Transportation (2021) 48:2735–2763
https://doi.org/10.1007/s11116-020-10146-4

Autonomous and conventional bus fleet optimization for fixed-route operations considering demand uncertainty

Qingyun Tian[1] · Yun Hui Lin[2] · David Z. W. Wang[1]

Published online: 28 October 2020

Abstract

The emerging technology of autonomous vehicles has been widely recognized as a promising urban mobility solution in the future. This paper considers the integration of autonomous vehicles into bus transit systems and proposes a modeling framework to determine the optimal bus fleet size and its assignment onto multiple bus lines in a bus service network considering uncertain demand. The mixed-integer stochastic programming approach is applied to formulate the problem. We apply the sample average approximation (SAA) method to solve the formulated stochastic programming problem. To tackle the nonconvexity of the SAA problem, we first present a reformulation method that transforms the problem into a mixed-integer conic quadratic program (MICQP), which can be solved to its global optimal solution by using some existing solution methods. However, this MICQP based approach can only handle the small-size problems. For the cases with large problem size, we apply the approach of quadratic transform with linear alternating algorithm, which allows for efficient solution to large-scale instances with up to thousands of scenarios in a reasonable computational time. Numerical results demonstrate the benefits of introducing autonomous buses as they are flexible to be assigned across different bus service lines, especially when demand uncertainty is more significant. The introduction of autonomous buses would enable further reduction of the required fleets and total cost.The model formulation and solution methods proposed in this study can be used to provide bus transit operators with operational guidance on including autonomous buses into bus services, especially on the autonomous and conventional bus fleets composition and allocation.

Keywords Autonomous bus service · Demand uncertainty · Stochastic programming · Conic programming · Quadratic transform

✉ David Z. W. Wang
wangzhiwei@ntu.edu.sg

Qingyun Tian
qytian@ntu.edu.sg

Yun Hui Lin
isemlyh@gmail.com

1 School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

2 Department of Industrial Systems Engineering and Management, National University of Singapore, 1 Engineering Drive 2, Blk E1A 06-25, Singapore 117576, Singapore

Introduction

Bus transportation companies usually operate a fleet of buses to serve multiple bus lines. A generic problem faced by transit operators is how to determine the optimal fleet size, as well as the best allocation of available bus fleet onto the service lines. This problem can be classified as a vehicle assignment problem (VAP) (Geetha and Nair 1993; Zak et al. 2009). On the tactic level, VAP deals with the interactive balance between bus service supply and demand. However, one intrinsic feature of public transit service demand is that it varies on a day-to-day basis. Therefore, how to handle the demand uncertainty is one of the most challenging questions for bus service operation design in terms of determining the optimal fleet size and vehicle assignment.

Generally, other than vehicles, bus drivers are also important resources for offering bus services with conventional human-driven buses. New bus captains have to complete a time-consuming training program before driving a scheduled bus (News 2017). The shortage of trained bus drivers poses another great challenge to the operation of the conventional bus transit services.

In presence of the emerging autonomous/driverless buses, transit operators will be able to consider using autonomous vehicles to offer bus transit services, in which case, the operation costs for hiring and training bus captains are not required. One more advantage of autonomous buses is the flexibility of being deployed across different bus service lines. Unlike the conventional buses that are usually fixed to certain service lines (mainly due to the reason that bus drivers are trained to operate with certain specific bus models on specific bus routes), autonomous buses can be assigned across the service lines in a more flexible manner. Using a centralized control, autonomous buses do not have to run on the fixed routes and can be dispatched to different bus routes according to the demand pattern. In reality, the transit demand is uncertain and may vary from day to day. When the buses are all central controlled autonomous buses, the idle buses can be used to run on other busy service lines. The flexible dispatching of an autonomous bus fleet has the potential to improve the efficiency of resource allocation and further decrease the required fleet size without compromising the level of service. In light of this, many countries, such as Singapore, China, UK, are seeking to implement trials on automated buses.

Resource allocation in traditional bus service has been well studied. However, similar research considering the inclusion of autonomous buses has not yet been done. This work aims to fill this gap by modeling and solving such a bus fleet optimization problem for a bus service system operating with a mixed fleet of conventional buses and autonomous buses, wherein the transit demand uncertainty is explicitly considered. Specifically, we aim to address the question that how to determine the optimal fleet size and fleet assignment for bus service operation with autonomous vehicles. This work contributes to the existing literature on the following aspects.

1. In this work, we optimize the required fleet size with mixed autonomous and conventional buses, as well as its allocation onto different service lines. Some existing studies have investigated the benefits of autonomous buses; however, they still considered the autonomous vehicles to be operated on fixed bus lines as human-driven buses. Indeed, the driverless attribute of the autonomous bus enables them to be assigned flexibly across the service lines so as to improve the bus operation efficiency and service quality. This is exactly the case when bus service demand uncertainty is significant and has to be considered in the operation strategy design. This study would develop a mathematical

modeling framework to explicitly capture the advantages of autonomous buses in bus service operation under demand uncertainty, which has not been taken into full account in the existing literature.

2. To solve the formulated stochastic programming problem, the sample average approximation (SAA) method is applied. The SAA problem is indeed a mixed integer nonlinear program (MINLP), which is hard to be solved due to its inherent nonlinear and nonconvex properties. To tackle this problem, we transform the MINLP into a mixed-integer conic quadratic program (MICQP), which admits the direct use of exact optimal solution approaches such as branch and bound method. For large-scale problems, we present a quadratic transform with linear alternating (QT-LA) algorithm to obtain a quality solution efficiently.

This paper is structured as follows: the literature review is presented in Sect. 2. Later, the model for the mixed autonomous and conventional buses is proposed in Sect. 3. The reformulation methods and solution approaches are given in Sect. 4, namely MICQP and QT-LA respectively. In Sect. 5, numerical examples are analyzed to test the model. At the end of this work, conclusions and future research are presented.

Literature review

This work studies the optimization of the bus fleet and the allocation in operation, which is related to vehicle assignment and service design. The vehicle assignment problem or vehicle fleet planning have been studied for decades (Salzborn 1972; Silman et al. 1974; Beaujon and Turnquist 1991). The required bus fleet is determined by the line frequency and cycling travel time, and thus, the bus fleet and the line frequency (headway) were usually optimized simultaneously in the relevant literature. The related works are summarized in the Table 1.

The aforementioned studies in Table 1 worked on traditional public transit services. Inspired by the emerging automation technology, many works studied the integration of the autonomous vehicle (AV) into public transit systems. Shen et al. (2018) simulated an integrated AV and public transportation (PT) system based on the Singapore environment. They proposed to preserve high demand bus routes and use AVs as an alternative in low-demand routes. Wen et al. (2018) considered the demand-supply interaction in transit-oriented autonomous vehicle operation. The agent-based simulation was used to design and evaluate the AV + PT system. Recently, Pinto et al. (2019) integrated shared autonomous vehicle service with conventional public transit. They applied a simulation approach to jointly determine autonomous vehicle fleet size and transit route frequencies. In these works, the AV is a promising service for the low-demand area or first/last-mile service. For a high demand area, the bus is still the more cost-effective mode. However, these works only focused on the AV, and the exact application and optimization of autonomous buses are understudied in existing research.

Recently, there is a proliferation of research on autonomous buses. Some research works focused on travelers' perception of autonomous buses. Dong et al. (2019) conducted a stated preference survey of users' willingness to ride and concerns about driverless buses. They found that the willingness to ride in driverless buses was influenced by gender, age, and knowledge of autonomous technology. Similar works have also been done by other researchers with survey-based methods in different areas (Salonen 2018; Wien 2019;

Table 1 Related studies

Literature	Modelling approach	Objective	Decision variable(s)	Solution method
Jansson (1980)	Analytical model	Min. operating, waiting and riding cost	Bus fleet, frequency and capacity	Analytical approach with square root formula
Beaujon and Turnquist (1991)	Continuum approximation	Max. profit	Vehicle fleet and allocation	Iterative procedure with Frank-Wolfe
Ker et al. (1995)	Multi-period model	Min. user and operator costs	Headway, capacity and fleet size	Quasi-Newton method with finite-difference gradient
Ceder (2005)	Deficit function model	Min. number of buses	Bus fleet size	Lower bound improvement
Li et al. (2008)	Analytical model	Max. total social welfare	Bus fleet size	Heuristic solution algorithm
Li et al. (2011, 2012a)	Binary IP	Authority: max social welfare Operator: max profit	Frequency, fleet size, fare	Heuristic solution algorithm
Kim and Schonfeld (2013)	Analytical model	Min. user and operator costs	Headway,capacity,bus fleet	Hybrid solution approach(analytic optimization with genetic algorithm)
Dandapat and Maitra (2015)	Simulation model	Max. benefit to users	Headway, fleet size, fare	Simulation
Jiménez and Román (2016)	MILP	Min. total pollutant emission	Bus fleet assignment	GAMS with Cplex
Nayan and Wang (2017)	MINLP	Min. operator cost, user waiting cost and package difference	Fleet assignment, frequency	Linearization into an MILP

IP interger programming, *MILP* mixed-integer linear programming, *MINLP* mixed-integer nonlinear programming;

Winter et al. 2019). Another direction focuses on the analysis of the benefits of autonomous buses. Bergqvist and Åstrand (2017) investigated how the introduction of automated minibuses into pre-existing public transport systems will affect their operation costs and the environment. A linear programming model is used to derive the best combination of traditional bus and autonomous minibus in terms of the cost. Zhang et al. (2019) evaluated the efficiency of autonomous bus services analytically in general trunk and branch networks. The results showed that fully automatic buses can significantly reduce user costs and operator costs.

