PERGAMON

Transportation Research Part B 36 (2002) 19–35

TRANSPORTATION
RESEARCH
PART B

www.elsevier.com/locate/trb

Benefit distribution and equity in road network design

Qiang Meng, Hai Yang *

Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Received 6 December 1999; received in revised form 7 June 2000; accepted 9 June 2000

Abstract

In the classical continuous network design problem, the optimal capacity enhancements are determined by minimizing the total system cost under a budget constraint, while taking into account the route choice behavior of network users. Generally the equilibrium origin–destination travel costs for some origin–destination (O–D) pairs may be increased after implementing these optimal capacity enhancements, leading to positive or negative results for network users. Therefore, the equity issue about the benefit gained from the network design problem is raised. In this paper, we examine the benefit distribution among the network users and the resulting equity associated with the continuous network design problem in terms of the change of equilibrium O–D travel costs. Bilevel programming models that incorporate the equity constraint are proposed for the continuous network design problem. A penalty function approach by embodying a simulated annealing method is used to test the models for a network example.

Keywords: Equity; Network design; Optimization; User equilibrium; Bilevel programming

1. Introduction

The conventional Network Design Problem (NDP) can be broadly classified into two types. The first type is the Discrete Network Design Problem (DNDP) that determines an optimal set of locations for constructing some new roads added to the current transportation network. The second type is the Continuous Network Design Problem (CNDP) that determines the optimal capacity enhancements for some existing roads. Both types of NDP attempt to minimize the total system costs under a budget constraint, while taking into account the route choice behavior of network users (for a recent comprehensive review, see, for example, Yang and Bell, 1998). Users'

* Corresponding author. Tel.: +852-2358-7178; fax: +852-2358-1534.
E-mail address: cehyang@ust.hk (H. Yang).

PII: S0191-2615(00)00036-9

route choice behavior is generally characterized either by the Deterministic User Equilibrium (DUE) or the Stochastic User Equilibrium (SUE) (Sheffi, 1985). Typical models for the NDP under DUE have been developed by Leblanc (1975); Abdulaal and LeBlanc (1979); Boyce and Janson (1980); Yang and Bell (1998). Davis (1994) extended the CNDP under the DUE to the SUE case. Mathematically, these NDP problems can be characterized by the bilevel programming model due to its containing two optimization problems. It is well known that bilevel programming is one of the hardest problems to solve in the optimization. This attracts considerable attention to design some efficient algorithms for different models in the NDP. The representative works include the approximate algorithm for the DNDP under DUE (Poorzahedy and Turnquist, 1982); the equilibrium decomposed optimization method for the CNDP under DUE (Suwansirikul et al., 1987); the sensitivity analysis-based method for the CNDP under DUE (Cho, 1988; Friesz et al., 1990); the simulated annealing approach for the CNDP under DUE (Friesz et al., 1992); the successive quadratic programming method for the CNDP under logit-based SUE (Davis, 1994). Apart from these bilevel formulations, there are also some equivalent single level optimization models and the corresponding methods for the CNDP using the different gap functions proposed by, for instance, Tan et al. (1979), Marcotte and Marquis (1992), Marcotte (1983), Fisk (1984), Friesz et al. (1992), Davis (1994) and Meng et al. (2001).

Most of the existing NDP models only emphasized minimizing the generalized system costs as an ultimate objective. As an exception, Friez et al. (1993) presented a multiobjective programming model for the CNDP under DUE, in which a number of system-wide objectives such as total system costs, total construction costs and total vehicle miles travelled are considered. The single or multiple objective functions adopted in existing models are meaningful from the perspective of social optimum. There is, however, an intriguing issue that has been ignored in the existing NDP models: the equilibrium O–D travel costs between some O–D pairs may be increased or decreased after implementing an optimal network design scenario obtained from the aforementioned CNDP models. This can lead to positive or negative results for the network users travelling among different O–D pairs. In other words, some network users cannot get any benefit from the network design project and further disburse the additional travel cost. Therefore, the social inequitable issue in terms of the changes of the equilibrium O–D travel cost is generated for the different groups of the network users classified by O–D pairs. As a consequence of such concerns about fairness, it may be difficult to rally public support and easy to evoke opposition to the implementation of a road construction project. Indeed, in reality, the public acceptability of a road project depends much on a positive net benefit; the distribution and visibility of the impact are also relevant. Thus, equity concerns must be addressed before a road construction project can be introduced into urban road networks with majority public support.

The purpose of this paper is to consider and model the benefit distribution and the equity issue for the CNDP under the DUE. The basic idea of the equity issue examined here is also suitable for other types of NDPs. We will define a critical O–D travel cost ratio as the maximal ratio of the equilibrium O–D travel cost after implementing a network design scenario divided by the equilibrium O–D travel cost before expanding the capacity for a set of specific O–D pairs, to reflect the greatest change of the equilibrium O–D travel cost for the CNDP. Under the budget constraint, this critical ratio must have a finite minimum value and a finite maximum value denoted by $\alpha_{\min}$ and $\alpha_{\max}$, respectively, for any given design scenario. If $\alpha_{\min} < 1$, there will exist a network design scenario such that all users in the network can benefit. Furthermore, the two extreme values, by $\alpha_{\min}$ and $\alpha_{\max}$ can be obtained by solving two bilevel programming problems. To deal with the equity issue, we can impose a constraint to the conventional CNDP that requires the travel cost ratio between after and before the road improvement project for any O–D pair to be less than a certain threshold selected in the interval $[\alpha_{\min}, \alpha_{\max}]$. In reality, this threshold can measure the degree of the equitability of benefit distribution derived from the CNDP. Moreover, threshold of the O–D travel cost ratio can also be treated as a decision variable in the problem. This results in a multiobjective programming model for the CNDP with the equitable issue.

All of the models proposed in this paper are in the form of bilevel programming, which makes it difficult to obtain a global optimal solution. On the other hand, we indeed require the global optimal value in determining effective interval of O–D travel cost ratio $[\alpha_{\min}, \alpha_{\max}]$ for a given budget. Although a globally convergent simulated annealing method proposed by Dekkers and Aarts (1991) has the capability to find a global optimum for the constrained optimization problem, it is only available for the problems with simple bound or side constraints in order to generate a feasible solution uniformly in its feasible set. Therefore, we have to use the penalty function technique to transfer the non-side constraints into the objective function as a penalty term and then use a simulated annealing method to solve this modified problem.

