

# **A Sustainable Road Network Design Problem with Land Use Transportation**

## **Interaction over Time**

W.Y. Szeto<sup>1\*</sup>, Y. Jiang<sup>1</sup>, D.Z.W. Wang<sup>2</sup>, A. Sumalee<sup>3</sup>

<sup>1</sup> Department of Civil Engineering, The University of Hong Kong, Hong Kong, P. R. China

<sup>2</sup> School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

<sup>3</sup> Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hong Kong, P. R. China

\* Phone: (852) 28578552, Fax: (852) 25595337, Email: ceszeto@hku.hk

### **ABSTRACT**

Sustainability has three dimensions, including social, economic, and environmental dimensions. However, existing road network design studies only focus on one or at most two dimensions, which do not allow decision makers to consider social, economic, and environmental impacts on human simultaneously. This paper proposes a multi-objective bilevel optimization model to consider all three dimensions in road network design. To examine the effect of road network design on landowner inequity and intergeneration inequity, land-use transport interaction over time is also captured in the model. The variance of discounted landowner profit and the variance of discounted generalized user cost over time are proposed as sustainability indicators of landowner inequity and intergeneration inequity respectively. Artificial bee colony algorithm (ABC) is proposed to search the network design solutions of the upper level problem, while the method of successive averages (MSA) and the Frank-Wolfe algorithm is adopted to solve the lower-level time-dependent land-use transport problem. Numerical studies are set up to illustrate the tradeoff between the three dimensions of sustainability objectives, the performance of the proposed algorithm, and the existence of landowner inequity and spatial inequity of residents.

## 1 Introduction

In Hong Kong like other major cities, many expensive road network improvement projects are still ongoing. Given the constrained government expenditure, especially for road network improvements, the government should carefully select cost-effective improvement projects to be implemented. Traditionally, the analysis involved belongs to the discipline of road network design.

In the past, many research works were done in this discipline during the last five decade (Zanjirani Farahani *et al.*, 2013). These research works can be roughly classified into several approaches. The earliest approach (e.g., LeBlanc 1975; Boyce and Janson 1980; Marcotte 1986; Chen and Alfa 1991; Friesz *et al.* 1993; Davis 1994; Meng *et al.* 2001; Chen and Yang 2004; Chen *et al.* 2006; Chen and Subprasom 2006; Chen *et al.* 2007; 2010a,b; Ban *et al.* 2006; Patil and Ukkusuri 2007; Ukkusuri *et al.* 2007; Chiou 2009; Long *et al.* 2010; Yang *et al.* 2010; Lin and Xie 2011; Farvaresh and Sepehri 2013; Miandoabchi *et al.* 2013) developed road network design models purely based on the static approach, which is based on Wardrop's principle or its stochastic extension for depicting the route choice behavior of travelers or for determining the best system performance. Some studies further considered departure time choice (e.g., Ukkusuri and Waller 2008; Chung *et al.* 2011; Lin *et al.* 2011). These two approaches allow analyzing the problem easily but cannot capture the realistic variations in demand over years, the changing network and land use pattern from time to time, and the interaction between land-use and transport.

In reality, transport interacts with land-use. When a new road is built or an existing road is widened, the travel costs between some zones decrease, and hence the accessibilities for those zones increase. Increases in the accessibilities lead to changes in population and employment distributions, and in turn a new travel demand pattern over time. The new travel demand pattern leads to a new traffic pattern and new bottlenecks, which may require further improvements in the future. As a result, the static and dynamic approaches may lead to underestimating the required capacity at some locations while overestimating the usage of transport infrastructure at other locations, because the travel demand level and distribution are altered as a result of the changes in the spatial distribution of employment and household after the improvement of transport networks is carried out. To address such issue, some network design studies (e.g., Smith and Liebman 1978; Los 1979; Feng and Lin 1999; Meng *et al.* 2000; Lee *et al.* 2006; Qiu and Chen 2007; Yim *et al.* 2011) took into account the interaction between transport and land use. They designed transport networks or modified existing ones while taking into account the spatial distribution and level of both employment and residential population, the traffic pattern, and the interaction between them. Although these works addressed the land use transport interaction, they still did not capture the changing demand over years and gradual network upgrades.

In the last 10 years, time-dependent frameworks (e.g., Lo and Szeto 2004; Szeto and Lo 2005, 2006, 2008; Kim *et al.* 2008; Ukkusuri and Patil 2009; Unnikrishnan *et al.* 2009) were proposed to deal with the changing demand and network over time. However, the land use component was still missing. Hence, this approach cannot study the impact of road network design on land use and landowners until recently, Szeto *et al.* (2010) proposed optimization frameworks for road network design considering the land-use transport interaction over time. They also focused on examining road network design's effects on the related parties under exact cost recovery, build-operate-transfer, and cross-subsidization scheme for road construction. Nevertheless, this work did not address one of the important issues directly, which is sustainability.

According to Transportation Research Board (1997), "Sustainability is not about threat analysis; sustainability is about systems analysis. Specifically, it is about how environmental,

economic, and social systems interact to their mutual advantage or disadvantage at various space-based scales of operation.” Clearly, this definition considers three dimensions of sustainability, namely social, economic, and environment dimensions. Examples of the issues involved in social, economic, and environment sustainability are, respectively, inequity, social welfare, and air pollution problems.

Concern about any dimension of sustainability is rooted in the growing awareness that human activities including transport activities that can have significant environmental impacts and can impose economic, social and ecological damages (Litman 2008). For example, exhausted gases from vehicles can lead to air pollution and global warming, which have negative impacts on people (Szeto *et al.* 2012). It is therefore important to consider sustainability into road network design to reduce the negative impacts of transport activities.

To our best knowledge, most of existing studies only focused on one or at most two dimensions of sustainability either implicitly or explicitly. In particular, a lot of static network design studies mentioned earlier (e.g., Ban *et al.* 2006; Kim *et al.* 2008; Lo and Szeto 2009) were related to the economic dimension of sustainability since they adopted total system travel time or consumer surplus in their objective function, which is an economic indicator of sustainability. Still, there were studies solely considered either the environmental or social dimension. For example, Zhou *et al.* (2008) and Huang *et al.* (2010) solely considered environmental sustainability and used total traffic emissions as an indicator. Lee *et al.* (2006) and Szeto and Lo (2008) only focused on the social dimension of sustainability. However, they differed in the social issue considered and the indicator adopted. For instance, Lee *et al.* (2006) used the relative change in origin-destination (OD) travel cost to measure spatial equity whereas Szeto and Lo (2008) adopted the consumer surplus of each OD pair at each time period before and after the improvement as an indicator of spatial equity. Although some of the above studies only considered one dimension (e.g., Chen and Subprasom 2006), they adopted more than one indicator such as mean and standard deviation of profit or social welfare, and determined Pareto optimal solutions.

For those studies only considering two dimensions of sustainability, multi-objective formulations were also proposed and different methods were adopted to deal with the tradeoff between various objectives. For example, Sharma and Mathew (2011) and Miandoabchi *et al.* (2012a,b) generated non-dominated (or Pareto) solution set while Yin and Lawphongpanich (2006), Qiu and Chen (2007), Ferguson *et al.* (2010), and Li *et al.* (2012) used the weighted sum method to combine different objectives. Others (Yang *et al.* 2010; Zhong *et al.* 2012) used side constraints, such as environmental capacity constraints, link emission constraints, and equity constraints, to handle one of the objectives or indicators. In these studies, the economic dimension of sustainability was always considered as one of the dimensions in those studies whereas either social or environment dimension was considered as the second dimension. The indicators used for measuring economic sustainability were mainly OD travel time (Miandoabchi *et al.* 2012b), total system travel time or cost (e.g., Ferguson *et al.* 2010; Sharma and Mathew 2011), and consumer surplus (Miandoabchi *et al.* 2012a), but sometimes investment cost was also included in the objective function (e.g., Qiu and Chen 2007). For environmental sustainability, the mainly indicators used were the emissions of one (Yin and Lawphongpanich 2006; Li *et al.* 2012) or more traffic pollutants (e.g., Ferguson *et al.* 2010) but energy consumption cost and land use impact cost were also used by Qiu and Chen (2007). For social dimension, one of the examples of the indicators adopted was service coverage (e.g., Miandoabchi *et al.* 2012a,b). Szeto and Lo (2006) used discounted OD generalized travel cost and discounted total social surplus per capita in each time period to measure intergeneration equity.

While it may be acceptable to use their models in some cases, their models do not allow decision makers to consider economic, social, and ecological damages simultaneously and

the resultant design does not lead to a good tradeoff between various sustainability objectives. Moreover, their models may not be flexible enough to allow decision makers to select appropriate sustainability indicators in each dimension in the model, given that there are many indicators in each sustainability dimension (see Litman (2008) for the example of indicators in each dimension).

This paper develops a flexible bilevel multi-objective road network design framework to address all the three dimensions. The framework proposed in this paper can consider multiple indicators and have multiple objectives, which can be used as a tool for the decision-making of sustainable network design. While it is possible to include as many indicators in the framework as possible, for the ease of result exposition, we consider the following four indicators in our numerical study:

1. the total vehicle emissions (a measure for the environmental dimension of sustainability),
2. the change in consumer surplus (a measure for the economic dimension of sustainability),
3. the variance of discounted landowner profit (which is a *new* measure for landowner inequity, representing one social dimension of sustainability), and
4. the variance of discounted generalized user cost over time (which is a *new* measure for the intergeneration inequity, representing another social dimension of sustainability).

These indicators are believed to be sufficient to illustrate the multi-objective nature of the problem and bring out some important issues. In particular, the numerical study shows that it may not be possible to simultaneously optimize all sustainability objectives. Tradeoffs must be carefully made between objectives.

To solve this problem, artificial bee colony algorithm (ABC) is proposed to search the network design solutions in the upper level problem, while MSA and the Frank-Wolfe algorithm are incorporated in ABC to solve the lower level time-dependent land-use transport problem.

Compared with Szeto *et al.* (2010), our paper has a different focus. Our paper focuses on the sustainability issue whereas that paper focuses examining road network design's effects on the related parties under exact cost recovery, build-operate-transfer, and cross-subsidization scheme for road construction. Hence, the upper level formulations of our paper and that paper, including constraints and the objective function, are different, although the two papers have a similar lower level formulation. Our paper further demonstrates that the lower level problem can be represented as a fixed point problem so that the solution algorithm for solving the fixed point problem can be used for solving our lower level problem. We also further prove the existence of solutions to the lower level problem, and propose a solution algorithm for solving the proposed bi-level problem that considers a general transportation network setting. Other than these, the major contributions of this paper include:

- proposing and formulating a new sustainable, multi-objective road network design problem with land use transport interaction over time,
- proposing two new measures of sustainability,
- developing a solution method for the new problem, which is NP hard,
- illustrating the tradeoff of various sustainability objectives and the possibility of having no feasible design to the problem.

The remainder of this paper is organized as follows: Section 2 describes the lower level formulation, while section 3 describes the upper level formulation and the sustainability indicators. The bilevel framework is given in section 4, followed by the proposed algorithm in section 5. Section 6 presents the numerical study. Finally, section 7 provides some concluding remarks.

## 2 The lower level problem

We consider a connected multi-modal transportation network with multiple Origin-Destination (OD) flows over the planning horizon  $[0, T]$ . The planning horizon, which can vary from 10-50 years in practice, is divided into  $N$  equal design periods, where  $N$  can also vary from 1-10 year depending on the modeling accuracy required. The network is further divided into  $M$  subnetworks, one for each mode, to account for the unique travel speed of each mode. The mode here can be an individual mode or a combined mode. With this consideration, we can formulate the proposed framework into a multi-objective bi-level problem.

The lower level problem of the bi-level problem is the time-dependent land-use transport problem and is similar to the lower level problem depicted in Szeto *et al.* (2010). There are two types of constraints for describing this problem, namely time-dependent Lowry-type constraints and time-dependent modal-split/assignment constraints.

### 2.1 Time-dependent Lowry-type constraints

The Lowry model was developed to predict the population and employment distribution (Lowry 1964). The Lowry model considers three sectors, namely *basic sector*, *non-basic sector*, and *household sector*. The basic sector represents industries that do not change their locations even there is a change in transport systems. Their clients are mainly outside the study area. The basic sector provides jobs to local residents and the number of jobs offered by the basic sector is referred to as basic employment. The non-basic sector provides services to residential community. Their locations highly depend on the location and the size of population. The number of jobs offered by the non-basic sector is referred to as service employment. The household sector consists of residential population. Their residential location depends on the location and number of basic and non-basic jobs available. Given model inputs, consisting of *composite travel cost* between zones and the spatial distribution of *basic employment* of base year (that can be obtained from travel characteristics survey (Wong *et al.* 1999)), the Lowry model determines model outputs such as the *service employment* and *residential population* of each zone.