While the study of the autonomous bus is in the rising stage, more research attention is required to the operation of autonomous buses. In this paper, we propose a mathematical modeling approach to investigate the optimal fleets and operation with mixed autonomous and conventional bus fleets. Being flexible and subject to central control, the autonomous bus could reduce the fleet size and operation cost by being assigned onto different bus lines. However, few studies discussed this point in detail. The integration of different operations in transit systems are discussed in many papers. For example, Aldaihani et al. (2004) developed an analytical model for designing a hybrid grid network that integrates a flexible demand-responsive service with fixed-route service. Li and Quadrifoglio (2009) studied the optimal zone design for feeder transit service. They derived the closed-form expressions for fixed-route and demand-responsive feeder services. Later, Li and Quadrifoglio (2010) developed analytical and simulation models to assist planners in determining how to choose between a demand responsive and a fixed-route operating policy. Kim and Schonfeld (2012, 2013, 2014) considered the integration of different types of services (e.g., conventional and flexible-route services) with mixed fleets can reduce total costs when demand varies over time and over regions. Chen and Nie (2018) proposed a new transit system that integrates the traditional fixed-route service with a demand-adaptive flexible-route service. These papers focused on the operation of conventional vehicles. Different from these studies, we incorporate the autonomous bus fleet in this work and consider the fact that the autonomous buses can be assigned to different fixed-route service lines in presence of demand uncertainty. This work fills this gap by comparing the optimal results (in terms of bus fleet and cost) of autonomous and conventional bus service in fixed-route transit service. In addition, transit demand fluctuates on a day-to-day basis. It is a challenge to decide the optimal fleet size with uncertain demand. Ker et al. (1995) considered the discrete distribution demand and they found that it is preferable to use mixed fleets on multiple routes when demand variation over time is significant. Yoo et al. (2010) proposed a methodology for modeling the transit frequency design problem with different demand patterns and fleet size constraints. Jara-Díaz et al. (2017) analyzed the problem of finding the optimal fleet for a single line to serve a pattern of transit users that changes during different periods. These works indicate the significance to consider the demand uncertainty, which coincides with the goal in this work.

Methodology

This work presents a mathematical modeling approach for solving the optimal design of fleet size and fleet assignment considering demand uncertainty. Both autonomous buses and conventional buses are considered. Consider a bus service network N with a set of bus lines L. The transit demand for each bus line is stochastic and subject to a specific distribution of U. Taking the operators' and passengers' costs into consideration, our objective is to find the optimal bus fleet with minimal total costs. Detailed assumptions and model formulations are made before the model formulation.

Assumptions

The proposed model is based on specific characteristics of the bus transit system. A few important assumptions are stated as follows.

A1: Bus service lines are pre-determined and regarded as fixed in this study. The demand for each service line is uncertain, subject to certain known distribution.

A2: All service lines are operated by one operator. Autonomous buses, due to the driverless nature, can be allocated to different bus lines.

A3: Conventional buses will stick to the assigned bus lines as bus captains are usually trained and scheduled to drive on specific lines.

A4: The in-vehicle travel time of both autonomous bus and conventional bus are assumed to be the same for the same bus line.

A5: Passengers are assumed to perceive the service quality for the autonomous buses and conventional buses equally. It means that the passengers have the same attitudes towards autonomous and conventional buses. If autonomous buses are scheduled to serve a bus line, the passengers on that line will use autonomous bus service as well.

Model formulation

Table 2 shows the notations that are used throughout this study.

Owning and operating autonomous buses and conventional buses involve different costs for operators. Without the need for hiring drivers, autonomous buses can reduce the operation cost significantly, because the driver cost accounts for a large proportion of the operation cost for bus service operation (Council 2006; Gentile and Noekel 2016). However, the autonomous driving technology constitutes the major cost components of the vehicle, including driving hardware, sensors, cameras, processing units, and V2X equipment et al.. Indeed, to own or hire an autonomous bus is more expensive compared to regular buses. According to the economic assessment of autonomous electric vehicles (AEV) for transit service done by Ongel et al. (2019) in Singapore, the acquisition costs of AEV are approximately 93% more than the Internal Combustion Engine Vehicles (ICEV). In terms of the operating cost, AEV is lower, mainly attributed to the savings from labour costs for hiring bus drivers. Depending on the bus type, the driver cost accounts for 40–70 % of total bus operation costs in Singapore (Ongel et al. 2019) and Australia (Council 2006). In Japan, drivers' salaries account for 53% of the total operating costs of buses (Abe 2019; Tirachini and Antoniou 2020). Based on the related works, we apply the operating cost reduction coefficient X and the ownership cost increase coefficient Y for autonomous buses to reflect the cost structure:

$$c_l^a = (1 - X)c_l^c \tag{1}$$

$$\alpha^a = (1 + Y)\alpha^c \tag{2}$$

where $X \in (0, 1)$ and $Y \geq 0$.

The total costs to be minimized in the model formulation consist of three main categories: operator's cost, passengers' waiting time cost and total penalty cost for unsatisfied

Table 2 Notation table

Sets	
L"	Set of transit routes $l \in L$
D:	Set of demand $d_l \in D$
R:	Set of demand samples $r \in R$
Parameters	
α^a:	Unit ownership cost of an autonomous bus
α^c:	Unit ownership cost of a conventional bus
c_l^c:	Unit conventional bus operation cost on line l
c_l^a:	Unit autonomous bus operation cost on line l
cap:	Bus capacity
k:	Parameter related to waiting time
c_p:	Penalty cost per unsatisfied demand
t_l:	In-vehicle transit time on line l
w_t:	Waiting time cost
p	The parameter to depict the uncertainty level
d_l:	Demand on line l
$d_{l,r}$:	Demand on line l under realization r
Dependent variables	
f_l:	Service frequency on line l, which is the summation of f_l^c and f_l^a
$f_{l,r}$:	Service frequency on line l under realization r
f_l^c:	Service frequency of conventional bus on line l, which depends on N_l
s_l:	Served demand of line l
$s_{l,r}$:	Served demand of line l under realization r
u_l:	Unsatisfied demand of line l
$u_{l,r}$:	Unsatisfied demand of line l under realization r
Decision variables	
M	Optimal number of autonomous buses, which is an integer variable
N_l:	Optimal number of conventional buses on line l, which is an integer variable
f_l^a:	Service frequency of autonomous bus on line l
$f_{l,r}^a$:	Service frequency of autonomous bus on line l under realization r

demand. Detailed cost functions are given below. Operator's cost includes the fixed bus ownership cost and the variable operating cost.

Bus ownership cost:

$$C_1 = \begin{cases} \alpha^a \cdot M & \text{for autonomous bus} \\ \sum_l \alpha^c \cdot N_l & \text{for conventional bus} \end{cases} \tag{3}$$

For the autonomous buses, the total fleet size is M, and the total ownership cost is the unit cost times the fleet size. While for the conventional buses, the vehicles are assigned to different lines. Therefore, the total ownership cost is the sum of the costs for buses deployed on different lines.

The bus operation cost is related to its operation frequency. The operation costs for the conventional service and autonomous service are:

$$C_2 = \begin{cases} \sum_l c_l^a f_l^a & \text{for autonomous bus} \\ \sum_l c_l f_l^c & \text{for conventional bus} \end{cases} \tag{4}$$

For passengers, their waiting time cost is to be minimized. The passenger waiting time is dependent on the arriving transit vehicles. The expected waiting cost can be expressed as:

$$W = w_t \cdot \sum_l s_l \cdot k/f_l \tag{5}$$

where k is determined by the distribution of headway. Here, k is set to be 0.5, as was done in many previous studies (Yu et al. 2012; An and Lo 2016).

Due to the demand fluctuation, there may be cases where the supply cannot meet the demand. We assume the unsatisfied demand will switch to other alternative transportation modes that could be more costly. To consider this, we introduce a high penalty cost in the cost function to represent the additional cost. The unsatisfied demand penalty is given as:

$$P = \sum_l c_p \cdot u_l \tag{6}$$

With the above cost components, we propose the following stochastic model P1 for the mixed bus fleet optimization problem. Let $d = (d_1, d_2, \ldots, d_l)$ represents the uncertain demand set. The problem can be formulated into a two-stage model, where the first stage is to make the decision on optimal fleet size; while at the second stage, after a realization of demand scenario, the fleet assignment onto different service lines is to be optimized.

[**P1**: *Two-stage stochastic programming model*]

$$\min_{M,N_l} \alpha^a M + \sum_l \alpha^c N_l + \sum_l c_l^c f_l^c + \mathbb{E}\left[Q(M, N_l, f_l^a, d_l)\right] \tag{7}$$

$$\text{st.} \quad M, N_l \geq 0, \in \mathbf{Z}, \forall l \in L \tag{8}$$

$$f_l^c = \frac{N_l}{2t_l} \tag{9}$$

where $Q(M, N_l, f_l^a, d_l)$ is the optimal objective value of the second-stage problem, upon the realization of the stochastic demand variable d_l:

$$Q = \min_{f_l^a} \sum_l c_l^a f_l^a + w_t \cdot \sum_l s_l \cdot k/(f_l^c + f_l^a) + \sum_l c_p u_l \tag{10}$$

$$\text{st.} \quad \sum_l (2t_l) f_l^a \leq M \tag{11}$$

$$s_l = \min\left(d_l, \text{cap} \cdot (f_l^c + f_l^a)\right), \forall l \in L \tag{12}$$

$$u_l = d_l - s_l, \forall l \in L \tag{13}$$

The objective function in (7) is to minimize the total cost. Here, the symbol $\mathbb{E}$ denotes the expectation. The first and second term are the bus ownership costs for the autonomous buses and conventional buses, respectively; the third term is the operation cost of the conventional buses; the final term is the expectation of the second stage problem. In (10), the objective function of the second stage problem consists of the total operation cost, waiting cost and unsatisfied demand penalty cost.

