-D flow estimation error term; if the sum of arc entropy impedances along some path is greater, then the flow rate along this path is zero. ∎

As for the optimality conditions of the problem with respect to the demand function parameters b_{rs} and e_{rs}, they can be readily evaluated by checking the first-order derivatives of the objective function, since no constraint restricts b_{rs} and e_{rs}:

$$\frac{\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial b_{rs}} = 0 \Rightarrow b_{rs}|I| + e_{rs} \sum_i \hat{t}_{rs}^i - \sum_i x_{rs}^i = 0 \tag{10}$$

$$\frac{\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial e_{rs}} = 0 \Rightarrow b_{rs} \sum_i \hat{t}_{rs}^i + e_{rs} \sum_i (\hat{t}_{rs}^i)^2 - \sum_i x_{rs}^i \hat{t}_{rs}^i = 0 \tag{11}$$

where $|I|$ is the cardinality of the network scenario set I.

The optimality conditions with respect to b_{rs} and e_{rs} do not imply any economic or behavior interpretation. However, this system of linear equations is particularly useful in the solution algorithm development, as we will see later in the next section, because the optimal solution values of b_{rs} and e_{rs} can be analytically determined by solving the linear system, if x_{rs}^i, $\forall r \in R$, $s \in S$, $i \in I$, is given.

Property 3. (Uniqueness of the optimal solution) The ME-LS problem's optimal solution is unique.

Proof. The solution uniqueness holds if both the objective function and the constraint set are convex. The convexity of the objective function is satisfied given that the Hessian matrix of the objective function with respect to decision variables $f_{k,rs}^i$, b_{rs} and e_{rs}, $\nabla^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})$, is positive definite (see the appendix). Plus the linear constraint set is convex, the solution uniqueness is thus proved. ∎

4. A convex combination method integrating optimality condition and column generation

Given the quadratic form of the objective function (in terms of x_{rs}^i, $\ln x_{rs}^i$, b_{rs} and e_{rs}) and the linear constraint set, the convex combination algorithm (Frank and Wolfe, 1956) seems particularly useful here for solving the proposed convex programming problem. The convex combination algorithm is a feasible descent direction method. Starting with a given initial solution, the algorithm iteratively finds a descent direction by deriving the first-order derivatives of the objective function with respect to the decision variables and searches for the optimal solution along this descent direction subject to the problem's constraints (i.e., auxiliary solution). A convex combination of the auxiliary solution and the current solution generates an updated solution, in which the combination coefficient is determined such that the value of the convex combination is optimized. This updated solution then becomes the current solution and the procedure is continuously executed until the solutions at two adjacent steps are sufficiently close.

The major concern with implementing the convex combination algorithm is to derive the auxiliary solution by solving a linearized version of the original optimization problem. This linearized problem in our case can be written as follows:

$$\min \sum_i \sum_{rs} \frac{\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial x_{rs}^i} x_{rs}^i + \sum_{rs} \left(\frac{\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial b_{rs}} b_{rs} + \frac{\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial e_{rs}} e_{rs} \right) \tag{12}$$

subject to (1.2)-(1.4)

However, this linearized problem is not easy to solve due to two algorithmic difficulties. First, note that no constraint bounds b_{rs} and e_{rs}, so the optimal solution values of b_{rs} and e_{rs} from solving this linear problem are $+\infty$ or $-\infty$, depending on the sign of $\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})/\partial b_{rs}$ and $\partial z(\mathbf{f}, \mathbf{b}, \mathbf{e})/\partial e_{rs}$. Without any external bounding rule, this result cannot be used to develop an appropriate move size for updating the current solution in the subsequent step of the convex combination algorithm. Second, the standard linear programming algorithm—the simplex method—may not be directly applied to solving this linear problem, because an explicit statement and manipulation of this linear problem requires an enumeration of path flows over the network, which in general is computationally prohibitive for problems with realistic network size.

In the following, we incorporate the optimality conditions of an unconstrained subproblem of b_{rs} and e_{rs} into the algorithmic procedure, which enables the optimal solution of b_{rs} and e_{rs} to be written in the form of x_{rs}^i and hence

makes the original problem collapse to an equivalent optimization problem of decision variables x_{rs}^i only. On the other hand, we devise a column generation approach for reducing the computational difficulty arising from path enumeration, by which path flows are only generated as and when needed within the solution framework of the revised simplex method (see Dantzig, 1963; Bazaraa et al., 1990). These algorithmic ideas are elaborated below and then a formal description of the implementation of the convex combination method is presented.

4.1. Problem equivalence and decomposition

It is readily known that the original optimization problem can be equivalently evaluated by solving the following optimization problem:

$$\min \quad \sum_i \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) + w \sum_{rs} \min \sum_i (b_{rs} + e_{rs} \hat{t}_{rs}^i - x_{rs}^i)^2 \tag{13}$$

subject to (1.2)-(1.4)

Note that if the solution value of x_{rs}^i is given, the subproblem in the objective function, $\min \sum_i (b_{rs} + e_{rs} \hat{t}_{rs}^i - xrsi2$, is an unconstrained optimization problem since no constraint bounds brs and ers. By solving the system of equations given in Equation (10)-(11), we can derive the optimal values of b_{rs} and e_{rs} for this subproblem (and for the original optimization problem) as follows,

$$b_{rs}^* = \frac{(\sum_i x_{rs}^i \hat{t}_{rs}^i) \sum_i \hat{t}_{rs}^i - (\sum_i x_{rs}^i) \sum_i (\hat{t}_{rs}^i)^2}{(\sum_i \hat{t}_{rs}^i)^2 - |I| \sum_i (\hat{t}_{rs}^i)^2} \tag{14}$$

$$e_{rs}^* = \frac{(\sum_i x_{rs}^i) \sum_i \hat{t}_{rs}^i - |I| \sum_i x_{rs}^i \hat{t}_{rs}^i}{(\sum_i \hat{t}_{rs}^i)^2 - |I| \sum_i (\hat{t}_{rs}^i)^2} \tag{15}$$

Inserting these results back into the objective function, a new equivalent optimization problem (with a more restricted feasible region) is resulted as follows,

$$\min \quad z_e(\mathbf{f}) = \sum_i \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) + w \sum_{rs} \sum_i (b_{rs}^* + e_{rs}^* \hat{t}_{rs}^i - x_{rs}^i)^2 \tag{16}$$

subject to (1.2)-(1.4)

where b_{rs}^* and e_{rs}^* can be referred to in Equation (14)-(15). Now, as we can see, this new equivalent optimization problem in Equation (16) is an optimization problem that only contains decision variable x_{rs}^i (or $f_{k,rs}^i$). In the subsequent discussion, our algorithmic design is all based on this equivalent optimization problem.

