Reliability-based equitable transit frequency design

Jiang, Yu

Published in:
Transportmetrica A: Transport Science

Link to article, DOI:
10.1080/23249935.2021.1902420

Publication date:
2022

Document Version
Early version, also known as pre-print

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Reliability-based Equitable Transit Frequency Design

Yu Jiang

DTU Management
Danmarks Tekniske Universitet (DTU)
DK-2800 Kgs. Lyngby, Denmark
ORCiD: 0000-0001-9461-633X
Email: yujiang@dtu.dk
Abstract

Fairness is an important criterion for achieving sustainable urban development, and thus, is of major concern in providing public transport services. While most existing studies focus on accessing or evaluating the fairness condition of a given transit network, this study explicitly incorporates fairness as an objective in the planning step to allocate a fleet to a set of bus lines. A multi-objective bilevel programming model is developed for the transit frequency setting problem. The lower level problem is the reliability-based transit assignment problem that captures the effects of supply-side uncertainty on passengers’ route choice behaviour via the concept of travel time budget referred to as effective travel cost. The upper level problem is to determine optimal frequency settings to simultaneously minimise the total effective travel cost and maximise the network fairness condition, measured by minimum reduction in effective travel cost of all Origin-Destination (OD) pairs. A multi-objective Artificial Bee Colony algorithm (MOABC) is developed to solve the bilevel model, where non-dominated solutions are maintained by a predefined external archive. Numerical studies find that: 1) increasing the frequency may not improve the fairness condition; 2) there is a tradeoff between the objective of minimising total effective travel cost and that of maximising the fairness measurement; 3) the two objectives can be improved simultaneously using a large fleet; 4) the effect of passengers’ risk aversion attitude on the fairness measurement depends on the frequency setting; it could either amplify the fairness measurements or have no impact.

Keywords: reliability-based transit assignment; transit frequency setting; transport equity; Artificial Bee Colony (ABC) algorithm; multiobjective bilevel optimization
1 Introduction

Equity is of major concern in designing public transport systems (Delbosc and Currie 2011), and it is of great importance to travellers and planners. If passengers are treated unfairly, they would perceive the public transport system to be less attractive, and might subsequently switch their travel mode from public transport to private transport, exacerbating urban traffic congestion problems. Hence, equity is increasingly considered a long-term objective, and as such, incorporated into sustainable urban transportation plans (Manaugh et al. 2015, El-Geneidy et al. 2016).

Equity was originally a social science concept (Sen 1973). Although there is no universally accepted definition, equity research in the field of transportation can be broadly classified into three categories:


(2) Assessing or evaluating the equity condition of a given network, scenario, or policy. In this realm, the equity condition can either be assessed by prevailing equity measures in social science, such as the Gini Index or Lorenz curves (Delbosc and Currie 2011, Welch and Mishra 2013, Lucas et al. 2016, Song et al. 2018, Ben-Elia and Benenson 2019) or by a proposed equity measure or framework fitted to the context of the transportation problems (Szeto et al. 2010, Martens et al. 2012, Foth et al. 2013, Kaplan et al. 2014, Wei et al. 2017, Gori et al. 2020, etc.) It warrants mention that most of the equity measures in the transportation field are based on accessibility measures, for which we refer to Carleton and Porter (2018) for a more comprehensive review. An extensive discussion of the linkage between accessibility and inequities can be found in Bocarejo and Oviedo (2012) and Lucas et al. (2016).
(3) Integrating the equity issue in the transportation planning process. Various models have been developed to approach equitable road network designs (Meng and Yang 2002, Yang and Zhang 2002, Chen and Yang 2004, Szeto and Lo 2006, Szeto et al. 2013, Mollanejad and Zhang 2014, Szeto et al. 2015, etc.). In contrast, only a few studies focus on equity issues in planning public transport services (Fan and MacHemehl 2011, Ferguson et al. 2012, Camporeale et al. 2016, Ruiz et al. 2017, Camporeale et al. 2019).

This study pertains to the latter category, focusing on incorporating equity consideration in designing transit network services.

Traditionally, public transport planning includes five steps: line planning, frequency setting, schedule design, vehicle scheduling, and crew scheduling (Ceder 2016). Among the five steps, the first three could induce equity problems. Changes in passengers’ travel costs, a key factor in assessing equity, could result from a change in any of the three steps, or a combination thereof. In general, line planning and frequency setting are integral to the strategic decision and have a higher long-term impact than schedule design, which involves a decision at the tactical level.

At the strategic level, Fan and MacHemehl (2011) published the first study explicitly considering equity in the public transportation network planning process, via a bi-level model solved by the genetic algorithm (GA). The upper level model determines the set of bus routes and frequencies of the routes for the purpose of minimising the sum of total user cost, operator cost, and cost resulting from unserved demand. The lower level model depicts travel using the route on which users incur minimum travel costs. The spatial equity constraint established in Meng and Yang (2002) is adopted in the upper level model. Ferguson et al. (2012) focused on the frequency setting problem. An accessibility-based equity objective is devised and explicitly minimised, taking demand distribution into account. In line with the modelling framework developed by Fan and MacHemehl (2011), Camporeale et al. (2016)
considered both the horizontal and vertical equity measures. They proposed a heuristic solution procedure like the general public network planning heuristic proposed by Ceder (2016). Horizontal equity is concerned with proportional distribution, while vertical equity focuses on the distribution of an attribute among specific groups (Welch 2013). Their work was extended in Camporeale et al. (2019) by proposing revised Gini indices for measuring horizontal and vertical equity. Comparatively speaking, horizontal equity focuses on efficiently moving a large number of people, while vertical equality concerns an individual’s accessibility needs (Xu et al. 2016). Ruiz et al. (2017) devised a bus frequency optimisation methodology built on a simple spreadsheet integrating geographic data about bus routes and socioeconomic information about the city, then solved by a commercial solver (i.e., Risk Optimiser). Kim et al. (2019) proposed a modal equality that captures the difference between transit time and car travel time. Table 1 summarises the preceding literature and presents the uniqueness of the study in terms of considering the equity issue in the transit network problem. Compared with the existing literature, the novelty of the study is twofold. One is the development of a bilevel, multi-objective model, and the other is proposing to maximise the minimum improvement in the spatial equity (see Section 3.2) as for the equity objective.

Table 1. Literature on the transit network design problem considering equity

<table>
<thead>
<tr>
<th>References</th>
<th>Bilevel/Single level</th>
<th>Multi/Single objective</th>
<th>Equity measurement</th>
<th>How equity is considered</th>
<th>Scale of the experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan and Machemehl 2011</td>
<td>Bilevel</td>
<td>Single*</td>
<td>Spatial equity</td>
<td>Constraint</td>
<td>Hypothetical; 28 zones, 3 bus routes.</td>
</tr>
<tr>
<td>Ferguson et al. 2012</td>
<td>Single</td>
<td>Single</td>
<td>Standard deviation accessibility</td>
<td>Minimize</td>
<td>Mid-sized US metropolitan area; 85 OD pairs, 26 existing and 3 new routes</td>
</tr>
<tr>
<td>Camporeale et al. 2016</td>
<td>Single</td>
<td>Single</td>
<td>Spatial equity</td>
<td>Constraint</td>
<td>Hypothetical; 9 nodes, 9 OD pairs, 1 route**</td>
</tr>
<tr>
<td>Ruiz et al. 2017</td>
<td>Single</td>
<td>Multi</td>
<td>Gini coefficient</td>
<td>Minimize</td>
<td>City of Palma; 88 districts, 31 bus routes</td>
</tr>
<tr>
<td>Camporeale</td>
<td>Single</td>
<td>Single</td>
<td>Revised Gini</td>
<td>Constraint</td>
<td>Hypothetical; 10 nodes, 9 OD</td>
</tr>
</tbody>
</table>
et al. 2019 | index | City of Molfetta | pairs; 28 zones, 6 routes
---|---|---|---
Kim et al. 2019 | Bilevel | Single | Modal and spatial equity | Constraint | Hypothetical | 15 nodes, 26 bus routes
This study | Bilevel | Multi | Spatial equity | Max-min | Hypothetical | 20 bus routes, 24 zones, 528 OD pairs

*This is considered a single objective model as it uses the weighted sum method to aggregate different costs in one singular value.

