Public Transp (2013) 5:53–78
DOI 10.1007/s12469-013-0067-7

ORIGINAL PAPER

Incremental bus service design: combining limited-stop and local bus services

Virot Chiraphadhanakul · Cynthia Barnhart

Published online: 6 July 2013

Abstract Long in-vehicle travel times resulting from frequent stops make bus service an unattractive choice for many commuters. Limited-stop bus services however have the advantage of shorter in-vehicle times experienced by passengers. In this work, we seek to modify a given bus service by optimally reassigning some number of bus trips, as opposed to providing additional trips, to operate a limited-stop service. We propose an optimization model to determine a limited-stop service route to be operated in parallel with the local service and its associated frequency to maximize total user welfare. A few theoretical properties of the model are established and used to develop a solution approach. As a proof of concept, we present numerical results obtained using real-world data together with comprehensive discussions of solution quality, computational times and the model's sensitivity to different parameters. Finally, we solve the optimization model for 178 real-world bus routes with different characteristics in order to demonstrate the impacts of some key attributes on potential benefits of limited-stop services.

Keywords Bus network design · Express service · Large-scale optimization · Limited-stop service · Mixed integer program · Public transit

This research was supported by the National Research Foundation Singapore through the Singapore-MIT Alliance for Research and Technology's Future Urban Mobility research program.

V. Chiraphadhanakul (✉)
MIT Operations Research Center, 77 Massachusetts Avenue Building E40-149, Cambridge, MA 02139, USA
e-mail: virot@alum.mit.edu

C. Barnhart
MIT School of Engineering, 77 Massachusetts Avenue Room 1-206, Cambridge, MA 02139, USA
e-mail: cbarnhart@mit.edu

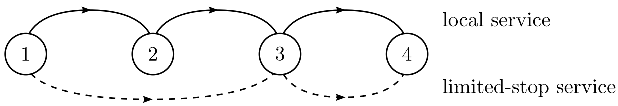


Fig. 1 An example of a 4-stop corridor with a local route overlapping with a limited-stop route serving stops 1, 3, and 4

1 Introduction

One of the major disadvantages of bus service, making it an unattractive choice for many commuters, is long in-vehicle travel times resulting from frequent stops. Limited-stop bus services on high-demand corridors, such as those successfully operated in cities such as Bogota, Chicago, Montreal, New York City, and Santiago, however, have the advantage of shorter in-vehicle times for passengers and shorter running times that enable bus operators to serve more demand with the same number of buses and reduced operating costs.

Given the benefits of limited-stop bus service, we are interested in finding an optimal way to introduce a limited-stop service to be operated in parallel with an existing local bus service, which serves every bus stop along the corridor (see Fig. 1). In this work, we focus on incremental changes to the existing schedule. In particular, we seek to modify a given bus service on a particular corridor by optimally reassigning some number of (local) bus trips, as opposed to providing additional trips, to operate a limited-stop service. This ensures that the new operation does not require additional buses and incurs no extra cost. Additionally, we consider introducing only one additional limited-stop route to facilitate adoption of the new service and avoid complicated operations.

The challenges in the incremental bus service design problem are as follows. There are trade-offs between in-vehicle time reduction and out-of-vehicle time increase. Specifically, while the passengers served by a limited-stop service experience shorter in-vehicle travel times, those served only by the reduced frequency local service have to wait longer for a bus. Additionally, we need to take into account passenger behavior in response to a limited-stop service. For passengers that are served by both local and limited-stop service, some may board the first bus to arrive; while the others may wait for the limited-stop service. A passenger's choice depends on the travel time savings he/she can get from the limited-stop service.