The organization of the reminder of this paper is as below. In Section 2, the equitable issue in the CNDP is discussed using two examples. Bilevel programming models for the CNDP with the equality issue are proposed in Section 3. In Section 4, we present a general framework of the penalty function approach by incorporating a simulated annealing method. In Section 5, the Sioux Falls network is used to verify our models. Conclusions are provided in the final section.

2. Benefit distribution and equity in the CNDP

This section will give two examples to show that the inequitable problem does exist in terms of the equilibrium origin–destination (O–D) travel costs after implementing a continuous network design project.

2.1. Example 1. An artificial example

Consider a simple network depicted in Fig. 1. It consists of two O–D pairs, $1 \to 4$ and $2 \to 4$. Let $q_{14} = 400$ and $q_{24} = 300$ be the demands from 1 to 4 and 2 to 4, respectively. The travel cost functions used are

$$t_1(v_1) = 2.25 + \frac{v_1}{400}, \tag{1}$$

$$t_2(v_2) = 1 + \frac{v_2}{200}, \tag{2}$$

$$t_3(v_3) = 1 + \frac{v_3}{400}, \tag{3}$$

$$t_4(v_4) = 0.5 + \frac{v_4}{400}. \tag{4}$$

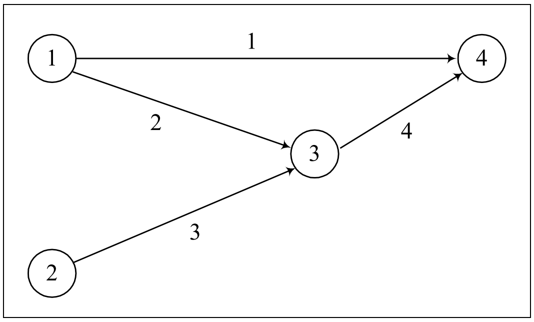

Fig. 1. A simple road network.

Assuming that the behavior of the network users' route choice follows the DUE, it is straightforward to obtain the equilibrium travel cost for each O–D pair as follows:

$$\bar{\mu}_{14} = 3.00 \quad \text{and} \quad \bar{\mu}_{24} = 3.25. \tag{5}$$

Suppose a network design project is to expand the capacity of link 2 from 200 to 800 and the capacity of link 3 from 400 to 450, respectively. After implementation of this network design project, the travel cost functions for these two links will become

$$t_2(v_2) = 1 + \frac{v_2}{800}, \tag{6}$$

$$t_3(v_3) = 1 + \frac{v_3}{450}, \tag{7}$$

and then the equilibrium travel cost for each O–D pair will be

$$\mu_{14} = 2.85 \quad \text{and} \quad \mu_{24} = 3.32. \tag{8}$$

In this case the corresponding ratios of the O–D travel costs between after and before the road capacity enhancement scheme is,

$$\frac{\mu_{14}}{\bar{\mu}_{14}} = 0.95 \quad \text{and} \quad \frac{\mu_{24}}{\bar{\mu}_{24}} = 1.02. \tag{9}$$

This implies that the equilibrium travel cost for the network users traveling from origin 2 to destination 4 is increased and the travel cost for the O–D pair from 1 to 4 is decreased. In other words, the network users from origin 2 to destination 4 cannot get any benefit from this network design scenario, but rather their travel times become longer than before expanding the network capacity.

2.2. Example 2. Sioux Falls network

The network of the Sioux Falls example is shown in Fig. 2. This example contains 24 nodes, 76 links and 552 O–D pairs. The 10 links 16, 17, 19, 20, 25, 26, 29, 39, 48 and 74 are the candidates for capacity expanding. The details of link travel cost functions, the construction cost functions and the other parameters are available in Suwansirikul et al. (1987). Let μ_w and $\bar{\mu}_w$ denote the equilibrium travel cost between O–D pair $w \in W$ after and before implementing the optimal

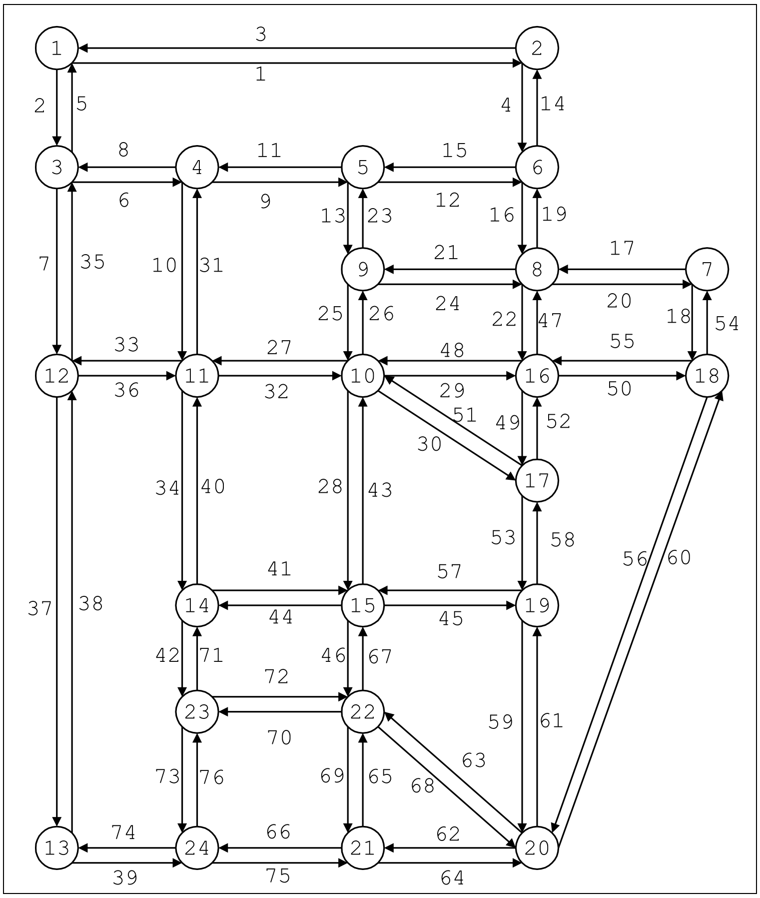

Fig. 2. The Sioux Falls network.

capacity enhancements, respectively, obtained by Friesz et al. (1992). Let α_w represent the ratio $\mu_w/\overline{\mu}_w$ for O–D pair $w \in W$. The distribution of the ratio α_w is illustrated in Fig. 3. From this figure, we can observe that there are more than 14% O–D pairs for which the equilibrium travel costs have become greater after implementing the network design scheme, although the total generalized system cost is minimized for the given CNDP. In addition, the biggest α_w is greater than 1.2 and the smallest one is less than 0.3. These results indicate some users can get benefit and some of them may suffer from the additional increasing travel costs from the same CNDP project.