**This study designs the signal settings on the intersections traversed by the bus route.

The last column of Table 1 briefly summarises the scale of the experiments. First, it can be seen that both hypothetical and realistic networks are used in the existing literature. In general, the works that use a hypothetical network focus more on the methodology (e.g., Fan and Machemehl, 2011), while the works that use a realistic network could provide more insight on considering fairness to real-life implications. For example, Ferguson et al. (2012) concluded that applying their methodology could benefit the areas with a high percentage of low-income persons. Secondly, it is interesting to notice that even in the works that use a real-life network, the number of bus routes considered is manageable and comparable to the numerical example in this study, meaning that the method developed in this study has the potential to be applied to solve a realistic network.

In the literature reviewed above, optimisation of frequency setting attracts more attention than schedule design or route structure. This is explained in the following two ways: first, as a strategic decision variable, changes in frequency will induce higher and longer impact than can be achieved by changes in the schedule design – a variable at the tactical decision level. Secondly, compared with line planning, adjusting frequency does not require changes in infrastructure or significant investments, and often results in improved service and reduction of operational costs (Ruiz et al. 2017).
Nevertheless, there are two critical issues that have not been addressed in the literature on optimising transit frequency setting to achieve equity. One is to use a transit assignment model that would allow for gaining insight into passengers' response to changes in the transit network. Although Camporeale et al. (2016) encapsulated equilibrium constraints, the unique feature of the transit assignment problem, i.e., the common line problem, was not considered. The common line problem refers to the route choice problem at transit stops served by several competing transit lines. To address this problem, there are two methods distinguished by their network representations: the hyperpath graph network representation (Nguyen and Pallottino 1988, Wu et al. 1994, Cominetti and Correa 2001, Hamdouch et al. 2004, Gentile et al. 2005, Hamdouch and Lawphongpanich 2008, Cortés et al. 2013, Sun et al. 2013, Hamdouch et al. 2014, Li et al. 2015, Chen and Nie 2015, Oliker and Bekhor 2018, 2020, Xu et al. 2020) and the route-section network representation (de Cea and Fernández 1993; Lam et al. 1999, Nielsen 2000, Lam et al. 2002, Teklu 2008, Szeto et al. 2011a, Leurent 2012, Szeto et al. 2013, Szeto and Jiang 2014, Jiang and Szeto 2016, Sun and Szeto 2018, 2019). The merit of the hyperpath graph representation is that the optimal set of attractive lines can be easily determined, while the route-section representation allows the development of a link-based formulation and the adoption of available algorithms developed for traffic assignment models to offer solutions. Speaking of the capacity constraints, both the soft capacity and hard capacity constraint method can be found in the literature. Their difference is that the former allows passengers to board a fully occupied vehicle, while the latter prohibits that. The hard capacity constraint approach could be more realistic in some situations; for example, under COVID-19 restrictions, the bus operator in Copenhagen restricts the number of passengers on a bus. Nevertheless, due to insufficient capacity there may not be a feasible solution to the problems. On the contrary, the soft capacity constraint approach ensures the existence of a solution or even the uniqueness of the solution under certain conditions such as
the compact solution set and monotonic mapping function (see, e.g., theorem 1.4 of Nagurney 1993) facing the risk of generating unrealistic link flows that exceed link capacities.

The other critical issue which begs addressing is the effect of travel time uncertainty. Due to various factors such as road accidents, signal breakdown, weather conditions, etc., the travel time components are indeed uncertain. The stochastic travel time will affect both the transit network design and passenger route choice, resulting in different values for the equity measures. Recently, Chen et al. (2019) found that there is a considerable bias towards traditional, deterministic, individual accessibility measures to evaluate accessibility equity under travel time uncertainty, and they developed two reliability-based equity measures.

To fill the identified reach niches, this study develops a bi-level modelling framework for the transit frequency setting problem, while considering the effect of supply-side uncertainty. The upper level problem is the transit frequency setting problem, while the lower level problem is the transit assignment problem. Taking the fairness issue into account, we first adopt the spatial equality metric as the equity measure for an OD pair.

The metric is defined as the ratio between the effective travel cost before and after the changes in the transit services. Afterwards, the equity objective in the upper level is formulated to maximise the minimum improvement in the spatial equality metrics of all OD pairs, following the Rawlsian principle stating that a just society maximises the welfare of those of its members in the worst state of welfare (Feldman and Kirman 1974, Karsu and Morton 2015). At the lower level, the reliability-based transit assignment model developed by Jiang and Szeto (2016) is adopted. Their model captures passengers’ risk-aversion attitudes over travel time uncertainty, such that it can be solved efficiently by the extragradient method.
In the existing literature, metaheuristics such as an ant colony algorithm, a genetic algorithm, Tabu search, simulated annealing, and ABC algorithms are widely used to solve the transit network design problem (Bielli et al. 2002, Chakroborty and Wivedi 2002, Tom and Mohan 2003, Chakroborty 2003, Zhao and Zeng 2006, Fan and Machemehl 2006, Fan and Machemehl 2008, Unnikrishnan et al. 2009, Szeto and Jiang 2012, 2014). This is due to the fact that the transit route design problem is NP-hard (Fan and Mumford 2010, Guihaire and Hao 2008). This study adopts the ABC algorithm proposed by Karaboga (2005). The ABC algorithm belongs to a class of evolutionary algorithms inspired by the intelligent behaviour of honeybees finding nectar sources around the hive. It has been developed to solve various problems. For example, Szeto et al. (2011b) improved the ABC algorithm to solve a capacitated vehicle routing problem. Szeto and Jiang (2012) enhanced the ABC algorithm to solve a single level transit network design problem without considering the in-vehicle congestion effect. Long et al. (2014) improved the ABC algorithm to solve a turn restriction design problem. Szeto and Jiang (2014) and Long et al. (2014) showed that their proposed ABC algorithm is better than the GA for solving their problems. Nevertheless, it has yet to be improved to the extent that it can solve multi-objective transit network design problems. This study extends the ABC algorithm to solve this problem.