One of the major goals of this work is to develop a tractable optimization model together with an efficient solution approach that can be used to solve the incremental bus service design problem for real-world bus services. We propose an optimization model to determine: (1) the bus stops along a route to be served by a limited-stop service; and (2) the frequencies of the limited-stop and the local services that maximize total user welfare, for a given route during a given time period. A few theoretical properties of the model are established and used to develop a solution approach. Despite the implementation of limited-stop services around the world and the extensive literature on bus network design, there are, to our knowledge, only two published works, one by Leiva et al. (2010) and the other by Ulusoy et al. (2010) that develop optimization models for the design of limited-stop services overlapping with local

Table 1 Comparison of papers by Leiva et al. (2010), Ulusoy et al. (2010), and this work

	Leiva et al. (2010)	Ulusoy et al. (2010)	This work
Assumptions			
O-D matrix	Fixed	Fixed	Fixed
Transfers	Allowed	Allowed	Not Allowed
Number of limited-stop routes allowed	Unlimited	Unlimited	1
Objective	Minimize social costs	Minimize social costs	Maximize user welfare
Constraints			
Capacity	✓(heuristic)	✓	✓
Fleet size	–	✓	✓
Passenger assignment	Proportional to the frequencies of each attractive line	A logit-based model considering wait, transfer, and in-vehicle times	A linear function of frequency share and in-vehicle travel time savings
Solution approach	Leiva et al. *(*2010*)*: Given a set of predefined limited-stop routes, find optimal frequencies using a nonlinear program without capacity constraints. Iteratively increase the frequencies of overcrowded lines until the capacity constraints are satisfied. Ulusoy et al. *(*2010*)*: Exhaustively search over all predefined limited-stop routes and all possible frequencies for an optimal solution to a mixed integer nonlinear model. *This work*: For each frequency allocation, find an optimal limited-stop route using a mixed integer program together with an algorithm for reducing problem size. Then, select the frequency allocation that yields the highest objective value. A heuristic is proposed to further improve computational times		

services. The major differences between earlier published works and our approach are summarized in Table 1, and discussed in detail in what follows. Using data from a bus operator in a major city, we present numerical results as a proof of concept. Through the results, we also examine solution quality, computational times and the model's sensitivity to different parameters.

Another goal of this paper is to provide insights into limited-stop bus service design. Solving the optimization model for 178 bus services with different characteristics, we investigate the impacts of some key attributes discussed in Scorcia (2010), Leiva et al. (2010), and Larrain et al. (2010) on potential benefits of limited-stop services.

The paper is organized as follows. In the next section, we review the literature related to the design of limited-stop service. In Sect. 3, we describe the incremental bus service design problem and present the optimization model. In Sect. 4, we present the solution approach consisting of three key parts: decomposition, problem size reduction, and a heuristic. The numerical results together with discussions on solution quality, computational times and sensitivity analyses are provided in Sect. 5. In Sect. 6, we discuss how some characteristics of a bus service impact the potential benefits of limited-stop services. Finally, in Sect. 7, we conclude and suggest future research on this topic.

2 Literature review

Despite the extensive literature on public transit network design (see Ceder and Wilson 1986; Desaulniers and Hickman 2007; Ceder 2007; Guihaire and Hao 2008), there are, to our knowledge, only two published works, one by Leiva et al. (2010) and the other by Ulusoy et al. (2010) that develop optimization models for the design of limited-stop services overlapping with local services.

Leiva et al. (2010) formulate a mixed integer nonlinear model to determine frequencies of a set of *predefined* limited-stop routes such that the social costs, comprising user and operator costs, are minimized. For passenger assignment, they assume that every passenger chooses a set of acceptable lines based on the expected total travel times and always takes the first bus to come. Consequently, demand for each acceptable line is proportional to its frequency and does *not* depend on its travel time savings. In order to obtain an optimal solution, they first limit the complexity of the model by omitting the capacity constraints and solve the resulting nonlinear model (no binary decision variables). If the optimal solution violates the capacity constraint, a heuristic is then applied to progressively increase the frequencies of the overcrowded lines until the capacity constraints are satisfied. They present numerical results for a bus service along a 19-stop corridor with 23 predefined limited-stop routes (including express services, short turning, and deadheading). Computational times however are not provided. Additionally, they examine the impacts of different demand profiles on the objective function value. This topic is further discussed in Larrain et al. (2010).

One major advantage of this work is the flexibility of the optimization model, which allows transfers, multiple limited-stop routes along a corridor, and non-homogeneous fleet (big and small buses). Nevertheless, the numerical results show that transfer do not occur if the transfer penalties are high, and the additional benefits from having more than two limited-stop routes operated along a corridor is minimal.