According to these two examples, the inequity issue indeed exists among the network users. To make a decision of equitable resource allocation acceptable to all network users, a decision-maker has to take into account the change of the equilibrium O–D travel cost between each individual O–D pair for the network design project. Based on these discussions, it is necessary and interesting to examine the benefit distribution associated with a given road network design project and the resulting equity issue for the design problem.

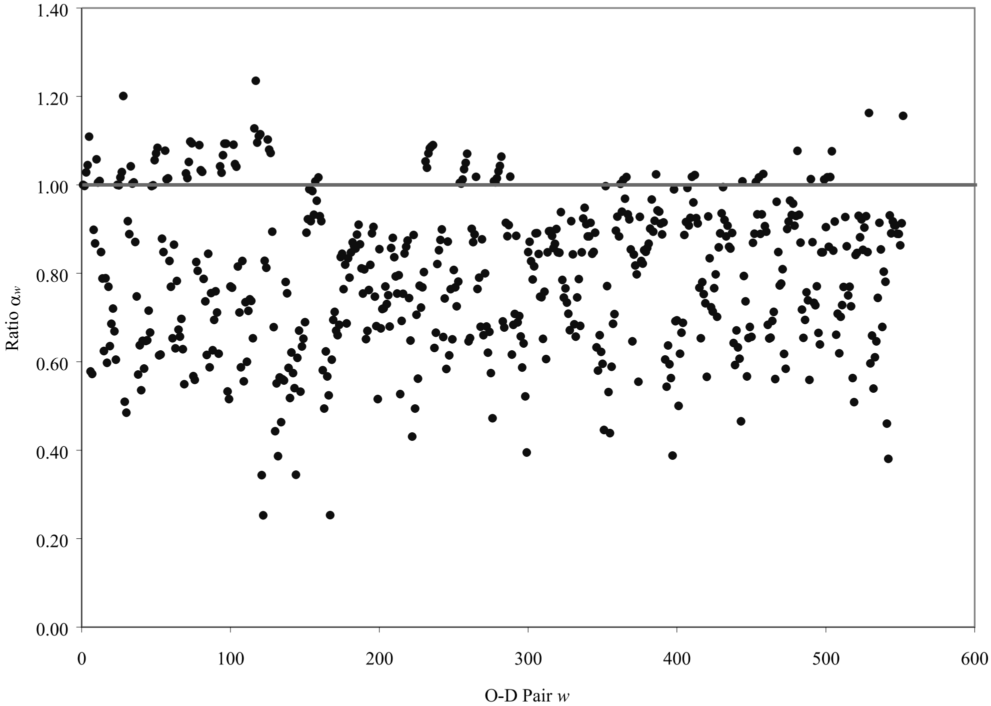

Fig. 3. The distribution of O–D travel cost ratio α_w between after and before implementing a CNDP project for the Sioux Falls network example.

3. Bilevel programming models

3.1. Notations, assumptions and definitions

This section gives some notations, assumptions and definitions used in the paper.

Notations

A	the set of links in the network
W	the set of O–D pairs
R_w	the set of routes between O–D pair $w \in W$
f_r^w	the path flow on route $r \in R_w$
v_a	the link flow on link $a \in A$
$\mathbf{v}$	the vector of link flow, $\mathbf{v} = (\ldots, v_a, \ldots)^{\mathrm{T}}$
y_a	the capacity enhancement of link $a \in A$
$\mathbf{y}$	the vector of link capacity enhancement $\mathbf{y} = (\ldots, y_a, \ldots)^{\mathrm{T}}$
$t_a(v_a, y_a)$	the travel cost on link $a \in A$ described as a function of link flow v_a and capacity enhancement y_a
u_a	an upper bound of $y_a,\ a \in A$, which is a constant
q_w	the fixed O–D demand between O–D pair $w \in W$

$\overline{\mu}_w$	the equilibrium travel cost between O–D pair $w \in W$ before any capacity expanding, which is a constant
$\mu_w(\mathbf{y})$	the equilibrium travel cost between O–D pair $w \in W$ after implementing the capacity enhancement $\mathbf{y}, \mu_w$ is an implicit function of $\mathbf{y}$
$g_a(y_a)$	the capacity expansion cost for link $a \in A$ as a function of capacity increment
δ_{ar}^w	1 if route r between O–D pair w uses link a, and 0 otherwise
B	a fixed and given capacity improvement budget

Assumptions

Assume that link travel cost function $t_a(v_a, y_a), a \in A$ is a strictly increasing, continuous and differentiable function with respect to link flow $v_a, a \in A$ for any fixed capacity enhancement $y_a, a \in A$ and that capacity expansion cost function $g_a(y_a), a \in A$ is a continuous and differentiable function with respect to y_a.

Definitions

Scenario: Any feasible point $\mathbf{y}$ in the following set is termed as a scenario for the CNDP.

$$S = \left\{ \mathbf{y} = (\ldots, y_a, \ldots)^{\mathrm{T}} \,\middle|\, \sum_a g_a(y_a) \leqslant B,\ 0 \leqslant y_a \leqslant u_a,\ a \in A \right\}. \tag{10}$$

The critical ratio: Let $\alpha(\mathbf{y})$ denote the critical O–D travel cost ratio for any scenario, $\mathbf{y}$, and define below

$$\alpha(\mathbf{y}) = \max_{w \in W} \left\{ \frac{\mu_w(\mathbf{y})}{\overline{\mu}_w} \right\}. \tag{11}$$

If $\alpha(\mathbf{y}) < 1$, then all users in the network benefit from implementation of the scenario $\mathbf{y}$; if $\alpha(\mathbf{y}) = 1$, then O–D travel cost between each O–D pair does not become greater; if $\alpha(\mathbf{y}) > 1$, then there is at least one O–D pair between which travel cost is greater than before, and in this case the inequity issue is revealed. Thus, the factor $\alpha(\mathbf{y})$ indeed can measure the equity issue of the O–D travel cost for the CNDP to a certain degree.