Given this equivalent optimization problem, we resort to solving the following linearized problem (instead of the problem given by Equation (12)) in the algorithmic framework of convex combination,

$$\min \quad \sum_i \sum_{rs} \frac{\partial z_e(\mathbf{f})}{\partial x_{rs}^i} x_{rs}^i \tag{17}$$

subject to (1.2)-(1.4)

where the first-order derivative of the equivalent objective function with respect to x_{rs}^i can be calculated as,

$$\frac{\partial z_e(\mathbf{f})}{\partial x_{rs}^i} = \ln x_{rs}^i + 2w \sum_j \left[\frac{\sum_i (\hat{t}_{rs}^i \sum_h \hat{t}_{rs}^h - \sum_h (\hat{t}_{rs}^h)^2 + \hat{t}_{rs}^j \sum_h \hat{t}_{rs}^h - |I| \hat{t}_{rs}^j \hat{t}_{rs}^i) x_{rs}^i}{(\sum_h \hat{t}_{rs}^h)^2 - |I| \sum_h (\hat{t}_{ts}^h)^2} \right.$$
$$\left. - x_{rs}^j \right] \left[\frac{\hat{t}_{rs}^i \sum_h \hat{t}_{rs}^h - \sum_h (\hat{t}_{rs}^h)^2 + \hat{t}_{rs}^j \sum_h \hat{t}_{rs}^h - |I| \hat{t}_{rs}^j \hat{t}_{rs}^i}{(\sum_h \hat{t}_{rs}^h)^2 - |I| \sum_h (\hat{t}_{ts}^h)^2} - \frac{\partial x_{rs}^j}{\partial x_{rs}^i} \right] \tag{18}$$

in which $\partial x_{rs}^j / \partial x_{rs}^i = 1$, if $j = i$; $\partial x_{rs}^j / \partial x_{rs}^i = 0$, if $j \neq i$.

Since the set of constraints for O-D flow rate x_{rs}^i (or path flow rate $f_{k,rs}^i$) under different network scenarios is separate from each other, we can solve the linearized problem for each network scenario separately. In other words, solving the linearized problem is equivalent to solving a set of scenario-specific linear problems, each of which is for a network scenario, and combining the optimal solutions of all these scenario-specific linear problems together. Such a scenario-specific linear problem for the network scenario i is then:

$$\min \quad \sum_{rs} \frac{\partial z_e(\mathbf{f})}{\partial x_{rs}^i} x_{rs}^i \tag{19}$$

subject to   (5.2)-(5.4)

4.2. Column generation

For discussion convenience, we rewrite the scenario-specific linear problem into the following matrix form of path flows:

$$\min \quad \mathbf{c}^{i\top} \mathbf{f}^i \tag{20.1}$$

subject to $\Delta \cdot \mathbf{f}^i = \hat{\mathbf{v}}^i$ (20.2)

$$\mathbf{f}^i \geq \mathbf{0} \tag{20.3}$$

where $\mathbf{c}^i$ is the first-order derivatives given by Equation (18), $\mathbf{c}^i = [z_e(\mathbf{f})/\partial x_{rs}^i]$, and $\mathbf{f}^i$ is the path flow vector, $\mathbf{f}^i = [f_{k,rs}^i]$, Δ is the arc-path incidence matrix, $\Delta = [\delta_{a,k}^{rs}]$, and $\hat{\mathbf{v}}^i$ is the given (measured or estimated) arc flow vector, $\hat{\mathbf{v}}^i = [\hat{v}_a^i]$.

Now assume that the simplex procedure reaches some iteration, where the current basic feasible solution contains $|A|$ basic paths of a positive flow rate, where $|A|$ is the cardinality of the arc set A. The sets of basic paths and nonbasic paths are labeled as P_B and $P_{\bar{B}}$, respectively. We also use $\mathbf{B}$ and $\mathbf{c}_B^i$ to represent the corresponding basis matrix and cost vector, where $\mathbf{B} = [\delta_{a,k}^{rs}]_{|A| \times |A|}$ and $\mathbf{c}_B^i = [z_e(\mathbf{f})/\partial x_{rs}^i]_{1 \times |A|}$. Given the simplex multiplier vector $\mathbf{w} = \mathbf{c}_B^i \mathbf{B}^{-1}$, we know that the reduced cost for a nonbasic path flow variable $f_{k,rs}^i$ is $c_{k,rs}^i - z_{k,rs}^i = z_e(\mathbf{f})/\partial x_{rs}^i - \mathbf{c}_B^i \mathbf{B}^{-1} \Delta_k^{rs}$, where Δ_k^{rs} is the corresponding column of Δ to the nonbasic path k. It is readily known that if all reduced costs $c_{k,rs}^i - z_{k,rs}^i \geq 0$, $\forall k \in P_{\bar{B}}$, the current basic feasible solution is optimal; otherwise, we may increase the path flow rate of a nonbasic path with $c_{k,rs}^i - z_{k,rs}^i < 0$, $k \in P_{\bar{B}}$ from 0 to some positive level so that the objective function value is decreased while the problem feasibility is maintained. In the latter case, a nonbasic path with the lowest reduced cost value may be chosen, according to Dantzig's rule (Dantzig, 1963). Without a need of enumerating all the nonbasic paths in the set $P_{\bar{B}}$, Dantzig's rule can be implemented by solving the following minimization problem:

$$\min_{k \in P} \{c_{k,rs}^i - z_{k,rs}^i\} \tag{21}$$

which can be decomposed into a set of smaller minimization problems, each of which is for an O-D pair:

$$\min_{r \in R, s \in S} \left\{ \cdots, \min_{k \in K_{rs}} \{c_{k,rs}^i - z_{k,rs}^i\}, \cdots \right\} \tag{22}$$

where $P = \bigcup_{r \in R, s \in S} K_{rs}$ is the set of all paths over the network.

Note that the above minimization problem for each O-D pair r-s is essentially a shortest path problem given below, given that $c_{k,rs}^i = z_e(\mathbf{f})/\partial x_{rs}^i$ is fixed for all paths connecting O-D pair r-s:

$$\min \quad -z_{k,rs}^i = -\mathbf{c}_B^i \mathbf{B}^{-1} \Delta_k^{rs} \tag{23.1}$$

subject to $k \in K_{rs}$ (23.2)

where Δ_k^{rs} is the arc-path incidence vector of path k connecting O-D pair r-s, which exists in Δ as a column. It is obvious that for this shortest path problem, the negative of the simplex multiplier vector $-\mathbf{w}^i = -\mathbf{c}_B^i \mathbf{B}^{-1}$ specifies the arc costs over the network. It should be noted that an arbitrary element in $\mathbf{w}^i$ (or $-\mathbf{w}^i$) (i.e., the cost of an arbitrary arc) may be positive or negative, so a shortest path algorithm that can detect and prevent negative cost loops is needed. For a most efficient implementation, following Chen et al. (2009) and Xie et al. (2010), we use a mixed strategy of shortest path search here, which is executed in a three-step process. First, the Bellman-Ford algorithm (Bellman, 1958) is used, to detect whether a negative loop exists in the network. Then, if any negative loop is found, all negative-cost arcs are set with a very small positive value; otherwise, skip this step. Third, the Floyd-Warshall algorithm (Floyd, 1962) is applied to find shortest paths between all O-D pairs.