Thus, the main contributions of this paper include the following:

1) Devising a bilevel, multi-objective model for the transit network design problem taking the equity issue into account;

2) Proposing a fairness objective that maximises the minimum improvement of all OD pairs in the spirit of the Rawlsian principle in the upper level model;

3) Adopting the reliability-based transit assignment model as the lower level model;

4) Developing a multi-objective ABC algorithm to solve the transit network design problem;
5) Examining the tradeoff between the objective of minimising the total effective travel cost and the objective of maximising the equity condition;

6) Demonstrating the effect of passengers’ risk aversive attitude on the fairness condition, the effect of fleet size on the Pareto frontiers, and the difference in the solutions on the Pareto frontier.

The remainder of this paper is structured as follows: Section 2 introduces the notations and assumptions used throughout this paper and the problem description. Section 3 develops the multi-objective bilevel formulation. The solution algorithm is depicted in Section 4. Section 5 presents the case studies and, finally, Section 6 concludes the paper and indicates directions for future research.

2 Network representation, assumptions, and problem description

2.1 Network representation

![Transit network representations](image)

Fig. 1. Transit network representations (de Cea and Fernandez 1993)

We consider a general transit network that consists of a set of transit lines and a set of transit stops (nodes), where passengers can board, alight, or transfer. The transit network is illustrated in Fig.1, where Fig.1(a) represents the transit network using transit lines, and
Fig. 1(b) is the corresponding route-section\(^1\) network representation following de Cea and Fernández (1993). In Fig. 1(a), there are 4 transit lines, L1 – L4. Fig. 1(b) contains 6 route sections, S1 – S6. The brace next to the section index states the set of common lines associated with the route section. For example, S3 (L2, L3) means that the set of common lines associated with route section S3 includes lines L2 and L3. This can be observed from Fig. 1(a) where both L2 (solid line) and L3 (dotted line) traverse nodes X and Y. The route-section network representation is used for addressing the common bus line problem. The common bus line problem is the route choice problem of passengers at stops served by several competing bus lines. The idea of the route-section network representation is to classify passengers waiting at transit stops (including origins or transfers) into different groups based on their next alighting node (that can be their next transfer location or destination), and then assign passengers according to the line’s relative frequency. There are two key concepts in the route-section approach:

1) Attractive lines: at a bus stop with more than one line which could take the passenger to his/her destination, s/he may only consider a subset of the lines, known as the set of attractive lines, and board the first arriving bus belonging to the set. By assuming that the passenger minimises his/her travel time, a hyperbolic model can be built and solved to find the set of attractive lines. Once the set of attractive lines is determined, passengers are distributed among different lines based on the relative frequency of the line.

2) Competing sections: whereas the route-section network representation groups passengers and constructs a route-section wherever there is direct service between two stops, the number of passengers travelling via one line includes passengers travelling via different route-sections. When adopting a BPR (Bureau of Public Roads) type congestion function

---

\(^1\) In this study, route-section, section, and link are used interchangeably.
to compute the congestion cost, the flow associated with one route-section depends not
only on the flow travelling through the route-section considered, but also on the flow
travelling through the competing sections sharing lines in common with the section
considered. For example, in Fig. 1(b) both S5 and S2 contain line L2. Thus, when
computing the congestion cost of S2, the flow travelling through section S5 should be
taken into account. For section S2, S5 is considered a competing section, while for section
S5, S2 is considered a competing section.

2.2 Assumptions

In accordance with the literature, the following assumptions are made.

A1) Passengers arrive randomly, consider a set of attractive lines, and board the first arriving
vehicle belonging to the set of attractive lines.

A2) Stochastic vehicle headways with the same distribution function (i.e., exponential
distribution) are assumed for vehicles serving different lines.

A3) The travel demand between each OD pair in the system is assumed to be known. This is
a reasonable assumption for the frequency setting problem, usually considered a strategic
decision.

A4) The passenger selects the route that minimises his/her effective travel cost, where the
effective travel cost is composed of expected travel cost and a safety margin computed based
on the variance of travel time and passengers’ risk-aversion attitudes. Calibrating the risk
aversion parameter is important but is beyond the scope of this study.

A5) For simplicity, changing the frequency will not induce changes in the set of attractive
lines associated with a route-section.

Assumptions A1) - A3) are the fundamental assumptions for the frequency-based transit
2011a, etc.). A4) is made to capture the effect of travel time uncertainty on passengers’ travel behaviour.

### 2.3 Problem description

Based on the preceding network representation and assumptions, the problem considered in this study is the development of an optimisation model to determine optimal transit frequency setting by allocating bus fleets to a set of bus lines considering 1) the minimum/maximum frequency requirement, 2) integrality of fleet allocation, 3) total effective travel cost and the fairness condition for all OD pairs, 4) passengers’ route choice behaviour under stochastic travel time.

The problem is formulated as a bilevel programming model. The first and second points are explicitly formulated as constraints in the upper level model. The third point is captured by considering the two objectives simultaneously in the upper level model. The last point is addressed by adopting the reliability-based transit assignment model as the lower level model.

### 2.4 Notations

The following key notations are used throughout this paper.

#### Sets

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Set of transit lines</td>
</tr>
<tr>
<td>$L_s$</td>
<td>Set of attractive transit lines associated with route-section $s$</td>
</tr>
<tr>
<td>$Q$</td>
<td>Set of OD pairs</td>
</tr>
<tr>
<td>$D$</td>
<td>Set of destination nodes</td>
</tr>
<tr>
<td>$N$</td>
<td>Set of nodes</td>
</tr>
<tr>
<td>$S$</td>
<td>Set of route-sections</td>
</tr>
<tr>
<td>$A_i^+$</td>
<td>Set of route-sections (approaches) leaving node $i$</td>
</tr>
</tbody>
</table>
\( A_i^- \) Set of route-sections (approaches) entering node \( i \)

**Parameters**

| \( t_i^l \) | Travel time through route-section \( s \) via line \( l \) |
| \( E[t_i^l] \) | Expected travel time of line \( l \) travelling through route-section \( s \) |
| \( Var[t_i^l] \) | Variance of travel time of line \( l \) travelling through route-section \( s \) |
| \( K_i \) | Capacity of a single vehicle of line \( l \) |
| \( g^{id} \) | Travel demand between nodes \( i \) and destination \( d \) |
| \( V_{\text{max}} \) | Maximum feet size |
| \( f_{\text{min}}, f_{\text{max}} \) | Minimum and maximum frequency |
| \( \mu_r, \mu_w, \mu_\epsilon \) | Parameters to convert travel time, waiting time, and passengers’ perceived congestion to monetary values |
| \( \beta_0, \beta_1, \beta_2, \beta_3 \) | Parameters used in the BPR type congestion cost function |

**Variables**

| \( f_l \) | Frequency of line \( l \) |
| \( \alpha_i^d \) | Proportion of passengers travelling toward destination \( d \) via route-section (approaches) \( s \) |
| \( v_i^d \) | Number of passengers travelling toward destination \( d \) via route-section (approaches) \( s \) |
| \( \pi^{rd} \) | Equilibrated effective travel cost between nodes \( r \) and \( d \) |
| \( q_i^d \) | Number of passengers leaving node \( i \) toward destination \( d \) |

**3 Bilevel formulation**

**3.1 Lower level problem**

The lower level problem models passengers’ route choice behaviour under a fleet allocation and corresponding frequency setting obtained from the upper level model. The reliability-
based transit assignment model developed by Jiang and Szeto (2016) is adopted. The merit of their model is twofold. First, it captures the effects of supply-side uncertainty on passengers’ route choice behaviour. Second, it uses a novel approach-based formulation, which can be efficiently solved by an extragradient method that only requires a mild condition to converge. The cost components and the lower level model are briefly introduced below.

