Abstract: Sustainability is a hot topic nowadays because there is growing awareness of human activities including transport activities that can have significant environmental impacts and can impose economic, social and ecological damages. It is therefore important to consider sustainability into network design to reduce the negative impacts of transport activities. This paper develops a multi-objective time-dependent network design model to consider the economic, social and ecological dimensions of sustainability. Land-use transport interaction over time is also captured to study the effect of network design on landowners. The variances of discounted landowner profit and discounted user generalized cost are proposed as indicators of two sustainability issues, namely landowner inequity and intergeneration inequity respectively. Numerical studies are set up to illustrate the properties of the problem. 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.

Keywords: Time-dependent road network design, land use, landowner inequity, sustainability, intergeneration inequity

1. INTRODUCTION

In Hong Kong like other Asian 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.

Currently, there are three main lines of research on road network design:

- Static or time-independent approach without lane use consideration (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; Chiou, 2009)
• Time-dependent approach without lane use consideration (e.g., Szeto and Lo, 2006; 2008; Lo and Szeto, 2009; Kim et al., 2008)

• Static approach with lane use consideration (e.g., Smith and Liebman, 1978; Los, 1979; Feng and Lin, 1999; Meng et al., 2000; Lin and Feng, 2003; Lee et al., 2006; Qiu and Chen, 2007)

The first approach is developed based on user equilibrium principle or its stochastic extension. This allows analyzing the problem easily but this approach cannot capture the realistic variation in demand, the land use pattern, the network upgrades and the land-use transport interaction over time. The second approach deals with some of the problems in the first approach by capturing the time-dimension into the analysis, but still the land use component is missing. Hence, this approach cannot study the impact of network design on land use. The third approach deals with the lane use component when compared with the first approach but the time dimension is missing. The changing demand and gradual network upgrades are therefore not captured. It is clear from the above that there is a need to develop an approach that considers both time-dimension and lane use. Therefore, this paper develops an optimization framework that considers both the aforementioned factors.

Other than network design, sustainability is a hot topic nowadays mainly because there is growing awareness of human activities including transport activities that can have significant environmental impacts and can impose economic, social and ecological damages. For example, exhausted gases from vehicles can lead to air pollution and global warming, which has negative impacts to people. It is therefore important to consider sustainability into network design to reduce the negative impacts of transport activities. Moreover, there is more than one indicator to measure sustainability. Hence, the framework proposed in this paper considers multiple indicators and hence multiple objectives, which can be used as a tool for decision-making considering sustainability. 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 vehicular emissions (for the environmental dimension of sustainability),
2. the change in consumer surplus (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 user generalized 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. The rest of the paper is organized as follows: Section 2 describes the formulation of the framework. Section 3 presents the numerical study. Finally, Section 4 provides some concluding remarks.

2. FORMULATION

We consider a strongly connected multi-modal transportation network with multiple Origin-Destination (OD) flows over the planning horizon \([0,T]\). The planning horizon is divided into \(N\) equal design periods. 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.
2.1 The Lower Level Problem

The lower level problem of the bi-level problem is the time-dependent land-use transport problem. There are two types of constraints for describing this problem: Time-dependent Lowry-type constraints and time-dependent modal-split/assignment constraints.

2.1.1 Time-dependent Lowry-type Constraints

The time-dependent Lowry-type constraints extend the Lowry-type equilibrium to a dynamic framework. In each design period, a Lowry-type equilibrium is assumed to be held. The equilibrium is depicted by a number of constraints. The first one describes how to allocate residents who work in employment zone \( i \) to residential zone \( j \) using the gravity-type model:

\[
R_{ij,\tau} = E_{i,\tau} B_{ij,\tau} W_{ij,\tau}^\alpha \left( -\beta' \left( 2c_{ij,\tau} + r_{ij,\tau} / 30 \right) \right), \forall i, j, \tau, \tag{1}
\]