The time-dependent Lowry-type constraints extend the Lowry model to a dynamic framework in a way that in each design period, a Lowry-type equilibrium is assumed to be held. The equilibrium conditions can be depicted by a number of constraints. The first key constraint is the following gravity-type constraint used to describe how to distribute residents to zones:

$$R_{ij,\tau} = E_{i,\tau} B_{i,\tau} W_{j,\tau}^{\alpha} \exp(-\beta^r c_{ij,\tau}), \forall i, j, \tau, \quad (1)$$

where

$$B_{i,\tau} = 1 / \sum_j W_{j,\tau}^{\alpha} \exp(-\beta^r c_{ij,\tau}), \forall i, \tau. \quad (2)$$

$R_{ij,\tau}$  is the number of work-to-home trips between OD pair  $ij$  in period  $\tau$ . This is the number of residents in zone  $j$  who work in zone  $i$  in period  $\tau$ .  $E_{i,\tau}$  is the total employment, including service employment and basic employment, in zone  $i$  in period  $\tau$ .  $W_{j,\tau}$  is the attractiveness of zone  $j$  in period  $\tau$ , which can be represented by the availability of floor space in zone  $j$  in period  $\tau$  for residential use.  $\alpha$  is the economy-of-scale parameter to regulate the attractiveness of each zone (Bureau of Transport Economics 1998).  $c_{ij,\tau}$  is the composite travel cost between OD pair  $ij$  in period  $\tau$ , representing the inter-zonal impedance and will be defined later.  $B_{i,\tau}$  is the balancing factor derived by Wilson (1970), and is used to ensure a

correct allocation of residents to zone  $j$  in period  $\tau$  so that  $\sum_j B_{i,\tau} W_{j,\tau}^\alpha \exp(-\beta^r c_{ij,\tau}) = 1$  and the total number of work-to-home trips from employment zone  $i$  in period  $\tau$  must equal the number of jobs available in that zone in period  $\tau$ ,  $E_{i,\tau}$ , or the number of people working in employment zone  $i$  in period  $\tau$  must equal the number of jobs available in that zone in the same period.  $\beta^r$  is a parameter to regulate the effect of transport cost on the distribution of residents. A high value of  $\beta^r$  will result in the residents being allocated close to their place of work.

To clearly illustrate how  $\beta^r$  regulates the effect of transport cost, we substitute (2) into (1) and re-express the resulting constraint to obtain:

$$R_{ij,\tau} = \frac{E_{i,\tau}}{1 + \sum_{j' \neq j} \frac{W_{j',\tau}^\alpha}{W_{j,\tau}^\alpha} \exp(-\beta^r (c_{ij',\tau} - c_{ij,\tau}))}, \forall i, j, \tau, i \neq j, \quad (3)$$

$$R_{ii,\tau} = \frac{E_{i,\tau}}{1 + \sum_{j' \neq i} \frac{W_{j',\tau}^\alpha}{W_{i,\tau}^\alpha} \exp(-\beta^r (c_{ij',\tau} - c_{ii,\tau}))}, \forall i, \tau.$$

By definition,  $c_{ii,\tau} < c_{ij,\tau}$ , as intrazonal travel impedance must be less than interzonal travel impedance. Hence,  $\exp(-\beta^r (c_{ij',\tau} - c_{ij,\tau}))$  tends to zero when  $\beta^r$  tends to infinity, and  $R_{ij,\tau}$  tends to  $E_{i,\tau}$ . Moreover, when  $\beta^r$  tends to infinity, some of the terms  $\exp(-\beta^r (c_{ij',\tau} - c_{ij,\tau}))$  tend to infinity, and hence  $R_{ij,\tau}$  tends to zero. These imply that if  $\beta^r$  tends to infinity, then all residents tends to live and work in the same zone. On the other hand, if  $\beta^r$  tends to zero, then the residents who work in zone  $i$  tend to live to all residential zones based on the ratio of their attractiveness to power  $\alpha$ , since  $\exp(-\beta^r (c_{ij',\tau} - c_{ij,\tau}))$  tends to one.

The second key constraint is the following gravity-type constraint used to allocate service employments to residents:

$$E_{ij,\tau}^S = sR_{j,\tau} A_{j,\tau} \tilde{W}_{i,\tau}^{\tilde{\alpha}} \exp(-\beta^s c_{ij,\tau}), \forall i, j, \tau, \quad (4)$$

where

$$A_{j,\tau} = 1 / \sum_i \tilde{W}_{i,\tau}^{\tilde{\alpha}} \exp(-\beta^s c_{ij,\tau}), \forall j, \tau. \quad (5)$$

$s$  is the service employment-to-population ratio, which is less than 1.  $E_{ij,\tau}^S$  is the number of service employees who work in zone  $i$  living in zone  $j$  or the number of service employment trips between OD pair  $ij$  in period  $\tau$ .  $R_{j,\tau}$  is the total number of residents in zone  $j$  in period  $\tau$ . The term  $sR_{j,\tau}$  in equation (4) is the total number of employees in zone  $j$  in period  $\tau$ . Similar as in Eq. (1),  $\tilde{W}_{i,\tau}$  is the attractiveness in zone  $i$  in period  $\tau$ , which can be represented by the availability of floor space in that zone at that period for commercial use.  $\tilde{\alpha}$  is the economy-of-scale parameter to regulate the attractiveness of each zone. The term  $A_{j,\tau}$  is to ensure a correct allocation of service employment to zone  $i$ , which has a similar function to the term  $B_{i,\tau}$  in (1). Similar to  $\beta^r$ ,  $\beta^s$  is a parameter to regulate the effect of transport cost but it is used to distribute service employees. A high value of  $\beta^s$  will result in service employment being allocated close to the residential location, and a small value will result in service

employment being allocated to all residential locations mainly based on the proportion of  $\tilde{W}_{i,\tau}^{\tilde{\alpha}}$ .

Other than the gravity-type constraints, the model also considers the conservation of total employment and service employment that are determined by,

$$E_{i,\tau} = E_{i,\tau}^B + E_{i,\tau}^S, \forall i, \tau, \quad (6)$$

$$E_{i,\tau}^S = \sum_j E_{ij,\tau}^S, \forall i, j, \tau, \quad (7)$$

where  $E_{i,\tau}^B$  and  $E_{i,\tau}^S$  are, respectively, the basic and service employments. In equation (6), the total employment in zone  $i$  in period  $\tau$ ,  $E_{i,\tau}$ , is the sum of the basic employment  $E_{i,\tau}^B$  and the service employment,  $E_{i,\tau}^S$ , in zone  $i$  in period  $\tau$ . In equation (7), the service employment in zone  $i$  in period  $\tau$ ,  $E_{i,\tau}^S$ , is equal to the number of service employment trips starting from zone  $i$  in period  $\tau$  or simply the number of service employees working there in that period.

In addition, the total number of residents in zone  $j$  in period  $\tau$ ,  $R_{j,\tau}$ , is defined as:

$$R_{j,\tau} = \mu \sum_i R_{ij,\tau}, \forall j, \tau, \quad (8)$$

where  $\mu$  is the population-to-employment ratio, which is intuitively not less than one. Moreover,  $\mu \leq 1/s$  because the product of  $s$  and  $\mu$  can be interpreted as the service employment-to-employment ratio and employment comprises service employment and basic employment. According to (8), in each period, the total number of employees in zone  $j$ ,  $\sum_i R_{ij,\tau}$ , multiplied by the population-to-employment  $\mu$  gives the total number of residents in that zone,  $R_{j,\tau}$ .

## 2.2 Time-dependent modal-split/assignment constraints

The time-dependent modal-split assignment constraints represent the transport model in this framework and describe the route and mode choices over time. These constraints are made up of Wardrop's conditions, travel cost constraints, as well as the modal split, flow conservation, and non-negativity conditions.

### 2.2.1 Wardrop's conditions

These conditions are supposed to be held in each design period for each mode. They require that for each mode  $k$  in each period  $\tau$ , the travel cost of each of the used routes between the same OD pair must be equal and minimal. This can be mathematically represented as:

$$f_{p,ij,\tau}^k [c_{p,ij,\tau}^k - \pi_{ij,\tau}^k] = 0, \forall p, i, j, k, \tau, \quad (9)$$

$$c_{p,ij,\tau}^k - \pi_{ij,\tau}^k \geq 0, \forall p, i, j, k, \tau, \quad (10)$$

where  $f_{p,ij,\tau}^k$  and  $c_{p,ij,\tau}^k$  are the representative hourly flow and travel cost of mode  $k$  on route  $p$  between OD pair  $ij$  in period  $\tau$ , respectively;  $\pi_{ij,\tau}^k$  is the lowest travel cost between OD pair  $ij$  by mode  $k$  in period  $\tau$ .

### 2.2.2 Travel cost constraints

Travel costs depend on flows, network characteristics such as the free flow travel times and capacities of links, and out-of-pocket costs such as tolls and fares. Route costs depend on route flows, in which the latter depends on link flows through:

$$v_{a,\tau}^k = \sum_{ij} \sum_p f_{p,ij,\tau}^k \delta_a^{p,k}, \forall a, k, \tau, \quad (11)$$

where  $v_{a,\tau}^k$  is the hourly flow of mode  $k$  on link  $a$  in period  $\tau$ , and  $\delta_a^{p,k}$  is a link-path incidence indicator for mode  $k$ , which equals one if link  $a$  is on route  $p$ , zero otherwise. Equation (11) states that for each mode  $k$ , the link flow in each period is obtained by summing the corresponding route flows on that link in that period.

The link time  $t_{a,\tau}^k$  (such as travel time, waiting time, or walking time) relates link flows through the link performance function:

$$t_{a,\tau}^k = t_{a,\tau}^k(\mathbf{v}_\tau), \quad (12)$$

where  $\mathbf{v}_\tau = [v_{a,\tau}^k]$  is the link flow vector in period  $\tau$ . This link performance function is non-separable as the travel time on link  $a$  depends on the flows on other links. Since the focus of this paper is on road network design, we give an example of the link performance function commonly used in road traffic assignment:

$$t_{a,\tau}^m = t_a^{0,m} \left[ 1 + \left( \frac{\sum_k \alpha_{a,k} v_{a,\tau}^k}{\bar{C}_{a,\tau}} \right)^{\gamma_{a,m}} \right], \forall a, m, \tau, \quad (13)$$

$$\bar{C}_{a,\tau} = C_a^0 + \sum_{\omega=1}^{\tau} y_{a,\omega}, \forall a, \tau, \quad (14)$$

where the superscript  $m$  stands for the mode that travels in the road network only (and is different from the superscript  $k$  that is used for representing any mode considered in this paper).  $t_a^{0,m}$  is the free flow travel time of mode  $m$  on link  $a$ .  $\alpha_{a,k}$  and  $\gamma_{a,k}$  are parameters of the performance function of link  $a$  for mode  $k$ .  $C_a^0$  is the initial capacity of link  $a$ .  $\bar{C}_{a,\tau}$  is the capacity of link  $a$  in period  $\tau$ .  $y_{a,\tau}$  is the capacity improvement on link  $b$  in period  $\tau$ , meaning that the capacity of link  $a$  is increased by  $y_{a,\tau}$  units at the beginning of period  $\tau$ .

Equation (13) is an asymmetric link performance function, which describes the monotonic relationship between the link travel time  $t_{a,\tau}^m$  and the link flows  $v_{a,\tau}^k$ . When  $\alpha_{a,m} = \sqrt[4]{0.15}$ ,  $\alpha_{a,k} = 0, m \neq k$ , and  $\gamma_{a,m} = 4$ , equation (13) is reduced to the typical Bureau of Public Roads (BPR) function. Equation (14) expresses the capacity of link  $a$  in period  $\tau$  as the sum of its initial capacity  $C_a^0$  and the total improvements up to period  $\tau$ .

The route cost  $c_{p,ij,\tau}^k$  in (8) and (9) is defined by summing up link attributes, which include the link tolls  $\rho_{a,\tau}^k$  and congestion-dependent attributes; for instance, travel time (and other costs such as fuel consumption) for road networks, or walking, on-board and boarding/alighting time for transit networks:

$$c_{p,ij,\tau}^k = \sum_a (\psi t_{a,\tau}^k + \rho_{a,\tau}^k) \cdot \delta_a^{p,k}, \forall p, i, j, k, \tau, \quad (15)$$



where  $\psi$  is the cost of unit (travel) time, and therefore  $\psi t_{a,\tau}^k$  is the (travel) time cost on link  $a$  by mode  $k$  in period  $\tau$ .  $\rho_{a,\tau}^k$  is the toll for mode  $k$  using link  $a$  in period  $\tau$ .