Similarly, Ulusoy et al. (2010) formulate a mixed integer nonlinear model to determine frequencies of a set of *predefined* limited-stop routes such that the social costs are minimized. The main advantage of this work is that they estimate demand for each service according to the in-vehicle, wait, and transfer times using a logit-based model. This however results in an intractable nonlinear model. They propose an exhaustive search algorithm for obtaining an optimal set of frequencies. Using data from a real-world rail transit line, they present numerical results for a 6-station service. Computational times are again not provided.

In addition to the papers discussed above, Scorcia (2010) extends the work of Schwarcz (2004) and proposes a comprehensive framework for the design and evaluation of limited-stop and BRT services overlapping with local services. The model evaluates limited-stop service configurations based on six measures of effectiveness including demand split between local and limited-stop services and change in average passenger travel time. Although the work does not involve any optimization, they are particularly useful for evaluating limited-stop services obtained from optimization models.

Finally, there is another set of closely related works that provide optimization models and algorithms for other bus route planning strategies for high-demand corridors such as deadheading (Ceder and Stern 1981; Furth 1985), short turning (Ceder 1989; Furth 1987) and zonal service (Jordan and Turnquist 1979; Furth 1986).

3 Incremental bus service design

Given the benefits of limited-stop bus services, we seek to modify a given bus schedule on a particular corridor by optimally reassigning some number of bus trips, as opposed to providing additional trips, to operate a limited-stop service in parallel with the local service, which serves every stop along the corridor. Consequently, the new operation does not require additional buses and incurs no extra costs. In this section, we present an optimization model to determine, for a given bus route, (1) the bus stops along the route to be served by a limited-stop service; and (2) the frequencies of the limited-stop and the local services that maximize total user welfare. It is important to note that we consider introducing only one additional limited-stop route, as implemented in many cities, to facilitate adoption of the new service and avoid complicated operations. Moreover, Leiva et al. (2010) find that additional benefits of having more than two limited-stop routes operated along a corridor are minimal.

3.1 Basic assumptions

In the proposed model, we assume the following.

1. The O-D demand is given and fixed. Specifically, we assume that passengers will continue to board and alight at the stops they previously prefer and not walk to nearby stops that are served by both local and limited-stop services. However, the demand split between the local and limited-stop services is captured in the model and determined according to the attractiveness of services, demand elasticity, and available capacities.
2. Passengers arrive randomly at their origins at a constant rate over the time period under consideration.
3. Passengers are assigned on each bus service according to a system-optimal assignment, rather than user-optimal assignment. The validity of this assumption will be discussed further in Sect. 5.2.
4. Transfers between local and limited-stop services are not allowed. The validity of this assumption depends on the cost of a transfer. Leiva et al. (2010) show in their case study that transfers do not occur with high transfer penalties.

3.2 Model formulation

Consider an existing bus service serving bus stops in a set $S = \{1, 2, \dots, |S|\}$. The bus route begins at stop 1 and ends at stop $|S|$. Operated by a homogeneous fleet of buses with capacity of C passengers, the service currently runs at a constant frequency[1] of

[1]Note that the term ‘frequency’ in this paper refers to *the number of bus trips* operated over a period under consideration of length T minutes.

f_0 trips over a period under consideration (e.g., AM peak or PM peak) of length T minutes. Let K denote the set of origin-destination (O-D) pairs served by this bus service, which is given by $\{k = (s^k, d^k) \mid s^k, d^k \in S, s^k < d^k\}$. The expected demand for an O-D pair $k \in K$ over the time period is p^k passengers. Lastly, an expected travel time saving from running express from stop i to stop $j (j > i)$ (i.e., a service stops at stops i and j and skips every stop between i and j) is c_{ij} minutes. Note that if stops i and j are adjacent, that is $j = i + 1$, the expected travel time saving c_{ij} equals zero.

3.2.1 Decision variables

The decision variables used in the model are summarized as follows. The first set of variables is related to limited-stop service, and the second set is related to passenger assignment.