The effective interval: The interval $[\alpha_{\min}, \alpha_{\max}]$ is defined as the effective interval for the CNDP, where

$$\alpha_{\min} = \min_{\mathbf{y} \in S} \alpha(\mathbf{y}) = \min_{\mathbf{y} \in S} \max_{w \in W} \left\{ \frac{\mu_w(\mathbf{y})}{\overline{\mu}_w} \right\}, \tag{12}$$

$$\alpha_{\max} = \max_{\mathbf{y} \in S} \alpha(\mathbf{y}) = \max_{\mathbf{y} \in S} \max_{w \in W} \left\{ \frac{\mu_w(\mathbf{y})}{\overline{\mu}_w} \right\}. \tag{13}$$

Based on our assumptions, these two values must exist and have a finite value. If $\alpha_{\min} < 1$, there exists at least one scenario such that all users in the network can positively benefit. If $\alpha_{\min} = 1$, there exists a scenario such that the benefit gained for each network user is at least zero or positive. If $\alpha_{\min} > 1$, then, for any scenario, there exists at least one O–D pair whose travel cost will become greater than before.

3.2. Determination of $\alpha_{\min}$ and $\alpha_{\max}$

We first consider how to determine the effective interval $[\alpha_{\min}, \alpha_{\max}]$ of the O–D travel cost ratio for a given budget. The value of $\alpha_{\min}$ can be obtained by solving the following bilevel programming model.

M1:

$$\min_{\mathbf{y}\in S} F_1(\mathbf{y}) = \max_{w\in W}\left\{\frac{\mu_w(\mathbf{y})}{\overline{\mu}_w}\right\}, \tag{14}$$

where $\mu_w(\mathbf{y}),\ w \in W$ is the equilibrium travel cost which can be obtained by solving the following network equilibrium problem

$$\min_{\mathbf{v}} \sum_{a\in A} \int_0^{v_a} t_a(\omega, y_a)\, \mathrm{d}\omega \tag{15}$$

subject to

$$\sum_{r\in R_w} f_r^w = q_w, \quad w \in W, \tag{16}$$

$$v_a = \sum_{w\in W}\sum_{r\in R_w} f_r^w \delta_{ar}^w, \quad a \in A, \tag{17}$$

$$f_r^w \geqslant 0, \quad r \in R_w,\ w \in W. \tag{18}$$

Note that the DUE problem can be alternatively formulated as the nonlinear complementarity problem (Aashtiani, 1979) or variational inequalities (Dafermos, 1980) under more general assumptions. For simplicity, here we adopt the convex optimization model to describe the DUE conditions. For given $\mathbf{y}$, the equilibrium travel cost $\mu_w(\mathbf{y}),\ w \in W$ and the equilibrium link flow $v_a(\mathbf{y}),\ a \in A$ are both continuous and directionally differentiable function with respect to $\mathbf{y}$ (Qui and Magnanti, 1989). They are also assumed to be unique for given $\mathbf{y}$ throughout the models introduced in this study.

Similarly, the value of parameter $\alpha_{\max}$ can be identified by the following bilevel programming model.

M2:

$$\max_{\mathbf{y}\in S} F_2(\mathbf{y}) = \max_{w\in W}\left\{\frac{\mu_w(\mathbf{y})}{\overline{\mu}_w}\right\}, \tag{19}$$

where $\mu_w(\mathbf{y}),\ w \in W$ is the equilibrium travel cost to be obtained by solving the optimization problem (15)–(18).

Let the optimal objective values for the above two bilevel programming models M1 and M2 be F_1^* and F_2^*, respectively, then we have $\alpha_{\min} = F_1^*$ and $\alpha_{\max} = F_2^*$. Since $y_a = 0, a \in A$ is a feasible point for the above two models, this yields

$$0 < \alpha_{\min} \leqslant 1.0 \leqslant \alpha_{\max} < +\infty. \tag{20}$$

3.3. The bilevel programming model for the CNDP with an equity constraint

In reality, it is almost impossible to guarantee that the benefit gained from a network design scenario will be identical for all users. Nevertheless, we can restrict the equilibrium O–D travel cost reduction for each O–D pair beyond a given level (or travel cost increase for each O–D pair below a given level) by requiring $\alpha(\mathbf{y})$ to be less than a desirable number. The following bilevel programming model can describe this type of equitable CNDP.

M3:

$$\max_{\mathbf{y}} F_3(\mathbf{y}) = \sum_{a \in A} t_a(v_a(\mathbf{y}), y_a) v_a(\mathbf{y}) \tag{21}$$

subject to

$$\max_{w \in W} \left\{ \frac{\mu_w(\mathbf{y})}{\overline{\mu}_w} \right\} \leqslant \beta, \tag{22}$$

$$\sum_{a \in A} g_a(y_a) \leqslant B, \tag{23}$$

$$0 \leqslant y_a \leqslant u_a, \quad a \in A, \tag{24}$$

where $\mu_w(\mathbf{y}),\ w \in W$ and $v_a(\mathbf{y}),\ a \in A$ are the equilibrium O–D travel cost and link flow obtained from the optimization problem (15)–(18), parameter β is a given appropriate positive constant, which measures the degree of equitability of benefit distribution. Thus, inequality (22) can be regarded as an equity constraint. A smaller value of parameter β means a more equitable distribution of benefit across network users. If β is set to be $\beta < 1.0$, then each user will enjoy a travel cost reduction at least by $100(1-\beta)\%$ derived from road capacity enhancements; if, however, β is set to be $\beta > 1.0$, it means that there may be users who suffer a travel cost increase induced by the road improvement scheme, but travel cost increase cannot be more than $100(\beta - 1)\%$. Note that β should be selected to be $\beta \in [\alpha_{\min}, \alpha_{\max}]$. When $\beta < \alpha_{\min}$ the above model does not have a solution; when $\beta > \alpha_{\max}$, constraint (22) becomes an abundant constraint. In the latter extreme case, the model is identical to the conventional bilevel programming model for the CNDP. Note that imposition of constraint (22) will result in increase in the equilibrium O–D travel cost between the O–D pairs that otherwise have greatest decrease. This means that our treatment of equity here is confined to both the greatest increase and greatest decrease of the equilibrium O–D travel costs.

3.4. The multiobjective equilibrium CNDP with an equity constraint

Since parameter β used in Section 3.3 reflects the degree of an equitable reduction of the equilibrium O–D travel cost before and after implementing a scenario and this value is selected by a decision-maker, it can be treated as a decision variable in the bilevel programming model. Thus, in this circumstance, there will be two objectives, the total system cost and the parameter β, to be minimized simultaneously. The following multiobjective programming model can characterize this requirement.