The composite travel cost between OD pair  $ij$  in period  $\tau$ ,  $c_{ij,\tau}$ , is defined as:

$$c_{ij,\tau} = \frac{-\ln \left[ \sum_k (\exp(-\bar{\beta}(\pi_{ij,\tau}^k + \theta^k))) \right]}{\bar{\beta}}, \forall i, j, \tau, \quad (16)$$

where  $\bar{\beta}$  is the parameter in the logit model to regulate the effect of the mode travel cost  $\pi_{ij,\tau}^k + \theta^k$ .  $\theta^k$  is the mode-specific cost. The composite cost is obtained by aggregating the mode travel cost  $\pi_{ij,\tau}^k + \theta^k$  over all modes  $k$ . The derivation of this composite cost can be found in Williams (1977). This composite cost will be utilized in the Lowry model (1), (2), (4) and (5) and is dependent on the lowest travel cost of each mode.

### 2.2.3 The Modal Split, Flow Conservation, and Non-Negativity Conditions

Modal split can be obtained by the logit model:

$$q_{ij,\tau}^k = R_{ij,\tau} \left( \frac{\exp(-\bar{\beta}(\pi_{ij,\tau}^k + \theta_{ij}^k))}{\sum_m \exp(-\bar{\beta}(\pi_{ij,\tau}^m + \theta_{ij}^m))} \right), \forall i, j, k, \tau, \quad (17)$$

where  $q_{ij,\tau}^k$  is the demand for mode  $k$  between OD pair  $ij$  in period  $\tau$ , and  $R_{ij,\tau}$  is the number of residents who work in zone  $i$  living in  $j$ .

The demand for mode  $k$  between OD pair  $ij$  in period  $\tau$ ,  $q_{ij,\tau}^k$ , in (17) is equal to the sum of the route flows of that mode between the OD pair in the same period so that route flows are conserved in each mode between each OD pair in each period:

$$q_{ij,\tau}^k = \sum_p f_{p,ij,\tau}^k, \forall i, j, k, \tau. \quad (18)$$

Moreover, route flow in (18) must be non-negative:

$$f_{p,ij,\tau}^k \geq 0, \forall p, i, j, k, \tau. \quad (19)$$

### 2.2.4 Lower level fixed point formulation

The lower level problem can also be formulated as a fixed-point problem. This will be shown in two major steps.

First, we will show that the time-dependent modal-split assignment problem is a fixed-point problem when  $\mathbf{R}$  is given. That is,  $\mathbf{f} = \mathbf{G}(\mathbf{R}, \mathbf{f})$ , where  $\mathbf{G}$  is a vector function. For the time-dependent modal-split traffic assignment problem, given the demand  $\mathbf{q} = [q_{ij,\tau}^k]$  and the equilibrium cost  $\boldsymbol{\pi} = [\pi_{ij,\tau}^k]$ , the flow vector  $\mathbf{f}$  can be represented as

$$\mathbf{f} = \mathbf{g}(\mathbf{q}, \boldsymbol{\pi}) \quad (20)$$

Then, based on (17),  $\mathbf{q}$  can be derived from the equilibrium cost  $\boldsymbol{\pi}$  and  $\mathbf{R} = [R_{ij,\tau}]$ , which is,

$$\mathbf{q} = \mathbf{h}(\mathbf{R}, \boldsymbol{\pi}), \quad (21)$$

where  $\mathbf{h}(\cdot)$  is the demand split function. Notice that the equilibrium cost  $\boldsymbol{\pi}$  is determined by

the flow vector  $\mathbf{f}$ , and can be represented as

$$\boldsymbol{\pi} = \mathbf{c}(\mathbf{f}), \quad (22)$$

where  $\mathbf{c}(\cdot)$  stands for cost map. Substituting (21) and (22) into (20), we have

$$\mathbf{f} = \mathbf{g}\{\mathbf{h}(\mathbf{R}, \mathbf{c}(\mathbf{f})), \mathbf{c}(\mathbf{f})\} \quad (23)$$

After simplifying, we obtain

$$\mathbf{f} = \mathbf{G}(\mathbf{R}, \mathbf{f}), \quad (24)$$

where  $\mathbf{G}(\mathbf{R}, \mathbf{f})$  contains cost map and demand split functions.

Second, we will prove that the time-dependent Lowry-type equilibrium problem is a fixed-point problem as well when  $\mathbf{f}$  is given. That is,  $\mathbf{R} = \mathbf{F}(\mathbf{R}, \mathbf{f})$ . By substituting (8) into (4), we have

$$E_{ij,\tau}^S = s\mu \sum_{i'} R_{i'j,\tau} \cdot p_{i,\tau}^S, \quad (25)$$

where  $p_{i,\tau}^S = \frac{W_{i,\tau}^\alpha \exp(-\beta^s c_{ij,\tau})}{\sum_{i'} W_{i',\tau}^\alpha \exp(-\beta^s c_{i'j,\tau})}$  and  $\sum_i p_{i,\tau}^S = 1$ .

Then, by substituting (25) into (7) and the resultant expression into (6), we get

$$E_{i,\tau} = E_{i,\tau}^B + \sum_{j'} \left( s \left( \mu \sum_{i'} R_{i'j,\tau} \right) \cdot p_{i,\tau}^S \right), \forall i, \tau. \quad (26)$$

Putting above equation into (1), it is easy to obtain

$$R_{ij,\tau} = p_{j,\tau}^R \cdot (E_{i,\tau}^B) + p_{j,\tau}^R s \cdot \mu \sum_{j'} \sum_{i'} R_{i'j',\tau} p_{i,\tau}^S. \quad (27)$$

where  $p_{j,\tau}^R = \frac{W_{j,\tau}^\alpha \exp(-\beta^r c_{ij,\tau})}{\sum_{j'} W_{j',\tau}^\alpha \exp(-\beta^r c_{ij',\tau})}$  and  $\sum_j p_{j,\tau}^R = 1$ .

Hence,

$$\mathbf{R} = \mathbf{F}(\mathbf{R}, \mathbf{f}), \quad (28)$$

where  $\mathbf{F}(\mathbf{R}, \mathbf{f}) = \left[ p_{j,\tau}^R \cdot (E_{i,\tau}^B) + p_{j,\tau}^R s \cdot \mu \sum_{j'} \sum_{i'} R_{i'j',\tau} p_{i,\tau}^S \right]$ , and  $p_{j,\tau}^R$  and  $p_{i,\tau}^S$  are functions of flows.

Therefore, by (24) and (28), we can conclude that the proposed lower level formulation is equivalent to the following fixed-point problem, which is,

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{f} \end{bmatrix} = \begin{bmatrix} \mathbf{F}(\mathbf{R}, \mathbf{f}) \\ \mathbf{G}(\mathbf{R}, \mathbf{f}) \end{bmatrix}.$$

*Proposition 1. A solution exists to the lower level problem.*

*Proof:* According to the Brouwer's fixed-point theorem, if the solution set is compact and convex, and  $\mathbf{F}(\mathbf{R}, \mathbf{f})$  and  $\mathbf{G}(\mathbf{R}, \mathbf{f})$  are continuous, then there exists at least one solution. The continuity is guaranteed because each function in the constraint is continuous over the solution space. To show that the solution set is compact and convex, we can prove that there exists a closed ball centered at  $\mathbf{0}$  with radius  $r$  that is bounded. It is because the ball is convex, and a bounded and closed set implies a compact set. This implies that we need to show that the radius of the ball is bounded but the ball includes all feasible solutions. This in turn

requires us to show that both  $\mathbf{R}$  and  $\mathbf{f}$  are bounded by the radius of the ball. Both  $\mathbf{R}$  and  $\mathbf{f}$  are bounded since

$$f_{ij,\tau}^k \leq q_{ij,\tau} \leq R_{ij,\tau} \leq \sum_j \sum_i R_{ij,\tau} = \frac{\sum_i E_i^B}{1-s\mu}, \quad (29)$$

where the first two inequalities are implied by (18) and (17) respectively; the third inequality holds by definition and the last equality is proved by the following.

By summing over  $i$  of equation (27), we have,

$$\sum_i R_{ij} = p_j^R \cdot \left( \sum_i E_i^B \right) + p_j^R \cdot \sum_i p_i^S \cdot s\mu \sum_{j'} \sum_{i'} R_{i'j'} = p_j^R \cdot \left( \sum_i E_i^B \right) + p_j^R \cdot s\mu \sum_{j'} \sum_{i'} R_{i'j'}. \quad (30)$$

Then, summing over  $j$ , we get

$$\sum_j \sum_i R_{ij} = \sum_j p_j^R \cdot \left( \sum_i E_i^B \right) + \sum_j p_j^R \cdot s\mu \sum_{j'} \sum_{i'} R_{i'j'} = \sum_i E_i^B + s\mu \sum_j \sum_i R_{ij}. \quad (31)$$

After simplifying the above equation, we get,

$$(1-s\mu) \sum_j \sum_i R_{ij} = \sum_i E_i^B. \quad (32)$$

By definition,  $s$  is the service employment-to-population ratio and  $\mu$  is the population-to-total employment ratio and they must be nonnegative. Moreover,  $s\mu$  is less than 1. Hence, we can divide both sides of equation (32) by  $(1-s\mu)$ , which gives

$$\sum_j \sum_i R_{ij} = \frac{\sum_i E_i^B}{1-s\mu}. \quad (33)$$

Hence, there exists a closed ball with  $r = \frac{\sum_i E_i^B}{1-s\mu}$  that is bounded. It is not difficult to see that

this ball allows  $\mathbf{R}$  and  $\mathbf{f}$  to be equal to the radius  $r$  because the equality sign in (29) can hold. This completes the proof.

However, the uniqueness of solution requires that the mapping function is strictly monotone, which may not be satisfied as the cost function  $\mathbf{c}(\mathbf{f})$  is not symmetric due to equation (13), which is an asymmetric cost function respect to path flow.

### 3 The upper level problem

The upper level problem considers sustainability indicators and design and financial constraints such as capacity constraints, toll constraints, and the cost recovery constraint.

#### 3.1 Design and financial constraints

##### 3.1.1 Capacity constraints

These constraints are included to address the fact that a link in a road network cannot be built or expanded beyond an upper limit due to space limitation:

$$\bar{C}_{a,\tau} \leq u_a, \forall a, \tau, \quad (34)$$

where  $u_a$  is the maximum allowable expansion capacity of link  $a$ . Equation (34) is the maximum allowable capacity constraint, which is to limit the total capacity of each link after road expansion or highway construction in period  $\tau$ ,  $\bar{C}_{a,\tau}$ , to be less than its maximum allowable capacity.

### 3.1.2 Toll constraints

These constraints cater for scenarios, such as for political reasons, that the toll  $\rho_{l,\tau}^m$  cannot be collected on certain links, or toll charges cannot be set too high. Mathematically, they can be stated as:

$$\rho_{l,\tau}^m = 0, \forall l, m, \tau, \quad (35)$$

$$\rho_{a,\tau}^m \geq 0, \forall a, m, \tau, \quad (36)$$

$$\rho_{a,\tau}^m \leq \rho_{\max}^m, \forall a, m, \tau, \quad (37)$$

where  $\rho_{\max}^m$  is the maximum allowable toll for mode  $m$  that travels on roads. The subscript  $l$  represents links where no tolls are allowed to be collected.

### 3.1.3 Cost recovery constraints

Cost recovery can be classified into three types: partial, exact, and profitable (Lo and Szeto 2009). Partial (exact) cost recovery occurs when the cost in a design period is partially (exactly) recovered by the revenue, adjusted to present value terms. Profitable cost recovery occurs when, in present value terms, the revenue is more than the cost, with a surplus or profit at the end of the planning horizon. These three cost recovery schemes can be mathematically formulated using one equation:

$$\sum_{\tau} \frac{T_{\tau}}{(1+\tilde{i})^{\tau-1}} + \sum_{\tau} \frac{S_{\tau}}{(1+\tilde{i})^{\tau-1}} - \sum_{\tau} \frac{K_{\tau}}{(1+\tilde{i})^{\tau-1}} = TOP, \quad (38)$$

where  $T_{\tau}$ ,  $S_{\tau}$ , and  $K_{\tau}$  are, respectively, the toll revenue, the subsidy (which is a model input), and the improvement and maintenance cost in period  $\tau$ .  $TOP$  is the profit or surplus of the toll road operator.  $\tilde{i}$  is the discount rate.