$f =$ number of limited-stop bus service trips over the period under consideration
$\alpha_{ij} =$ 1 if the limited-stop service runs express from stop i to stop j (i.e., stops consecutively at stop i and j), where $i, j \in S$ and $j > i$
$\beta_i =$ 1 if the limited-stop service serves stop $i \in S$
$\gamma^k =$ 1 if the limited-stop service serves O-D pair $k \in K$. Specifically, an O-D pair $k = (s^k, d^k) \in K$ is served by a limited-stop service if and only if both origin s^k and destination d^k are served by the limited-stop service.

$x_{ij}^k =$ portion of passengers of O-D pair $k \in K$ assigned to the express segment from stops i to j ($i, j \in S$ and $j > i$) of limited-stop service
$y^k =$ portion of passengers of O-D pair $k \in K$ assigned to local service
$z^k =$ portion of passengers of O-D pair $k \in K$ preferring the limited-stop service
$w_i =$ number of passengers on the local service traveling from stops i to $i + 1$

3.2.2 Objective function

The objective of our model is to maximize total user welfare, which is defined as total in-vehicle time savings for the passengers served by the limited-stop service, minus the total increase in wait time, weighted by disutility of wait time relative to in-vehicle time μ_w, for those served by the reduced frequency local service and those preferring the limited-stop service. Because we fix the total number of bus trips operated by local and limited-stop services, the new operation incurs no extra cost, and we omit operator cost from the objective function. Mathematically, the objective function is given by

$$\begin{aligned} \text{Maximize} \quad & \sum_{k\in K} p^k \sum_{(i,j)\in \Gamma^k} c_{ij} x_{ij}^k \\ & - \mu_w \sum_{k\in K} p^k (1-\gamma^k) \frac{1}{2}\left(\frac{T}{f_0 - f} - \frac{T}{f_0}\right) \\ & - \mu_w \sum_{k\in K} p^k z^k \frac{1}{2}\left(\frac{T}{f} - \frac{T}{f_0}\right), \end{aligned} \tag{1}$$

where Γ^k denotes a set of segments that can be used to serve O-D pair $k \in K$. Mathematically, for an O-D pair $k = (s^k, d^k) \in K$, Γ^k is given by $\{(i, j) \mid i, j \in S, s^k \leq i < j \leq d^k\}$.

The first term is the total in-vehicle time savings for the passengers served by the limited-stop service. For an O-D pair served only by local service, no passengers can be assigned on the limited-stop service (i.e., x_{ij}^k's are zero), and hence this term is zero. The second term corresponds to the total increase in the expected wait time (in equivalent in-vehicle minutes) for passengers served by the reduced frequency local service. For an O-D pair served by both local and limited-stop services, we have that $1 - \gamma^k$ equals 0, and this term is zero. Finally, the last term represents the total increase in the expected wait time (in equivalent in-vehicle minutes) for passengers preferring the limited-stop service.

Note that in Eq. (1), we have the expected wait times equal half the headway, assuming vehicle arrivals are equally spaced with perfect headway. In general, for a random headway H, the expected waiting time for a randomly arriving passenger is equal to $\frac{\sigma_H^2 + E^2[H]}{2E[H]}$, where $E[H]$ and σ_H^2 are the mean and variance of the headway H. Therefore, the perfect headway assumption can be relaxed by replacing a factor of $\frac{1}{2}$ with an appropriate value. Additionally, this form of wait time function may not be appropriate for low-frequency service, for which passengers tend to time their arrivals according to the published schedule. Nonetheless, because we focus on incremental changes to high-frequency service, operating every 15 minutes or less, passengers are presumably accustomed to not timing their arrivals and will continue this practice even after the limited-stop service is introduced.