M4:

$$\min_{\mathbf{y},\beta} \mathbf{Z}(\mathbf{y},\beta) = \begin{pmatrix} \left(\sum_{a\in A} t_a(v_a(\mathbf{y}),y_a)v_a(\mathbf{y})\right)/\left(\sum_{w\in W} q_w\bar{\mu}_w\right) \\ \beta \end{pmatrix} \tag{25}$$

subject to

$$\max_{w\in W}\left\{\frac{\mu_w(\mathbf{y})}{\bar{\mu}_w}\right\} \leqslant \beta, \tag{26}$$

$$\sum_{a\in A} g_a(y_a) \leqslant B, \tag{27}$$

$$\alpha_{\min} \leqslant \beta \leqslant \alpha_{\max}, \tag{28}$$

$$0 \leqslant y_a \leqslant u_a, \quad a \in A, \tag{29}$$

where $\mu_w(\mathbf{y}),\ w \in W$ and $v_a(\mathbf{y}),\ a \in A$ are again the equilibrium O–D travel cost and link flow to be obtained from the optimization problem (15)–(18).

A mathematically well-defined optimal solution does not exist for the multiobjective programming problem. Therefore, the decision-maker has to choose a compromise solution among the set of available non-dominated solutions. Selection of the most preferred non-dominated solution is subjective and depends upon the decision-maker. There are a number of methods to find a non-dominated solution, for instance, the utility function method and the interactive algorithms (Weistroffer and Narula, 1991). Here we use a simple utility function method to transfer the two objectives in the upper level optimization problem into a single objective optimization problem. Namely, we transfer the objective vector (25) into the following utility function to be minimized

$$F_4(\mathbf{y},\beta) = w_1\frac{\sum_{a\in A} t_a(v_a(\mathbf{y}),y_a)v_a(\mathbf{y})}{\sum_{w\in W} q_w\bar{\mu}_w} + w_2\beta, \tag{30}$$

where w_1 and w_2 are the weights for the two objectives which are determined by the decision maker based on his/her preference. In this case, the model becomes a standard bilevel programming model.

4. A penalty function approach embodying a simulated annealing method

In the aforementioned section, we have presented the bilevel programming models to characterize the CNDP problem with the equity constraint. The equilibrium link flow $v_a(\mathbf{y}),\ a \in A$ and O–D travel time $\mu_w(\mathbf{y}),\ w \in W$ generally are non-convex, continuous and non-differentiable functions with respect to $\mathbf{y}$. This means that the CNDP problem is a non-convex minimization problem and might be difficult to solve by the standard optimization method. In addition, we need the global optimal objective value for model M1 and M2 in order to find $\alpha_{\min}$ and $\alpha_{\max}$. Hence it is necessary to employ a global optimization method such as a simulated annealing method developed by Dekkers and Aarts (1991) to solve the variant models proposed. Simulated annealing method is particularly attractive for the problem examined here, because it can find the global optimum for the optimization problem without the differentiable requirement.

Simulated annealing originated from an analogy with the physical annealing process to finding low energy states of a solid in a heat bath (Metropolis et al., 1953), it is a stochastic method to avoid getting stuck in a local, non-global optimum, when searching for a global optimum. This is made by accepting, in addition to transitions corresponding to a decrease in function value, transitions corresponding to an increase in function value. The latter is done in a limited way by means of a stochastic acceptance criterion. In the course of the minimization process, the probability of accepting deteriorations descends slowly towards zero by using the cooling schedule. These 'deteriorations' make it possible to climb out of the local optimum and explore the feasible region of the problem entirely. This procedure will lead to a (near) global optimum.

Application of simulated annealing to the minimization of a continuously valued function has been addressed by a number of authors (Kirkpatrick et al., 1983; Vanderbilt and Louie, 1984; Dekkers and Aarts, 1991; Romeijin and Smith, 1994; Jones and Forbes, 1995). We note that the simulated annealing method proposed by Vanderbilt and Louie (1984) was also used as a solution method for the conventional CNDP by Friesz et al. (1992). Nevertheless, the convergence of this simulated annealing method can not be guaranteed in principle (Boender and Romejin, 1995). Here we use the simulated annealing method introduced by Dekkers and Aarts (1991), which combines the pure random search (Rubinstein, 1981; Rinnooy and Timmer, 1984) and local search, as one part of our solution method. In the algorithm by Dekkers and Aarts (1991), a candidate point is, with probability p, generated from the uniform distribution over the feasible region of the problem. With probability $1-p$, a few iterations of a local search routine are performed from the current point. A basic requirement for the simulated annealing method is that a candidate point should be generated uniformly over the feasible region of the problem.

In view of the fact that for any given enhancement of capacity $\mathbf{y}$, the equilibrium link flows and O–D travel costs can be obtained by implementing a DUE traffic assignment procedure, we can regard $v_a(\mathbf{y})$ and $\mu_w(\mathbf{y})$ as implicit functions of $\mathbf{y}$ and regard the bilevel programming problem as single level optimization problems with respect to variable $\mathbf{y}$. In this case there remain nonlinear budget and equity constraints in the implicit single level optimization model, it is impossible to generate a feasible point uniformly over the feasible region of the problem expressed by nonlinear and implicit functions. Hence we have to use the penalty technique to transfer these complicated constraints into a penalty function and add it to the objective function. Then we can increase the penalty multiplier step by step to find a global solution of the problem (Bazaraa et al., 1993, Theorem 9.2.2). In this manner we only need to solve a minimization problem with simple bound constraints for any given penalty multiplier λ. Table 1 shows the objective functions with the penalty terms for models M1–M4. For simplicity of notation, let

$$\Omega = \{\mathbf{y} \mid 0 \leqslant y_a \leqslant u_a,\ a \in A\}. \tag{31}$$

The penalty function approach for models M1–M3 is stated below.