Constraint (8) defines the feasible region. Constraint (9) reflects the relationship between fleet size and service frequency for conventional bus fleets (Huang et al. 2013). Constraint (11) is the fleet size constraint of autonomous buses. It imposes that the used autonomous buses cannot exceed the total available buses. We use "≤" instead of "=" to capture the fact that not the entire fleet will be required to satisfy the passenger demand if the demand is not large. Constraints (12) determines the satisfied demand, which is the minimal value between demand and total bus occupancy for each line. Constraint (13) is the difference between demand and satisfied demand, which calculates the unsatisfied demand. If all demand can be satisfied, the value will become 0; otherwise, it is a positive value.

Remark 1 In this work, we use the $2t_l$ to stand for the cycling time on service line l, which can be modified as $2(t_l + e_l)$ to include the recovery time e_l at the terminal station (Huang et al. 2013). The parameter e_l is the constant for each line and will not influence the property of the model itself. Hence, we simplify $e_l = 0$ in this work.

Remark 2 We focus on the fixed-route bus transit service and did not consider the correlation between passenger demands on different bus lines. In reality, multiple bus lines may overlap with each other and passengers will transfer between different lines. Therefore, the passenger demands among different lines may be correlated with each other. In this work, we focus on the framework of the autonomous and conventional bus fleet optimization under demand uncertainty and simply assume that the correlation among different service lines are negligible to simplify the problem. This assumption is expected to be relaxed in the future studies.

Remark 3 It is assumed that the demand realization on each day can be accurately predicted by using the historical data and existing statistical approach such as time series analysis, so that the operator can determine the optimal operation of the autonomous buses in priori.

In the model formulation, the demand d_l is a stochastic variable. The model is indeed a stochastic programming problem. In most settings, the closed-form solutions to stochastic programming problems are unavailable (Shapiro and Philpott 2007). Here, we transform the stochastic program into a deterministic optimization problem, by estimating the cost using the sample average approach, i.e., the sample average approximation (SAA).

Sample average approximation (SAA) method

The basic idea of SAA method is to use random samples to approximate the expected objective function of a stochastic programming problem. In general, not all distributions have a closed-form solution. We consider the SAA procedure which can be used to deal with general distributions. With random samples, our original formulation is reduced to a deterministic model. Assume that a random sample set of $|R|$ demand realizations (scenarios) are selected and each scenario r has equal probability. Therefore, the model formulation with scenarios can be expressed as below.

[**P2**: *Sample Average Approximation Problem*]

$$
\begin{aligned}
\min \quad & \alpha^a M + \sum_l \alpha^c N_l + \sum_l c_l^c f_l^c \\
& + \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + w_t \cdot \sum_r \sum_l s_{l,r} \cdot k/(f_l^c + f_{l,r}^a) + \sum_r \sum_l c_p \cdot u_{l,r}\right]
\end{aligned} \tag{14}
$$

$$
\text{st.} \quad f_l^c = N_l/(2t_l), \quad \forall l \in L \tag{15}
$$

$$
\sum_l 2t_l f_{l,r}^a \leq M, \quad \forall r \in R \tag{16}
$$

$$
s_{l,r} = \min\left(d_{l,r}, \text{cap} \cdot (f_l^c + f_{l,r}^a)\right), \quad \forall l \in L, r \in R \tag{17}
$$

$$
u_{l,r} = d_{l,r} - s_{l,r}, \quad \forall l \in L, r \in R \tag{18}
$$

$$
N_l, M \geq 0, \in \mathbf{Z}, \quad \forall l \in L, r \in R \tag{19}
$$

The original stochastic problem P1 is transformed into a deterministic problem P2 by SAA. In P2, the fleet size of autonomous bus M and the number of conventional fleets for each line N_l must keep unchanged across all scenarios. The optimal frequencies of autonomous buses operated for each line f_l^a are optimized under different demand realizations. The objective function (14) is to minimize the expected total cost, which has the same cost components as P1, but represented by the sample average. Most of the constraints resembles those in P1. The difference is that some variables in P2 have one more dimension r. Autonomous buses are flexible in that the number of buses serving line l under each scenario r can be different. The operation can be controlled by a central manager with respect to the available fleets.

The detailed SAA procedures are described in the following procedure of Algorithm 1 (Kleywegt et al. 2002; Patil and Ukkusuri 2011; Long et al. 2012). According to Patil and Ukkusuri (2011), the sample size used in SAA algorithm could be much smaller than the possible number of scenarios in real life. The realizations of a stochastic variable in practice can be very large. Through SAA approach, a proper sample size less than the practical realizations can be used to obtain a reasonable result with a small optimal gap. The following procedure shows how to determine the sample size and derive the expectation result.

Algorithm 1: SAA algorithm

0. Initialize sample sizes $|R|$ and n, replication size m.
1. For k = 1, 2 , ... , m
Sample iid observations $r_1, r_2, ..., |R|$, and solve proposed model P_2 to obtain solution Z_k and its corresponding objective $f(Z_k)$
2. For k = 1, 2 , ... , m
   2.1. Sample iid observations $n_1, n_2, ..., n$ from the demand distribution
   2.2. Evaluate P_2 using Z_k, and get the value $f'(Z_k)$
3. Take $\hat{Z}$ as the optimal solution of these SAA problems which provides the smallest estimated objective value $\hat{Z} \in \arg\min f'(Z_k) : Z_k \in Z_1, ..., Z_m$ and calculate the optimal gap: $G_k = f'(\hat{Z}) - \bar{f}$, where $\bar{f} = \frac{1}{m}\sum_{k=1}^{m} f(Z_k)$
4. Check stopping criteria. If satisfied, stop; else increase $|R|$ and go to step 1.

Step 0 is used to initialize the parameters used in the SAA procedure. $|R|$ is the sample size used to derive the optimal solution Z of a sample and m is the replication times. Another sample size n (which is usually larger than $|R|$) is used to evaluate the derived solutions $(Z_1, Z_2, \ldots, Z_m)$. In step 1, the model P_2 is solved m times. The candidate solutions Z_k and objective functions $f(Z_k)$ can be obtained each time. Step 2 is to evaluate the solutions derived in Step 1. A larger sample size n is used to evaluate m candidate solutions and get the value $f'(Z_k)$. Step 3 shows that the candidate solution with the smallest objective function value is taken as the optimal solution. Then the optimal gap can be computed. If the optimal gap can meet the stopping criteria, then the sample size $|R|$ is acceptable. Otherwise, increase the sample size and repeat the process.

Solution methods for solving SAA problem

The proposed model formulation in P2 is a mixed-integer nonlinear problem (MINLP), which is difficult to solve due to its intrinsic nature of non-convexity. Heuristic methods, such as genetic algorithm (GA) and artificial bee colony (ABC), have been widely applied to solve the MINLP problem in transportation research (Szeto and Wu 2011; Chen et al. 2011; Szeto and Jiang 2012; Liu et al. 2013; Szeto and Jiang 2014; Chen et al. 2015). These heuristic methods typically perform stochastic searching, and the drawbacks are obvious, i.e., they generally requires a large number of function evaluations; it could be trapped in local optimal solution, and the global optimal solution cannot be guaranteed; and the solutions could be different across different runs. To deal with these issues, we adopt reformulation techniques to ensure the global optimal solution can be obtained. Linearization and convexification techniques (Wang and Lo 2008, 2010; Luathep et al. 2011; Liu and Wang 2015; Wang et al. 2015, 2018) are applied in this section to reformulate the problem into an equivalent mixed-integer conic quadratic program (MICQP), which can be solved to its global optimal solution by using many existing solution methods. While this method can guarantee the global optimal solution, its computational efficiency may not be sufficiently high to solve large-size problems. To further improve the efficiency of the solving the proposed model, we also propose a quadratic transform with linear alternating (QT-LA) approach, which can be applied to solve a problem with larger size.

Conic quadratic program reformulation

In this section, we would introduce how to apply conic reformulation techniques (CRT), to transform P2 into a problem that can be directly solved by using the existing solution methods for MICQP. An MICQP takes the form:

$$\min_{x} \quad c^T x \tag{20}$$

$$st. \quad ||A_i x + b_i||_2 \leq g_i^T x + j_i, \forall i \tag{21}$$

where $||\cdot||_2$ is the Euclidean norm and all parameters are rational. When $A_i = 0$, it reduces to a linear programming. When $g_i = 0$, it reduces to a convex quadratic programming. For more detailed introductions, see Alizadeh and Goldfarb (2003) and Ben-Tal and Nemirovski (2001). The conic quadratic inequality (21) is often used to represent a rotated cone/hyperbolic inequality, i.e., for $x, y, z \geq 0$,

$$x^2 \leq yz \iff ||(2x, y-z)||_2 \leq y+z \tag{22}$$

We will use the rotated cone inequity in our problem. To ensure that P2 can be transformed into an MICQP, we need to fulfill the following two steps of reformulation first.

1. Reformulation of minimization operator in the constraints

We first deal with the minimization operator in (12). As we assume a large unit penalty cost, the model will tend to serve as many as passengers as possible in the optimal solution. Therefore, constraint (11) can be replaced by the following equations:

$$\begin{cases} s_l \leq d_l \\ s_l \leq f_l \cdot \mathrm{cap}\ \forall l \in L \\ s_l \geq 0 \end{cases} \tag{23}$$

2. Reformulation of non-linear term in the objective function

For illustration purposes, we introduce $f_{l,r} = f_l^c + f_{l,r}^a$. Then we have the term $s_{l,r} \cdot k/f_{l,r}$ in the objective function, which involves the division of two decision variables. The objective is a minimization problem. Hence, we rewrite the objective as

$$\min \quad \alpha^a M + \sum_l \alpha^c N_l + \sum_l c_l^c f_l^c + \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + w_t \cdot \sum_r \sum_l \varphi_{l,r} + \sum_r \sum_l c_p \cdot u_{l,r}\right] \tag{24}$$

where $\varphi_{l,r} \geq s_{l,r} \cdot k/f_{l,r}$. The linear objective (24) is equivalent to the original objective function (14).