3.2.3 Limited-stop service route constraints

In order to allow a limited-stop service route to begin and end at any bus stops in S, we introduce dummy stops s^+ and s^- at which a limited-stop service route virtually begins and ends, respectively. The following set of constraints ensures that the values of α_{ij}'s constitute a valid route.

$$\sum_{i \in S \setminus \{|S|\}} \alpha_{s^+,i} = 1 \tag{2}$$

$$\sum_{j \in S: j < i} \alpha_{ji} + \alpha_{s^+,i} = \sum_{j \in S: j > i} \alpha_{ij} + \alpha_{i,s^-} \quad \forall i \in S \tag{3}$$

$$\sum_{i \in S \setminus \{1\}} \alpha_{i,s^-} = 1 \tag{4}$$

$$\begin{aligned} \alpha_{ij} &\in \{0, 1\} && \forall (i, j) \in \left\{(i, j) \mid i, j \in S, i < j\right\} \\ \alpha_{s^+,i}, \alpha_{i,s^-} &\in \{0, 1\} && \forall i \in S \end{aligned} \tag{5}$$

Note that a limited-stop route serving exactly one stop (i.e., all α_{ij}'s are zeros except $\alpha_{s^+,i}$ and α_{i,s^-} for some $i \in S$) is also valid according to constraints (2)–(5). Such routes however cannot serve any passengers, while increasing wait times of all passengers. Therefore, their corresponding objective function values are negative, and they cannot be an optimal solution.

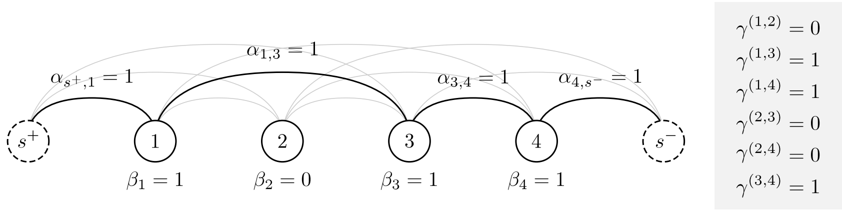


Fig. 2 An example of a 4-stop corridor with a limited-stop route serving stops 1, 3, and 4

Given the values of α_{ij}'s, the values of β_i's and γ^k's can then be obtained through the following constraints.

$$\beta_i = \sum_{j \in S: j>i} \alpha_{ij} + \alpha_{i,s^-} \quad \forall i \in S \tag{6}$$

$$\gamma^k \leq \beta_{s^k} \quad \forall k = \left(s^k, d^k\right) \in K \tag{7}$$

$$\gamma^k \leq \beta_{d^k} \quad \forall k = \left(s^k, d^k\right) \in K \tag{8}$$

$$\beta_i \in \{0, 1\} \quad \forall i \in S \tag{9}$$

$$\gamma^k \in \{0, 1\} \quad \forall k \in K \tag{10}$$

Note that a stop $i \in S$ is served by a limited-stop service ($\beta_i = 1$) if there exists an express segment starting at stop i in the limited-stop service route. According to constraints (7) and (8), for each O-D pair $k = (s^k, d^k) \in K$, if both origin and destination are served by the limited-stop service ($\beta_{s^k} = \beta_{d^k} = 1$), the value of γ^k in an optimal solution must be 1 in order to maximize the objective function value. Figure 2 illustrates the values of α_{ij}'s, β_i's and γ^k's for a 4-stop corridor with a limited-stop service serving stops 1, 3, and 4.

3.2.4 Passenger flow constraints

The validity of passenger flows is captured by the following constraints.

$$x_{ij}^k \leq \alpha_{ij} \quad \forall k \in K, (i, j) \in \Gamma^k \tag{11}$$

$$y^k + \sum_{j \in S: s^k < j \leq d^k} x_{s^k,j}^k = 1 \quad \forall k = \left(s^k, d^k\right) \in K \tag{12}$$

$$\sum_{j \in S: s^k \leq j < i} x_{ji}^k - \sum_{j \in S: i < j \leq d^k} x_{ij}^k = 0 \quad \forall k = (s^k, d^k) \in K, \; i \in \{i \in S \mid s^k < i < d^k\} \tag{13}$$

$$0 \leq x_{ij}^k \leq 1 \quad \forall k \in K, (i, j) \in \Gamma^k \tag{14}$$

$$0 \leq y^k \leq 1 \quad \forall k \in K \tag{15}$$

In words, constraints (11) ensure that each passenger can be assigned on an express segment only if the segment is included in the limited-stop service route. Imposed

Fig. 3 Example flows of passengers on O-D pair $(1, 4)$ on both local service and limited-stop service serving stops 1, 3, and 4