By executing the shortest path search for each O-D pair, we can then obtain the entering path (to the basis matrix) with the lowest $c_{k,rs}^i - z_{k,rs}^i$ value over all O-D pairs, which generates a new column for the basis matrix, $\Delta_{k'}^{r's'}$. The remaining algorithmic task is then to determine the flow rate of the entering path and accordingly identify a leaving path (from the basis matrix). Suppose that the entering path is k' between O-D pair r'-s' and its flow rate and the arc-path incidence vector are $f_{k'}^{r's'}$ and $\Delta_{k'}^{r's'}$, respectively. Then the leaving path is the one that maximizes $f_{k',r's'}^i$ while maintaining the problem feasibility (i.e., the flow rates of all the basic feasible paths must be greater than or equal to 0):

$$\max\{f_{k',r's'}^i: \mathbf{f}_B^i = \mathbf{B}^{-1}\hat{\mathbf{v}}^i - (\mathbf{B}^{-1}\Delta_{k'}^{r's'})f_{k',r's'}^i \geq \mathbf{0}\} \tag{24}$$

where $\mathbf{f}_B^i$ is the path flow vector corresponding to the current basis matrix and $\hat{\mathbf{v}}^i$ is the arc flow vector. Let us write $\mathbf{B}^{-1} = [\mathbf{p}_m^{-1}]_{|A|\times 1}$, where $\mathbf{p}_m^{-1} = [p_{m,n}^{-1}]_{1\times|A|}$ is the mth row of $\mathbf{B}^{-1}$. Then, the above inequality in Equation (21) is reduced to $\max\{f_{k',r's'}^i: f_{k',r's'}^i \leq \mathbf{p}_m^{-1}\hat{\mathbf{v}}^i/\mathbf{p}_m^{-1}\Delta_{k'}^{r's'}, \forall m\}$, which in turn results in:

$$(f_{k',r's'}^i)_{\max} = \min\{\mathbf{p}_m^{-1}\hat{\mathbf{v}}^i/\mathbf{p}_m^{-1}\Delta_{k'}^{r's'}, \forall m\} \tag{25}$$

Following this, the path flow rate in the current basis matrix can be accordingly updated as $\mathbf{f}_B^i = \mathbf{B}^{-1}\hat{\mathbf{v}}^i - (\mathbf{B}^{-1}\Delta_{k'}^{r's'})(f_{k',r's'}^i)_{\max}$, by which the leaving path can also be chosen as the one whose flow rate is decreased to 0.

4.3. Algorithmic procedure

So far the most complex computational components for the algorithm implementation have been discussed in detail. As a summary, we combine all these details and formally write the convex combination procedure as follows:

Step 0. (Initialization) Find a feasible solution as an initial point of the algorithm. Set the iteration counter $n := 0$. A feasible O-D flow pattern can be obtained by setting $x_{rs}^{i(n)} = \hat{v}_a^i, \forall a \in A, i \in I$, where nodes r and s are the head and tail nodes of arc a, i.e., $a = (r, s)$, and setting $x_{rs}^{i(n)} = 0, \forall i \in I$, for all other O-D pairs; and a feasible set of demand function parameters, $b_{rs}^{(n)}$ and $e_{rs}^{(n)}, \forall r \in R, s \in S$, can be calculated by Equation (14)-(15).

Step 1. (Direction finding) Find an auxiliary O-D flow pattern $\mathbf{x}' = \{\mathbf{x}'^i\} = \{[x_{rs}'^i]\}$ by solving such a linearized problem for each network scenario i,

$$\min \sum_{rs} \left(\frac{\partial z_e(\mathbf{f})}{\partial x_{rs}^i}\bigg|_{x_{rs}^i = x_{rs}^{i(n)}}\right) x_{rs}'^i$$

subject to   (5.2)-(5.4)

through the following substeps:

Step 1.1. (Linearized problem initialization) Find an initial solution for the scenario-specific linearized problem. This can be done by setting $f_{k,rs}^i = \hat{v}_a^i$ for path k connecting O-D pair r-s that nodes r and s are the head and tail nodes of arc a, i.e., $a = (r, s)$, and path k contains arc a only, i.e., $\delta_{a,k}^{rs} = 1$ and $\delta_{b,k}^{rs} = 0, \forall b \neq a$, and setting $f_{k',rs}^i = 0, \forall k' \neq k$ connecting O-D pair r-s. All other path flow rates are set to be 0.

Step 1.2. (Entering path choosing) Solve a shortest path problem defined in Equation (23) for each O-D pair and identify the entering path k connecting O-D pair r-s with the minimum $c_{k,rs}^i - z_{k,rs}^i$ value over all O-D pairs. If the minimum $c_{k,rs}^i - z_{k,rs}^i$ value is greater than or equal to 0, the current basic feasible solution is optimal, go to step 1.5; otherwise, go to step 1.3.

Step 1.3. (Leaving path choosing) Compute the flow rate of entering path k' connecting O-D pair r'-s' by $\left(f^i_{k',r's'}\right)_{\max} = \min\{\mathbf{p}_m^{-1}\hat{\mathbf{v}}^i/\mathbf{p}_m^{-1}\Delta^{r's'}_{k'}, \forall m\}$ and identify the leaving path whose flow rate is decreased to 0.

Step 1.4. (Basis matrix updating) Update the flow rates of the basic feasible paths by $\mathbf{f}^i_B = \mathbf{B}^{-1}\hat{\mathbf{v}}^i - \left(\mathbf{B}^{-1}\Delta^{r's'}_{k'}\right)f^i_{k',r's'}$, and update the basis matrix by inserting the arc-path incidence vector of the entering path and removing the arc-path incidence vector of the leaving path. Go to step 1.2.

Step 1.5. (Demand function parameter updating) Find an auxiliary set of demand function parameters $\mathbf{b}' = [b'_{rs}]$ and $\mathbf{e}' = [e'_{rs}]$ by using Equation (14)-(15) based on the O-D flow result obtained in step 1.

Step 2. (Line search) Find an optimal α value for $0 \le \alpha \le 1$ by solving the following line search problem:

$$\min \quad \sum_i \sum_{rs} \left[\left(x^{i(n)}_{rs} + \alpha\left(x'^i_{rs} - x^{i(n)}_{rs}\right)\right) \ln\left(x^{i(n)}_{rs} + \alpha\left(x'^i_{rs} - x^{i(n)}_{rs}\right)\right) - \left(x^{i(n)}_{rs} + \alpha\left(x'^i_{rs} - x^{i(n)}_{rs}\right)\right) \right]$$

$$+ w \sum_i \sum_{rs} \left[\left(b^{(n)}_{rs} + \alpha\left(b'_{rs} - b^{(n)}_{rs}\right)\right) + \left(e^{(n)}_{rs} + \alpha\left(e'_{rs} - e^{(n)}_{rs}\right)\right)\hat{t}^i_{rs} - \left(x^{i(n)}_{rs} + \alpha\left(x'^i_{rs} - x^{i(n)}_{rs}\right)\right) \right]^2$$

subject to $\quad 0 \le \alpha \le 1$

Step 3. (Solution update) Set $x^{i(n+1)}_{rs} = x^{i(n)}_{rs} + \alpha\left(x'^i_{rs} - x^{i(n)}_{rs}\right)$, $b^{(n+1)}_{rs} = b^{(n)}_{rs} + \alpha\left(b'_{rs} - b^{(n)}_{rs}\right)$ and $e^{(n+1)}_{rs} = e^{(n)}_{rs} + \alpha\left(e'_{rs} - e^{(n)}_{rs}\right)$.