where \( B_{ij,\tau} = 1 / \sum_j W_{ij,\tau} \exp \left( -\beta' \left( 2c_{ij,\tau} + r_{ij,\tau} / 30 \right) \right), \forall i, \tau. \tag{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 \( j \) that work in zone \( i \) in period \( \tau \). \( E_{i,\tau} \) is the total employment in zone \( i \) in period \( \tau \). \( W_{ij,\tau} \) is the attractiveness of zone \( j \) in period \( \tau \), which can be represented by the availability of floor space for residential use. \( \alpha \) is the economy-of-scale parameter to regulate the attractiveness of each zone (Bureau of Transport Economics, 1998). \( \beta' \) is a parameter to regulate the effect of transport cost on distribution of residents. A high value of \( \beta' \) will result in the residents being allocated close to their place of work; if \( \beta' \) tends to infinity, all residents will live and work in the same zone. On the other hand, if \( \beta' \) tends to zero, the residents whose work in zone \( i \) will locate to all residential zones equally. \( c_{ij,\tau} \) is the composite travel cost between OD pair \( ij \) in period \( \tau \), representing the inter-zonal impedance and will be defined later. \( r_{ij,\tau} \) is house rent per resident in residential zone \( j \) in period \( \tau \). The balancing factor \( B_{ij,\tau} \) is derived by Wilson (1970), which is to ensure a correct allocation of residents to zone \( j \) in period \( \tau \) so that \( \sum_j B_{ij,\tau} W_{ij,\tau}^\alpha \exp \left( -\beta' \left( 2c_{ij,\tau} + r_{ij,\tau} / 30 \right) \right) = 1 \) 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 (i.e., \( \sum_j R_{ij,\tau} = E_{i,\tau} \)).

The total employment in zone \( i \) in period \( \tau \), \( E_{i,\tau} \), in (1) is the sum of the basic employment \( E_{i,\tau}^B \) and the service employment or non-basic employment, \( E_{i,\tau}^S \), in zone \( i \) in period \( \tau \):

\[
E_{i,\tau} = E_{i,\tau}^B + E_{i,\tau}^S, \forall i, \tau. \tag{3}
\]

The service employment in zone \( i \) in period \( \tau \), \( E_{i,\tau}^S \), in (3) 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:

\[
E_{i,\tau}^S = \sum_j E_{ij,\tau}^S, \forall i, j, \tau. \tag{4}
\]
where \( E_{ij}^S \) 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 \( \tau \).

The number of service employment trips \( E_{ij}^S \) in (4) is obtained by:

\[
E_{ij}^S = sR_{j,\tau}A_{j,\tau}W_{i,\tau}^{\alpha} \exp\left(-\beta^c\left(2c_{g,\tau} + r_{j,\tau}/30\right)\right), \forall i, j, \tau, \tag{5}
\]

where \( A_{j,\tau} = 1/\sum_i W_{i,\tau}^{\alpha} \exp\left(-\beta^c\left(2c_{g,\tau} + r_{j,\tau}/30\right)\right), \forall j, \tau, \tag{6} \]

\( s \) is a service employment-to-population ratio. \( W_{i,\tau} \) is the attractiveness in zone \( i \) in period \( \tau \), which can be the availability of floor space for commercial use. \( \alpha \) is the economy-of-scale parameter to regulate the attractiveness of each zone. \( \beta^c \) is a parameter to regulate the effect of transport cost on distribution of service employees, and its function is similar to \( \beta^c \) in (1). A high value of \( \beta^c \) 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 equally. 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 (2). The term \( sR_{j,\tau} \) in (5) is the total number of employees in zone \( j \) in period \( \tau \).

The total number of residents in zone \( j \) in period \( \tau \), \( R_{j,\tau} \), in (5) is defined as:

\[
R_{j,\tau} = \mu \sum_i R_{y,\tau}, \forall j, \tau, \tag{7}
\]

where \( \mu \) is a population-to-employment ratio. According to (7), in each period, the total number of employees in zone \( j \), \( \sum_i R_{y,\tau} \) multiplied by the population-to-employment \( \mu \) gives the total number of residents in that zone \( R_{j,\tau} \).