The first term on the left hand side (LHS) of (38) is the total discounted toll revenue for the entire planning horizon. Similarly, the second (third) term is the total discounted government subsidy (the total discounted improvement and maintenance cost). The cost recovery equation (38) requires that, in present value terms, the total toll revenue plus the total subsidy minus the total improvement and maintenance cost equals the surplus or profit. Depending on the values of  $S_{\tau}$  and  $TOP$ , equation (38) reduces to a) the partial cost recovery equation if  $S_{\tau}$  is positive and  $TOP$  is zero; b) the exact cost recovery equation if all  $S_{\tau}$  and  $TOP$  are zero; or c) the profitable cost recovery equation if all  $S_{\tau}$  are zero and  $TOP$  is positive.

The toll revenue  $T_{\tau}$  and the improvement and maintenance cost  $K_{\tau}$  in period  $\tau$  can be expressed in terms of the equilibrium link flow  $v_{a,\tau}^m$ , the toll  $\rho_{a,\tau}^m$ , and the improvement  $y_{b,\tau}$  as follows:

$$T_{\tau} = \sum_m \sum_a n v_{a,\tau}^m \rho_{a,\tau}^m, \forall \tau, \quad (39)$$

$$K_\tau = \sum_b (h_{b,\tau} + w_{b,\tau}), \forall \tau, \quad (40)$$

$$h_{b,\tau} = (1 + \tilde{j})^{\tau-1} \bar{b}_{b,0} l_b (y_{b,\tau})^{\bar{b}_{b,1}}, \forall b, \tau, \quad (41)$$

$$w_{b,\tau} = (1 + \tilde{j})^{\tau-1} \left[ \beta_{b,0} + \beta_{b,1} \left( \sum_m (nv_{b,\tau}^m) \right)^{\beta_{b,2}} \right], \forall b, \tau, \quad (42)$$

where  $h_{b,\tau}$  and  $w_{b,\tau}$  are the improvement (or construction) and maintenance cost functions of link  $b$  in period  $\tau$  respectively;  $\bar{b}_{b,0}, \bar{b}_{b,1}, \beta_{b,0}, \beta_{b,1}, \beta_{b,2}$  are the parameters of these cost functions;  $n$  converts link flows from an hourly basis to a period basis;  $\tilde{j}$  is the inflation rate;  $l_b$  is the length of link  $b$ .

Equation (39) calculates the toll revenue in period  $\tau$ , which is the sum of the product of the link flow and toll in that period. Equation (40) computes the improvement and maintenance cost in period  $\tau$  by adding the improvement and maintenance cost of all links. Equation (41) is the time-dependent improvement cost function. The term  $(1 + \tilde{j})^{\tau-1}$  represents the inflation factor: for the same capacity enhancement, the improvement cost increases by  $\tilde{j}\%$  each period. The term  $\bar{b}_{b,0} l_b (y_{b,1})^{\bar{b}_{b,1}}$  models the improvement cost of link  $b$  in period 1 (i.e., the base period). Equation (41) depicts the general relationship that the improvement cost of a link is proportional to the extent of the widening (and hence capacity gain,  $y_{b,1}$ ) and its length. This function is adopted for illustration and simplicity; other functional forms can be adopted in this framework without difficulty. Equation (42) is the time-dependent maintenance cost function, which is set to be:  $\beta_{b,0} + \beta_{b,1} \left( \sum_m (nv_{b,1}^m) \right)^{\beta_{b,2}}$  in the base period, consisting of the fixed cost  $\beta_{b,0}$  and the variable cost  $\beta_{b,1} \left( \sum_m (nv_{b,1}^m) \right)^{\beta_{b,2}}$ , in which  $\sum_m (nv_{b,1}^m)$  is the total link flow passing through link  $b$  in period 1. Again, the maintenance cost depends on the inflation factor  $(1 + \tilde{j})^{\tau-1}$ .

In this study, we consider profitable cost recovery:

$$TOP \geq 0 \text{ and } S_\tau = 0, \forall \tau. \quad (43)$$

### 3.2 Sustainability indicators

There are many sustainability indicators. Litman (2008) provides a recent list on them. This paper considers four indicators, namely total vehicle emissions, the change in consumer surplus, the variance of discounted landowner profit, and the variance of generalized user cost.

### 3.2.1 Total vehicle emissions

There are two types of vehicle emissions: link and network (or overall). For the link vehicle emissions, we consider carbon monoxides (CO) as representative major pollutants and adopt the following emission factor function:

$$e(v_{a,\tau}^m) = 0.2038 \cdot t_{a,\tau}^m \cdot \exp^{0.7962(l_a/t_{a,\tau}^m)}, \forall a, \tau, m, \quad (44)$$

where  $l_a$  is the length of link  $a$ . The above emission function has been used and discussed in Yin and Lawphongpanich (2006) and Ng and Lo (2013). According to equation (43), the vehicle emission function for mode  $m$  on a particular link depends on the traffic conditions, including travel speed and travel time, on link  $a$ . Given the emission function of link  $a$  in period  $\tau$ , the link emission  $Q_{a,\tau}^m$  and the overall vehicle emission  $Q$  can be obtained by:

$$Q_{a,\tau} = \sum_m Q_{a,\tau}^m = \sum_m e(v_{a,\tau}^m) v_{a,\tau}^m, \forall a, \tau, \quad (45)$$

$$Q = \sum_{\tau} \sum_a Q_{a,\tau}. \quad (46)$$

Equation (44) states that the total vehicle emissions on this link are given by the sum of vehicle emissions for all modes traveling on this link and for each mode, the total link vehicle emissions is the product of the emission factor and traffic flow. Equation (45) states that the overall vehicle emissions  $Q$  are the sum of the vehicle emissions on each link over time.

### 3.2.2 The change in consumer surplus

Consumer surplus (CS) measures the difference between what consumers would be willing to pay for travel and what they actually pay. It internalizes the effect of network congestion and the public's propensity to travel. For the same network and demand characteristics, a higher CS implies a better performing system. Hence, when we plan for the network improvement, the network improvement implemented should increase CS. This increase implies that the change in CS should be positive, because the CS before improvement is fixed. Therefore, we can adopt the change in CS as an alternative measure. Here, a common and simple approximation to this change, denoted as  $\Delta CS$ , is employed (Williams 1976) and can be expressed in *its present value* as follows:

$$\Delta CS = \sum_{\tau} \sum_{ij} \sum_k \frac{\Delta CS_{ij,\tau}^k}{(1+\tilde{i})^{\tau-1}}, \quad (47)$$

$$\Delta CS_{ij,\tau}^k = (1/2)(q_{ij,\tau}^{k,before} + q_{ij,\tau}^{k,after})(\pi_{ij,\tau}^{k,before} - \pi_{ij,\tau}^{k,after}), \forall i, j, k, \tau, \quad (48)$$

where the superscripts 'before' and 'after' denote before and after improvement and construction project implementations, respectively.  $\tilde{i}$  is the interest rate.  $(1+\tilde{i})^{-(\tau-1)}$  is the discount factor for period  $\tau$ . According to (47), the change in CS is the sum of the change in CS for all modes and for all OD pairs over time, discounted to its present value terms. Equation (48) is the rule-of-half definition for CS. Note that maximizing  $\Delta CS$  gives the same design as maximizing CS after network improvement as the CS before improvement is a constant. Hence, unlike other measures, we define this sustainability indicator based on two conditions, namely before and after improvements, but not just based on after network improvement.

### 3.2.3 The variance of discounted landowner profit

The variance of discounted landowner profit,  $Var(LOP)$ , is proposed to measure landowner inequity. The smaller is the variance, the lower is the degree of inequity. In the extreme case, when the variance is zero, all landowners have the same discounted profit.

Similar to the classical definition of profit, the discounted profit of land owner  $j$ ,  $LOP_j$ , is defined as follows:

$$LOP_j = \sum_{\tau} \frac{LOP_{j,\tau}}{(1+\tilde{i})^{\tau-1}}, \forall j, \quad (49)$$

where  $LOP_{j,\tau}$  represents the profit of land owner  $j$  in period  $\tau$ . The profit is the difference between the total revenue and total maintenance cost in that period:

$$LOP_{j,\tau} = R_{j,\tau} r_{j,\tau} - M_{j,\tau}^H, \forall j, \tau, \quad (50)$$

where  $R_{j,\tau}$ ,  $r_{j,\tau}$ , and  $M_{j,\tau}^H$  are the total number of residents, the rent, and the total maintenance cost on houses in zone  $j$  in period  $\tau$ , respectively. Equation (49) implies that a higher number of residents, a higher rent results or a lower value of maintenance cost results in a higher profit.

The total maintenance cost in (49) can be formulated as a linear function as follows:

$$M_{j,\tau}^H = \tilde{M}_{j,\tau} + m^h R_{j,\tau}, \forall j, \tau, \quad (51)$$

where  $\tilde{M}_{j,\tau}$  is the fixed maintenance cost on houses in residential zone  $j$  in period  $\tau$ .  $m^h$  is a parameter. Equation (50) describes the relationship that a more population leaving in the same zone results in the higher maintenance cost due to a higher utilization rate of each property in that zone.

The rent in (49) is assumed to be linearly increasing with the total population in a zone:

$$r_{j,\tau} = r_{j,\tau,\min} + \gamma_j R_{j,\tau}, \quad (52)$$

where  $r_{j,\tau,\min}$  is the minimum rent in residential zone  $j$  in period  $\tau$ , and  $\gamma_j$  is a parameter. The rent  $r_{j,\tau,\min}$  in (52) is assumed to increase over time due to inflation:

$$r_{j,\tau+1,\min} = r_{j,\tau} (1 + \tilde{j}), \quad (53)$$

where  $\tilde{j}$  is the inflation rate.

Equation (51) is indeed defined based on the situation like Hong Kong that residential land is limited. Given limited residential land, the higher the population is, the higher rent the landowner receives. We just adopted a linear function to model this characteristic. Indeed, we could adopt nonlinear functional forms in (50) and (51) but we believe that the main conclusion on the existence of an inequity issue on the profit distribution between the landowners will not be changed, as long as the function used in (50) can capture the monotonic relationship between zonal population and maintenance cost and the function used in (51) can capture the monotonic property between zonal population and rent.

Note that the profit will not affect the residential choice. The profit is only considered by landowners, not residents. Moreover, in the classical Lowry model, residents are only concerned with travel cost.

### 3.2.4 The variance of discounted generalized user cost

The sum of the variance of discounted generalized user costs of all OD pairs is proposed to measure intergeneration equity and is expressed as:

$$II = \sum_{\tau} \text{Var} \left[ \frac{c_{ij,\tau}}{(1+\tilde{i})^{\tau-1}} \right]. \quad (54)$$

The lower is the  $II$  value, the more intergeneration equitable is the design.

This social sustainability indicator,  $II$ , and social sustainability indicator for measuring the landowner inequity,  $\text{Var}(LOP)$ , are defined using the concept of variance as it is quite easy to understand. The larger the value, the more the dispersion, and the higher the degree of inequity. In social science, a similar indicator for income levels is defined using the variance of log income. However, income is not our social sustainability focus. Only intergeneration inequity of travelers and landowner inequity are our focus. Hence, we proposed two new measures for social sustainability.

#### 4 Bi-level model

The proposed bi-level model is formulated as follows:

$$\min y_i(\mathbf{x}), \quad (55)$$

subject to

$$y_j(\mathbf{x}) \leq \varepsilon_j, \forall j \neq i; \quad (56)$$

capacity, toll and cost recovery constraints (34)-(43), and;  
the lower level problem (1), (2), (4)-(19),

where  $\mathbf{x}$  is the vector of decision variables including tolls and capacity enhancements.  $\varepsilon_j$  is the aspiration level or the maximum acceptable value of the  $j$ -th sustainability indicator.  $y_i(\mathbf{x})$  is the  $i$ -th sustainability indicator which we want  $y_i(\mathbf{x})$  to be as small as possible. For example, when considering economic sustainability, we can set  $y_i(\mathbf{x}) = -\Delta CS$  since we want  $\Delta CS$  to be as large as possible. For environmental sustainability, we just simply set  $y_i(\mathbf{x}) = Q$ . Similarly,  $y_i(\mathbf{x})$  can be set to equal  $II$  or  $\text{Var}(LOP)$ . In the above framework, condition (56) is the performance constraint (or  $\varepsilon$ -constraint), which considers the indicators that do not include in (55). The  $j$ -th indicator is set to be less than the corresponding maximum acceptable value  $\varepsilon_j$  to ensure that at optimality, the  $j$ -th indicator value is at most equal to the maximum acceptable value.