Now, by introducing a quadratic term $S_{l,r}^2$ such that $S_{l,r}^2 = s_{l,r}$, we can transform P2 into P3:

$$[\mathbf{P3}]\ \min\ \alpha^a M + \sum_l \alpha^c N_l + \sum_l c_l^c f_l^c + \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + w_t \cdot \sum_r \sum_l \varphi_{l,r} + \sum_r \sum_l c_p \cdot u_{l,r}\right] \tag{25}$$

$$\text{st.}\quad kS_{l,r}^2 \le \varphi_{l,r} \cdot f_{l,r} \tag{26}$$

$$S_{l,r}^2 \le d_{l,r} \tag{27}$$

$$S_{l,r}^2 \le cap \cdot f_{l,r} \tag{28}$$

$$S_{l,r}^2 = d_{l,r} - u_{l,r} \tag{29}$$

$$(15) - (16), (19) \tag{30}$$

where constraints (26)–(28) are all conic quadratic inequalities. The equality constraint (29) is, in effect, a non-convex function that does not satisfy the conic structure in (21). However, the left-hand-side is a quadratic function of $S_{l,r}$, which belongs to the class of general constraints that can be handled by using piecewise linear approximation to any desired accuracy[1] as in many commercial softwares such as Gurobi. Overall, the problem has been transformed into a standard MICQP, which can be solved by using many existing solution methods embedded in commercial softwares.

Quadratic transform with linear alternating (QT-LA) algorithm

In Sect. 4.1, we present a solution method wherein P2 is transformed into an MICQP. However, when applying this solution approach, the piecewise linear approximation of the quadratic term requires additional binary variables and constraints for each bus line and demand scenario. The solution time for large-scale problems could be prohibitive. To solve large-scale problems, we propose another solution method based on the techniques of Quadratic Transform with Linear Alternating (QT-LA), which can produce high quality solutions efficiently.

To introduce this solution method, we start with the following proposition.

Proposition 1 *P2 is equivalent to the following maximization problem.*
[***P4***: *Equivalent maximization problem*]

$$\begin{aligned} \max\quad \Phi(M, N, f^c, f^a, s, u, \beta, y) = {} & \frac{w_t}{|R|}\left(\sum_r \sum_l 2y_{l,r}\beta_{l,r} - \sum_r \sum_l y_{l,r}^2 (f_l^c + f_{l,r}^a)\right) \\ & - \alpha^a M - \sum_l \alpha^c N_l - \sum_l c_l^c f_l^c - \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + \sum_r \sum_l c_p \cdot u_{l,r}\right] \end{aligned} \tag{31}$$

$$\text{st.}\quad \beta_{l,r} \le h(f_l^c, f_{l,r}^a, s_{l,r}) \tag{32}$$

[1] https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf.

$$y_{l,r} \geq 0, \quad \forall l \in L, r \in R \\ (15)–(19) \tag{33}$$

where $h(f_l^c, f_{l,r}^a, s_{l,r}) = \sqrt{\delta(f_l^c + f_{l,r}^a) - s_{l,r}k}$ *and* δ *is a parameter which is chosen so that* $\delta(f_l^c + f_{l,r}^a) - s_{l,r}k > 0, \forall l \in L, r \in R$.

Proof See Appendix 1. □

P4 is called the Quadratic Transform of P2 (Benson 2004; Shen and Yu 2018; Lin et al. 2020). Obviously, P4 is a mixed-integer bi-concave maximization problem: On one hand, when y is fixed, P4 reduces to a mixed-integer concave maximization program with variables $(M, N, f^c, f^a, s, u, \beta)$; on the other hand, when $(M, N, f^c, f^a, s, u, \beta)$ are fixed, P4 becomes a quadratic program with variable y where the optimal y can be found in closed form as

$$y_{l,r}^* = \frac{\sqrt{\delta(f_l^c + f_{l,r}^a) - s_{l,r}k}}{f_l^c + f_{l,r}^a} \tag{34}$$

To solve P4, one can alternate between updating y and solving the mixed-integer concave program until some predetermined stopping condition is met. However, this requires solving the mixed-integer concave program for multiple iterations, which can be computationally prohibitive when the problem size is large. To speed up the solution procedure, we apply the sequential outer linearization approach to linearize the nonlinear function h in P4 (Lin et al. 2020).

One can observe that, given any point $(\bar{f}_{l,r}^c, \bar{f}_{l,r}^a, \bar{s}_{l,r})$, since h is a concave function, its upper bound can be constructed by the first-order linear approximation on $(\bar{f}_{l,r}^c, \bar{f}_{l,r}^a, \bar{s}_{l,r})$, i.e, the linear function v

$$v(\bar{f}_{l,r}^c, \bar{f}_{l,r}^a, \bar{s}_{l,r}) = \frac{\partial h_{l,r}}{\partial f_l^c}(f_l^c - \bar{f}_l^c) + \frac{\partial h_{l,r}}{\partial f_{l,r}^a}(f_{l,r}^a - \bar{f}_{l,r}^a) + \frac{\partial h_{l,r}}{\partial s_{l,r}}(s_{l,r} - \bar{s}_{l,r}) + h(\bar{f}_{l,r}^c, \bar{f}_{l,r}^a, \bar{s}_{l,r}), \quad \forall l \in L, r \in R \tag{35}$$

where $\frac{\partial h_{l,r}}{\partial f_l^c}$, $\frac{\partial h_{l,r}}{\partial f_{l,r}^a}$ and $\frac{\partial h_{l,r}}{\partial s_{l,r}}$ are the partial derivatives of h with respect to f_l^c, $f_{l,r}^a$ and $s_{l,r}$, respectively. Then, the following linear inequality is valid for P4:

$$\beta_{l,r} \leq v(\bar{f}_{l,r}^c, \bar{f}_{l,r}^a, \bar{s}_{l,r}) \tag{36}$$

This is because any feasible point satisfying (32) is also feasible in the region defined by (36), meaning that (36) does not eliminate any feasible region of P2. We can therefore model P4 by using the following MILP formulation:

$$\max \quad \Phi(M, N, f^c, f^a, s, u, \beta, y) \tag{37}$$

$$\text{st.} \quad \beta_{l,r} \leq v(f_l^{c,t}, f_{l,r}^{a,t}, s_{l,r}^t), \quad \forall l \in L, r \in R, (f_l^{c,t}, f_{l,r}^{a,t}, s_{l,r}^t) \in T \tag{38}$$

$$y_{l,r} \geq 0, \quad \forall l \in L, r \in R \\ (15)–(19) \tag{39}$$

where T is the set of recorded points.

With the above program, we present the Quadratic Transform with Linear Alternating (QT-LA) in Algorithm 2. In Step 0, we initialize the parameters and y. Throughout this paper, we initialize $y_{l,r} = 0$. i is the iteration count and I_{max} is the maximum number of iterations.

In Step 1, we solve a spare MILP. This can be done by using advanced MILP solvers. As the algorithm proceeds, new points (tentative solutions to MILP in previous iterations) are added to set T and thus, the MILP contains more constraints and yields better approximation for the original formulation.

Step 2 is to check whether the stopping condition is satisfied. The algorithm terminates when the difference between two consecutive objective cost is smaller than tolerance or the number of iterations exceeds the given maximal iteration times.

In Step 3, we update set T and y and compute the partial derivative for defining new valid inequalities that will add to the MILP in the next iteration. Here, η serves as the step-size. It is a value between 0 and 1, which controls the intensity of moving y towards the value defined by (34) at current solution $(f_l^{c,i}, f_{l,r}^{a,i}, s_{l,r}^i)$. When computing the partial derivative, we add a small number to the denominator to enhance the numerical stability.

Algorithm 2: QT-LA Algorithm

Step 0: Initialize $y^0 = 0$; Iteration number $i = 0$; $\epsilon = 0.001$; $T = \emptyset$; choose $\eta \in (0, 1]$; maximal number of iteration $I_{max} \in Z_+$.

Step 1: Solve the following MILP:

$$\begin{aligned} \max \quad & \Phi(M, N^c, f^c, f^a, s, u, \beta, y) \\ \text{st.} \quad & \beta_{l,r} \le v(f_l^{c,t}, f_{l,r}^{a,t}, s_{l,r}^t), \ \forall l \in L, r \in R, (f_l^{c,t}, f_{l,r}^{a,t}, s_{l,r}^t) \in T \\ & y_{l,r} \ge 0, \forall l \in L, r \in R \\ & (15) - (19) \end{aligned}$$

to obtain the solution $(f_l^{c,i}, f_{l,r}^{a,i}, s_{l,r}^i)$.

Step 2: If $i = I_{max}$ or the difference between two consecutive objective cost is smaller than tolerance, stop. Else, go to Step 3.

Step 3: $T := T \cup \{(f_l^{c,i}, f_{l,r}^{a,i}, s_{l,r}^i)\}$. $i := i + 1$ Update parameters:

$$\begin{aligned} y_{l,r}^i &:= (1 - lr) y_{l,r}^{i-1} + \eta \cdot \frac{\sqrt{\delta(f_l^{c,i} + f_{l,r}^{a,i}) - s_{l,r}^i k}}{f_l^{c,i} + f_{l,r}^{a,i}} \\ \frac{\partial h_{l,r}^i}{\partial f_l^c} &:= \frac{\delta}{2\sqrt{\delta(f_l^{c,i} + f_{l,r}^{a,i}) - s_{l,r}^i k} + \epsilon}, \ \forall l \in L, r \in R \\ \frac{\partial h_{l,r}^i}{\partial f_l^a} &:= \frac{\delta}{2\sqrt{\delta(f_l^{c,i} + f_{l,r}^{a,i}) - s_{l,r}^i k} + \epsilon}, \ \forall l \in L, r \in R \\ \frac{\partial h_{l,r}^i}{\partial s_{l,r}} &:= \frac{-k}{2\sqrt{\delta(f_l^{c,i} + f_{l,r}^{a,i}) - s_{l,r}^i k} + \epsilon}, \ \forall l \in L, r \in R \end{aligned}$$

and return to Step 1

Numerical studies

Comparison of the two solution methods

A toy example

We first conduct numerical experiments on a small toy network. Three transit lines are considered, and the related parameters are given in Table 3. We assume bus capacity is 30 pax/bus; waiting time cost is $w_c = \$15/h$; penalty is set as \$20/h per unmet demand. The demands for each route are assumed to follow uniform distribution between the given lower bound and upper bound. The lower bound and upper bound of demand on each line is set as $[(1-p)\bar{d}_l, (1+p)\bar{d}_l]$ with mean value $\bar{d}_l$. The demand uncertainty level is set as $p = 0.4$. The parameters $X = 0.5$ and $Y = 1$ are used. The computational experiments are done using Python on a 16 GB memory Windows computer with a 1.90 GHz Intel Core i7 processor. We use Gurobi 9 under default settings as the MIP solver.