2.1.2 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.

2.1.2.1 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:

\[
f_{p,ij,\tau}^k \left[c_{p,ij,\tau}^k - \pi_{g,ij,\tau}^k\right] = 0, \forall p, i, j, k, \tau. \tag{8}
\]

\[
c_{p,ij,\tau}^k - \pi_{g,ij,\tau}^k \geq 0, \forall p, i, j, k, \tau, \tag{9}
\]

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; \( \pi_{g,ij,\tau}^k \) is the lowest travel cost between OD pair \( ij \) by mode \( k \) in period \( \tau \).
2.1.2.2 Travel Cost Constraints

Travel costs depend on flows, network characteristics such as 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:

\[ v_k^{a,\tau} = \sum_{ij} \sum_p f_{p,ij}^k \delta_{a}^{p,k}, \forall a, k, \tau, \]

(10)

where \( v_k^{a,\tau} \) 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 (10) 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:

\[ t_{a,\tau}^k = t_{a,\tau}^{k,0} (v_{\tau}), \]

where \( v_{\tau} = [v_k^{a,\tau}] \) 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:

\[ t_{a,\tau}^m = t_{a,\tau}^{0,m} \left[ 1 + \left( \frac{\sum_a \sum_k \alpha_{a,k} v_k^{a,\tau}}{\tau_{a,\tau}^m} \right)^{\gamma_{a,k}} \right], \forall a, m, \tau, \]

(12)

\[ \tau_{a,\tau}^m = \tau_{a,\tau}^{m,0} + \sum_b \sum_{\omega=1}^{\tau} \delta_{a,b,\omega}^{m} y_{b,\omega}, \forall a, m, \tau, \]

(13)

where the superscript \( m \) stands for the mode that travels in the road network only (which is different from the superscript \( k \) that is used for representing any mode considered in this paper), \( t_{a,\tau}^{0,m} \) and \( \tau_{a,\tau}^{m,0} \) are the free flow travel time and initial capacity of link \( a \) used by mode \( m \) that travels on roads. \( \alpha_{a,k} \) and \( \gamma_{a,k} \) are parameters of the performance function of link \( a \) for mode \( m \). \( \tau_{a,\tau}^m \) is the capacity of link \( a \) used by mode \( m \) in period \( \tau \). \( y_{b,\tau} \) is the capacity improvement on link \( b \) in period \( \tau \), meaning that the capacity of link \( b \) is increased by \( y_{b,\tau} \) units at the beginning of period \( \tau \). \( \delta_{a,b}^{m} \) is the simultaneous improvement/construction indicator, which equals one if the improvement/construction on link \( b \) (in which \( b \) stands for a particular road) will also result in improvement/construction on link \( a \) used by mode \( m \) or \( a = b \), and zero otherwise. This indicator is used in (13) because a street or road may be used by two different modes, say buses and cars, with different travelling speeds and we model the street or road by two different links with the same capacity.

Equation (12) is an asymmetric link performance function, which describes the monotonic relationship between the link travel time \( t_{a,\tau}^m \) and the link flows \( v_k^{a,\tau} \). When \( \alpha_{a,m} = 0.15, \alpha_{b,k} = 0, a \neq b, m \neq k \), and \( \gamma_{a,m} = 4 \), equation (12) is reduced to the typical Bureau of Public Roads (BPR) function. Equation (13) expresses the capacity of link \( a \) used by mode \( m \) in period \( \tau \) as the sum of its initial capacity \( c_{a}^{m,0} \) and the total improvements up to period \( \tau \).
The route cost $c_{p,ij,\tau}^k$ is the sum of the link-wise additive costs $g_{p,ij,\tau}^k$ and the route specific costs $\theta_{p,ij,\tau}^k$:

$$c_{p,ij,\tau}^k = g_{p,ij,\tau}^k + \theta_{p,ij,\tau}^k, \forall p,i,j,k,\tau.$$  \hspace{1cm} (14)