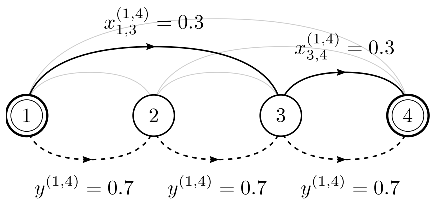
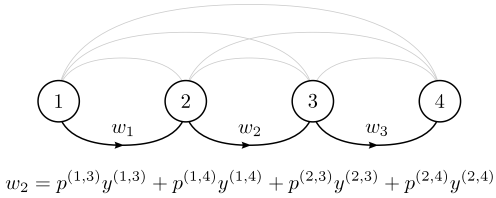


Fig. 4 Number of passengers on the local service traveling between adjacent stops



by constraint (12), the model assigns every passenger to either local or limited-stop service. Additionally, for a given O-D pair $k = (s^k, d^k) \in K$, the flow of passengers on the limited-stop service is conserved at each stop between s^k and d^k by constraint (13). Figure 3 depicts flows of passengers on O-D pair $(1, 4)$ (i.e., traveling from stop 1 to 4) on both local service ($y^{(1,4)}$) and limited-stop service ($x_{ij}^{(1,4)}$'s).

3.2.5 *Capacity constraints*

The total number of passengers on each service cannot exceed its total capacity, defined as frequency multiplied by bus capacity. Note that although the total capacity of both services is greater than the given travel demand, this set of constraints is still needed to ensure that the numbers of limited-stop service passengers and local service passengers each do not exceed their respective capacities. While the number of passengers on each segment of the limited-stop service route can be computed directly from the x_{ij}^k's, the number of passengers on each segment of the local service has to be derived through the w_i's (see Fig. 4). Thus, the capacity constraints can be written as follows.

$$w_1 = \sum_{\substack{k \in K, \\ s^k = 1}} p^k y^k \tag{16}$$

$$w_i = \sum_{\substack{k \in K, \\ s^k \le i}} p^k y^k - \sum_{\substack{k \in K, \\ d^k \le i}} p^k y^k \qquad \forall i \in S \setminus \{1, |S|\} \tag{17}$$

$$0 \le w_i \le (f_0 - f)C \qquad \forall i \in S \setminus \{|S|\} \tag{18}$$

$$\sum_{k \in K : \Gamma^k \ni (i,j)} p^k x_{ij}^k \le (fC)\alpha_{ij} \qquad \forall (i, j) \in \big\{(i, j) \mid i, j \in S, i < j\big\} \tag{19}$$

$$f \in \{1, 2, \ldots, f_0 - 1\} \tag{20}$$

Because our definition of frequency is the number of bus trips operated over a period under consideration, we are only interested in integral values of frequency. Additionally, we can ignore the cases where f equals 0, and f equals f_0. Clearly, when f is zero, the solution yields an objective function value of zero as there is no change to the original operation. When f equals f_0, that is, all buses operate a limited-stop service, the limited-stop service route must contain every stop along the corridor; otherwise, some O-D demand would not be satisfied. As a result, the limited-stop service becomes the local service, and again, there is no change to the original operation.

3.2.6 Demand split constraints

One of the challenges for this optimization problem formulation is to capture passenger behavioral changes in response to a new limited-stop service. We model demand for a limited-stop service as follows:

$$\sum_{j \in S: s^k < j \leq d^k} x^k_{s^k,j} \leq \frac{f}{f_0} + a^k \left(\sum_{(i,j) \in \Gamma^k} \alpha_{ij} c_{ij} \right) \quad \forall k \in K, \tag{21}$$

where a^k's are constants representing the incremental proportion of passengers preferring the limited-stop service per minute of travel time reduction.

From the equation, the demand for a limited-stop service is a linear function of frequency share and travel time reduction, which is given by the term in parentheses. If a limited-stop service does not provide any travel time reduction, passengers are indifferent between the local and limited-stop services and board the first bus to arrive. In this case, the demand for the limited-stop service is proportional to its frequency relative to the local service. As the reduction in travel time increases, the demand for the limited-stop service increases linearly at the rate of a^k because some passengers are willing to wait longer for the limited-stop service, as opposed to boarding the first arriving bus. One possible choice of a^k, which we use in this work, is the negative of travel time elasticity divided by the expected travel time of O-D pair k on the local service.