3.1.1 Cost components

The travel cost associated with a route-section \( s \), \( C_s \), is defined by

\[
C_s = \mu_T T_s + \mu_W W_s + \mu_\Phi \Phi_s, \quad \forall s \in S, \tag{1}
\]

where \( T_s \), \( W_s \), and \( \Phi_s \) denote the in-vehicle travel time, passengers’ waiting time for the first arriving vehicle, and perceived congestion cost incurred from insufficient vehicle capacity, respectively. \( \mu_T \), \( \mu_W \), and \( \mu_\Phi \) are the corresponding coefficients to convert these values into monetary values. Eq. (1) states that the travel cost associated with route-section \( s \) is comprised of the in-vehicle travel cost, waiting cost, and passengers’ perceived congestion cost.

Due to supply-side uncertainties, the cost components are stochastic and modelled as random variables. The expectation and variance of link travel costs are expressed as

\[
E[C_s] = \mu_T E[T_s] + \mu_W E[W_s] + \mu_\Phi E[\Phi_s], \quad \forall s \in S \quad \text{and} \tag{2}
\]

\[
Var[C_s] = \mu_T^2 Var[T_s] + \mu_W^2 Var[W_s] + \mu_\Phi^2 Var[\Phi_s], \quad \forall s \in S \tag{3}
\]

where \( E[\cdot] \) and \( Var[\cdot] \) denote the expected value and variance of the term in the bracket. It is worth noting that Eq. (3) implies that the travel costs among different sections are independent. However, in practice, the travel costs between different sections are likely to be dependent. For example, a delay occurring in an upstream section could propagate the
downstream section. In such a case, a covariance term should be added to the equation. Nevertheless, to a certain extent, the covariance could be mitigated, making the independent travel cost assumption reasonable, if the bus operator applies different operation tactics during real time bus operation, such as bus holding, stop skipping, etc.

Mathematically, the notations in Eqs. (2) and (3) are expressed by:

\[ E[T_s] = \sum_{i \in L_s} w_i E[t'_{s,i}], \forall s \in S \] (4)

\[ E[W_s] = \frac{1}{\sum_{i \in L_s} f_i}, \forall s \in S \] (5)

\[ E[\Phi_s] = \beta_0 \beta_3 \left( \frac{\beta_3 V_s + \beta_5 V_s}{\sum_{i \in L_s} f_i K_i} \right)^{\beta_3}, \forall s \in S \] (6)

\[ \text{Var}[T_s] = \sum_{i \in L_s} \left( w_i \right)^2 \text{Var}[t'_{s,i}] + \sum_{i \in L_s} \sum_{j \in L_s} \text{Cov}[t'_{s,i}, t'_{s,j}], \forall s \in S \] (7)

\[ \text{Var}[W_s] = \left( \frac{1}{\sum_{i \in L_s} f_i} \right)^2, \forall s \in S \] (8)

\[ \text{Var}[\Phi_s] = (\beta_0)^2 \left( (2\beta_3)^2 - (\beta_3)^2 \right) \left( \frac{\beta_3 V_s + \beta_5 V_s}{\sum_{i \in L_s} f_i K_i} \right)^{2\beta_3}, \forall s \in S \] (9)

The details of the derivations and discussions on the means and variance can be found in Szeto et al. (2011a) and Szeto and Jiang (2016). The following remarks are noteworthy:

1) Every cost component is expressed as a function of frequency, meaning that once the frequency is changed, all cost components will be affected. In general, a higher frequency...
induces a lower expected travel cost, waiting cost, and passenger perceived congestion cost.

2) In Eqs. (4) and (7), expected travel time $E[t'_s]$, variance of travel time $\text{Var}[t'_s]$, and covariance $\text{Cov}[t'_s,t''_s]$, are assumed to be given and treated as input parameters.

3) In Eqs. (4) and (7), $w'_l$ denotes the relative frequency of line $l$ on route-section $s$ and is calculated by

$$w'_l = \frac{f'_l}{\sum_{i \in L_s} f'_i}, \forall s \in S, l \in L_s.$$  \hspace{1cm} (10)

4) In Eqs. (6) and (9), $v_s$ and $\overline{v}_s$, respectively, represent the flow travelling on route-section $s$ and the flow travelling on the competing sections of route-section $s$ (See Section 2.1. for the concept of competing section). Mathematically, they are obtained by

$$v_s = \sum_{d \in D} \sum_{l \in L_s} v'_d, \forall s \in S \text{ and}$$

$$\overline{v}_s = \sum_{d \in D} \sum_{m \in S, m \neq s} \sum_{j \in L_c \cap L_m} v'_d, \forall s \in S,$$  \hspace{1cm} (11)

$$\text{where}$$

$$v'_d = v'_s w'_l, \forall s \in S, d \in D, l \in L_s.$$  \hspace{1cm} (12)

Eqs. (11) and (12) mean that $v_s$ and $\overline{v}_s$ are computed by aggregating the flow travelling on the lines associated with section $s$ and its competing sections. $\delta'_s$ is the competing section indicator. It equals 1 if section $m$ is a competing section of section $s$. Otherwise, it is 0. Eq. (13) means that the flows travelling on route-section $s$ determined by the transit assignment model are distributed to the transit lines on that section based on relative frequencies.
3.1.2 Effective travel cost

Due to the variability in the in-vehicle travel time and waiting time, a passenger cannot accurately anticipate the exact trip time/cost required to complete his/her journey. To counter such variability, the passenger may choose to depart early, adding additional buffer time to the trip time. Such additional time is commonly referred to as the safety margin that depends on both the purpose of the trip and the individual’s risk-taking behaviour. A prevailing approach to capturing the safety margin is to define an effective travel cost that includes both the expected trip time and the safety margin. The effective travel cost can be formulated either via a mean-standard deviation formula or a mean variance formula (Lo et al. 2006, Shao et al. 2008, Lam et al. 2008, Siu and Lo 2008). In this study, we adopted the latter, since it allows us to apply the dynamic programming method to conduct network loading without path generation. Mathematically, the effective travel cost associated with section \( s \) can be expressed as

\[
\mu_s = E[C_s] + \rho \text{Var}[C_s], \forall s \in S
\]

where \( \rho \) represents a parameter that captures travellers’ risk averse attitude. When \( \rho = 0 \), it means that travellers are risk-neutral and when \( \rho > 0 \), it means that travellers are risk-aversive.

3.1.3 Reliability-based transit assignment

Given the effective travel cost of a link defined by Eq. (14) and a link-path incidence matrix, the effective travel cost associated with a path can be obtained. Accordingly, we can define the path-based, reliability-based user equilibrium as follows: For each OD pair, the effective travel cost on used paths are equal to each other and do not exceed those on unused paths. Based on this, Szeto et al. (2011a) developed a path-based formulation. Nevertheless, the path-based formulation requires path generation (i.e., providing the link path incidence
matrix) and path storage, which may not be computationally efficient in large network applications. Accordingly, Jiang and Szeto (2016) reformulated the path-based formulation using the concept of approach proportion, in which an approach of a node is defined by the links coming out from that node, and an approach proportion is defined as the proportion of passengers leaving a node via the link considered.