The link-wise additive costs $g_{p,ij,\tau}^k$ are 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. The link-wise additive cost $g_{p,ij,\tau}^k$ can be written as:

$$g_{p,ij,\tau}^k = \sum_a \left( \psi_{t_a}^k + \rho_{a,\tau}^k \right) \cdot \delta_{p,ij,\tau}^k, \forall p,i,j,k,\tau,$$  \hspace{1cm} (15)

where $\psi$ is the cost of unit (travel) time, and therefore $\psi_{t_a}^k$ 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 route specific costs $\theta_{p,ij,\tau}^k$ are non-linear and/or nonadditive over links; for instance, some types of tolls in road networks (e.g., non-linearly proportional to distance), or waiting time and some fare structures for transit networks (e.g., zone-wise prices).

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

$$c_{ij,\tau} = -\ln \left[ \sum_k \left( \exp(-\bar{\beta} \left( \pi_{ij,\tau}^k + \theta^k \right)) \right) \right] \bar{\beta}, \forall i, j, \tau,$$  \hspace{1cm} (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).

2.1.2.3 The Modal Split, Flow Conservation, and Non-Negativity Conditions

Modal split can be obtained by the logit model:

$$q_{ij,\tau}^k = R_{ij,\tau} \left( \frac{\exp(-\bar{\beta} \left( \pi_{ij,\tau}^k + \theta^k \right))}{\sum_m \exp(-\bar{\beta} \left( \pi_{ij,\tau}^m + \theta^m \right))} \right), \forall i, j, k, \tau,$$  \hspace{1cm} (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 $j$ defined in (1).

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:

$$q_{ij,\tau}^k = \sum_p f_{p,ij,\tau}^k, \forall i, j, k, \tau.$$  \hspace{1cm} (18)

Moreover, route flow in (18) must be non-negative.
\[ f_{p,j \tau}^k \geq 0, \forall p,i,j,k,\tau. \]  

(19)

2.1.3 Lower Level Optimization Model

2.1.3.1 Gap Function

The lower level problem can indeed form an optimization problem by employing a gap function. Many gap functions can serve this purpose. In this paper, we adopt

\[ G = \sum_p \sum_j \sum_{\tau} f_{p,j \tau}^k \left[ c_{p,j \tau}^k - \pi_{j \tau}^k \right]. \]

(20)

This gap function must have non-negative values if \( f_{p,j \tau}^k \geq 0 \) and \( c_{p,j \tau}^k - \pi_{j \tau}^k \geq 0 \) because the sum of non-negative numbers must be non-negative. One property of this gap function is that if \( f_{p,j \tau}^k \geq 0 \), \( c_{p,j \tau}^k - \pi_{j \tau}^k \geq 0 \) and the gap function attains its minimum value of zero, the time-dependent Wardrop’s condition (8) is satisfied. This property will be used in developing our lower level optimization model.

2.1.3.2 Lower Level Optimization Formulation

Given the time-dependent tolls and capacity improvements, the lower level problem is formulated as follows:

\[ \min_{f,E,R} G \]

subject to

- time-dependent Lowry-based land-use model constraints (1)-(7), and;
- time-dependent traffic assignment constraints (10)-(19).

where \( f,E,R \) represent, respectively, the vector of path flows, service employment trips, and residential trips. Since \( f_{p,j \tau}^k \geq 0 \) and \( c_{p,j \tau}^k - \pi_{j \tau}^k \geq 0 \) are ensured by the time-dependent traffic assignment constraints (9) and (19), path flows, service employment trips, and residential trips will satisfy (1)-(19) when the gap \( G \) is zero.