To be more specific, when the negative value of consumer surplus is adopted in the objective function of the minimization program and other sustainability indicators are used in the performance constraints, the above model can be reduced to the following Consumer Surplus Maximization model, which will be used in the numerical example in Section 6.

#### Consumer surplus maximization model

$$\begin{aligned} & \min_{E,R,f,y,\rho} -\Delta CS = \max_{E,R,f,y,\rho} \Delta CS \\ \text{subject to} & \quad \text{capacity, toll, and cost recovery constraints (34)-(43);} \\ & \quad \text{the lower level problem (1), (2), (4)-(19);} \\ & \quad \text{Var}(LOP) \leq \varepsilon_{\max, \text{var} LOP}; \end{aligned} \quad (57)$$

$$II \leq \varepsilon_{\max II}, \text{ and} \quad (58)$$



$$Q \leq \varepsilon_Q, \quad (59)$$

where  $\varepsilon_{\max, \text{var } LOP}$  is the maximum acceptable variance of LOP and  $\varepsilon_{\max II}$  is the maximum acceptable variance of intergeneration equity.  $\varepsilon_Q$  is the maximum allowable total vehicle emissions. Conditions (57) and (58) are called the landowner and intergeneration inequity constraints, respectively. Inequality (58) is the total emission constraint.

## 5 Solution method

It is known that a general network design problem is *NP*-hard and extremely difficult to solve. Hence, various meta-heuristic methods have been proposed in the literature. ABC has been recently used for solving transit network design problems and is shown to perform better than GA (Szeto and Jiang 2012). In general, ABC belongs to a class of evolutionary algorithms, and are inspired by the intelligent behavior of the honeybees in finding nectar sources around their hives (Karaboga and Basturk 2008). In this paper, a newly developed hybrid Artificial Bee Colony algorithm (ABC) is adopted to solve our proposed bi-level problem, where ABC is used to solve the upper level problem, while the MSA and the Frank-Wolfe method are used to solve the lower level fixed-point problem.

In ABC, the bees are divided into three types: employed bees, onlookers and scouts. Employed bees are responsible for exploiting available food sources (i.e. solutions) and gathering required information. They also share the information with the onlookers, and the onlookers select existing food sources to be further explored. When a food source is abandoned by an employed bee, the employed bee becomes a scout and starts to search for a new food source in the vicinity of the hive. The abandonment happens when the quality of the food source is not improved after performing a maximum allowable number of iterations called *limit*. The overall algorithmic step is given below.

Step 0. Generate initial solutions and assign each of them to one employed bee; Set iteration = 0.

Step 1. Iteration = iteration + 1.

Step 2. Employed bee phase.

For  $i = 1$  to the number of employed bees.

Step 2.1. Employed bee  $i$  generates a neighbor solution.

Step 2.2. Solve the lower level problem and evaluate the fitness value of the neighbor solution.

Step 2.3. Replace the solution associated with employed bee  $i$  by its neighbor if the latter is better. Otherwise, increase the limit counter of the solution associated with employed bee  $i$  by 1.

Next

Step 3. Onlooker phase.

For  $i = 1$  to the number of onlookers

Step 3.1. Apply the roulette selection method to assign onlooker  $i$  to one of the solutions associated with employed bees and generates its neighbor solution.

Step 3.2. Solve the lower level problem and evaluate the fitness of the neighbor solution.

Step 3.3. Replace the solution associated with employed bee  $i$  by its neighbor if the latter is better. Otherwise, increase the limit counter of the solution associated with employed bee  $i$  by 1.

Next

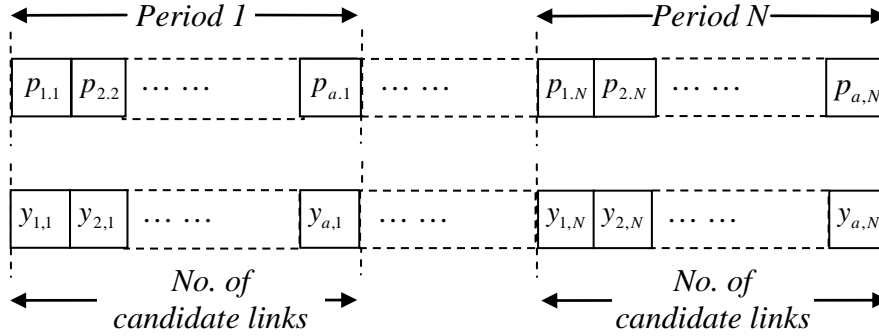
Step 4. Scout phase.

Step 4.1. Identify the abandoned solutions, whose limit counters are greater than limit.

Step 4.2. Generate new solutions and replace the abandoned ones.

Step 5. If iteration is less than a predefined maximum value, then return to step 1, otherwise, stop and output best solutions.

In order to implement the proposed hybrid ABC to solve our problem, we specifically designed the solution representation. Figure 1 illustrates the representation scheme. In one solution, two chromosomes are used to represent the toll charging plan,  $p_{a,\tau}^k$ , and the capacity improvements,  $y_{a,\tau}$ , at period  $\tau$ , respectively. Each chromosome is divided into  $N$  parts, while the number of genes in each part equals the number of candidate links considered to expand.



**Figure 1 Solution representation scheme**

To generate initial solutions in ABC, a link based generating procedure is proposed. For each candidate link, a random number between 1 and  $N + 1$  is produced to determine the timing and magnitude of capacity expansion, where  $N + 1$  means no expansion. Next, the expansion plan is obtained period by period based on the remaining allowable capacity defined as the difference between the total allowable capacity and the cumulative improved capacity. Then, we set the toll level on the links before capacity expansion to be zero. The toll in the first year of the expansion is set randomly by choosing a random between 0 and the maximum allowable toll. In the following periods, the toll either remains constant or is increased by a fixed increment, while maintaining the toll to be less than the maximum allowable value. Following these procedures, the toll constraints and capacity constraints are satisfied automatically.

The other constraints are handled by introducing penalty terms in the fitness function  $f$  defined as

$$f = \Delta CS + \max(-TOP, 0.0) - \sum_i p_i \frac{\max((y_i - \varepsilon_i), 0.0)}{\varepsilon_i} + C ,$$

where  $C$  is a large constant to avoid producing negative fitness values, and  $p_i$  is the penalty value for sustainability constraint  $i$ . If the  $i$ -th indicator  $y_i$  violates the constraint  $y_i(\mathbf{x}) \leq \varepsilon_i$ , then the penalty term  $\max((y_i - \varepsilon_i), 0.0)$  becomes a positive value to lower the fitness value. Since the magnitudes of different indicator values vary a lot, each violation is normalized by  $\varepsilon_i$  to scale its value. If  $TOP$  is negative (i.e., the cost recovery constraint is violated), then the penalty term  $\max(-TOP, 0.0)$  becomes positive. If no constraint is violated, then the fitness value will not be adjusted.

The neighbour solutions are generated as follows: first, a candidate link is randomly selected; then its toll charging and capacity expansion plan are redesigned by adopting the same method in the solution generation procedure mentioned in the last paragraph. In such condition, no specific remedy procedures are required to ensure the toll and capacity constraints being satisfied.

To solve the lower level problem, the Method of Successive Averages (MSA) is applied twice for each period. The algorithm procedures for each period  $\tau$  are described as follows, in which  $C_R$  and  $C_q$  are defined as iteration counters and their reciprocals are adopted as step sizes, respectively.

Step 0. Obtain upper level solutions from ABC.

Step 1. Initialize  $R_{ij,\tau}, \forall ij$  by setting them to be equal to corresponding base year number of residents and set  $C_R = 0$ .

Step 2. Set  $C_R = C_R + 1$ .

Step 3. Solve the modal split/assignment problem

Step 3.0. Initialize  $q_{ij,\tau}^k, \forall ij, k$  by evenly distributing  $R_{ij,\tau}$  to each mode, and set  $C_q = 0$ .

Step 3.1. Set  $C_q = C_q + 1$ .

Step 3.2. Solve UE assignment for each sub-network by a classical method such as the Frank-Wolfe algorithm and obtain  $\pi_{ij,\tau}^k, \forall ij, k$ .

Step 3.3. Calculate the composite cost  $c_{ij}, \forall ij$  according to Eq. (16).

Step 3.4. Calculate  $q_{ij,\tau}^k, \forall ij, k$  based on  $c_{ij}$  obtained in Step 3.3 according to Eq. (17).

Step 3.5. Set  $q_{ij,\tau}^k = q_{ij,\tau}^k + \frac{1}{C_q} (q_{ij,\tau}^k - q_{ij,\tau}^k), \forall ij, k$ .

$$\frac{\sqrt{\sum_{ij} \sum_k (q_{ij,\tau}^k - q_{ij,\tau}^k)^2}}{\sum_{ij} \sum_k q_{ij,\tau}^k} \leq \varepsilon$$

Step 3.6. If  $\frac{\sqrt{\sum_{ij} \sum_k (q_{ij,\tau}^k - q_{ij,\tau}^k)^2}}{\sum_{ij} \sum_k q_{ij,\tau}^k} \leq \varepsilon$ , where  $\varepsilon$  is an acceptable error, then go to step 4. Otherwise go to step 3.1.

Step 4. Calculate  $R_{ij,\tau}, \forall ij$  based on  $c_{ij}$  obtained in Step 3.3 and  $E_{i,\tau}^B$  by Eqs. (1)-(2), (4)-(7).

Step 5. Set  $R_{ij,\tau} = R_{ij,\tau} + \frac{1}{C_R} (R_{ij,\tau} - R_{ij,\tau}), \forall ij$ .

$$\frac{\sqrt{\sum_{ij} (R_{ij,\tau} - R_{ij,\tau})^2}}{\sum_{ij} R_{ij,\tau}^k} \leq \varepsilon$$

Step 6. If  $\frac{\sqrt{\sum_{ij} (R_{ij,\tau} - R_{ij,\tau})^2}}{\sum_{ij} R_{ij,\tau}^k} \leq \varepsilon$ , then return to ABC. Otherwise, go to step 2.

## 6 Numerical studies

Two examples are created to demonstrate the properties of the proposed model and the effectiveness of the proposed algorithm. First, a small network is used to illustrate the trade-off between different objectives. Then, a modified Sioux-Falls network is employed to investigate the distribution pattern of residents, the spatial equity issue, and the change in

consumer surplus with and without capacity expansion. The computation performance is also depicted. The algorithm was coded and compiled by CVF 6.5 and ran on a computer with 2.66GH CPU.

### 6.1 Example 1

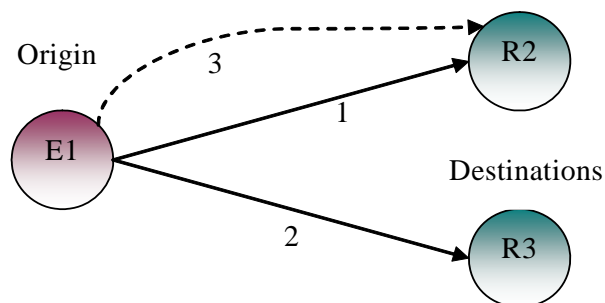
For the ease of exposition, a simple network is adopted as shown in figure 2. There are 3 links in this network: link 1, link 2, and link 3. Links 1 and 2 are links whose travel times are given by the BPR functions. Link 3 is a separate transit link, as represented by a dash line in the figure. There are 3 zones too: E1, R2, and R3, in which 'E' stands for an employment zone whereas 'R' stands for a residential zone. The attractiveness of each zone is assumed to follow the following function:

$$W_{i,\tau+1} = W_{i,\tau} (1 + \tilde{h}_{w,i}),$$

where  $\tilde{h}_{w,i}$  is the growth rate of attractiveness of zone  $i$  over time. The basic employment in the employment zone is supposed to grow linearly over time:

$$E_{i,\tau+1}^B = E_{i,\tau}^B (1 + \tilde{h}_{E,i}),$$

where  $\tilde{h}_{E,i}$  is the growth rate of basic employment. The three zones form two OD pairs: E1-R2 and E1-R3. Both OD pairs are connected by highways but only OD pair E1-R2 has a segregated transit connection. In other words, there are two modes for OD pair E1-R2 but there is only one mode for OD pair E1-R3. It is proposed that link 1 is widened and collects toll to recover the improvement and maintenance cost. Mode 1 is car mode and mode 2 is transit mode.