Denote the node as $i$ and the set of links leaving node $i$ as $A_i^+$. Then the proportion of passengers leaving node $i$ and going to destination $d$ via section $s$ is denoted as $\alpha_s^d$, $s \in A_i^+$, and must satisfy the following conditions.

$$0 \leq \alpha_s^d \leq 1, \forall s \in S, d \in D$$  \hspace{1cm} (15)

$$\sum_{s \in A_i^+} \alpha_s^d = 1, \forall i \in N, d \in D$$  \hspace{1cm} (16)

where $D$ denotes the set of destination nodes and $N$ is the set of nodes in the network. Eqs. (15) and (16) are the definitional constraints, stating that the value of one approach proportion is nonnegative and no greater than 1, and the sum of the total approach proportions coming out from node $i$ equals 1. In addition, the following flow conservation constraint is required in the approach-based formulation.

$$\sum_{s \in A_i^+} \alpha_s^d q_i^{(s)} + g_i^d = \sum_{s \in A_i^+} \alpha_s^d q_i^d, \forall i \in N, d \in D.$$  \hspace{1cm} (17)

Eq. (17) is the flow conservation constraint, meaning that the sum of the number of passengers traversing node $i$ and the demand generated at node $i$ toward destination $d$ equals the total number of passengers leaving node $i$ toward destination $d$ via all approaches emitting from node $i$. 
Using the concept of the approach proportion, the approach-based, reliability-based equilibrium condition can be expressed as the following nonlinear complementarity problem (NCP).

\[
\begin{align*}
\left( \pi_s^{(s) \text{d}} - \pi_s^{(s) \text{d}} \right) \alpha_i^d &= 0, \forall s \in A^+, i \in N, d \in D \\
\pi_s^{(s) \text{d}} - \pi_s^{(s) \text{d}} &\geq 0, \forall s \in A^+, i \in N, d \in D \\
\alpha_i^d &\geq 0, \forall s \in S, d \in D
\end{align*}
\]

(18)

where

\[
\pi_s^{(s) \text{d}} = \pi_s^{h(s) \text{d}} + u_s, \forall s \in S, d \in D, \text{ and}
\]

\[
\pi_s^{(s) \text{d}} = \min_{s \in d(i)} \pi_s^{h(s) \text{d}}, \forall s \in S, d \in D,
\]

(19) (20)

In the above equations, \( t(s) \) and \( h(s) \), respectively, represent the tail and head nodes of section \( s \). \( \pi_s^{(s) \text{d}} \) is the minimum effective travel cost between the tail node of section \( s \) and destination \( d \), and \( \alpha_i^d \) is the effective travel cost between the tail node of section \( s \) and destination \( d \) via section/approach \( s \). Eq. (19) states that the effective travel cost via approach \( s \) is computed by adding the effective travel cost associated with section \( s \) to the minimum effective travel cost between the head node of section \( s \) and destination \( d \).

Eq. (18) implies that if an approach proportion carries a flow (i.e., \( 0 < \alpha_i^d \leq 1 \)), then the effective travel cost from the tail node of section \( s \) to destination \( d \) via section \( s \) must equal the minimum effective travel cost from the tail node of section \( s \) to destination \( d \) (i.e., \( \pi_s^{(s) \text{d}} = \pi_s^d \)). Eq. (18) can be reformulated as a Variational Inequality (VI) problem, which is to determine \( \alpha^* = \left[ \alpha_s^{ds} \right] \) such that

\[
(\alpha - \alpha^*)^T \pi(\alpha^*) \geq 0, \forall \alpha \in \Omega_\alpha,
\]

(21)
where $\mathbf{a}^* = \begin{bmatrix} \alpha_s^d \end{bmatrix}$ is the optimal solution vector and $\Omega_\alpha$ is the solution space of the approach proportion satisfying Eqs. (15) - (17). $\mathbf{\pi}(\mathbf{a}) = \begin{bmatrix} \pi_s^d \end{bmatrix}$ is the mapping function.

The VI formulation can be solved by the extragradient method, which only requires a mild condition to converge. Since the focus of this study is to examine the effect of considering fairness in the transit frequency setting problem, the details of the derivation of the VI formulation and the solution method are not presented in this paper. Nevertheless, the following remarks warrant mention for the completeness and clarity of this paper.

1) For each iteration of the extragradient method, the values of the proportion variables, $\mathbf{a} = \begin{bmatrix} \alpha_s^d \end{bmatrix}$, are updated, since the computation of the perceived congestion cost, i.e., Eqs. (11) – (13) require the input of flow. Therefore, a flow update algorithm is embedded in the extragradient method in order to obtain the flow values from the values of the proportion variables.

2) The approach-based formulation obviates path storage and enumeration, thus potentially reducing the computation burden caused by path enumeration and storage in large network applications. Although the method does not provide a unique path flow solution, it offers the equilibrated effective travel cost associated with each node, which is sufficient for the proposed strategic frequency setting problem.

### 3.2 Upper level problem

Two objectives are considered in the upper level formulation. One is the total effective travel cost for passengers, $\varepsilon^{\text{eff}}$, representing the efficiency of the system. The other is the fairness objective, $\varepsilon^{\text{fair}}$, defined by the minimum improvement in the effective travel cost among all OD pairs. Mathematically, they are expressed as

$$
\varepsilon^{\text{eff}} = \sum_{(r,d) \in Q} \pi^{rd}(\mathbf{f})g^{rd} \quad \text{and} 
$$

(22)
\[ f_i = \frac{V_i}{2E[F_l]} \quad \forall l \in L \] (24)

\[ \sum_{l \in L} V_i \leq V_{\text{max}}, \forall l \in L \] (25)

\[ f^l \leq f_{\text{max}}, \forall l \in L \] (26)

\[ f^l \geq f_{\text{min}}, \forall l \in L \] (27)

\[ V_i \in \mathbb{Z}^+ \] (28)

Eq. (24) defines the relationship between the frequency, \( f_i \), and the number of buses allocated to line \( l \), \( V_i \), where \( E[F_l] \) is the expected trip time associated with line \( l \).

Following Li et al. (2008) and Szeto et al. (2013), the expected trip time considers layover time, dwell time, travel time, and variance of travel time. Eq. (25) is the fleet size constraint requiring the total number of buses allocated to the bus lines not be greater than the available fleet size. Eqs. (26) and (27) together restrict the upper and lower bounds of the frequency.

Eq. (28) defines that \( V_i \) must be a nonnegative integer value.
3.3 Overview of the bilevel model

Based on the preceding sections, we can formulate the proposed transit frequency setting problem as a bilevel programming problem, as follows.

Upper level:

\[ Z = \{ \min z^{\text{eff}}, \max z^{\text{fair}} \} \]

s.t. Eqs. (24) - (28)

Lower level: Eq. (21)

The upper level problem is the transit frequency setting problem of simultaneously minimising total effective travel cost and maximising the fairness measurement, while the lower level problem is the VI formulation of the reliability-based transit assignment problem.