Step 4. (Convergence test) If a set of convergence criteria is met, e.g.,

$$\frac{\sum_i \sum_{rs} \frac{\left|x^{i(n+1)}_{rs} - x^{i(n)}_{rs}\right|}{x^{i(n+1)}_{rs}}}{|I||R||S|} < \epsilon_x$$

$$\frac{\sum_{rs} \frac{\left|b^{(n+1)}_{rs} - b^{(n)}_{rs}\right|}{b^{(n+1)}_{rs}}}{|R||S|} < \epsilon_b$$

$$\frac{\sum_{rs} \frac{\left|e^{(n+1)}_{rs} - e^{(n)}_{rs}\right|}{e^{(n+1)}_{rs}}}{|R||S|} < \epsilon_e$$

stop the procedure; otherwise, set $n := n + 1$ and go to step 1. Here ϵ_x, ϵ_b and ϵ_e are prespecified convergence error bound associated with $\{[x^i_{rs}]\}$, $[b_{rs}]$ and $[e_{rs}]$, respectively.

5. Numerical evaluation and comparison

The solution algorithm presented above involves some sophisticated computational procedures. For the implementation efficiency, we coded the algorithm in C++. The code is then tested on a set of networks with different sizes and topologies. Through a couple of selected examples, this section presents some application results from the test. All the evaluation results obtained from applying the elastic-demand estimator are compared to a fixed-demand estimator, which was developed earlier by the authors (see Xie et al., 2010).

5.1. The Sioux Falls network

The first example we adopt here is the widely used Sioux Falls network. A small rectangle area in the middle of this network is confined as a subnetwork (see Figure 1). The subnetwork has 12 nodes and 34 arcs. We separately, randomly increase or decrease the capacity of each arc of the subnetwork between -50 to +200 percent and conduct a full-network traffic assignment on each of these network scenarios, generating 34 network flow samples. The parameters of the elastic demand functions for this Sioux Falls subnetwork are estimated based on these 34 samples.

The estimator's performance is then indirectly evaluated by carrying out an elastic-demand traffic assignment with the estimated demand function parameters and examining the generated network flows.

For the illustration purpose, two example network change scenarios are used here for the Sioux Falls subnetwork:

- Scenario SF-1: Increasing the arc capacity of segment 11-10-16 from 0 to 300 percent;
- Scenario SF-2: Increasing the arc capacity of segment 4-15-19 from 0 to 300 percent.

Then we evaluate how different the traffic flow patterns generated from the subnetwork are from those generated from the full network. Two arcs, namely, arc (10, 16) and arc (14, 15) are selected here for demonstration. First of all, we can see that the flow variation patterns with capacity changes are vastly different for different arcs under different network change scenarios, indicating the complexity of the network's supply-demand relationship. Despite these different arc flow variation patterns, the elastic-demand approach evidently outperforms the fixed-demand approach in replicating traffic flow patterns from the full network. This is especially apparent, in the case of, for example, arc (10, 16) when segment 11-10-16 is expanded in capacity by 3 times, or arc (14, 15) when the capacity of segment 14-15-19 is increased by 3 times, which attracts large amount of traffic going through these roadway segments. In these cases, the fixed-demand approach cannot take into account such a traffic attraction phenomenon and as a result, seriously underestimates relevant O-D flows and arc flows.

5.2. The Austin network

An aggregate evaluation for the elastic-demand estimator is further conducted by applying it to a larger-size subnetwork extracted from the Austin network (see Figure 3). The subnetwork consists of 60 nodes and 178 arcs, covering a small urban area of the Austin city. Both of the elastic-demand and fixed-demand approaches are used to generate trip matrices for the subnetwork and then produce network flows by elastic-demand and fixed-demand traffic assignments, respectively. A set of six network upgrading scenarios are used here:

- Scenario A-1: Adding one northbound lane to Burnet Rd;
- Scenario A-2: Adding one northbound and one southbound lanes to Burnet Rd;
- Scenario A-3: Adding one eastbound and one westbound lanes to Anderson Ln;
- Scenario A-4: Adding one westbound lane to RM 2222;
- Scenario A-5: Adding one eastbound and westbound lanes to RM 2222;
- Scenario A-6: Adding one northbound and one southbound lanes to Lamar Blvd.

Under these network scenarios, the aggregated results from evaluating these arc flow estimates obtained from the subnetwork against the results from the full network are analyzed by regression analysis below (see Figure 4).

Again, the comparison outcome from the regression analysis shows that the elastic-demand estimator more accurately produces O-D flows and in turn arc flows than the fixed-demand estimator. We use the root mean square error (RMSE) as the performance measure to assess the closeness between the full-network and subnetwork traffic assignment results. In the Austin subnetwork case, the RMSE of arc flow rates estimated from the elastic trip matrix is on the level of 7.4-7.6 percent across the set of network scenarios, which is significantly lower than the RMSE level from the fixed trip matrix, which is around 11.7-13.3 percent.

6. Conclusions and research extensions

This paper presents a joint ME-LS estimator for the subnetwork O-D trip matrix estimation problem. A convex optimization model is suggested to characterize the problem and a convex combination algorithm is adapted for problem solutions. Unconstrained optimization and column generation techniques are incorporated into the convex combination procedure, which converts the original optimization problem to an equivalent decomposable problem with the O-D flow variables only and relaxes the path enumeration difficulty encountered in solving the linearized ME subproblem. This estimator can be used not only to capture the traffic attraction/diversion effects arising from subnetwork analysis, but also is potentially useful in evaluating demand elasticities for general urban or regional networks, when the presence of elastic demands cannot be ignored. Numerical results from applying the elastic O-D trip matrix estimator to subnetworks from the Sioux Falls and Austin networks show that its performance is consistently better than the fixed O-D trip matrix estimator.

The elastic O-D trip matrix estimation model and solution method may be improved or enhanced in several ways. The present model assumes the prevalence of origin and destination nodes in a network and hence avoids an

explicit treatment of the input data inconsistence issue caused by traffic count errors or noises. In a more general setting, we should allow the existence of non-origin and non-destination nodes in the network and incorporate data errors into the model estimation process. For example, we can add a new term of minimizing the sum of squared traffic count errors and form an extended optimization model as follows:

$$\min \quad z(\mathbf{f}, \mathbf{b}, \mathbf{e}) = \sum_i \sum_{rs} (x_{rs}^i \ln x_{rs}^i - x_{rs}^i) + w \sum_i \sum_{rs} \left(d_{rs}(\hat{t}_{rs}^i | b_{rs}, e_{rs}) - x_{rs}^i \right)^2$$

$$+ u \sum_i \sum_a \left((y_a^{i+})^2 + (y_a^{i-})^2 \right) \tag{26.1}$$

$$\text{subject to} \quad \sum_{rs} \sum_k f_{k,rs}^i \delta_{a,k}^{rs} + y_a^{i+} - y_a^{i-} = \hat{v}_a^i \qquad \forall a \in A, i \in I \tag{26.2}$$

$$f_{k,rs}^i \geq 0 \qquad \forall k \in K_{rs}, r \in R, s \in S, i \in I \tag{26.3}$$

$$y_a^{i+}, y_a^{i-} \geq 0 \qquad \forall a \in A, i \in I \tag{26.3}$$

with the following definitional relationships for O-D flow rate x_{rs}^i,

$$x_{rs}^i = \sum_k f_{k,rs}^i \qquad \forall r \in R, s \in S, i \in I \tag{26.4}$$

$$\sum_s x_{rs}^i = 0 \qquad \forall r \in N/R, i \in I \tag{26.5}$$

$$\sum_r x_{rs}^i = 0 \qquad \forall s \in N/S, i \in I \tag{26.6}$$

In this extended model, the added artificial variables y_a^{i+} and y_a^{i-} denote the difference between the given arc flow rate on arc a under network scenario i and the corresponding estimated arc flow rate consistent with the estimated O-D flow matrix. The newly added LS term, $u \sum_i \sum_a ((y_a^{i+})^2 + (y_a^{i-})^2)$, aims at prompting the optimal link flow pattern (i.e., the one corresponding to the optimal solution of the model) as close as possible to the given link flow pattern, where u is the weighting coefficient for this term.