**Figure 2 The scenario network**

The parameters in this study include: 1) Land use parameters:  $E_{1,1}^B = 5000$  jobs;  $W_{1,1} = 3000$  jobs;  $W_{2,1} = W_{3,1} = 3000$  houses;  $\mu = 5$ ;  $\alpha = 1$ ;  $s = 0.1$ ;  $\beta^r = 0.04 \text{€}^{-1}$ ;  $\beta^s = 0.03 \text{€}^{-1}$ ;  $\tilde{M}_{j,\tau} = \text{€}100$ ;  $m^h = \text{€}0.01$  /household;  $\gamma_2 = \gamma_3 = \text{€}$ ;  $\tilde{h}_{w,1} = \tilde{h}_{w,2} = \tilde{h}_{w,3} = 0.05$ ;  $h_{E,1} = 0.04$ ;  $r_{2,\tau,\min} = r_{3,\tau,\min} = \text{€}1000$ ; 2) Transport network parameters:  $c_1^0 = c_2^0 = 3000$  vph;  $u_1^1 = 7500$  vph;  $u_2^1 = 0$  vph;  $t_1^0 = t_2^0 = 5$  hours;  $t_3^0 = 4$  hours;  $l_1 = l_2 = 400$  km per hour 3)  $Y_\tau^2 = 1000000 \text{€}$  4) Parameters of improvement cost functions:  $b_{1,1} = b_{2,1} = 1$ ;  $b_{1,0} = b_{2,0} = \text{€}2000$ ; 5) Parameters of maintenance cost functions:  $\beta_{1,0} = \text{€}1200$ ;  $\beta_{1,1} = \text{€}0.001$ ;  $\beta_{1,2} = 1$ ; 6) Parameters in travel cost functions:  $\psi = \text{€}15/\text{h}$ ;  $\theta^1 = \text{€}16$ ;  $\theta^2 = \text{€}30$ ;  $\bar{\beta} = 0.05 \text{€}^{-1}$ ; 7) Interest and inflation rates:  $\tilde{i} = 0.03$ ;  $\tilde{j} = 0.01$ ; 8) Converting factor:  $n = 365$  days  $\times$  24

hours  $\times$  1 year = 8760 hours/period; 9) Length of each period: 1 year; 10) Planning horizon and franchised period:  $[0,5]$ ; 11) The transit fare on link 3 = €40; the toll on link 2,  $\rho_{2,\tau} = \text{€}$ ; and  $\rho_{\max}^1 = \text{€}$ . These values are chosen for illustrative purposes.

To clearly illustrate the tradeoff between objectives graphically, in this example, we restrict that the toll is constant through the planning horizon and the capacity improvement is done in the first period. The lower level model was solved for each combination of toll and capacity improvement to obtain the total vehicle emissions, the variance of discounted generalized user cost, the variance of LOP, and the change in consumer surplus. The results are plotted in Figure 3, respectively. The x-axis is for toll charge whereas the y-axis is for the capacity improvement. The large shaded triangle region represents the infeasible region where cost recovery cannot be achieved or the profit is less than zero. The small shaded triangle region represents the infeasible region where the objective value, the consumer surplus, after implementing the design is worse than that before implementation. The inclined straight line is the zero profit line where exact cost recovery can be achieved. The cross represents the optimal objective value in the feasible domain. Although the example setting is quite specific, Figure 3 does illustrate that there may be a tradeoff between optimizing different sustainability indicators as can be seen later.

As you can see from figure 3a, for a given toll charge, the total vehicle emissions decrease with increasing capacity improvement. Moreover, for a given capacity improvement, the total vehicle emissions decrease with an increasing toll level. If our objective is minimizing the total vehicle emissions, then from the graph, the optimal solution point is (6, 350). This is because higher toll or lower capacity improvement means higher travel cost on link 1, resulting in decreasing number of trips on that link, and increasing the number of trips on link 2. Meanwhile, the reduction of the emission factor of link 1 is larger than the increment of the emission factor of link 2 under our scenario setting. Accordingly, the emission reduction of link 1 is larger than the emission increment of link 2. Hence, the overall network emission reduces.

According to Figure 3b, decreasing capacity improvement or increasing toll charge reduces the variance of discounted LandOwner Profit (LOP). The optimal solutions for minimizing the variance of LOP is at (6, 350), which is identical to the optimal solution of minimizing total vehicle emissions. This implies that if we minimize total vehicle emissions, we can achieve the minimum variance of LOP simultaneously for *this particular case*. However, in general, there may be a tradeoff between these two objectives. This is left for future studies.

Figure 3c shows that decreasing toll charge or capacity improvement lowers the variance of discounted generalized user cost in general. However, it is beneficial to have a small increase in capacity. Therefore, the minimum variance of generalized user cost occurs at (0.4, 1500). This optimal solution is far away from (6, 350), implying that minimizing the variance of discounted generalized user cost can be in conflict with minimizing the variance of LOP and total vehicle emissions.

From Figure 3d, we can see that decreasing capacity improvement or increasing toll charge lowers the change in consumer surplus. The optimal solution is at (1.7, 7500), which is different from previous optimal solutions. This means that the objective of maximizing the change in CS can be in conflict with the objectives of minimizing total vehicle emissions and the variances of discounted generalized user cost and LOP. Simultaneous satisfying all objectives may not be possible in general situations. Tradeoffs must be made between these objectives.

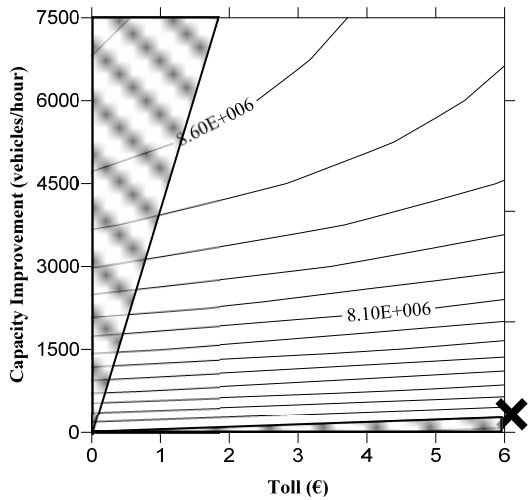


FIGURE 3(a) Total vehicle emissions.

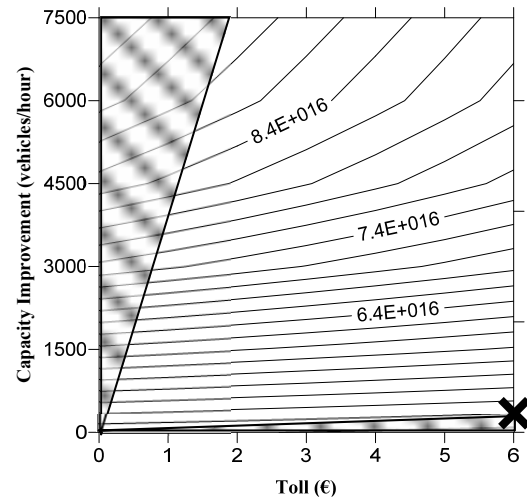


FIGURE 3(b) Variance of LOP.

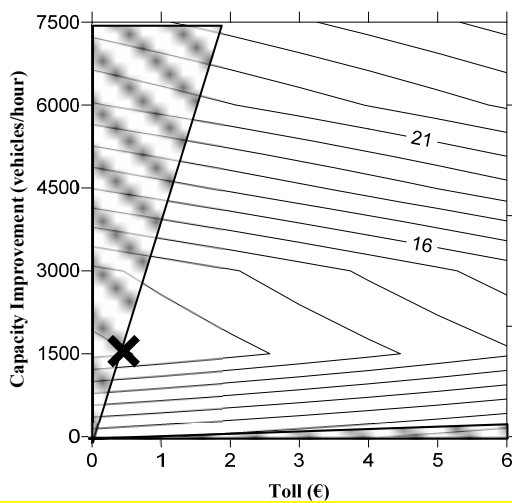


FIGURE 3(c) Variance of discounted generalized user cost

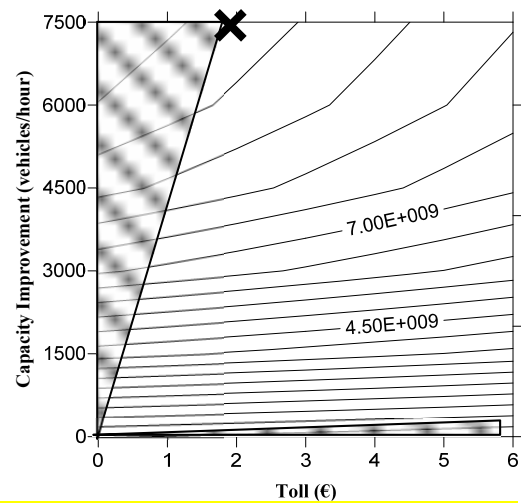


FIGURE 3(d) The change in consumer surplus.

Figure 3 Measures of sustainability.

To determine the optimal network design that accounts for the tradeoffs, we need to solve multi-objective optimization models. For illustrative purposes, we consider the two conflicting objectives: maximizing the change in CS and minimizing the variance of LOP. The consumer maximization model was solved many times by varying maximum acceptable variance of LOP while the parameters in the other two performance constraints were set to be very large (i.e. were inactive). Figure 4 shows the optimal change in CS for various maximum acceptable variance of LOP. From this figure, we can observe that there are three portions of the curve: 1) no feasible design region, 2) tradeoff region, and 3) no tradeoff region. The first region represents the case where there is no design that can give a variance equal to or lower than the maximum acceptable variance of LOP because the change in CS is negative or the CS after improvement is lower than that before. The second region represents the case where the landowner inequity constraint is binding and a tradeoff is required between the two objectives. In this region, the change in CS increases nonlinearly with an increasing variance, indicating that one sustainability objective value is improved when the other sustainability criterion is relaxed. The third region represents the case where the landowner equity constraint is non-binding but the cost recovery constraint is binding. The

change in CS in this region is always at its maximum value when the maximum acceptable variance of LOP increases because no further capacity improvement can be achieved.

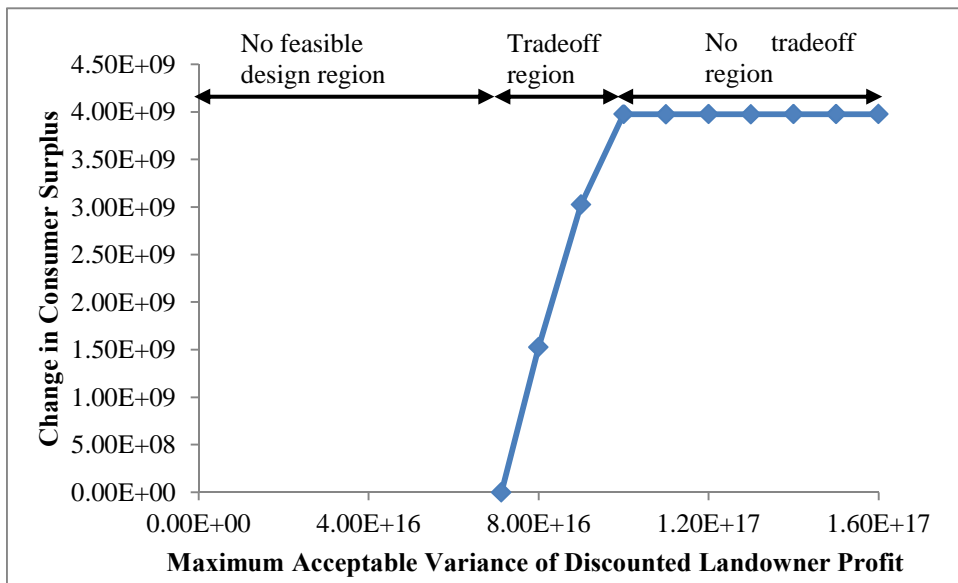


Figure 4 Trade-off between the change in CS and maximum acceptable variance of discounted LOP.