2.2 The Upper Level Problem

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

2.2.1 Design and financial Constraints

2.2.1.1 Capacity Constraints

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

\[ \overline{u}_{a \tau}^m \leq u_a^m, \forall a,m,\tau, \]

(21)

where \( u_a^m \) is the maximum allowable capacity of link \( a \) for mode \( m \) that uses the highways or roads. Equation (21) is the maximum allowable capacity constraint, which is to limit the total capacity of each link after road expansion or highway construction in period \( \tau \), \( \overline{u}_{a \tau}^m \), to be less than its maximum allowable capacity.
2.2.1.2 Toll Constraints

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

\[ \rho_{\tau l}^m = 0, \forall l, m, \tau, \] \hspace{1cm} (22)
\[ \rho_{\tau a}^m \geq 0, \forall a, m, \tau, \] \hspace{1cm} (23)
\[ \rho_{\tau a}^m \leq \rho_{\max}^m, \forall a, m, \tau, \] \hspace{1cm} (24)

where $\rho_{\max}^m$ is the maximum allowable toll for mode $m$ that travels on roads. The subscript $l$ represents links without tolls.

2.2.1.3 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 more than covers 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:

\[ \sum_{\tau} T_{\tau} \frac{r^{\tau - 1}}{(1 + i)^{\tau - 1}} + \sum_{\tau} S_{\tau} \frac{r^{\tau - 1}}{(1 + i)^{\tau - 1}} - \sum_{\tau} K_{\tau} \frac{r^{\tau - 1}}{(1 + i)^{\tau - 1}} = TOP, \] \hspace{1cm} (25)

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

The first term on the left hand side (LHS) of (25) 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 (25) 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 (25) 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:

\[ T_{\tau} = \sum_m \sum_a n v_{a,\tau}^m \rho_{a,\tau}^m, \forall \tau, \] \hspace{1cm} (26)
\[ K_{\tau} = \sum_b \left( h_{b,\tau} + w_{b,\tau} \right), \forall \tau, \] \hspace{1cm} (27)
\[ h_{b,\tau} = \left( 1 + j \right)^{\tau - 1} b_{b,\tau} \lambda_{b} \left( y_{b,\tau} \right)^{\beta_{b}}, \forall b, \tau, \] \hspace{1cm} (28)
\[ w_{b,\tau} = (1 + \hat{j})^{-1} \left[ \beta_{b,0} + \beta_{b,1} \left( \sum_m (n_{v_{h,\tau}}^m) \right) \right], \forall b, \tau. \] (29)

where \( h_{b,\tau} \) and \( w_{b,\tau} \) are the improvement (or construction) and maintenance cost functions of link \( b \) in period \( \tau \) respectively; \( \overline{h}_{b,0}, \beta_{b,0}, \beta_{b,1}, \beta_{b,2} \) are parameters of these cost functions; \( n \) converts link flows from an hourly basis to a period basis; \( \hat{j} \) is the inflation rate; \( l_b \) is the length of link \( b \). Equation (26) calculates the toll revenue in period \( \tau \), which is the sum of the product of the link flow and toll in that period. Equation (27) computes the improvement and maintenance cost in period \( \tau \) by adding the improvement and maintenance cost of all links. Equation (28) is the time-dependent improvement cost function. The term \( (1 + \hat{j})^{-1} \) represents the inflation factor: for the same capacity enhancement, the improvement cost increases by \( \hat{j} \% \) each period. The term \( \overline{h}_{b,0}, \beta_{b,1} \) models the improvement cost of link \( a \) in period 1 (i.e., the base period). Equation (28) 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 (29) is the time-dependent maintenance cost function, which is set to be: \( \beta_{b,0} + \beta_{b,1} \left( \sum_m (n_{v_{h,\tau}}^m) \right) \) in the base period, consisting of the fixed cost \( \beta_{b,0} \) and the variable cost \( \beta_{b,1} \left( \sum_m (n_{v_{h,\tau}}^m) \right) \), in which \( \sum_m (n_{v_{h,\tau}}^m) \) is the link flow on link \( b \) in period \( \tau \). Again, the maintenance cost depends on the inflation factor \( (1 + \hat{j})^{-1} \).