From the basic and extended models, we have observed that the major modeling component of the ME-LS estimator is the objective function. We arbitrarily specified a linear demand function for the estimator, which results in a relatively simple LS term in the objective function. The linear structure of the demand function implies a constant demand elasticity, which may not be ideal in many cases. Other functional forms of the elastic demand function should be investigated in the present modeling framework, especially those decreasing, convex and continuously differentiable functions. On the other hand, the prediction performance of the estimator is heavily dependent on the weighting coefficient w in the objective function. An excessively low value of w may promote each scenario-specific O-D trip matrix to closely match its ME pattern but possibly result in O-D demand functions poorly estimated; an excessively high value of w may lead to a fine fitting criterion in estimating O-D demand functions but discourage the maximization of the entropy of scenario-specific O-D trip matrices. The results from both the extreme sides could lower down the prediction performance of the estimated elastic O-D flow table when used as input for subnetwork flow evaluations. A sensitivity analysis or an optimization-oriented method should be subsequently pursued to determine the best value or range of w.

Under the convex combination framework, the presented solution procedure decomposes the linear approximation problem at each iteration into a set of linear ME subproblems and an LS subproblem. An alternative solution approach, originally proposed by Carey et al. (1981), may be applied here for problem solutions as well. This alternative decomposes the original optimization problem into an ME subproblem and an LS subproblem. The

resulting ME subproblem can be solved by the convex combination method, as shown in Xie et al. (2010). Iteratively solving the ME subproblem and the LS subproblem in this alternative manner also leads the solution to converge to the optimal point but poses a different problem decomposition strategy from the one we present in this text. Which of the two strategies is more effective and efficient is still a question at this point. A further numerical analysis and comparison on the performance of these two different decomposition strategies is of great interest and deserves an investigation in the future.

The subnetwork trip matix estimation problem presented in this paper is entirely based on static network states and flows. In principle, time-dependent network flows can better reflect the variable nature of a subnetwork trip matrix, where it is a result of both route and departure time choice changes of individual travelers. Researchers so far only have had a very limited experience in analyzing and estimating dynamic subnetwork trip matrices (see Zhou et al., 2006). Our approach provides an alternative promising direction. Adding the time dimension into the already complicated ME-LS estimator, however, resembles more challenging modeling and solution tasks. It is anticipated that such an attempt would bring greater theoretical and practical values to subnetwork flow analysis.

7. Acknowledgements

References

Bar-Gera, H. (2006). Primal method for determining the most likely route flows in large road networks. *Transportation Science*, 40(3), 269-286.

Bazaraa, M.S., Jarvis, J.J., & Sherali, H.D. (1990). *Linear Programming and Network Flows*. Wiley, New York, NY.

Bell, M.G.H. (1983). The estimation of an origin-destination matrix from traffic counts. *Transportation Science*, 17(2), 198-217.

Bell, M.G.H. (1985). Variances and covariances for origin-destination flows when estimated by log-linear models. *Transportation Research*, 19B(6), 497-507.

Bell, M.G.H. (1991). The estimation of an origin-destination matrix from traffic counts. *Transportation Science*, 17(1), 198-217.

Boyles, S.D. (2010). Subnetwork trip table generation with bush-based sensitivity analysis. Submitted to *the 90th Transportation Research Board Annual Meeting*, Washington, DC.

Carey, M., Hendrickson, C., & Siddharthan, K. (1981). A method for direct estimation of origin/destination trip matrices. *Transportation Science*, 15(1), 32-49.

Carey, M., & Revelli, R. (1986). Constrained estimation of direct demand functions and trip matrices. *Transportation Science*, 20(3), 143-152.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least square estimator. *Transportation Research*, 18B(4/5), 289-299.

Chan, Y. (1976). A method to simplify network representation in transportation planning. *Transportation Research*, 10(3), 179-191.

Chan, Y., Shen, T.S., & Mahaba, N.M. (1989). Transportation network design problem: Application of a hierarchical search algorithm. *Transportation Research Record*, 1251, 24-34.

Chen, A., Chootinan, P., & Recker, W. (2009). Norm approximation method for handling traffic count inconsistencies in path flow estimator. *Transportation Research*, 43B(8-9), 852-872.

Dantzig, G.B. (1963). *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ.

Dowling, R.G., & May, A.D. (1985). Comparison of small-area O-D estimation techniques. *Transportation Research Record*, 1045, 9-15.

Eash, R.W., Chon, K.S., Lee, Y.J., & Boyce, D.E. (1983). Equilibrium traffic assignment on an aggregated highway network for sketch planning. *Transportation Research Record*, 944, 30-37.

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2), 95-110.

Fisk, C.S. (1988). On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research*, 22B(1), 69-79.

Kaplan, M.P., Gur, Y.J., & Vyas, A.D. (1984). Sketch planning model for urban transportation policy analysis. *Transportation Research Record*, 952, 32-39.

Gajewski, B.J., Rilett, L.R., Dixon, M.P., & Spiegelman, C.H. (2002). Robust estimation of origin-destination matrices. *Journal of Transportation and Statistics*, 5(2/3), 37-56.

Haghani, A.E., & Daskin, M.S. (1984). Network design application of an extraction algorithm for network aggregation. *Transportation Research Record*, 944, 37-46.

Haghani, A.E., & Daskin, M.S. (1986). Aggregation effects on the network design problem. *Journal of Advanced Transportation*, 20(3), 239-258.

Hearn, D.W. (1984). Practical and theoretical aspects of aggregation problems in transportation planning methods. *Transportation Planning Models*, M. Florian, Eds., North-Holland, Amsterdam, The Netherlands, 257-287.

Hendrickson, C., & McNeil, S. (1984). Matrix entry estimation errors. *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, Delft, The Netherlands, 413-430.

Janson, B.N. (1993). Most likely origin-destination link uses from equilibrium assignment. *Transportation Research*, 27B(5), 333-350.

Larsson, T., Lundgren, J.T., Rydergren, C., & Patriksson, M. (2002). Most likely traffic equilibrium route flows analysis and computation. *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, F. Giannessi Eds., Springer, New York, NY, 129-159.

LeBlanc, L.J., & Farhangian, K. (1982). Selection of a trip table which reproduces observed link flows. *Transportation Research*, 16B(2), 83-88.

Lo, H.P., Zhang, N., & Lam, W.H.K. (1996). Estimation of an origin-destination matrix with random link choice proportions: A statistical approach. *Transportation Research*, 30B(4), 309-324.

Maher, N.J. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. *Transportation Research*, 17B(6), 435-447.

McNeil, S., & Hendrickson, C. (1985). A regression formulation of the matrix estimation problem. *Transportation Science*, 19(3), 278-292.

Nguyen, S. (1981). *Modèles de Distribution Spatial Tenant Compte des Itinéraries*. Publication No. 225, Centre de Recherché sur les Transports, Univerité de Montréal, Montreal, Canada.