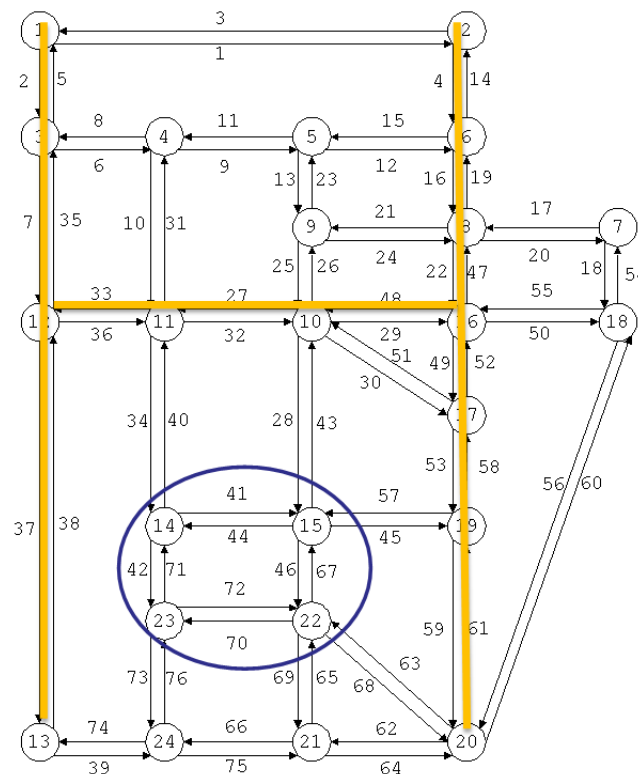
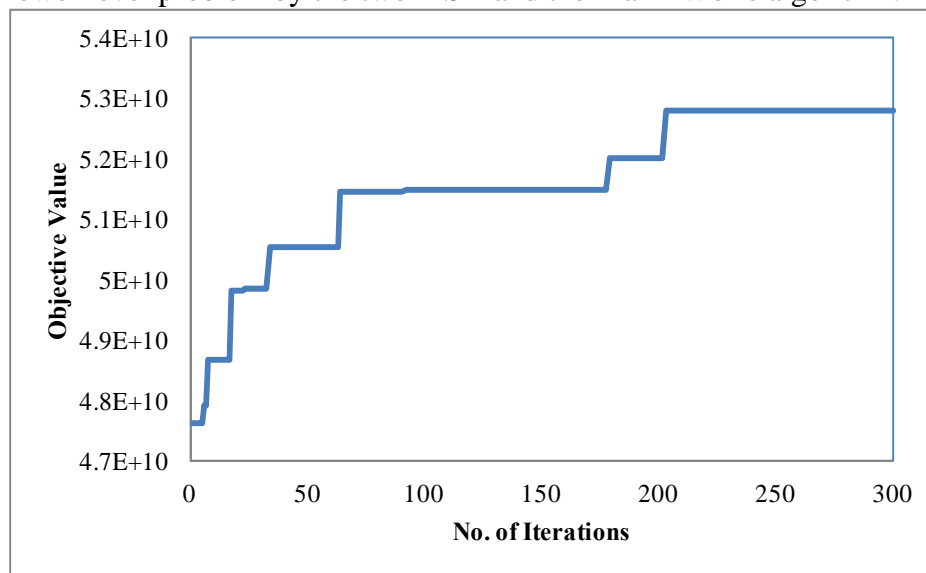


Figure 5 Modified Sioux-Falls network with transit lines

## 6.2 Example 2

The Sioux Falls network is used in this example but to cater the multimodal feature, a public transit network is created on the Sioux Falls network and marked by the yellow lines as shown in figure 4. The blue circle is the proposed developed area, where eight candidate links are included. For simplicity, the attractiveness of each zone is assumed to be equal, so as the emission parameter of every link. The other parameters are the same as those in example 1, except  $\mu = 2.0$ ,  $\beta^r = 0.02$ , and  $\beta^s = 0.01$ .

The parameters used in ABC were set as follows: The colony size was set to 20, composing of 10 employed bees and 10 onlookers. The stopping criteria of both MSA were  $\varepsilon = 0.001$  and the maximum number of iterations in MSA was 500. The performance of the ABC algorithm is shown in figure 6, from which, we can see that the objective value converged in a stepwise manner. It took 6169.95 seconds to conduct 300 major iterations in ABC. Such as long computation time is mainly due to the computation time required to solve the lower level problem by the two MSA and the Frank-Wolfe algorithm.



**Figure 6 Convergence of ABC**

Figure 7 plots the relative increase of consumer surplus, in which, for each OD pair, the relative increase in CS is obtained by dividing the maximum increase of CS among all OD pairs. Then, the maximum relative increase of CS must be equal to 1 or 100%. According to the results, OD pair 22-15 is found to have such value. For such case, the route (i.e., link 67) was highly congested before expansion, and hence, after improvement, travelers or residents will benefit significantly from the travel cost reduction.

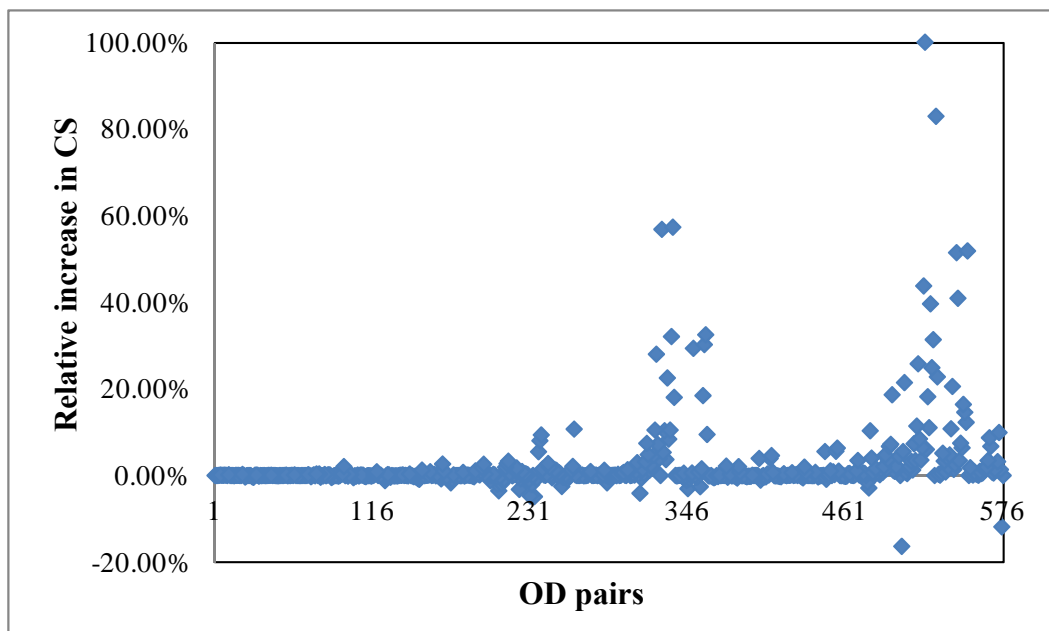
Negative values in Figure 7 represent a reduction of CS. The maximum reduction occurs at OD pair 21-22. This is because link 65, one of the entrance corridors connecting the circled area, carries much more flow after expansion as more residents are attracted to the selected area. Therefore, the travelers of OD pair 21-22 directly suffer from the large cost increment of link 65. Similar scenarios are observed at OD pairs 10-15, 11-14, 19-15, 20-22, and 24-23 as well. This phenomenon indicates that the expansion of the selected area can deteriorate the traffic condition on its corridor links, such as link 65, and cause CS reduction of certain OD pairs. However, such lost is minor when compared to the total CS increase obtained by aggregating the change in CS over all OD pairs so that such phenomenon can occur. Moreover, even for those OD pairs with positive increase, the magnitudes of increment vary



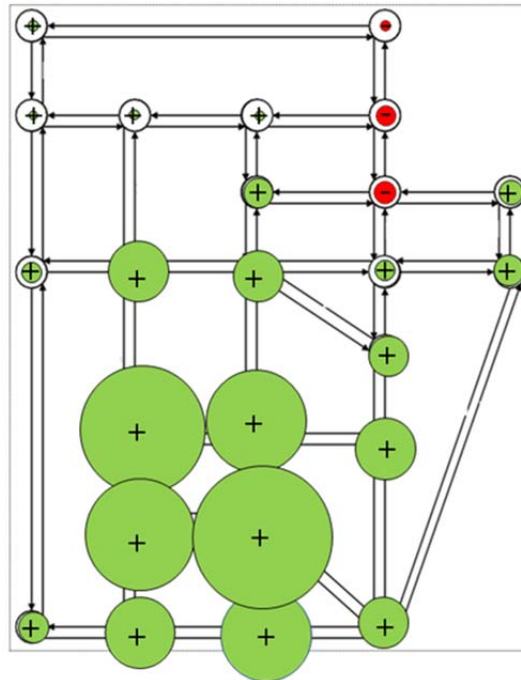
substantially. This spatial inequity issue, as measured by the relative change in CS of each OD pair, should also be considered in network design problems.

More importantly, there is one more spatial inequity issue, which is at the node level. This inequity can be measured by the total change in CS with respect to a node (residential zone) where the total change in CS for the node is the sum of the change in CS over destination, time, and mode, discounted to present value terms ( $\sum_{\tau} \sum_j \sum_k \frac{\Delta CS_{ij,\tau}^k}{(1+\tilde{i})^{\tau-1}}$ ). It represents the

change in the benefit of residents of that zone. This inequity can represent the realistic situation better as the residents do have a choice of destination or workplace, in addition to mode and route choice. Figure 8 illustrates the existence of spatial inequity at the node level. The green circles with “+” sign indicate an increment of CS, while the red ones with “-” represent the decrement. Meanwhile, the radius of the circle represents to the magnitude of the change. Despite the fact that most of the residents are benefitted from network improvement, the residents living at nodes 2, 6, and 8 are suffered. Even though most of the residents are benefitted, they are not benefitted to the same extent. To handle these spatial inequity issues, we can add suitable inequity indicators or constraints in our formulation without conceptual difficulty.



**Figure 7** Relative increase of different OD pairs



**Figure 8 Total change in CS at each node**

Figures 9 and 10 plot the differences of  $R_j$  and  $E_i^S$  at the final period between the scenario with expansion and that without. Similar to figure 8, the green circle with “+” sign indicate an increment of  $R_j$  or  $E_i^S$ , while the red ones with “-” represent the decrement; the radius of the circle represents to the extent of the change. Figure 9 illustrates that after expansion, residents are attracted to work in the selected areas. In general, the reason is the decrease of generalized travel cost due to capacity improvement. In addition, the change of the number of residents varies a lot across different nodes, which depends on the influence of generalized travel cost for each OD pair case by case. As results of residents’ redistribution, the profit of landowner changes according to equations (49) to (53). Practically speaking, figure 9 can also be used to demonstrate the equity issue of landowners. The landowners of green nodes are expected to gain profit, while the landowners of red nodes are expected to lose profit. This may raise the issue that most of landowners oppose to the network improvement. The impact of loss can be handled by incorporating another constraint in the proposed framework to ensure that the difference of LOP after and before network improvement cannot be too negative. Similar patterns can be observed in figure 10, which shows that the service employment opportunities increase around the proposed developed area. This is consistent with the Lowry model that an increment of residents requires more service employee to match up.

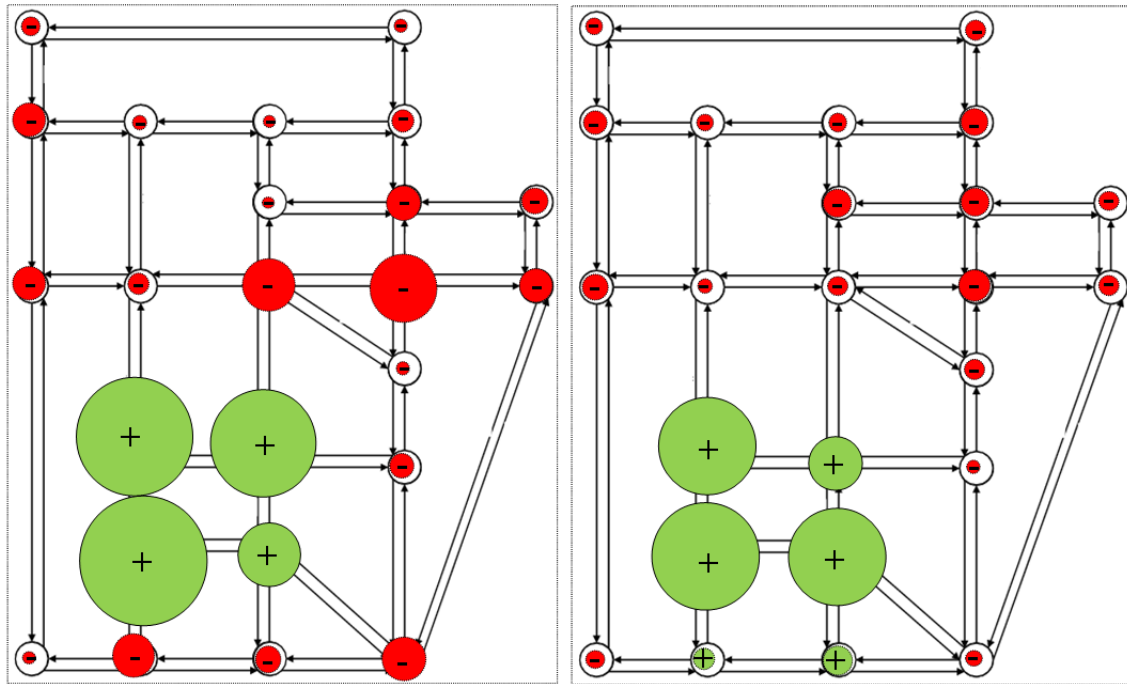


Figure 9 Change of  $R_j$

Figure 10 Change of  $E_i^S$

## 7 Concluding remarks

This paper develops a multi-objective time-dependent road network design model to consider land-use transport interaction over time and sustainability. The variance of discounted landowner profit is proposed as one of the measure of landowner inequity while the variance of discounted generalized user cost is proposed as a measure of intergeneration inequity. A hybrid artificial bee colony algorithm is proposed to solve the network design problem, while MSA and Frank-Wolfe algorithm is adopted to handle lower level constraints. Numerical studies are set up to illustrate the multi-objective nature of the problem and the pattern of residential distribution with capacity improvement. In particular, the result shows that it may not be possible to simultaneously optimize all sustainability objectives. Tradeoffs must be carefully made between these objectives. The results also show that the proposed developed area tends to absorb more travel demand and produce more service employment opportunities, which agree with our expectation. More importantly, we found that although capacity expansion is expected to increase the overall consumer surplus, the benefits of different OD pairs and at different locations varies a lot. This raises the issue of spatial inequity.