This example is used to demonstrate the validity of the proposed MICQP and QT-LA methods and compare the efficiency and accuracy of them. For demonstration purpose, we only perform a single calculation of a random sample set of R demand realizations. The maximal running time is set as 3600s. The results are reported in Table 4.

On one hand, the cost derived by these two methods are almost the same. This implies the proposed QT-LA algorithm generates the optimal (at least high-quality) solution. On the other hand, when sample size is 10 (small), both MICQP and QT-LA can solve the instance efficiently. However, when the sample size grows, the computation time of MICQP increases. In particular, when the sample size is 50, MICQP fail to solve the instance and terminate with a MIPGap 0.21%. In contrast, the QT-LA approach remains efficient. In conclusion, MICQP method can only work well for small-scale problems with a small number of scenarios. QT-LA outperforms MICQP in terms of computational time.

More experiments

To further test the application of the proposed two solution methods in the large transit network, we select an idealized network with more bus lines to verify the algorithm efficiency. The input data and parameters for this example are described as follows:

1. The in-vehicle transit time for each line is randomly generated within [30,60];
2. The mean passenger demand for each bus line is randomly generated within [100,400];
3. The operating cost for each bus line is randomly generated within [15,30];
4. The other related parameters are set as in Sect. 5.1.

Table 3 Network characteristics

Parameter	Line1	Line2	Line3
Mean demand (pax/h)	225	100	150
t_l	35	45	50
c_l^c	30	36	40

Table 4 Solving method results

Method	MICQP	QT-LA	MICQP	QT-LA	MICQP	QT-LA
Sample size	10	10	30	30	50	50
Cost	1678.0	1678.1	1612.6	1612.7	1643.1	1642.7
AB	1	1	5	5	3	4
CB	[10 6 11]	[10 6 11]	[9 4 8]	[9 4 8]	[10 5 10]	[9 5 9]
Operator cost	1112.4	1112.5	1045.3	1045.3	1118.4	1081.3
Waiting cost	564.4	564.4	555	555.1	520.2	536.4
Penalty cost	1.2	1.2	12.3	12.3	4.5	25
CPU (s)	0.89	0.10	541.4	0.27	3600 (gap = 0.21%)	0.39

AB autonomous bus, *CB* conventional bus

We consider different number of bus lines and scenarios in this numerical experiment. When the number of bus lines increases, larger sample size is required to obtain a reasonable solution. Here, we run a single calculation of one random sample with different number of scenarios. The computational time (within 3600s) and cost for all the cases using MICQP and QT-LA are concluded in Tables 5 and 6 respectively. The MIPgap in MICQP and the stopping tolerance of the QT-LA are set as 10^{-4}.

We can see that the proposed solution method MICQP can solve to the global optimal solution, but only when both the number of bus service lines and sample sizes are relatively small. The computational time increases dramatically with the dimension of variables. When the number of bus lines is 9, all the cases will terminate at 3600s with a small MIPGap. In terms of the QT-LA algorithm, it can solve the mixed bus fleet problem in a reasonable time with a large network and sample size. We can observe that the numerical cases (30 bus lines with 1000 sample size and 40 bus lines with 500 sample size) can be solved within 3600s. Comparing the same cases in Tables 5 and 6, we find that the obtained solutions from these two algorithms are almost the same. This finding indicates that both methods can be used to solve the problem. However, in terms of the computational performance, MICQP can only work on small problem size, while QT-LA is able to solve problems in larger size. Therefore, we will apply QT-LA solution method to solve the numerical example in this study hereafter.

Numerical examples with practical transit service network

In this section, we apply the proposed model formulation and solution algorithms into a numerical example, in which a Singapore bus transit network with 6 transit routes is considered (Nayan and Wang 2017). The example network is shown in Fig. 1. Detailed bus route information and demand data are given in Tables 7 and 8. The uniform distribution is also adopted for the passenger demand. The operation cost is given for conventional bus service. We assume bus capacity is 40 pax/bus and other parameters are set as given in Sect. 5.1.

Table 5 Computational results of MICQP

Lines	20		40		60		80	
	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost($)	CPU (s)	Cost ($)
Sample size								
3	2.2	1785.1	189.7	1786.2	1408.9	1792.9	3600 (gap 0.56%)	1800.6
6	20.0	3743.5	464.9	3729.2	3600 (gap 0.12%)	3766.2	3600 (gap 0.11%)	3780.7
9	3600 (gap 0.16%)	6065.3	3600 (gap 0.20%)	6062.2	3600 (gap 0.27%)	6123.2	3600 (gap 0.24%)	6143.1

Table 6 Computational results of QT-LA

Lines	20		40		60		80		100		200		500		1000	
	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)	CPU (s)	Cost ($)
Sample size																
3	0.2	1785.1	0.9	1786.3	0.9	1793.1	1.1	1799.1	1.5	1812.7	2.7	1802.8	13.9	1797.7	17.8	1802.3
6	0.7	3743.7	1.3	3729.7	2.3	3766.3	2.8	3780.4	3.8	3831.2	9.5	3799.5	31.9	3769.6	52.8	3825.8
9	1.6	6065.0	3.2	6062.2	5.7	6122.7	6.0	6140.7	7.3	6199.8	27.7	6139.1	80.1	6082.5	108.2	6136.9
15	4.3	9784.2	6.1	9774.9	7.4	9870.1	16.5	9910.1	20.9	10007.4	36.2	9922.4	267.6	9787.9	708.2	9888.7
20	5.8	12422.4	10.3	12424.5	11.3	12528.3	23.0	12583.8	36.1	12718.4	80.7	12614.2	478.4	12447.1	1626.3	12580.6
30	8.8	17684.8	18.4	17730.2	41.9	17832.7	51.6	17984.6	69.7	18154.4	216.7	18019.2	884.2	17807.6	2900.0	17836.5
40	14.1	24937.1	34.2	24906.4	64.2	25084.1	83.1	25198.9	171.7	25389.3	479.5	25227.5	1549.7	24954.2	>3600	24982.2

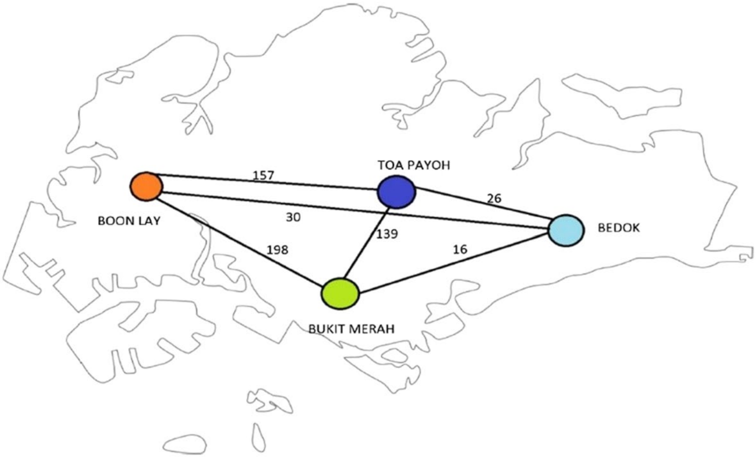


Fig. 1 Singapore bus transit network

Table 7 Network characteristics

Origin	Destination	Bus line	Transit time (min)	Operation cost ($/h/veh)
Boon Lay	Toa Payoh	157	35	21
Boon Lay	Bedok	30	45	24
Boon Lay	Bukit Merah	198	50	27
Toa Payoh	Bukit Merah	139	40	21
Toa Payoh	Bedok	26	30	21
Bedok	Bukit Merah	16	50	30

Application of SAA method

When solving the stochastic programming problem, SAA method is applied. The numerical experiment is conducted by Gurobi Optimizer interfaced with Python. The replication number m is set to be 20. The number of scenarios n to evaluate the solution is set to be 1000. Note that the accuracy of the approximation of the SAA method depends on the sample size R. Therefore, we analyze the performance of SAA method by changing the sample size R. The examples with sample size 20, 50 and 100 are evaluated respectively.

The results of the SAA method are examined with the key results shown in Table 9. One can observe that the estimated objective value of the three cases are close to each other and the maximal estimated optimality gaps is 0.59%, which implies that even small sample size can derive a quality solution. Detailed results are plotted in Fig. 2. When the sample size is 20, a good solution can be obtained. However, when the sample size is small, the estimated variance of the objective value is large. To derive a more reliable solution, it is desirable to increase the sample size so that the variance of the estimated objective value can be reduced. For this consideration, we choose a larger sample size $R = 100$ when applying the SAA method.