Nguyen, S. (1977). *Estimation an O-D Matrix from Network Data: A Network Equilibrium Approach*. Publication No. 60, Centre de Recherché sur les Transports, Université de Montréal, Montreal, Canada.

Nguyen, S. (1984). Estimating origin-destination matrices from observed flows. *Transportation Planning Models*, F. Florian Eds., Elsevier, Amsterdam, The Netherlands, 363-380.

Nie, Y., & Lee, D.H. (2002). Uncoupled method for equilibrium-based linear path flow estimator for origin-destination trip matrices. *Transportation Research Record*, 1783, 72-79.

Nie, Y., Zhang, H.M., & Recker, W.W. (2005). Inferring origin-destination trip matrices with a decoupled GLS path flow estimator. *Transportation Research*, 39B(6), 497–518.

Nie, Y., & Zhang, H.M. (2010). A relaxation approach for estimating origin-destination trip tables. *Networks and Spatial Economics*, 10(1), 147-172.

Taylor, M.A.P., Young, W., & Newton, P.W. (1988). PC-based sketch planning methods for transport and urban applications. *Transportation*, 14(4), 361-375.

Rogus, M.J. (1996). *Building Sketch Networks: The Development of Sketch Zones Boundaries and Aggregation Process of Detailed Highway Networks*. Working Paper No. 96-11, Chicago Area Transportation Study, Chicago, IL.

Rossi, T.F., McNeil, S., & Hendrickson, C. (1989). Entropy model for consistent impact-fee assessment. *Journal of Urban Planning and Development*, 115(2), 51-63.

Sherali, H.D., Sivanandan, R., & Hobeika, A.G. (1994). A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transportation Research*, 28B(3), 213-233.

Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices. *Transportation Research*, 21B(5), 395-412.

Turnquist, M.A., & Gur, Y. (1979). Estimation of trip tables from observed link volumes. *Transportation Research Record*, 730, 1-6.

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

Willumsen, L.G. (1981). Simplified transport models based on traffic counts. *Transportation*, 10(3), 257-278.

Wilson, A.G. (1970). *Entropy in Urban and Regional Modeling*. Pion, London, England.

Xie, C., Kockelman, K.M., & Waller, S.T. (2010). Maximum entropy method for subnetwork origin-destination trip matrix estimation. *Transportation Research Record*. (In press)

Yang, H., Iida, Y., & Sasaki, T. (1991). An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. *Transportation Research*, 25B(5), 351-363.

Yang, H., Iida, Y., & Sasaki, T. (1994). The equilibrium-based origin-destination matrix estimation problem. *Transportation Research*, 28B(1), 23-33.

Yang, H., Sasaki, T., Iida, Y., & Asakura, Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research*, 26B(6), 417-434.

Zhou, X., Erdogan, S., & Mahmassani, H.S. (2006). Dynamic origin-destination trip demand estimation for subarea analysis. *Transportation Research Record*, 1964, 176-184.

Appendix: The convexity of the objective function

The convexity of the objective function is satisfied if the Hessian matrix of the objective function with respect to decision variables $f_{k,rs}^i$, b_{rs} and e_{rs}, $\nabla^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})$, is positive definite. In particular, elements of the Hessian, i.e., the second-order derivatives of the objective function with respect to a pair of decision variables, include,

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial f_{k,rs}^i \partial f_{k,r's'}^{i'}} = \begin{cases} \frac{1}{x_{rs}^i} + 2w & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial f_{k,rs}^i \partial b_{r's'}} = \begin{cases} -2w & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial f_{k,rs}^i \partial e_{r's'}} = \begin{cases} -2w\hat{t}_{rs}^i & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial b_{rs} \partial b_{r's'}} = \begin{cases} 2w|I| & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial b_{rs} \partial e_{r's'}} = \begin{cases} 2w \sum_i \hat{t}_{rs}^i & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial b_{rs} \partial f_{k,r's'}^i} = \begin{cases} -2w & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial e_{rs} \partial e_{r's'}} = \begin{cases} 2w \sum_i (\hat{t}_{rs}^i)^2 & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial e_{rs} \partial b_{r's'}} = \begin{cases} 2w \sum_i \hat{t}_{rs}^i & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})}{\partial e_{rs} \partial f_{k,r's'}^{i'}} = \begin{cases} -2w\hat{t}_{rs}^i & r = r', s = s' \\ 0 & \text{otherwise} \end{cases}$$