This paper opens up many research directions. First, this paper does not consider heterogeneous values of time, mixed routing strategies, and the elasticities of housing demand and supply. One can incorporate them into the proposed bi-level framework in future studies. Second, in reality, demand and supply are uncertain. Extending the proposed framework to capture the uncertainties is an important future research direction. Finally, most of computation time was used in solving the lower-level problem. How to improve the proposed solution scheme to avoid solving many lower-level problems is another worthy direction.

## Acknowledgments

The research was jointly supported by a grant (200902172003) from the Hui Oi Chow Trust Fund, grants (201011159026 and 201211159009) from the University Research Committee and Faculty of Engineering Top-up Grant of the University of Hong Kong, and a grant from National Natural Science Foundation of China (71271183). The authors are grateful for the constructive comments of the two reviewers.

## References

- Ban JX, Liu, HX, Ferris MC, Ran B (2006) A general MPCC model and its solution algorithm for continuous network design problem. *Math Comput Model* 43:493-505
- Boyce DE, Janson, BN (1980) A discrete transportation network design problem with combined trip distribution and assignment. *Transp Res Part B* 14(1):147-154
- Bureau of Transport Economics (1998) *Working paper 39: Urban Transport Models: A Review*. Department of Transport and Regional Services, Canberra, Australia.
- Chen A, Kim J Lee S, Kim Y (2010a) Stochastic multi-objective models for network design problem. *Expert Syst Appl* 37(2):1608-1619
- Chen A, Kim J, Zhou Z, Chootinan P (2007) Alpha reliable network design problem. *Transp Res Rec* 2029:49-57
- Chen A, Pravinongvuth S, Chootinan P (2010b) Scenario-based multi-objective AVI reader location models under different travel demand patterns. *Transp* 6(1):53-78
- Chen A, Subprasom K, Ji Z (2006) A simulation-based multi-objective genetic algorithm (smoga) for build-operate-transfer network design problem. *Optim Eng J* 7(3):225-247
- Chen A, Yang C (2004) Stochastic transportation network design problem with spatial equity constraint. *Transp Res Rec* 1882:97-104
- Chen MY, Alfa AS (1991) A network design algorithm using a stochastic incremental traffic assignment approach. *Transp Sci* 25(3):215-224
- Chiou SW (2009) A subgradient optimization model for continuous road network design problem. *Appl Math Model* 33(3):1386-1396
- Davis GA (1994) Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. *Transp Res Part B* 28(1):61-75
- Do Chung B, Yao T, Xie C, Thorsen A (2011) Robust Optimization Model for a Dynamic Network Design Problem Under Demand Uncertainty. *Netw Spat Econ* 11(2):371-389
- Farvaresh H, Sepehri MM (2013) A branch and bound algorithm for bi-level discrete network design problem. *Netw Spat Econ* 13: 67-106
- Feng CM, Lin JJ (1999) A land use-network design model to generate alternative sketch maps for urban planning. *J East Asia Soc Transp Stud* 3(4):71-86
- Ferguson EM, Duthie J, Waller ST (2010) Network methods for project selection based on optimizing environmental impact. Research Report SWUTC/10/161026-1 (Austin: University of Texas at Austin).
- Friesz TL, Anandalingam G, Mehta NJ, Nam K, Shah SJ, Tobin RL (1993) The multiobjective equilibrium network design problem revisited: A simulated annealing approach. *Eur J of Oper Res* 65(1):44-57
- Huang K, Zhang J, He M, Liao W (2010) An optimal model and solution algorithm of urban traffic network considering exhaust emission control. In *Proceedings of the 2010 International Conference of Logistics Engineering and Management*: 528-534
- Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. *Appl Soft Comput* 8(1):687-697

- Kim BJ, Kim WK, Song BH (2008) Sequencing and scheduling highway network expansion using a discrete network design model. *Ann Reg Sci* 42(3):621-642
- LeBlanc LJ (1975) An algorithm for discrete network design problem. *Transp Sci* 9(3):183-199
- Lee DH, Wu L, Meng Q (2006) Equity based land-use and transportation problem. *J Adv Transp* 40(1):75-93
- Li ZC, Lam WHK, Wong SC, Sumalee A (2012) Environmentally sustainable toll design for congested road networks with uncertain demand. *Int J Sustain Transp* 6(3): 127-155
- Lin DY, Karoonsoontawong A, Waller ST (2011) A Dantzig-Wolfe decomposition based heuristic scheme for bi-level dynamic network design problem. *Netw Spat Econ* 11(1):101-126
- Lin DY, Xie C (2011) The Pareto-optimal solution set of the equilibrium network design problem with multiple commensurate objectives. *Netw Spat Econ* 11(4):727-751
- Litman T (2008) *Well Measured: Developing Indicators for Comprehensive and Sustainable Transport Planning*. Victoria Transport Policy Institute ([www.vtpi.org](http://www.vtpi.org)). [www.vtpi.org/wellmeas.pdf](http://www.vtpi.org/wellmeas.pdf).
- Lo HK, Szeto WY (2009) Time-dependent transport network design under cost-recovery. *Transp Res Part B* 43(1):142-158
- Long JC, Gao ZY, Zhang HZ, Szeto WY (2010) A turning restriction design problem in urban road networks. *Eur J of Oper Res* 206(3):569-578
- Los M (1979) A discrete-convex programming approach to the simultaneous optimization of land use and transportation. *Transp Res Part B* 13(1):33-48
- Marcotte P (1986) Network design problem with congestion effects: a case of bilevel programming. *Math Program* 34(2):142-162
- Meng Q, Yang H, Bell MGH (2001) An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. *Transp Res Part B* 35(1):83-105
- Meng Q, Yang H, Wong SC (2000) A combined land-use and transportation model for work trips. *Environ Plan B* 27(1):93-103
- Miandoabchi E, Daneshzand F, Szeto WY, Zanjirani Farahani R (2013) Multi-objective discrete urban road network design. *Comput Oper Res*. doi:<http://dx.doi.org/10.1016/j.cor.2013.03.016>
- Miandoabchi E, Zanjirani Farahani R, Dullaert W, Szeto WY (2012b) Hybrid evolutionary metaheuristics for concurrent multi-objective design of urban road and public transit networks. *Netw Spat Econ* 12(3):441-480
- Miandoabchi E, Zanjirani Farahani R, Szeto WY (2012a) Bi-objective bimodal urban road network design using hybrid metaheuristics. *Cent Eur J Oper Res* 20(4):583-621
- Nagurney A (2000) *Sustainable Transport Networks*. Edward Elgar Publishers, UK.
- Nagurney A, Ramanujam P, Dhanda KK (1998) A multimodal traffic network equilibrium model with emission pollution permits: compliance versus noncompliance. *Transp Res Part D* 3(5):349-374
- Ng M, Lo HK (2013) Regional air quality conformity in transportation networks with stochastic dependencies: A theoretical copula-based model. *Netw Spat Econ*: 1-25
- Oryani K, Harris B (1997) Review of Land Use Models: Theory and Application. In Sixth TRB Conference on the Application of Transportation Planning Methods.
- Patil GR, Ukkusuri SV (2007). System-optimal stochastic transportation network design. *Transp Res Rec* 2029:80-86
- Qiu Y, Chen S (2007) Bi-level programming for continuous network design of comprehensive transportation system based on external optimization. *Proceedings of 2007*

- IEEE International Conference on Grey Systems and Intelligent Services, November 18-20, China.
- Sharma S, Mathew TV (2011) Multi-objective network design for emission and travel time tradeoff for sustainable large urban transportation network. *Environ Plan B* 38(3):520-538
- Smith JM, Liebman JS (1978) A zero-one integer-programming formulation of the problem of land-use assignment and transportation-network design. *Environ Plan B* 5(1):101-115
- Szeto WY, Jaber XQ, O'Mahony M (2010) Time-dependent discrete network design frameworks considering land use. *Comput-Aided Civ Infrastruct Eng* 25(6):411-426
- Szeto WY, Jaber XQ, Wong SC (2012) Road network equilibrium approaches to environmental sustainability. *Transp Rev* 32(4): 491-518
- Szeto WY, Jiang Y (2012) A hybrid artificial bee colony for transit network design problem. *Transp Res Rec* 2284:47-56.
- Szeto WY, Lo HK (2005) Strategies for road network design over time: robustness under uncertainty. *Transp* 1(1):47-63
- Szeto WY, Lo HK (2006) Transportation network improvement and tolling strategies: The issue of intergeneration equity. *Transp Res Part A* 40(3): 227-243
- Szeto WY, Lo HK (2006) Transportation network improvement and tolling strategies: the issue of intergeneration equity. *Transp Res Part A* 40(3):227-243
- Szeto WY, Lo HK (2008) Time-dependent transport network improvement and tolling strategies. *Transp Res Part A* 42(2):376-391
- Transportation Research Board. (1997) *Toward a Sustainable Future*. Special Report 251, Washington, D. C.
- Ukkusuri SV, Mathew TV, Waller ST (2007) Robust transportation network design under demand uncertainty. *Comput-Aided Civ Infrastruct Eng* 22(1):6-18
- Ukkusuri SV, Patil G (2009) Multi-period transportation network design under demand uncertainty. *Transp Res, Part B* 43(6):625-642
- Ukkusuri SV, Waller ST (2008) Linear programming models for the user and system optimal dynamic network design problem: Formulations, comparisons and extensions. *Netw Spat Econ* 8(4):383-406
- Unnikrishnan A, Valsaraj V, Damnjanovic I, Waller ST (2009) Design and management strategies for mixed public private transportation networks: A meta-heuristic approach. *Comput-Aided Civ Infrastruct Eng* 24(4):266-279
- Williams H (1976) Travel demand models, duality relations and user benefit analysis. *J Reg Sci* 16(2):147-166
- Williams H (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environ Plan A* 9(3):285-344
- Wilson AG (1970) *Entropy in Urban and Regional Modelling*. Pion, London.
- Wong CK, Tong CO, Wong SC (1999) The development and calibration of a Lowry model with multiple market segments. *Environ Plan A* 31:1905-1918
- Yang H, Xu W, He B, Meng Q (2010) Road pricing for congestion control with unknown demand and cost functions. *Transp Res Part C* 18(2):157-175
- Yim KKW, Wong SC, Chen A, Wong CK, Lam WHK (2011) A reliability-based land use and transportation optimization model. *Transp Res Part C* 19(2):351-362
- Yin Y, Lawphongpanich S (2006) Internalizing emission externality on road networks. *Transp Res Part D* 11(4):292-301
- Yin Y, Lu H (1999) Traffic equilibrium problems with environmental concerns. *J East Asia Soc Transp Stud* 3:195-206
- Zanjirani Farahani R., Miandoabchi E, Szeto WY, Rashidi H (2013) A review of urban transportation network design problems. *Eur J Oper Res*.doi: <http://dx.doi.org/10.1016/j.ejor.2013.01.001>

- Zhong RX, Sumalee A, Maruyama T (2012) Dynamic marginal cost, access control, and pollution charge: a comparison of bottleneck and whole link models. *J of Adv Transp* 46(3):191–221
- Zhou S, Yan X, Wu C (2008) Optimization model for traffic signal control with environmental objectives. In *Proceedings of the 4th International Conference on Natural Computation* (pp. 530–534), Jinan, China, October 18–20