Table 8 Route demand (pax/h)

Bus line	157	30	198	139	26	16
Mean demand	450	200	300	400	320	250

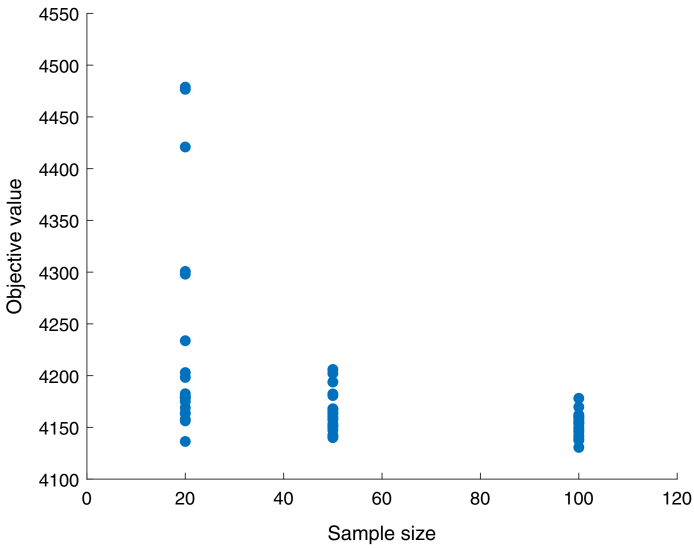


Fig. 2 Improvement with the increase of sample size

Table 9 Results of the SAA method

Sample size (R)	$\bar{f}$	$f'(\hat{S}^*)$	Optimal gap	% gap
20	4111.94	4136.38	– 24.44	0.59
50	4137.89	4140.1	– 2.21	0.05
100	4131.53	4130.65	0.88	0.02

The impact of demand uncertainty

The optimal solutions under different uncertainty levels are investigated. In the parameter settings, as p grows, the demand variance becomes larger, and the mean value remains unchanged. The solution results are demonstrated in Table 10. Detailed cost components are depicted in Fig. 3 for comparison purpose.

The total cost increases as uncertainty level rises. For the mixed fleets operation, the operation costs increase as p grows (see blue bar in Fig. 3). This is because more services are required to tackle with the demand uncertainty and to reduce unsatisfied demand. This can explain why the waiting costs decrease slightly. Besides, the possibility of fail-to-board for passengers will increase when the demand uncertainty is larger, which could lead to a higher penalty cost (see yellow bar in Fig. 3). The additions of operator's cost and penalty cost overweight the reduction of waiting cost. Hence, the total cost will increase with the increase of p.

Table 10 Optimal mixed bus fleet

Demand uncertainty level (p)	AB fleet	CB fleet	Total cost
0.2	4	$[14\ 8\ 13\ 14\ 12\ 11]_{72*}$	4022.26
0.4	9	$[15\ 8\ 13\ 15\ 12\ 11]_{74}$	4130.65
0.6	17	$[14\ 8\ 13\ 14\ 12\ 11]_{72}$	4293.74
0.8	23	$[15\ 8\ 14\ 14\ 12\ 9]_{72}$	4444.71

**AB* autonomous bus, *CB* conventional bus, The number at the foot indicates the total CB fleets

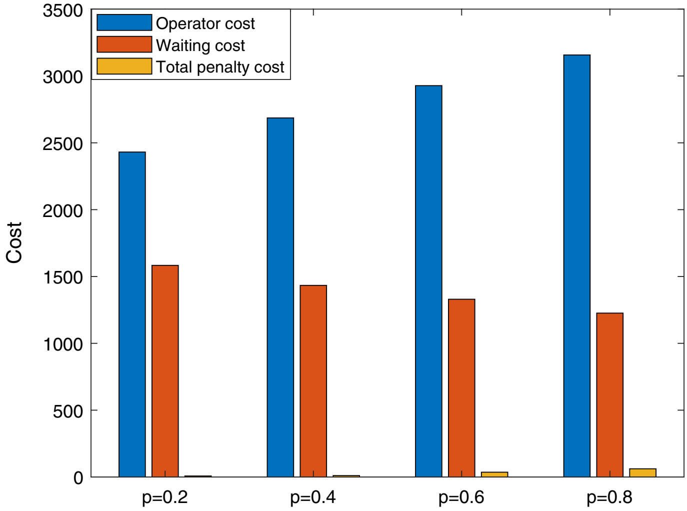


Fig. 3 Cost components under different p

When the variance of demand is large, more services should be provided to satisfy the uncertain travel demand. For conventional buses, the total required fleet size in the optimal solution is relatively stable when demand variation becomes larger. The explanation could be that the conventional buses can only operate on fixed routes due to the constraints of drivers training and scheduling. Therefore, the conventional vehicle fleet on one bus line is more determined by the demand density of that line. As for the demand uncertainty, it can be served by the complementary autonomous buses. We can see that the optimal fleet size of the autonomous bus is increasing to provide more service when the demand uncertainty is more significant. The results clearly demonstrate the advantage of deploying autonomous buses, that is the allocation flexibility of autonomous buses can offset the negative effects of demand uncertainty.

To further investigate to what extent the benefits of autonomous buses could offer, we compare the optimal operation of mixed fleets with conventional-bus-only operation. The reductions of fleet size and total cost under different uncertainty level are summarized in Table 11. One can notice that when values of p are 0.2, 0.4, 0.6 and 0.8, the cost reductions are 1.38%, 3.61% , 5.02% and 7.45% respectively. Similarly,the required total mixed fleets in each case are lower than these of conventional fleets, and it can reduce up to 19.49% of the conventional bus fleets when $p = 0.8$. In conclusion, using autonomous buses can

Table 11 Optimal bus fleet comparison

p	CB only		AB + CB		Reduction (%)	
	Fleets	Cost	Fleets	Cost	Fleets	Cost
0.2	82	4078.62	76	4022.26	7.32	1.38
0.4	93	4285.2	83	4130.65	10.75	3.61
0.6	106	4520.88	89	4293.74	16.04	5.02
0.8	118	4802.57	95	4444.71	19.49	7.45

**AB* autonomous bus, *CB* conventional bus

substantially reduce the required fleet size and the total costs. The benefits of using autonomous buses are more significant when the level of demand uncertainty is higher.

The impact of penalty cost

In the model formulation, penalty costs are imposed when travel demands cannot be fully satisfied. If the penalty costs are set to be high, the system will be optimized to avoid unsatisfied demand. Table 12 indicates how the required fleet size and total system costs change with the settings of the unit penalty cost. The values of unit penalty cost are set as 2, 5, 10, 15 and 20 respectively.

One can notice that, as penalty cost per unit demand loss increases, the total required bus fleet size tends to increase (for both autonomous and conventional buses). The explanation is that, when penalty cost becomes larger, more buses are needed to reduce demand loss and bus fleet size goes up. The detailed cost components under different values of c_p are compared in Table 12. The results imply that the operator cost is increasing, while the waiting cost and the total penalty cost are reduced as c_p increases. The reason is that the price of demand loss is growing with c_p. Considering the trade-off among the three cost components, the operator will spend more costs to serve as many passengers as possible. As a consequence, the operator cost will increase to provide services with larger fleet size, and the waiting time cost will be reduced. Since more passengers will be served as well (seen from the expected unserved demand), the level of service of the system is improving when c_p increases.

The impact of cost parameters

In this work, both X and Y have a significant impact on the total cost and mixed bus fleets. When vehicle production technology advances, X and Y may change over time. A sensitivity analysis is conducted to investigate the impact of cost parameter X and Y. Different combinations of these parameters are considered and the results of model solutions on total cost and fleet size are plotted in Figs. 4 and 5.

Figure 4 shows that, when Y is fixed, the expected total cost tends to decrease as X rises, which indicates that X has negative effect on the total cost. On the other hand, Y has positive effect on the total cost because large value of Y will induce higher total cost. This is reasonable as lower ownership cost and operation cost will lead to smaller value of total cost in the mixed fleet situation. When the value of Y is large (i.e., 1 or 0.8), the decreasing of X can only reduce the total cost by a small margin. On the contrary, when the Y is small, especially at the value of 0.4, increasing X can induce significant cost-saving. This result implies the importance of the

Table 12 Results under different penalty cost

c_p (\$)	AB fleet	CB fleet	Operator cost (\$)	Waiting cost (\$)	Penalty cost (\$)	Unserved demand (pax)
2	13	[12,8,11,13,11,9]	2520.9	1490.3	94.3	47.15
5	14	[13,8,13,14,11,10]	2718.0	1411.2	80.9	16.18
10	15	[13,8,13,14,12,11]	2819.4	1368.5	55.9	5.59
15	15	[15,7,13,15,11,11]	2853.5	1352.2	52.3	3.49
20	17	[14,8,13,14,12,11]	2927.7	1329.9	36.1	1.81

AB autonomous bus, *CB* conventional bus

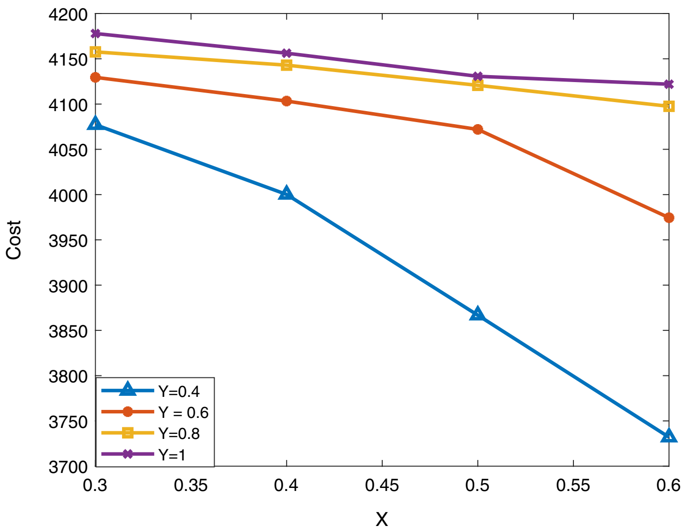
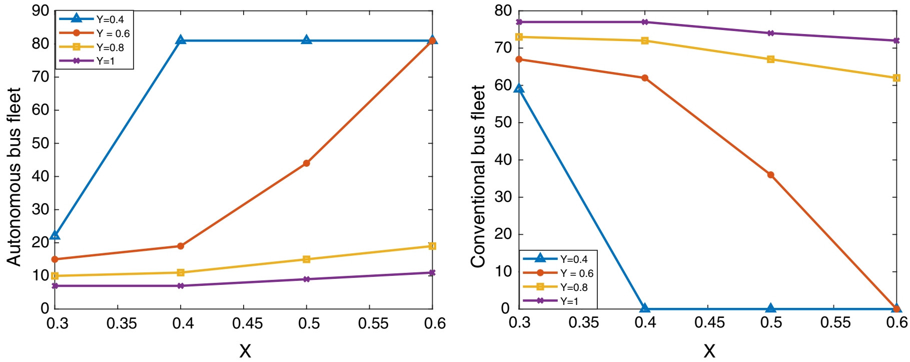


Fig. 4 The impact of X and Y on cost

technology investment and design on autonomous buses. Basically, if the acquisition cost of autonomous vehicle can be substantially reduced, then it could be deployed more widely and contribute more in reducing the total costs of the transit system.