4 Solution method

To solve the proposed bilevel model, the ABC algorithm is adopted. The ABC algorithm is an evolutionary algorithm inspired by the foraging behaviours of honeybees. A single objective ABC algorithm involves three phases: employed bee, onlooker, and scout bee phases. The ABC algorithm iterates three phases until a predefined stopping condition is reached. Initially, a population of solutions is generated by the employed bees. Afterwards, in each iteration, the employed bees start to find a new neighbourhood solution. If the new neighbourhood solution is better than the existing one, then the better one is kept. When all employed bees have finished this process, the quality information associated with solutions are shared with onlookers. Each of the onlookers then selects a solution according to the probability proportional to the solution quality and finds a new neighbourhood solution. If the newly found one is better, it is kept. Finally, if a solution has not been improved within a certain limited number of times, the employed bee associated with the solution abandons the
solution and becomes a scout bee to generate a new solution. To solve the bilevel problem, an MOABC algorithm is developed by utilising a fixed sized archive to maintain good solutions. In what follows, we first provide an overview of the MOABC algorithm, and then explain the details of each phase.

4.1 Overview MOABC Algorithm

Fig. 2. Overview of MOABC

An overview of the proposed MOABC algorithm is given in Fig. 2. The algorithm starts by initialising all parameters, including population size, number of employed bees, number of onlookers, and number of scouts. Then, the algorithm iterates the employed bee, onlooker, and scouts phases sequentially until a predefined maximum number of iterations, $I_{\text{max}}$, is reached. At the end of every iteration, an external multi-objective archive is updated to keep nondominated solutions.

4.2 Solution representation and initialisation

Each solution is comprised of $|L|$ elements, where each element contains an integer number representing the number of vehicles assigned to the corresponding transit line. An initial
feasible solution is generated by three steps. First, for each line, the maximum and minimum feasible fleet size are calculated. A feasible fleet size is the number of vehicles resulting in a feasible frequency according to Eqs. (24), (26), and (27). Secondly, an integer number within the maximum and minimum feasible fleet size is generated. Lastly, a solution repair operator (see Section 4.4) is performed to ensure the generated initial solution is feasible.

4.3 Neighbourhood operators

The neighbourhood operators are designed to generate new neighbourhood solutions in the employed bee and onlooker phases. Four operators are devised as described below:

1) Swap: This operator randomly selects two lines and swaps their allocated fleet size.

2) n-Transfer: This operator randomly selects two lines and transfers a random fleet number from one line to the other.

3) Increase: This operator randomly selects one bus line and increases its fleet by 1 if the result does not violate the maximum frequency constraint and maximum fleet constraint.

4) Reduce: This operator randomly selects one bus line and reduces its fleet by 1 if the result does not violate the minimum frequency constraint.

One of the above operators is randomly selected to generate a neighbourhood solution. Since all operators rely on random operations, the feasibility of the resultant solution may not be guaranteed. Therefore, a repair operator is devised and performed after each neighbourhood operator.

4.4 Repair operator

The repair operator is designed to ensure that the solution to be evaluated in the lower level is feasible. A feasible solution is a solution that satisfies integrality constraint (28), fleet size constraint (25), and frequency boundary constraints (26) and (27). Constraint (28) is
automatically satisfied since all random numbers generated are integer numbers. The other
two constraints are not necessarily guaranteed. To remedy this, a repair operator is devised. It
contains two steps. The first step is to ensure that the frequency boundary constraints (26) and
(27) are satisfied by adding (removing) fleet from the line with frequency less (more) than the
minimum (maximum) frequency. The second step is executed if the resultant total fleet size is
more than the maximum fleet size after the first step. In the second step, the excessive fleet is
removed from the bus lines, the frequency of which, after removal, does not violate the
minimum frequency constraint.

4.5 Employ bee phase

In this phase, a neighbourhood solution is generated, associated with each solution in the
existing population using one randomly selected neighbourhood operator followed by the
repair operator. The existing solution is replaced if it is dominated by the new solution.
Otherwise, the newly generated solution is abandoned, and a limit counter is increased by 1.

4.6 Onlooker phase

In this phase, a roulette wheel selection is conducted to determine which solution obtained by
an employed bee is selected by an onlooker. Then, a neighbourhood search is conducted for
each solution selected by an onlooker, followed by evaluating the fitness of each
neighbourhood solution. Similar to the employed bee phase, the newly generated solution is
kept if it dominates, or is nondominated, by the existing solution. If the newly generated
solution is abandoned, then increase the limit counter by one. Otherwise the limit counter
should be reset to zero.
4.7 Scout bee phase

In the scout bee phase, all of the solutions are scanned and the solution that fails to improve within limit successive iterations is abandoned and replaced by a newly generated random solution. Then, set the limit counter associated with the solution to zero.

4.8 Fitness evaluation

The fitness value represents the quality of a solution generated. For the proposed multi-objective optimisation problem, the quality of a solution is measured by the number of solutions dominated by the generated solution. The higher the number, the better the quality of the solution.

4.9 Update multi-objective archive

To maintain the non-dominated solutions, this study uses a fixed-sized external archive in a grid environment. For the bi-objective optimisation model, it is assumed that a two-dimensional space exists, corresponding to the two objectives considered. The space of the objective values is sliced into a $\varepsilon_{\text{eff}} \times \varepsilon_{\text{fair}}$ grid, where $\varepsilon_{\text{eff}}$ and $\varepsilon_{\text{fair}}$ denote the number of grid lines for the space of the efficiency and fairness objectives, respectively. As a result, $\varepsilon_{\text{eff}} \times \varepsilon_{\text{fair}}$ small boxes are created on the grid and all solutions can be placed in the grid based on their objective values. Then, if one box contains more than one solution, only the nondominated solutions are kept. If there is more than one nondominated solution, the one with the minimum distance with respect to the utopian point is maintained.

5 Numerical examples

To investigate the properties of the problem, the four-line network in de Cea and Fernández (1993) is adopted. Fig. 3 presents the network together with the path and OD pair setting in the example. Two OD pairs are considered, A-B and X-B. Without further specification, the
minimum and maximum frequencies are set at \( f_{\text{min}} = 2 \) buses/hour, \( f_{\text{max}} = 14 \) buses/hour, and \( \rho \) is set at 0.05. In order to clearly illustrate the effect of frequency by changing the frequency continuously within its feasible region, the integrality constraint and total fleet size constraints are not considered in the experiments in Sections 5.1 and 5.2. All the experiments were coded in Fortran and complied with Intel Fortran Compiler and run on a personal computer with Intel(R) Core (TM) i5-3380M Cpu@2.90Ghz and 8G RAM.

![Diagram](image)

Fig. 3. Small network data

### 5.1 Tradeoff between efficiency and fairness

The experiment is designed to demonstrate the tradeoff between the two objectives. In the test, the frequency of line L2 was increased from 2.0 to 10.0 buses/per hour. The average computation time for executing the frequency-based transit assignment model and evaluate the objectives is 0.0063 seconds for this small network. To compute the fairness objective, the frequencies set in the base scenario are 6.0, 4.0, 8.0, and 15 buses/per hour associated with the four transit lines. Fig. 4 plots the changes in the values of the two objective functions with respect to the changes in the frequency of line L2. To facilitate the explanation, the equilibrated effective travel costs associated with the two OD pairs are plotted in Fig. 5.