4.1. The penalty function method

Step 0. *Initialization*. Given a stop tolerance ε, an initial penalty multiplier λ_0, a scale parameter $\rho > 1$ and an initial point $\mathbf{y}^{(0)} \in \Omega$. Set $k = 0$.
Step 1. *Finding an optimal solution*. Starting with $\mathbf{y}^{(k)}$, solving the following problem by using a simulated annealing method described subsequently and let $\mathbf{y}^{(k+1)}$ denote an optimal solution

Table 1
The penalty terms for the bilevel programming models M1–M4

Model	Decision variable	Objective function $F(\mathbf{y})$	Constraints
M1	$\mathbf{y}$	$F_1(\mathbf{y}) + \lambda\alpha(\mathbf{y})$, where $\alpha(\mathbf{y}) = \max\left\{0, \frac{\sum_{a\in A} g_a(y_a)}{B} - 1.0\right\}$	$0 \leqslant y_a \leqslant u_a,\ a \in A$
M2	$\mathbf{y}$	$-F_2(\mathbf{y}) + \lambda\alpha(\mathbf{y})$, where $\alpha(\mathbf{y}) = \max\left\{0, \frac{\sum_{a\in A} g_a(y_a)}{B} - 1.0\right\}$	$0 \leqslant y_a \leqslant u_a,\ a \in A$
M3	$\mathbf{y}$	$F_3(\mathbf{y}) + \lambda\alpha(\mathbf{y})$, where $\alpha(\mathbf{y}) = \max\left\{0, \frac{\sum_{a\in A} g_a(y_a)}{B} - 1.0\right\} + \max\left\{0, \max_{w\in W}\{\frac{\mu_w(\mathbf{y})}{\bar{\mu}_w}\} - \beta\right\}$	$0 \leqslant y_a \leqslant u_a,\ a \in A$
M4	$\mathbf{y}, \beta$	$F_4(\mathbf{y}) + \lambda\alpha(\mathbf{y})$, where $\alpha(\mathbf{y}) = \max\left\{0, \frac{\sum_{a\in A} g_a(y_a)}{B} - 1.0\right\} + \max\left\{0, \max_{w\in W}\{\frac{\mu_w(\mathbf{y})}{\bar{\mu}_w}\} - \beta\right\}$	$0 \leqslant y_a \leqslant u_a, a \in A$, $\alpha_{\min} \leqslant \beta \leqslant \alpha_{\max}$

$$\min_{\mathbf{y}\in\Omega} \hat{F}(\mathbf{y}) = F(\mathbf{y}) + \lambda_k\alpha(\mathbf{y}). \tag{32}$$

Step 2. *Verifying the stop criterion*. If $\lambda_k\alpha(\mathbf{y}^{(k+1)}) < \varepsilon$, then stop. Otherwise, set, $\lambda_{k+1} = \rho\lambda_k$ and $k := k+1$, go to Step 1.

For any given vector $\mathbf{y}$ of the capacity enhancements, we can use the DUE traffic assignment procedure to find the equilibrium link flow $v_a(\mathbf{y}),\ a \in A$ and O–D travel cost $\mu_w(\mathbf{y}),\ w \in W$, and thus we can calculate the objective function $\hat{F}(\mathbf{y})$. With this in mind, the simulated annealing method for solving problem (32) with any given penalty λ_k is outlined below.

4.2. Simulated annealing method for problem (32) for given penalty multiplier λ_k

Step 0. *Initialization*. Given an initial point $\mathbf{y}^{(0)} \in \Omega$ and the parameter $0 < \chi_0 < 1.0$, $0 < \delta < 1,\ 0 < t < 1$, integer L_0, m_0 and T_s (a stop tolerance of temperature). Set $k = k_1 = 0$.
Step 1. *Finding an initial temperature*. Uniformly generate at random m_0 points denoted by $\mathbf{z}^{(i)}$ $(i = 1, \ldots, m_0)$ over the feasible set Ω. For each $\mathbf{z}^{(i)}$, use DUE traffic assignment to obtain the equilibrium link flow and O–D travel cost associated with $\mathbf{z}^{(i)}$ and then calculate the corresponding function value $\hat{F}(\mathbf{z}^{(i)})$. Let m_2 denote the number of points $\mathbf{z}^{(i)}$ with $\hat{F}(\mathbf{z}^{(i)}) - \hat{F}(\mathbf{y}^{(0)}) \geqslant 0$ and $\overline{\Delta\hat{F}^+}$ the average value of those $\hat{F}(\mathbf{z}^{(i)}) - \hat{F}(\mathbf{y}^{(0)})$, for which $\hat{F}(\mathbf{z}^{(i)}) - \hat{F}(\mathbf{y}^{(0)}) \geqslant 0$. Then the initial temperature T_0 is calculated as below

$$T^{(0)} = \overline{\Delta\hat{F}^+}\left(\ln\frac{m_2}{m_2\chi_0 + (1-\chi_0)(m_0 - m_2)}\right)^{-1}. \tag{33}$$

Step 2. *Verifying the termination*. If $T^{(k)} < T_s$ then stop. Otherwise, go to Step 3.

Step 3. *Checking the termination of a Markov chain*. If $k_1 > L_0N$ (N represents the number of decision variables), then go to Step 6. Otherwise, go to Step 4.

Step 4. *Generation of points*. Uniformly generate at random a number denoted by t_{random} from the internal $[0, 1)$. If $t_{\text{random}} > t$ then use the method of Hooke and Jeeves with the discrete step (Bazaraa et al., 1993, pp. 288) from the point $\mathbf{y}^{(k_1)}$ as a local search procedure to find a local solution for problem (32) denoted by $\mathbf{x}$. If $t_{\text{random}} \leqslant t$, then uniformly generate at random a point denoted by $\mathbf{x}$ over Ω.

Step 5. *Metropolis' rule*. If $\hat{F}(\mathbf{x}) < \hat{F}(\mathbf{y}^{(k_1)})$ then $\mathbf{y}^{(k_1+1)} = \mathbf{x}$, set $k_1 := k_1 + 1$ and go to Step 3. If $\exp(-(\hat{F}(\mathbf{y}^{(k_1)}) - \hat{F}(\mathbf{x}))/T^{(k)}) >$ random $[0, 1)$ then $\mathbf{y}^{(k_1+1)}) = \mathbf{x}$, set $k_1 := k_1 + 1$ and go to Step 3. Otherwise, $\mathbf{y}^{(k_1+1)} = \mathbf{y}^{(k_1)},\ k_1 := k_1 + 1$, go to Step 3.

Step 6. *Cooling schedules*. Calculate the standard derivation of the values of the objective function $\hat{F}(\mathbf{y}^{(k_1)})$ $(k_1 = 0, \ldots, L_0N)$, denoted by $\sigma(T^{(k)})$. Set the temperature as follows:

$$T^{(k+1)} = T^{(k)}\left(1 + \frac{T^{(k)} \ln(1+\delta)}{3\sigma(T^{(k)})}\right)^{-1}, \tag{34}$$

$k := k + 1,\ \mathbf{y}^{(0)} = \mathbf{y}^{(L_0N)},\ k_1 = 0$ and go to Step 2.