In terms of bus fleet, we can see from Fig. 5 that the autonomous bus fleet size tends to increase when X goes up while the number of conventional buses tapers. When $Y < 0.6$, the fleet sizes of autonomous bus and conventional bus are very sensitive to X, which can be seen from the steep slopes. For the other cases, fleet sizes vary slightly. Other interesting results include that, with the combination of $X = 0.6, Y = 0.6$ and $X > 0.4, Y = 0.6$, the optimal solution is to provide all service withautonomous buses and the number of conventional buses is 0. Figures 4 and 5 give the optimal solutions under different values of X and Y. The result would offer guidance to bus operators on how to determine more cost-effective service operations.



(a) The impact of X and Y on autonomous bus fleets

(b) The impact of X and Y on conventional bus fleets

Fig. 5 The impact of X and Y on autonomous and conventional bus fleets

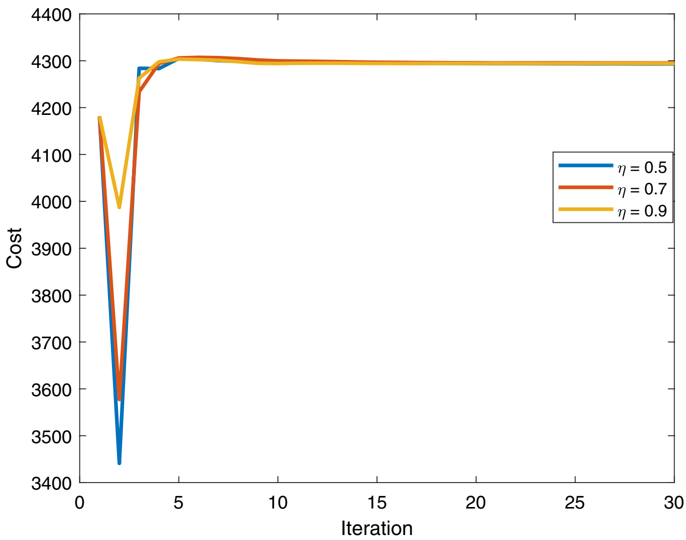


Fig. 6 The impact of η on QT-LA algorithm

The impact of η

To investigate the effects of step size η onto the performance of the algorithm, a brief discussion of its impacts on QT-LA algorithm is presented in this section. Here, we run the QT-LA algorithm under 3 different values of η, 0.5,0.7 and 0.9 respectively in the practical network. The case with demand uncertainty level $p = 0.6$ and the sample size $R = 1000$ is used for the three cases. The iteration number is set as 30.

In Fig. 6, one can notice that the algorithm can converge after about 10 iterations under all three different values of η. Meanwhile, under different η values, QT-LA algorithm almost converge to the same total cost. In our problem, η does not have much impact on the algorithm.

Conclusion

In this work, we developed a mathematical model formulation to find the optimal bus fleet size and vehicle allocation for a bus transit service network when autonomous buses are introduced. Stochastic demand was explicitly considered. To efficiently solve the proposed model, we transformed the SAA problem in its MINLP form into an MICQP, which can be solved to guarantee the global optimal solution. For large-scale problems, we proposed a more efficient QT-LA algorithm. In addition, we tested the application of the proposed QT-LA algorithm with large networks and sample size. Numerical results showed the optimal fleet size of autonomous buses and conventional buses under different scenarios. The flexibility of autonomous service can offset the uncertainty in demand patterns. Using autonomous vehicles can reduce the total cost and required fleet size. The impact analysis has been done in terms of cost parameters (X, Y). Nowadays, more and more major cities in the world are planning to deploy autonomous buses. The model formulation and solution algorithms proposed in this study would assist the bus service operators to determine the optimal operation strategies if autonomous buses are to be introduced into the bus service network.

There are some limitations and future directions. For example, we assumed that demands on different bus lines are independent; however, demands could be correlated in practice because of the potential line transfer and the existence of common lines. In this case, Clark's approximation (Clark 1961) or the multivariate random variate generation procedure can be adopted to tackle the issue (Li et al. 2012b). Nevertheless, we point out that our proposed SAA framework is still applicable because once the scenarios have been generated, we can approximate the stochastic program as a deterministic optimization problem. In this paper, the demand uncertainty is modeled on a day-to-day base. For future research, the within-day dynamics can be incorporated by separating the one-day operation into multiple periods, leading to a multi-stage stochastic program that requires more advanced solution techniques. Furthermore, one can consider the joint optimization of the fleet size and vehicle capacity. This will bring more flexibility to the operation and improve the system's ability to hedge against demand uncertainty. However, the line capacity will then depend on both the line frequency and the vehicle capacity, which may result in some bilinear or biconvex terms that impose challenges for solving the problem by our SAA framework. Finally, the driver shifting for conventional buses and vehicle charging for autonomous buses in operation is also worth investigating in future studies.

Acknowledgements This work is supported by Singapore Ministry of Education Academic Research Fund MOE2017-T2-2-093.

Author Contributions QT Drafting of manuscript, model formulation and solution method design. YHL Computational experiments and solution analysis. DZWW Study conception and critical revision.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix 1

Minimizing the objective function is equivalent to maximizing the following function:

$$\begin{aligned} \max \quad & \frac{\delta w_t |R||L|}{|R|} - Z(P2) \\ & = -\alpha^a M - \sum_l \alpha^c N_l - \sum_l c_l^c f_l^c - \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + \sum_r \sum_l c_p \cdot u_{l,r}\right] \\ & + \frac{w_t}{|R|} \sum_r \sum_l \frac{\delta(f_l^c + f_{l,r}^a) - k s_{l,r}}{f_l^c + f_{l,r}^a} \end{aligned} \tag{40}$$

where $Z(P2)$ is the objective function of P2. Define a variable y such that $y_{l,r} = \frac{\sqrt{\delta(f_l^c + f_{l,r}^a) - k s_{l,r}}}{f_l^c + f_{l,r}^a}$. (40) is equivalent to

$$\begin{aligned} \max \quad & \frac{w_t}{|R|}\left(\sum_l \sum_r 2 y_{l,r} \sqrt{\delta(f_l^c + f_{l,r}^a) - k s_{l,r}} - \sum_l \sum_r y_{l,r}^2 (f_l^c + f_{l,r}^a)\right) \\ & - \alpha^a M - \sum_l \alpha^c N_l - \sum_l c_l^c f_l^c - \frac{1}{|R|}\left[\sum_r \sum_l c_l^a f_{l,r}^a + \sum_r \sum_l c_p \cdot u_{l,r}\right] \end{aligned} \tag{41}$$

since plugging $y_{l,r}$ into (41) will lead to (40). Finally, define function $h(f_l^c, f_{l,r}^a, s_{l,r}) = \sqrt{\delta(f_l^c + f_{l,r}^a) - k s_{l,r}}$, we can rewrite the formulation in the hypograph form of h, i.e., $\beta \le h$, leading to the formulation in P4. Note that, to ensure the square root function is well-defined, δ should be chosen such that $\delta(f_l^c + f_{l,r}^a) - k s_{l,r} > 0, \forall l \in I, r \in R$.

References

Abe, R.: Introducing autonomous buses and taxis: quantifying the potential benefits in japanese transportation systems. Transp. Res. Part A Policy Pract. **126**, 94–113 (2019)

Aldaihani, M.M., Quadrifoglio, L., Dessouky, M.M., Hall, R.: Network design for a grid hybrid transit service. Transp. Res. Part A Policy Pract. **38**(7), 511–530 (2004)

Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**(1), 3–51 (2003)

An, K., Lo, H.K.: Two-phase stochastic program for transit network design under demand uncertainty. Transp. Res. Part B Methodol. **84**, 157–181 (2016)

Beaujon, G.J., Turnquist, M.A.: A model for fleet sizing and vehicle allocation. Transp. Sci. **25**(1), 19–45 (1991)

Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, vol. 2. Siam, Peninsula (2001)

Benson, H.: On the global optimization of sums of linear fractional functions over a convex set. J. Optim. Theory Appl. **121**(1), 19–39 (2004)

Bergqvist, O., Åstrand, M.: Bus line optimisation using autonomous minibuses (2017). https://www.divaportal.org/smash/get/diva2:1120402/FULLTEXT01.pdf

CAN News: Singapore: More Bus Drivers Needed in Next Five Years. https://www.channelnewsasia.com/news/singapore (2017). Accessed 26 June 2019

Ceder, A.: Estimation of fleet size for variable bus schedules. Transp. Res. Rec. **1903**(1), 2–10 (2005)

Chen, P.W., Nie, Y.M.: Optimal design of demand adaptive paired-line hybrid transit: case of radial route structure. Transp. Res. part E Logist. Transp. Rev. **110**, 71–89 (2018)

Chen, A., Zhou, Z., Chootinan, P., Ryu, S., Yang, C., Wong, S.: Transport network design problem under uncertainty: a review and new developments. Transp. Rev. **31**(6), 743–768 (2011)

Chen, J., Liu, Z., Zhu, S., Wang, W.: Design of limited-stop bus service with capacity constraint and stochastic travel time. Transp. Res. part E Logist. Transp. Rev. **83**, 1–15 (2015)