2.2.2. Sustainability Indicators.

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

2.2.2.1 Total Vehicular Emissions

There are two types of vehicular emissions: link and network (or overall). The link vehicular emissions are defined through the link emission factor approach:

\[ Q_{a,\tau} = \sum_m Q_{a,\tau}^m = \sum_m h_{a,\tau}^m v_{a,\tau}^m, \forall a, \tau, \] (30)

where \( Q_{a,\tau}^m \) is the vehicular emissions for traffic mode \( m \) on link \( a \) in period \( \tau \); \( v_{a,\tau}^m \) represents the hourly traffic flow for mode \( m \) on link \( a \) in period \( \tau \); \( h_{a,\tau}^m \) is the emission factor for mode \( m \) on link \( a \) in period \( \tau \), which is assumed to be given for all links. The factors affecting the value of \( h_{a,\tau}^m \) are discussed in Nagurney (2000). This link emission factor approach has been adopted by Nagurney et al. (1998) and others. According to (30), the vehicular emissions for mode \( m \) on a particular link is the product of the link flows of mode \( m \) and the corresponding emission factor. Thus, the total vehicular emissions on this link are
given by the sum of vehicular emissions for all modes travelling on this link. The overall vehicular emissions $Q$ are the sum of the vehicular emissions on each link over time:

$$Q = \sum_{\tau} \sum_{a} Q_{a,\tau}.$$  \hspace{1cm} (31)

### 2.2.2.2 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 (positive change in CS) implies a better performing system. Alternatively, we adopted the change in CS as a measure because the CS before improvement is fixed. Here, an approximation to this change $\Delta CS$, in present value terms, is employed (Williams, 1976) and can be expressed as follows:

$$\Delta CS = \sum_{\tau} \sum_{j} \sum_{k} \Delta CS^{\mu}_{ij,\tau},$$  \hspace{1cm} (32)

$$\Delta CS^{\mu}_{ij,\tau} = (1/2)(q_{ij,\tau}^{k,\text{before}} + q_{ij,\tau}^{k,\text{after}})(\pi_{ij,\tau}^{k,\text{before}} - \pi_{ij,\tau}^{k,\text{after}}), \forall i, j, k, \tau,$$  \hspace{1cm} (33)

where the superscripts ‘before’ and ‘after’ denote before and after improvement and construction project implementations, respectively. $\bar{i}$ is the interest rate. $(1 + \bar{i})^{-(\tau - 1)}$ is the discount factor for period $\tau$. According to (32), the change in CS is the sum of the change in CS for all modes and for all OD pairs over time, discounted to present value terms. Equation (33) is the rule-of-half definition for CS.

### 2.2.2.3 Variance of Discounted Landowner Profit

The variance of discounted landowner profit is used 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.

The discounted profit of land owner $j$ is defined as:

$$LOP_j = \sum_{\tau} \frac{LOP_{j,\tau}}{(1 + \bar{i})^{\tau-1}}, \forall j,$$  \hspace{1cm} (34)

where $LOP_{j,\tau}$ represent the profits of land owner $j$ in period $\tau$, which is the difference between its total revenue and maintenance cost in that period:

$$LOP_{j,\tau} = R_{j,\tau} r_{j,\tau} - M^H_{j,\tau}, \forall j, \tau,$$  \hspace{1cm} (35)

where $R_{j,\tau}$ is the total number of residents in zone $j$ in period $\tau$. The maintenance cost, $M^H_{j,\tau}$, on houses can be formulated as a linear function as follows:

$$M^H_{j,\tau} = \tilde{M}_{j,\tau} + m^h R_{j,\tau}, \forall j, \tau,$$  \hspace{1cm} (36)