Given this list of second-order derivatives, we can readily show that $\mathbf{y}^\top \nabla^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})\mathbf{y} > 0$, where $\mathbf{y}$ is an arbitrary vector whose length is equal to the size of $\nabla^2 z(\mathbf{f}, \mathbf{b}, \mathbf{e})$. This proves the strict convexity of the objective function.
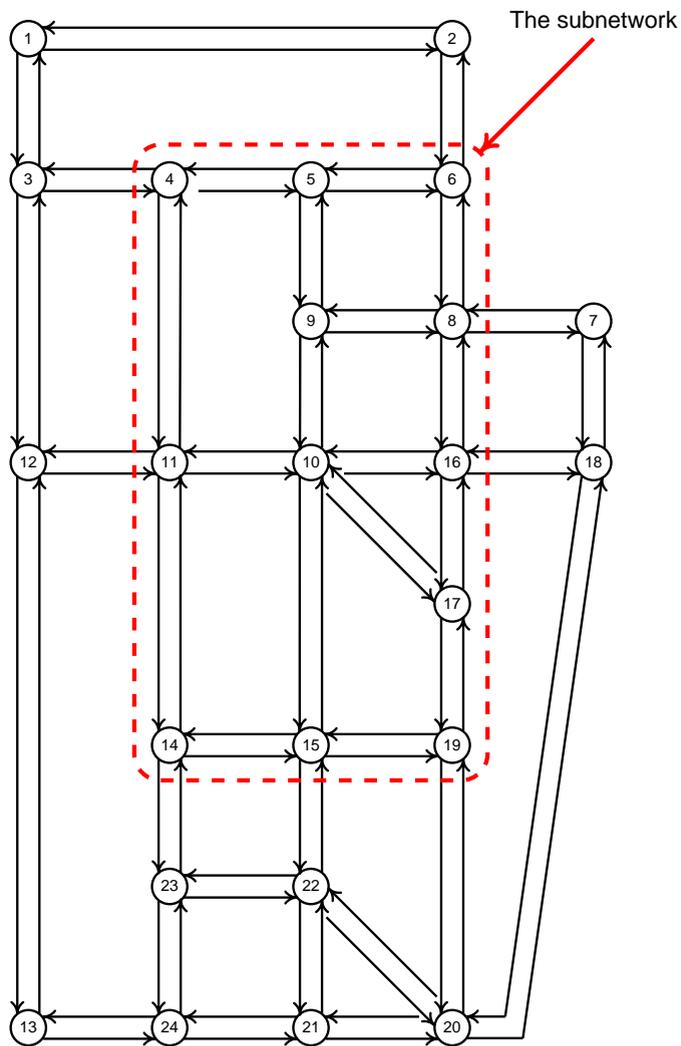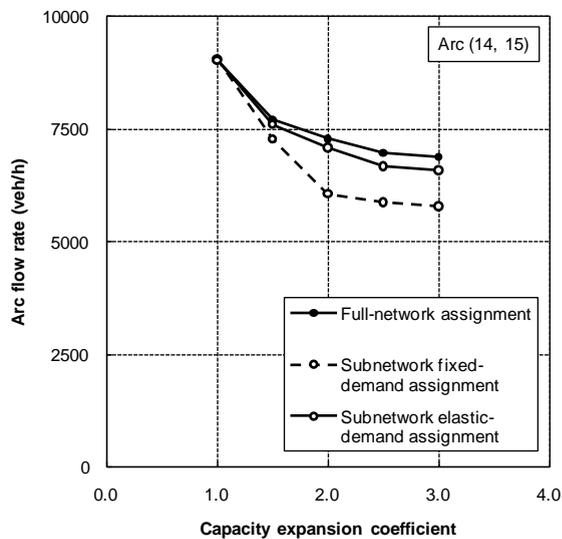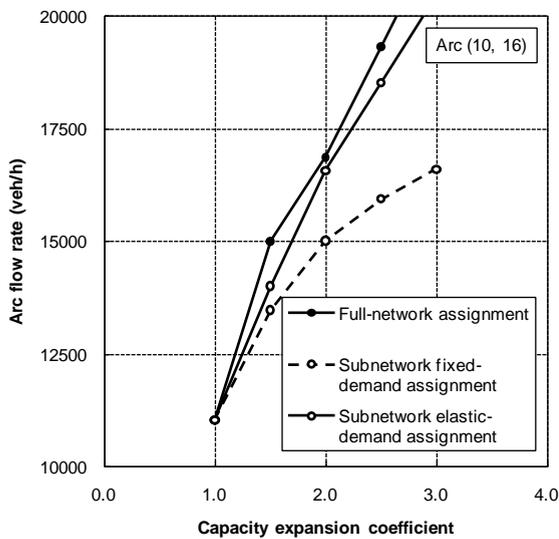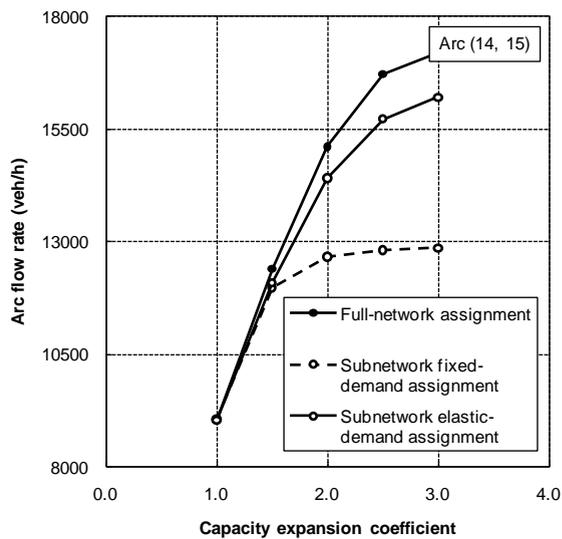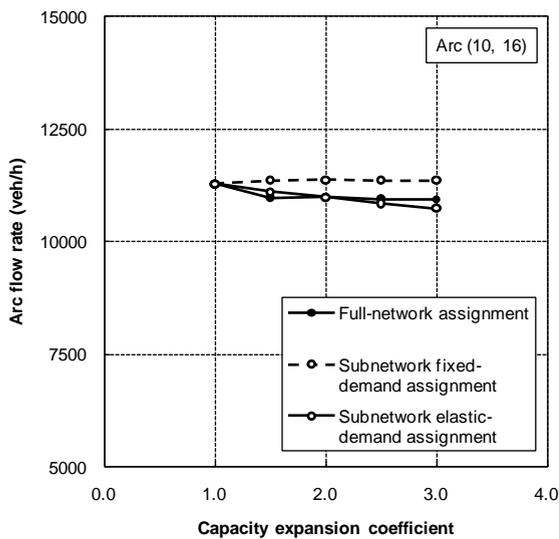
Figure 1 The Sioux Falls network and its subnetwork