Clark, C.E.: The greatest of a finite set of random variables. Oper. Res. **9**(2), 145–162 (1961)

Council, A.T.: National guidelines for transport system management in Australia. Backgr. Mater. 5, Volume 4: Urban Transport, pp. 179–211 (2006)

Dandapat, S., Maitra, B.: An approach for identifying optimal service for rural bus routes. Case Stud. Transp. Policy **3**(3), 287–294 (2015)
Dong, X., DiScenna, M., Guerra, E.: Transit user perceptions of driverless buses. Transportation **46**(1), 35–50 (2019)
Geetha, S., Nair, K.: A variation of the assignment problem. Eur. J. Oper. Res. **68**(3), 422–426 (1993)
Gentile, G., Noekel, K.: Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems. Springer, Gewerbestrasse (2016)
Huang, Z., Ren, G., Liu, H.: Optimizing bus frequencies under uncertain demand: case study of the transit network in a developing city. Math. Prob. Eng. (2013). https://doi.org/10.1155/2013/375084
Jansson, J.O.: A simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy **14**(1), 53–80 (1980)
Jara-Díaz, S., Fielbaum, A., Gschwender, A.: Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. Transp. Res. Part A Policy Pract. **106**, 65–74 (2017)
Jiménez, F., Román, A.: Urban bus fleet-to-route assignment for pollutant emissions minimization. Transp. Res. Part E Logist. Transp. Rev. **85**, 120–131 (2016)
Ker, K., Kuo, H., Schondeld, M.: Optimal mixed bus fleet for urban operations. Transp. Res. Rec. **1503**, 39–48 (1995)
Kim, M., Schonfeld, P.: Conventional, flexible, and variable-type bus services. J. Transp. Eng. **138**(3), 263–273 (2012)
Kim, M.E., Schonfeld, P.: Integrating bus services with mixed fleets. Transp. Res. Part B Methodol. **55**, 227–244 (2013)
Kim, M.E., Schonfeld, P.: Integration of conventional and flexible bus services with timed transfers. Transp. Res. Part B Methodol. **68**, 76–97 (2014)
Kleywegt, A.J., Shapiro, A., Homem-de Mello, T.: The sample average approximation method for stochastic discrete optimization. SIAM J. Optim. **12**(2), 479–502 (2002)
Li, X., Quadrifoglio, L.: Optimal zone design for feeder transit services. Transp. Res. Rec. **2111**(1), 100–108 (2009)
Li, X., Quadrifoglio, L.: Feeder transit services: choosing between fixed and demand responsive policy. Transp. Res. Part C Emerg. Technol. **18**(5), 770–780 (2010)
Li, Z.-C., Lam, W.H., Sumalee, A.: Modeling impact of transit operator fleet size under various market regimes with uncertainty in network. Transp. Res. Rec. **2063**(1), 18–27 (2008)
Li, Z.-C., Lam, W.H., Wong, S.: On the allocation of new lines in a competitive transit network with uncertain demand and scale economies. J. Adv. Transp. **45**(4), 233–251 (2011)
Li, Z.-C., Lam, W.H., Wong, S.: Optimization of number of operators and allocation of new lines in an oligopolistic transit market. Netw. Spat. Econ. **12**(1), 1–20 (2012a)
Li, Z.-C., Lam, W.H., Wong, S., Sumalee, A.: Environmentally sustainable toll design for congested road networks with uncertain demand. Int. J. Sustain. Transp. **6**(3), 127–155 (2012b)
Lin, Y.H., Wang, Y., He, D., Lee, L.H.: Last-mile delivery: Optimal locker location under multinomial logit choice model. Transp. Res. Part E Logist. Transp. Rev. **142**, 102059 (2020)
Liu, H., Wang, D.Z.: Global optimization method for network design problem with stochastic user equilibrium. Transp. Res. Part B Methodol. **72**, 20–39 (2015)
Liu, Z., Yan, Y., Qu, X., Zhang, Y.: Bus stop-skipping scheme with random travel time. Transp. Res. Part C Emerg. Technol. **35**, 46–56 (2013)
Long, Y., Lee, L.H., Chew, E.P.: The sample average approximation method for empty container repositioning with uncertainties. Eur. J. Oper. Res. **222**(1), 65–75 (2012)
Luathep, P., Sumalee, A., Lam, W.H., Li, Z.-C., Lo, H.K.: Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach. Transp. Res. Part B Methodol. **45**(5), 808–827 (2011)
Nayan, A., Wang, D.Z.: Optimal bus transit route packaging in a privatized contracting regime. Transp. Res. Part A Policy Pract. **97**, 146–157 (2017)
Ongel, A., Loewer, E., Roemer, F., Sethuraman, G., Chang, F., Lienkamp, M.: Economic assessment of autonomous electric microtransit vehicles. Sustainability **11**(3), 648 (2019)
Patil, G.R., Ukkusuri, S.V.: Sample average approximation technique for flexible network design problem. J. Comput. Civ. Eng. **25**(3), 254–262 (2011)
Pinto, H.K., Hyland, M.F., Mahmassani, H.S., Verbas, I.Ö.: Joint design of multimodal transit networks and shared autonomous mobility fleets. Transp. Res. Part C Emerg. Technol. **111**, 255–293 (2020)
Salonen, A.O.: Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. Transp. Policy **61**, 106–110 (2018)
Salzborn, F.J.: Optimum bus scheduling. Transp. Sci. **6**(2), 137–148 (1972)

Shapiro, A., Philpott, A.: A tutorial on stochastic programming. www2.isye.gatech.edu/ashapiro/publications.html, 17 (2007)
Shen, K., Yu, W.: Fractional programming for communication systems–part i: power control and beamforming. IEEE Trans. Signal Process. **66**(10), 2616–2630 (2018)
Shen, Y., Zhang, H., Zhao, J.: Integrating shared autonomous vehicle in public transportation system: a supply-side simulation of the first-mile service in Singapore. Transp. Res. Part A Policy Pract. **113**, 125–136 (2018)
Silman, L.A., Barzily, Z., Passy, U.: Planning the route system for urban buses. Comput. Oper. Res. **1**(2), 201–211 (1974)
Szeto, W.Y., Jiang, Y.: Hybrid artificial bee colony algorithm for transit network design. Transp. Res. Rec. **2284**(1), 47–56 (2012)
Szeto, W.Y., Jiang, Y.: Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transp. Res. Part B Methodol. **67**, 235–263 (2014)
Szeto, W.Y., Wu, Y.: A simultaneous bus route design and frequency setting problem for tin Shui Wai, Hong Kong. Eur. J. Oper. Res. **209**(2), 141–155 (2011)
Tirachini, A., Antoniou, C.: The economics of automated public transport: effects on operator cost, travel time, fare and subsidy. Econ. Transp. **21**, 100151 (2020)
Wang, D.Z., Lo, H.K.: Multi-fleet ferry service network design with passenger preferences for differential services. Transp. Res. Part B Methodol. **42**(9), 798–822 (2008)
Wang, D.Z., Lo, H.K.: Global optimum of the linearized network design problem with equilibrium flows. Transp. Res. Part B Methodol. **44**(4), 482–492 (2010)
Wang, D.Z., Liu, H., Szeto, W.: A novel discrete network design problem formulation and its global optimization solution algorithm. Transp. Res. Part E Logis. Transp. Rev. **79**, 213–230 (2015)
Wang, D.Z., Nayan, A., Szeto, W.: Optimal bus service design with limited stop services in a travel corridor. Transp. Res. Part E Logis. Transp. Rev. **111**, 70–86 (2018)
Wen, J., Chen, Y.X., Nassir, N., Zhao, J.: Transit-oriented autonomous vehicle operation with integrated demand-supply interaction. Transp. Res. Part C Emerg. Technol. **97**, 216–234 (2018)
Wien, J.: An assessment of the willingness to choose a self-driving bus for an urban trip: a public transport user's perspective. http://resolver.tudelft.nl/uuid:8064cc17-dc0e-4c0c-9a9c-6efca8564d94 (2019)
Winter, K., Wien, J., Molin, E., Cats, O., Morsink, P., van Arem, B.: Taking the self-driving bus: a passenger choice experiment. In: 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp 1–8. IEEE (2019)
Yoo, G.-S., Kim, D.-K., Chon, K.S.: Frequency design in urban transit networks with variable demand: model and algorithm. KSCE J. Civ. Eng. **14**(3), 403–411 (2010)
Yu, B., Yang, Z.-Z., Jin, P.-H., Wu, S.-H., Yao, B.-Z.: Transit route network design-maximizing direct and transfer demand density. Transp. Res. Part C Emerg. Technol. **22**, 58–75 (2012)
Zak, J., Jaszkiewicz, A., Redmer, A.: Multiple criteria optimization method for the vehicle assignment problem in a bus transportation company. J. Adv. Transp. **43**(2), 203–243 (2009)
Zhang, W., Jenelius, E., Badia, H.: Efficiency of semi-autonomous and fully autonomous bus services in trunk-and-branches networks. J. Adv. Transp. (2019). https://doi.org/10.1155/2019/7648735

Qingyun Tian is a Ph.D. candidate at the School of Civil and Environmental Engineering, Nanyang Technological University, Singapore. She received the bachelor degree from Southeast University, China in 2016. Her research involves transport planning, network modeling, transit service operation and optimization.

Yun Hui Lin is a Ph.D. candidate at Department of Industrial Systems Engineering and Management, National University of Singapore. He received his bachelor degree in 2016 from Southeast University, China. His research interests include Facility Location, Last-mile Logistics, Supply Chain Optimization, and Urban Transportation Service.

Dr. David Z. W. Wang is currently an Associate Professor in the School of Civil and Environmental Engineering at Nanyang Technological University, Singapore. He received his Ph.D. degree from Hong Kong University of Science and Technology in 2008. His research area includes transport network modeling and optimization, public transport operations and planning, and integrated transportation planning.