The convergence of this simulated annealing algorithm is proved by Dekkers and Aarts (1991). Note that the original Hooks–Jeeves method is designed for the unconstrained optimization problem. In the problem examined here, there are simple bound constraints, $0 \leqslant y_a \leqslant u_a,\ a \in A$ only, we can modify the Hooks–Jeeves method slightly to deal with this situation by projecting the trial point in the method onto the region Ω defined by the bound constraints. Furthermore, the objective function evaluation is necessarily required in the Hookes–Jeeves method. This means we need to do a DUE traffic assignment procedure at each trial point in this local search method.

Note that the above penalty function approach and the corresponding simulated annealing method is also available for model M4 although there is another decision variable β. The heavy computation burden is a main disadvantage of the simulated annealing method. With the development of powerful computer technology such as parallel computation, the simulated annealing method can solve relatively large-scale problem.

5. A numerical example

In order to demonstrate the models and the solution algorithm proposed here, we use the Sioux Falls network shown in Fig. 2 as a test example. This example has been used in Example 2 (Section 2.2) to show the inequity problem for the CNDP. The network consists of 24 nodes and 76 arcs and 552 O–D pairs. A subset containing 10 links is subject to expanding capacity, in which the link numbers are 16, 17, 19, 20, 25, 26, 29, 39, 48, 74. The link cost function, $t_a(v_a, y_a),\ a \in A$, the capacity expanding cost function of quadratic form, $g_a(y_a),\ a \in A$ and the O–D demands are presented in Suwansirikul et al. (1987). The capacity expanding costs and the equilibrium travel costs at the optimal network design solution without consideration of equity issue are also presented in Suwansirikul et al. (1987). All those data are used here.

We assume the budget available is $B = 5500$. This number is a fluctuation of the total capacity expanding costs, 5486, for a set of optimal capacity enhancements obtained by Friesz et al. (1992)

by incorporating the construction cost in the total system travel cost using an appropriate weighting factor. With these data, we obtain $\alpha_{\min} = 1.00$ and $\alpha_{\max} = 1.99$ by solving the corresponding bilevel programming models M1 and M2, respectively. Note that the travel cost ratio for the O–D pair with zero demand is not considered in the example and '$\alpha_{\min} = 1.00$' corresponds to the do-nothing alternative. This implies that any feasible scenario will generate at least one O–D pair with deteriorate equilibrium O–D travel cost than before expanding the capacity.

The optimal design solutions for different upper bounds β are illustrated in Table 2, which is obtained from model M3. By trial and error, we find that constraint (22) in model M3 becomes useless when $\beta \geqslant 1.24$ which is the critical O–D travel cost ratio α corresponding to the optimal design scenario without equity constraint. As aforementioned, a smaller value of β implies a more equitable network design. This can be seen from the coefficient of variation of O–D travel cost ratios among the O–D pairs with non-zero demand. With decreasing value of parameter β, this coefficient of variation decreases, which means a more uniform reduction in O–D travel times among all O–D pairs derived from the design scenario.

Fig. 4 shows the change of the total system costs versus the value of parameter β. From this figure, we can observe that with the increase of parameter β, the total system costs will decrease monotonically to a fixed value that is the optimal objective value for the conventional CNDP without an equity constraint. This is expected because a larger value of β means that less emphasis is placed on the equity issue and a sufficiently large β makes the equity constraint inactive.

For different combinations of the weights in model M4, the optimal solution including the threshold of O–D travel cost ratios β and the value of the weighted objective function (30) are shown in Table 3. From this table, we can see that the optimal design solution and the corresponding total system cost will vary less and less and become identical as the weight allocated to the equity objective function becomes smaller. In this case, the equity constraint becomes less influential and the resulting solution will tend to that for the conventional CNDP without incorporating the equity issue.

Table 2
Numerical results with various values of parameters β for model M3

Variable	Parameter β				
	$\beta = 1.05$	$\beta = 1.10$	$\beta = 1.15$	$\beta = 1.20$	$\beta = 1.25$
y_{16}	1.068	3.339	3.524	4.521	5.434
y_{17}	2.765	1.720	1.291	2.474	1.822
y_{19}	1.215	3.961	4.507	4.765	4.620
y_{20}	3.236	2.435	1.493	1.774	2.025
y_{25}	3.314	5.780	5.878	2.813	2.225
y_{26}	5.136	5.415	4.949	4.128	4.339
y_{29}	6.805	3.492	3.204	3.856	3.403
y_{39}	0.531	3.702	4.501	4.922	5.144
y_{48}	5.612	5.215	4.572	4.722	4.714
y_{74}	0.925	3.914	4.945	4.777	4.734
Average O–D cost ratio	0.864	0.812	0.809	0.804	0.804
Standard deviation	0.133	0.157	0.161	0.167	0.169
Coefficient of variation	0.154	0.194	0.199	0.208	0.210
Total system cost	80.939	75.456	74.933	74.346	74.234

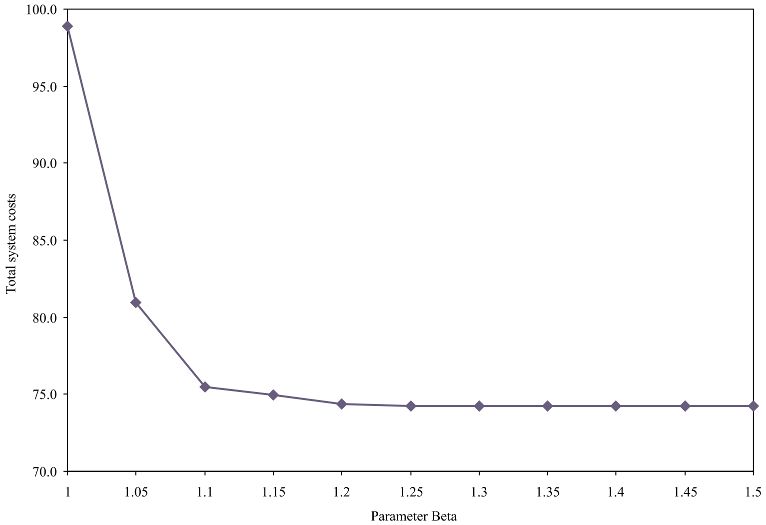

Fig. 4. The change of the total system cost versus the value of parameter β.