(a) Scenario SF-1: Increasing the arc capacity of segment 11-10-16

(b) Scenario SF-2: Increasing the arc capacity of segment 4-15-19

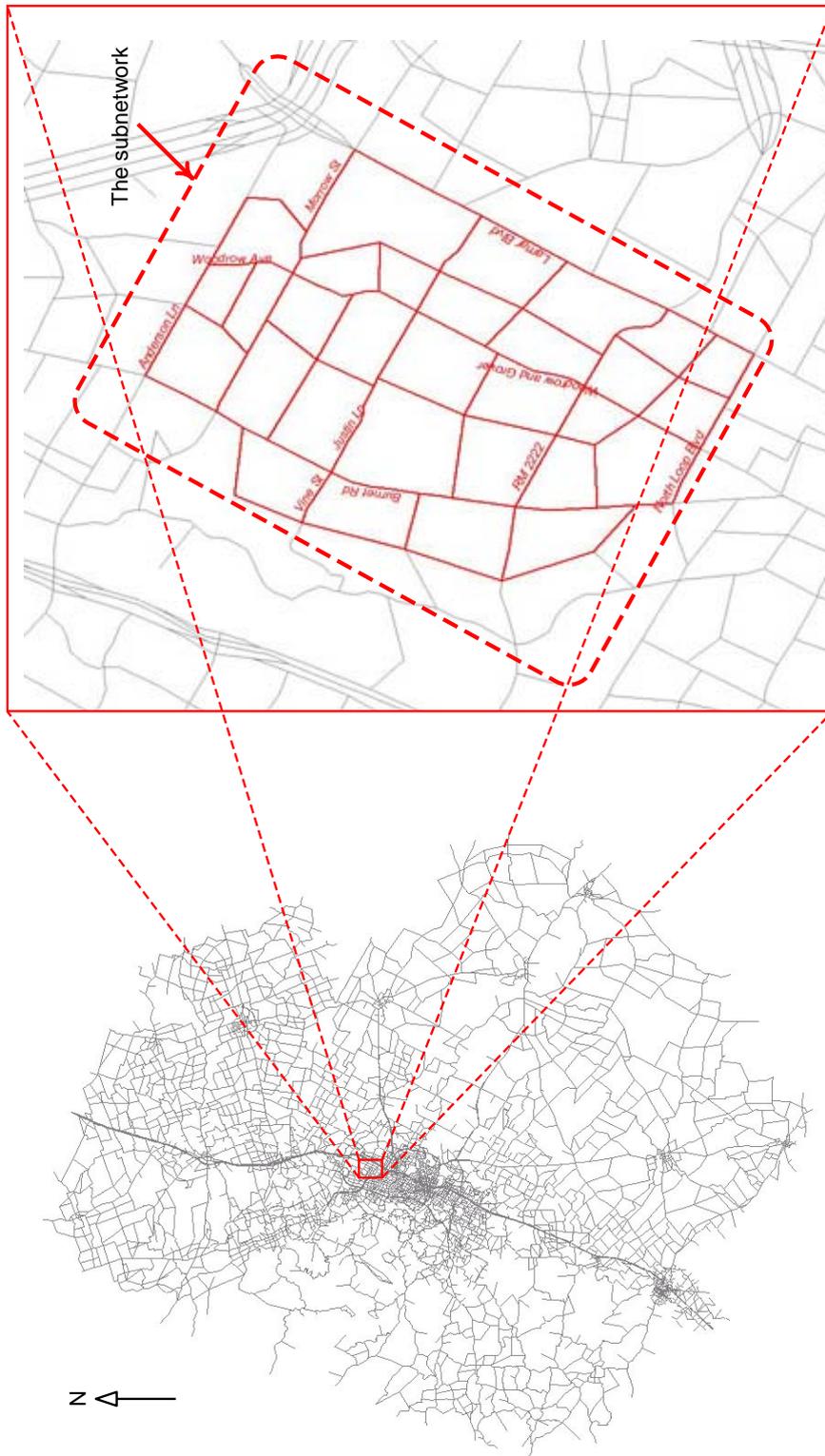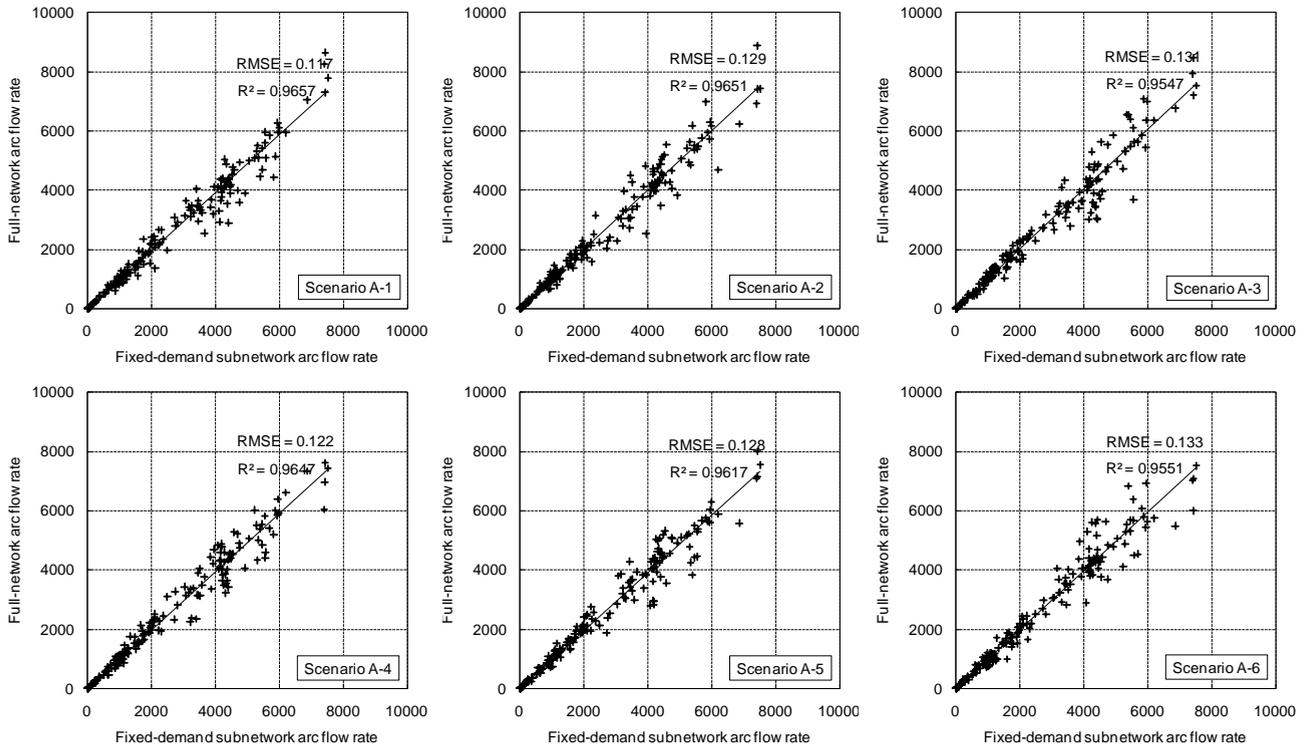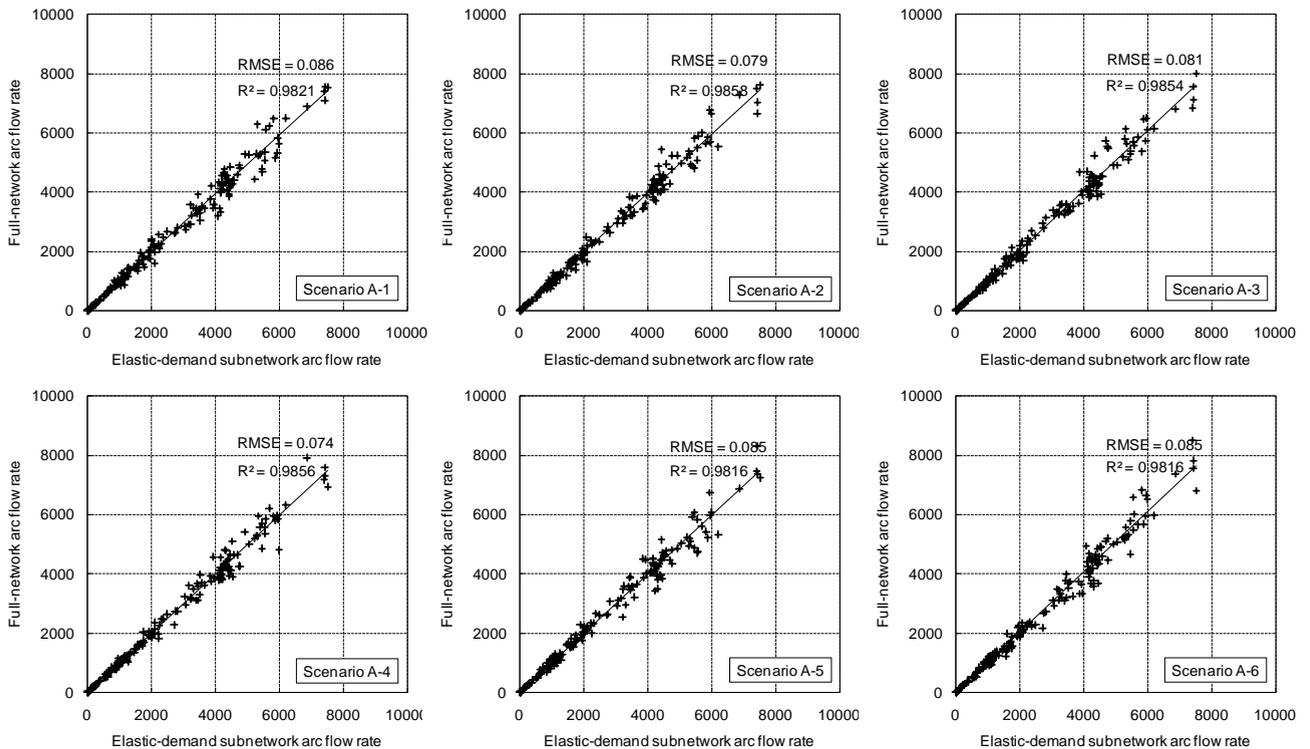Figure 2 Example individual arc flow estimates of the Sioux Falls network

Figure 3 The Austin network and its subnetwork

(a) Arc flow rates estimated by the subnetwork fixed-demand assignment against the full-network assignment

(b) Arc flow rates estimated by the subnetwork elastic-demand assignment against the full-network assignment

Figure 4 Regression analysis of arc flow estimates of the Austin network