It is obvious in Fig. 4(a) that the total effective travel cost is generally reduced as the frequency increases, since a higher frequency means a shorter waiting time and lesser perceived congestion costs. The dotted lines in the figure mark the knot points where the
reducing rate changes. In contrast, the fairness objective values fluctuate. When the frequency is below 3.40 buses/per hour, the fairness objective value is less than 0.0 due to the fact that the designed frequency is less than the frequency in the base scenario, which is 4.0 buses/per hour. Between 2.0 buses/per hour and 3.40 buses/per hour, the increment in the frequency of line L2 does not induce any changes in the effective travel cost of the two OD pairs as shown in Fig. 5, since line L2 is not utilised by the two OD pairs. When line L2’s frequency is greater than 3.40 buses/per hour, L2 becomes attractive to the travellers travelling between OD pair X-B, leading the reduction in the effective travel cost of the OD pair (see red line in Fig. 5). When the frequency is within 4.0 to 4.9 buses/per hour, although the total effective travel cost drops, the values of the fairness objective do not change at all. This is because only the effective travel cost associated with OD pair X-B is reduced, while that associated with OD pair A-B remains constant as shown in Fig. 5. Therefore, the fairness measurement for OD pair A-B is zero, as per the fairness objective computed by Eq. (23). The reason that the effective travel cost does not change is that all the passengers travelling between OD pair A-B travel via Path 1 and that only utilises line L1. Although increasing the frequency of line L2 reduces the effective travel cost using paths that utilises line L2, their effective travel costs are still not attractive enough.

The value of the fairness objective begins to increase when the frequency exceeds 4.90 buses/per hour and below 5.20 buses/per hour (see Fig. 4(b)). The improvement is mainly due to reduction in the effective travel cost associated with OD pair A-B (see blue line in Fig. 5 between 4.90 and 5.20) as some travellers switch to use line L2, which inevitably increases the perceived congestion cost for passengers travelling between OD pair X-B, making their effective cost increase (see red line in Fig. 5 between 4.90 and 5.20). When the frequency of line L2 is greater than 5.20 buses/per hour, more and more passengers travelling between OD pair A-B begin to use line L2, inducing a higher perceived congestion cost to passengers
travelling between OD pair X-B. Such an increase triggers the reduction in the fairness objective in Fig. 4(b). Nevertheless, the increment in the effective travel cost of OD pair X-B ceases when the frequency is above 6.15 buses/per hour since the increment in perceived congestion cost is counterbalanced by the reduction in waiting time and in-vehicle travel time. Therefore, the overall effective travel cost of OD pair X-B begins to decline, leading the improvement in the fairness objective.

In general, the changes in the effective travel cost and fairness are attributed to passengers’ route choices. Passengers travelling between OD A-B could reduce their own effective travel cost via using line L2, without considering the fact that such behaviour induces a higher perceived congestion cost to the passengers on the downstream flow of the bus services. This reveals the inequality between the different OD pairs under certain transit frequency settings.

It can be concluded, from this example, that within a certain range of frequency settings, there is a tradeoff between the total effective travel cost and the fairness objective. Nevertheless, the two objectives could be improved simultaneously, implying that a Pareto transit frequency setting can be achieved. It should be noticed that in the example such concurrent improvement occurs if the frequency is above 6.15 buses/hour or between 4.90 and 5.20 buses/per hour. If the operator does not have sufficient resources, i.e., fleet size, to provide this frequency, the improvement will not be achieved. This indeed demonstrates the importance and value of the proposed model.
(a) Changes in the total effective travel cost

(b) Changes in the fairness objective

Fig. 4. Changes in the values of the two objectives with respect to the increase in the frequency of line L2
Fig. 5. Changes in equilibrated effective travel costs with respect to the increase in the frequency of line L2

Fig. 6. Effect of $\rho$ on the total effective travel cost

5.2 Effect of $\rho$ on the objectives

One vital contribution of this study is its capturing of passengers’ risk aversive attitude via a parameter $\rho$ in the transit frequency setting problem. In this experiment, the value of $\rho$ was increased from 0.0 to 0.20 and the changes in the total effective travel cost and fairness objectives are plotted in Figs. 6 and 7, respectively. It is obvious that the total effective travel cost is higher at a larger value of $\rho$ in Fig. 6. This is as expected since a higher value of $\rho$
induces a larger safety margin, taken into account in the effective travel cost. In Fig. 7, when
the frequency is below 4.0 buses/per hour or above 6.65 buses/per hour, the fairness objective
value obtained under $\rho = 0.20$ is much lower or higher than that obtained under $\rho = 0.00$
and $\rho = 0.10$, implying that a higher value of $\rho$ could exaggerate the unfairness and fairness
condition. When the frequency is between 4.0 and 5.0 buses/per hour, the fairness objective is
not affected by the value of $\rho$. When the frequency is between 5.0 and 6.65 buses/per hour,
the value of $\rho$ does not have a monotonic effect on the fairness objective value. The reason
that $\rho$ has different effects on the fairness objective value is that effective travel costs vary
with the changes in the value of $\rho$, resulting in different route choices and objective values.

![Fig. 7. Effect of $\rho$ on the value of the fairness objective](image)

### 5.3 Effect of fleet size on the Pareto frontier

The effect of the fleet size on the Pareto frontier is illustrated in Fig. 8. In this experiment, the
maximum fleet size is increased from 7 to 9. In each case, a brute force method is used to
enumerate all feasible fleet allocations under the given fleet size. The resultant Pareto
frontiers are plotted in Fig. 8. The corresponding fleet allocation solutions and objective
values are presented in Table 2. Fig. 8 shows that the Pareto frontier moves towards the left
as the fleet size increases, meaning that the two objectives can be improved simultaneously by increasing the fleet size.

**Fig. 8.** Effect of the maximum fleet size on the Pareto frontier

**Table 2.** Objectives and fleet allocations of the points marked in Fig. 8.

<table>
<thead>
<tr>
<th>Total Fleet Size</th>
<th>Line L1</th>
<th>Line L2</th>
<th>Line L3</th>
<th>Line L4</th>
<th>Total effective Travel cost</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>12766</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Point A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>11913</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Point B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>11397</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Point D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>11053</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Point F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the far left point on the Pareto frontier corresponds to the model that only minimises the total effect on travel cost, while the far right point corresponds to the model that only maximises the fairness objective, the effect of considering the fairness objective on the fleet allocation can be concluded by observing Table 2. It can be seen that although the far left and far right points have different fleet allocation solutions, not all the fleet line allocations change. For example, the fleet allocations for bus lines L1 and L4 remain constant in all cases. Meanwhile, it shows that the additional fleet is not always allocated to the same line. For example, from point A to point C, the additional fleet is added to line L3, while from point C
to point E, the additional fleet is allocated to line L2. The above observations verify that considering fairness objective results in different solutions, and, thus, developing an optimisation model for determining the fleet allocation with respect to different objectives, is necessary.

5.4 Sioux Falls network

In this experiment, the Sioux Falls network is adopted to examine the effect of considering fairness. As the Sioux Falls network originally is built for a road network, 20 hypothetical transit lines are created such that all OD pairs in the network are connected via transit services. 30 runs of MOABC algorithm were executed, where each run used a different seed for generating random numbers used in the algorithm. The algorithm stops after 1000 iterations and the average computation time for the 30 runs is 446.32 seconds and the standard deviation of the computation time is 11.93 seconds. All nondominated solutions obtained are merged to plot one Pareto frontier in Fig. 10. The setting of bus lines is given in Table 3. The detail data of the network is available at the author’s GitHub repository (https://github.com/hkuwj/Reliability-based-Equitable-Transit-Frequency-Design-).