where $\tilde{M}_{j,\tau}$ is the fixed maintenance cost on houses in residential zone $j$ in period $\tau$. $m^h$ is a parameter. This rent is assumed to be linearly increasing with the total population in a zone:

$$r_{j,\tau} = r_{j,\tau,\text{min}} + \gamma_j R_{j,\tau},$$  \hspace{1cm} (37)

where $r_{j,\tau,\text{min}}$ is the minimum monthly rent in residential zone $j$ in period $\tau$, and $\gamma_j$ is a parameter. The rent $r_{j,\tau}$ in (35) is assumed to increase over time due to inflation:
\[ r_{j,\tau+1} = r_{j,\tau} (1 + \bar{j}), \quad (38) \]

where \( \bar{j} \) is the inflation rate.

### 2.2.2.4 The Variance of Discounted User Generalized Cost

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

\[ II = \sum_{\gamma} Var \left[ \sum_{\tau} \left( \frac{2c_{y,\tau} + r_{j,\tau}}{1 + \bar{i}} \right) \right]. \quad (39) \]

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

### 2.3 General Multi-objective Bi-level Model

The proposed bi-level model is formulated as follows:

\[ \max y_j(x), \quad (40) \]

subject to

\[ y_j(x) \leq \varepsilon_j, \forall j \neq i; \quad (41) \]

- design and financial constraints (21)-(29), and;
- lower level problem (1)-(19),

where \( y_j(x) \) is the \( i \)-th objective function or desirable sustainability indicator; \( x \) is the vector of decision variables including tolls and capacity enhancements; \( \varepsilon_j \) is the aspiration level or the desirable value or the \( j \)-th sustainability indicator. In the above framework, condition (41) is the performance constraint (or \( \varepsilon \)-constraint), which considers the objective that does not include in (40). The \( j \)-th objective function is set to be greater than the corresponding desirable or satisfactory objective value to ensure that at optimality, the \( j \)-th objective value is at least equal to the satisfactory value \( \varepsilon_j \).

The proposed bi-level model is formulated as a single-level minimization program where the lower level problem is expressed as constraints (1)-(19) instead of the lower level model described in Section 2.1.3.2. The advantage of formulating the bi-level problem as a single-level problem is that we can use the existing single optimization algorithms and packages to solve for solutions. In this study, the bi-level model is solved by PREMIUM SOLVER PLATFORM.

### 2.4 Consumer Surplus Maximization Model

The above model can be reduced to the following:

\[ \max \Delta CS \]

subject to

- design and financial constraints (21)-(29);
- lower level problem (1)-(19);
- \( TOP \geq 0 \) and \( S_\tau = 0, \forall \tau \), and;

\[ Var(LOP) \leq \varepsilon_{\text{max, var} LOP}, \quad (43) \]

where \( \varepsilon_{\text{max, var} LOP} \) is the maximum acceptable variance of landowner profit. (43) is called the landowner inequity constraint.
3. NUMERICAL STUDIES

For the ease of exposition, a simple network is adopted as shown in figure 1. There are 3 links in this network: link 1, link 2, and link 3. Links 1 and 2 are links whose travel time is 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:

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

where \(\tilde{h}_{i,t}\) 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:

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

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.