Table 3
Numerical results with various combinations of the weights for model M4

Variable	Combination of weights				
	$w_1 = 0.1$, $w_2 = 0.9$	$w_1 = 0.3$, $w_2 = 0.7$	$w_1 = 0.5$, $w_2 = 0.5$	$w_1 = 0.7$, $w_2 = 0.3$	$w_1 = 0.9$, $w_2 = 0.1$
y_{16}	2.590	2.590	3.923	3.923	3.923
y_{17}	1.887	1.887	2.539	2.539	2.539
y_{19}	2.688	2.688	3.544	3.544	3.544
y_{20}	4.465	4.465	1.759	1.759	1.759
y_{25}	5.697	5.697	6.774	6.774	6.774
y_{26}	7.526	7.526	4.839	4.839	4.839
y_{29}	3.524	3.524	4.616	4.616	4.616
y_{39}	3.638	3.638	4.623	4.623	4.623
y_{48}	3.385	3.385	3.224	3.224	3.224
y_{74}	3.658	3.658	3.462	3.462	3.462
β	1.078	1.078	1.111	1.111	1.111
Weighted objective function value	1.048	0.988	0.936	0.867	0.797
Total system cost	77.130	77.130	75.333	75.333	75.333

6. Conclusions

In this paper, we have posted the equity issue in the CNDPs in terms of the change of the equilibrium O–D travel cost between each O–D pair before and after implementing a network design project. We then developed bilevel programming models to deal with the CNDP with an equity constraint. In order to find the global optimal solution for these bilevel programming models, we used the penalty function approach by incorporating a simulated annealing method as a subroutine. The proposed models and algorithm have been demonstrated with the Sioux Falls network example and prove to be a meaningful tool for equity-based network design. The equity issue for the network design problem does exist in reality and should be considered to make a network improvement plan acceptable to all network users.

Acknowledgements

This study was supported by the Hong Kong Research Grants Council through a RGC-CERG Grant (HKUST6203/99E).

References

Aashtiani, H., 1979. The muti-modal traffic assignment problem, Ph.D thesis. Operation Research Center, MIT, Cambridge, MA.

Abdulaal, M., LeBlanc, L.J., 1979. Continuous equilibrium network design models. Transportation Research B 13, 19–32.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 1993. Nonlinear Programming: Theory and Algorithms. Wiley, New York.

Boender, C.G.E., Romejin, H.W., 1995. Stochastic methods. In: Horst, R., Pardalos, P.M. (Eds.), Handbook of Global Optimization. Kluwer Academic Publishers, London.

Boyce, D.E., Janson, B.N., 1980. A discrete transportation network design problem with combined trip distribution and assignment. Transportation Research B 14, 147–154.

Cho, H.J., 1988. Sensitivity analysis of equilibrium network flows and its application to the development of solution methods for equilibrium network design problems, Ph.D dissertation. University of Pennsylvania, Philadelphia.

Dafermos, S., 1980. Traffic equilibrium and variational inequalities. Transportation Science 14, 42–54.

Davis, G.A., 1994. Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. Transportation Research B 28, 61–75.

Dekkers, A., Aarts, E., 1991. Global optimization and simulated annealing. Mathematical Programming 50, 367–393.

Fisk, C.S., 1984. Optimal signal controls on congested networks. In: Proceedings of the Ninth International Symposium on Transportation and Traffic Theory. VNU Science Press, Netherlands, pp. 197–216.

Friesz, T.L., Tobin, R.L., Cho, H.J., Mehta, N.J., 1990. Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. Mathematical Programming 48, 265–284.

Friesz, T.L., Anandalingam, G., Mehta, N.J., Nam, K., Shah, S.J., Tobin, R.L., 1993. The mutiobjective equilibrium network design problem revisited: a simulated annealing approach. European Journal of Operational Research 65, 44–57.

Friesz, T.L., Cho, H.J., Mehta, N.J., Tobin, R.L., Anandalingam, G., 1992. A simulated annealing approach to the network design problem with variational inequality constraints. Transportation Science 26, 18–26.

Jones, A.E.W., Forbes, G.W., 1995. An adaptive simulated annealing algorithm for global optimization over continuous variables. Journal of Global Optimization 6, 1–37.

Kirkpatrick, C.D., Grlatt, J., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 671–680.

Leblanc, L.J., 1975. An algorithm for the discrete network design problem. Transportation Science 9, 183–199.

Marcotte, P., 1983. Network optimization with continuous control parameters. Transportation Science 17, 181–197.

Marcotte, P., Marquis, G., 1992. Efficient implementation of heuristic for the continuous network design problems. Annals of Operations Research 34, 163–176.

Meng, Q., Yang, H., Bell, M.G.H., 2001. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. Transportation Research B in press.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. The Journal of Chemical Physics 21, 2087–2092.

Poorzahedy, P., Turnquist, M.A., 1982. Approximate algorithms for the discrete network design problem. Transportation Research B 16, 45–56.

Qiu, Y., Magnanti, T.L., 1989. Sensitivity analysis for variational inequality defined on polyhedral sets. Mathematics of Operations Research 14, 410–432.

Rinnooy, K.A.H., Timmer, G.T., 1984. Stochastic methods for global optimization. American Journal of Mathematics and Management Science 4, 7–40.

Romeijin, H.E., Smith, R.L., 1994. Simulated annealing for constrained global optimization. Journal of Global Optimization 5, 101–126.

Rubinstein, R.Y., 1981. Simulation and the Monte Carlo Method. Wiley, New York.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, NJ.

Suwansirikul, C., Fiesz, T.L., Tobin, R.L., 1987. Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem. Transportation Science 21, 254–263.

Tan, H.N., Gershwin, S.B., Athans, M., 1979. Hybrid optimization in urban traffic networks. Report No. DOT-TSC-RSPA-79-7, Laboratory for Information and Decision System, MIT, Cambridge, MA.

Vanderbilt, D., Louie, S.G., 1984. A Monte Carlo simulated annealing approach to optimization over continuous variables. Journal of Computational Physics 56, 259–271.

Weistroffer, H.R., Narula, S.C., 1991. The current states of nonlinear multiple criteria decision making. In: Fandel, G., Gehring, H. (Eds.), Operation Research. Springer, Berlin, pp. 109–119.

Yang, H., Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments. Transport Review 18, 257–278.