Fig. 10 shows that the fairness objective changes significantly, from -0.36 (left-most point) to 0.05 (right-most point). On the contrary, the variation in the total effective cost is minor, which is only 5% increment from 363168.46 (left-most point) to 379549.06 (right-most point). Moreover, the Pareto frontier reveals that the substitution rate between the two objectives varies. For example, at the left-hand side of the Pareto frontier, with a small sacrifice in the total effective travel cost, the fairness objective can be improved significantly, while in the middle and right-hand side of the Pareto frontier, the changes in the fairness are minor.
Fig. 9 Sioux Falls network with 20 bus lines

**Table 3.** Transit line setting of the Sioux Falls Network

<table>
<thead>
<tr>
<th>Line</th>
<th>Stop sequence</th>
<th>Line</th>
<th>Stop sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1, 3, 12, 11, 4</td>
<td>11</td>
<td>20, 21, 24, 13, 12, 3, 1</td>
</tr>
<tr>
<td>2</td>
<td>2, 6, 8, 7, 18, 20, 21, 24, 13</td>
<td>12</td>
<td>23, 22, 21, 20, 19, 17, 16, 8, 6, 2</td>
</tr>
<tr>
<td>3</td>
<td>7, 8, 16, 17, 19, 20, 21, 22, 23</td>
<td>13</td>
<td>13, 24, 21, 20, 18, 7</td>
</tr>
<tr>
<td>4</td>
<td>13, 24, 21, 20, 22, 23, 14, 15, 19</td>
<td>14</td>
<td>14, 15, 19, 20, 21, 22, 23, 24, 13</td>
</tr>
<tr>
<td>5</td>
<td>24, 13, 12, 11, 10, 15, 22, 21</td>
<td>15</td>
<td>4, 11, 14, 23, 24</td>
</tr>
<tr>
<td>6</td>
<td>21, 22, 15, 10, 9, 5, 6, 2</td>
<td>16</td>
<td>5, 9, 10, 15, 22, 21</td>
</tr>
<tr>
<td>7</td>
<td>20, 19, 17, 16, 8, 9, 10, 11, 12</td>
<td>17</td>
<td>2, 6, 8, 16, 17, 19, 20</td>
</tr>
<tr>
<td>8</td>
<td>18, 16, 10, 11, 12, 3, 1</td>
<td>18</td>
<td>12, 11, 10, 16, 18</td>
</tr>
<tr>
<td>9</td>
<td>1, 3, 4, 5, 9, 8, 7, 18</td>
<td>19</td>
<td>20, 19, 15, 10, 9, 5, 4, 3, 1</td>
</tr>
<tr>
<td>10</td>
<td>2, 6, 8, 9, 5, 4, 3, 1</td>
<td>20</td>
<td>12, 11, 10, 9, 8, 6, 2</td>
</tr>
</tbody>
</table>
Fig. 10. The Pareto frontier of the Sioux Falls network

Fig. 11. Fleet allocation of the three points marked in Fig. 10
Fig. 12. Distributions of the relative changes in the effective travel cost of all OD pair
Three points are marked on the Pareto frontier. The solutions of the fleet allocation and the distributions of the relative changes in the effect travel cost (i.e., \( \left( \pi_{\text{eq}}^{rd} - \pi^{rd}(f) \right) / \pi_{\text{eq}}^{rd} \)), the term within the brace in Eq. (23) of all OD pairs associated with the three points are plotted in Fig. 11 and Fig. 12, respectively. Fig. 11 illustrates that there does not exist a unique effect of the change in the fleet allocation, in addition to showing that different fairness level corresponds to different fleet allocation solution. For example, the number of fleets keeps either increasing or decreasing for bus lines 12 and 7 when the solution movies from point A to point C. On the contrary, the fleet allocations for lines 15 and 16 maintain constant, and those for lines 4 and 5 varies. In Fig. 12, the scatter plots on the left-hand side show that from point A to point B, the mean of the relative changes does not change much and the standard deviation reduces, while from point B to point C, the mean reduces, and the standard deviation does not change much. The histogram figures on the right-hand side reveal a trend of moving left, implying that the number of OD pairs with a higher reduction in the effective travel cost reduces to achieve a better fairness condition defined by the minimum improvement all OD pairs.

6 Conclusions

Motivated by considering the fairness issue in the transit frequency setting problem, this study develops a multi-objective bilevel programming model. The upper level problem is to determine the frequency setting to simultaneously minimise the total effective travel cost and maximise the fairness measurement which is defined as the minimum reduction in the equilibrated effective travel cost of all OD pairs in the network. The lower level problem is the reliability-based transit assignment problem that captures the uncertainty of travel time on passengers’ route choices. To solve the model, a multi-objective artificial bee colony (MOABC) algorithm is developed. Numerical studies demonstrate the tradeoff between the
total effective travel cost and the fairness measurement. It is shown that by using a large fleet size, both objectives can be improved simultaneously. The effect of passengers’ risk attitude has different effects on the fairness objective, depending on the frequency setting.

This study opens various paths to further research. First, the study only designs the frequency setting for the transit network assuming the network structure is given. It is possible to consider such an objective in designing the route structure (i.e., Wei et al. 2019), transit fare (i.e., Wang and Qu 2017; Wang et al. 2018), or timetable (Zhao et al. 2013; Zhan et al. 2020). It is expected that an equitable route structure could have a more significant effect on the fairness condition. Secondly, regarding the fairness measurement, this study based on Rawls’ theory of justice maximises the minimum improvement in the relative effective cost. It would be interesting to examine other fairness measures and compare the resultant designs and the fairness issue in other dimensions, such as mobility and health (Jiang and Szeto 2015). Thirdly, when formulating the effective travel cost in the lower level transit assignment model, this study assumes that the cost components between different route sections are independent. Explore a methodology that could relax the assumption without compromising the simplicity of the formulation and the algorithmic efficiency would be a challenging yet significant research direction. Fourthly, the reliability-based transit assignment model is adopted as the lower level model. It would be interesting to examine the effect of using a different assignment model such as a schedule-based model (Tong and Wong 1999, Nuzzolo et al. 2001, 2012, Poon et al. 2004, Wilson and Nuzzolo 2013, Gentile et al. 2016, Cats et al. 2016, Cats and West 2020, Xie et al. 2021, etc.) or a personalised assignment model (Ceder and Jiang 2019, 2020; Jiang and Ceder, 2021). More importantly, when a schedule-based transit assignment model is adopted. It would allow us to incorporate other factors that can affect passengers’ route choice, such as the departure time window, expected arrival time, and schedule delay, etc., and examine the fairness condition associated
with these factors. Last but not least, the paper develops a modelling framework for the transit network design problem with equity consideration. It will be interesting in the future to apply the method on a realistic network and examine the effect of the spatial distribution of demand on the transit services as well as the changes in the equity condition of the network.

Acknowledgement

The author thanks four anonymous reviewers for their valuable suggestions and constructive comments for improving the quality of the paper. The opinions expressed in this paper reflect the author’s own views. This research has been supported by the National Natural Science Foundation of China (Project No. 71971038 and 71701030).

References