The parameters in this study include: 1) Land use parameters: \(E_{1,1} = 5000\) jobs; \(W_{1,1} = 3000\) jobs; \(W_{2,1} = W_{3,1} = 3000\) houses; \(\mu = 5;\) \(\alpha = 1, s = 0.1;\) \(\beta' = 0.04€^{-1};\) \(\beta'' = 0.03€^{-1};\) \(\tilde{M}_{j,t} = €100;\) \(m^t = €0.01\)/household; \(\gamma_{2,r,\text{min}} = \gamma_{3,r,\text{min}} = €2;\) \(\tilde{h}_{w,1} = \tilde{h}_{w,2} = \tilde{h}_{w,3} = 0.05;\) \(h_{E,1} = 0.04,\) \(r_{2,r,\text{min}} = r_{3,r,\text{min}} = €1000;\) 2) Transport network parameters: \(c_1^0 = c_2^0 = 3000\) vph; \(u_1 = 7500\) vph, \(u_2 = 0\) vph; \(t_{1}^0 = t_{2}^0 = 5\) hours; \(t_{3}^0 = 4\) hours; 3) \(Y_2 = 1000000€;\) 4) Parameters of improvement cost functions: \(b_{1,1} = b_{2,1} = 1, b_{1,0} = b_{2,0} = €2000;\) 5) Parameters of maintenance cost functions: \(\beta_{1,0} = €1200,\) \(\beta_{1,1} = €0.001,\) \(\beta_{1,2} = 1;\) 6) Parameters in travel cost functions: \(\psi = €15/h;\) \(\theta^1 = €16;\) \(\theta^2 = €30;\) \(\beta\) = 0.05€^{-1}; 7) Interest and inflation rates: \(\tilde{i} = 0.03;\) \(\tilde{j} = 0.01;\) 8) Converting factor: \(n = 365\text{days} \times 24\text{hours} \times 1\text{ year} = 87600\text{ hours/period};\) 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, \(\rho_{2,r} = €0;\) \(\rho_{3}^1 = €6,\) and; 12) Emission parameters: \(h_1 = 1752\) litres/year; \(h_2 = 876\) litres/year. 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 vehicular emissions, the variance of user generalized cost, the variance of landowner profit, and the increase in consumer surplus. These results are plotted
in Figures 2-5, respectively. The x-axis is toll charge whereas the y-axis is the capacity improvement. The shaded region represents the infeasible region where cost recovery cannot be achieved or the objective value 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.

As you can see from figure 2, for a given toll charge, the total vehicular emissions decrease with decreasing capacity improvement. Moreover, for a given capacity improvement, the total vehicular emissions decrease with increasing toll. If our objective is minimizing the total vehicular emissions, then from the graph, the optimal solution point is (6, 0). This observation is consistent with our understanding that high toll and low capacity means higher travel cost on the high emission link of link 1, resulting in decreasing number of trips on that link and increasing the number of trips on the low emission link of link 2.

According to Figures 3 and 4, decreasing capacity improvement and increasing toll charge reduce the variances of user cost and landowner profit. The optimal solutions for minimizing these variances is at (6, 0), which is identical to the optimal solution of minimizing total vehicular emissions. This implies that if we minimize total vehicular emissions for this
example network, we can achieve the minimum variances of user cost and landowner profit simultaneously.

From Figure 5, we can see that decreasing capacity improvement and increasing toll charge lower the change in consumer surplus. The optimal solution is at (1.8, 4500), which is different from previous optimal solutions. This means that the objective of maximizing the change in CS is in conflict with minimizing total vehicular emissions, and with minimizing the variances of user generalized cost and landowner profit. Simultaneous satisfying all objectives is not possible. Tradeoffs must be made between these objectives.

Figure 6 Tradeoff between the change in CS and maximum acceptable variance of LOP

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 discounted landowner profit (LOP). The consumer maximization model with the landowner inequity constraint was solved many times by varying maximum acceptable variance of LOP. Figure 6 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 landowner profit as the change in CS is negative. 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 the increasing variance, indicating that one sustainable 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.
4. CONCLUDING REMARKS

This paper develops a multi-objective time-dependent 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. The variance of user generalized cost is proposed as a measure of intergeneration inequity. Numerical studies are set up to illustrate the multi-objective nature of the problem. 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.

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, the proposed framework is path-based, which is not suitable for large networks involving many paths. Moreover, the multi-objective problem is highly non-convex, which is difficult to solve for global solutions efficiently. Developing a link-based formulation and an efficient global optimisation technique for this problem represents another worthy research direction.

ACKNOWLEDGMENT

This research is jointly funded by the start-up grant R-264-000-229-112 from the National University of Singapore.

REFERENCES