We refer to the portion of passengers assigned to a limited-stop service beyond its frequency share (f/f_0) as those *preferring* the limited-stop service. Mathematically, the portion of passengers on O-D pair k *preferring* the limited-stop service (z^k) can be obtained through the following constraints.

$$z^k \geq \sum_{j \in S: s^k < j \leq d^k} x^k_{s^k,j} - \frac{f}{f_0} \quad \forall k \in K \tag{22}$$

$$0 \leq z^k \leq 1 \quad \forall k \in K \tag{23}$$

Because the objective function improves as z^k decreases, the constraints ensure that z^k is equal to $\max(0, \sum_{j \in S: s^k < j \leq d^k} x^k_{s^k,j} - \frac{f}{f_0})$ in an optimal solution.

Lastly, note that if an O-D pair $k \in K$ is not served by a limited-stop service, x^k_{ij}'s must be zero, and hence the constraint (21) is redundant.

We close this section by establishing one important property of the model:

Proposition 1 *The integrality of β_i's and γ^k's can be relaxed.*

Proof Because constraints (2)–(4) together with (5) ensure that the right hand side of constraint (6) is either 0 or 1, we can simply omit the integrality constraint (9) of β_i's. Now consider the value of γ^k associated with O-D pair $k = (s^k, d^k)$. From constraints (7) and (8), if β_{s^k} and/or β_{d^k} take the value 0, γ^k must also be 0—an integral value. If both β_{s^k} and β_{d^k} equal 1, without the integrality constraint (10), γ^k may take any real value from 0 to 1, while all the constraints are still satisfied. In the optimal solution, however, γ^k has to take the value 1 in order to maximize the objective function, provided that p^k is positive. If p^k is 0, then γ^k can take any value without affecting the optimal solution. □

4 Solution approach

In this section, we present a solution approach to the mixed integer nonlinear model described earlier. The solution approach, consisting of three key parts, allows us to solve the incremental bus service design problem for real-world bus services efficiently.

4.1 Decomposition

Note that the nonlinearity in our model is caused by the limited-stop service frequency variable f in the capacity constraint (19) and the objective function (1). If a value of f is fixed, the model will become linear and can be solved for an optimal limited-stop service route more easily. We therefore decompose the original optimization problem into two stages. First, we repeatedly solve the optimization model assuming different limited-stop service frequencies. Then, given the set of optimal limited-stop routes for different limited-stop service frequencies, we select the limited-stop service frequency that yields a limited-stop route with the highest objective function value. This can be done because we are interested in values of f from a finite set $\{1, 2, \dots, f_0 - 1\}$. In particular, let z be the optimal objective function value of the optimization model in Sect. 3 and $z(f)$ be the optimal objective function value for a fixed f, we have that

$$z = \underset{f \in \{1,2,\dots,f_0-1\}}{\text{Maximize}}\; z(f).$$

When it is optimal to have no limited-stop service, the optimal limited-stop service routes for every f in $\{1, 2, \dots, f_0 - 1\}$ will be identical to local service, yielding an objective function value of zero.

One might attempt to establish a systematic way to search for the optimal limited-stop route frequency. However, we empirically found that $z(f)$ is not necessarily a unimodal function of f, and therefore search algorithms might find only a local optimum. This occurs because of the discrete nature of the limited-stop service route decisions. Consequently, we have to exhaustively search over the set of possible values of f in oder to ensure optimality. Nevertheless, there are usually a small number of possible frequency allocations for a period under consideration (e.g., AM peak or PM peak), and only frequency shares within a certain range are likely to be of interest to transit agencies.

4.2 Problem size reduction

The mixed-integer linear model resulting from fixing the value of limited-stop service frequency remains difficult to solve for large instances, that is, problems with many stops along the routes. It has some parallels to the facility location problem, which is commonly known as a hard problem. One possible way to limit computational complexity is to reduce problem size. To do so, we derive upper bounds on the contributions of α_{ij}'s to the objective function value, and then use these to eliminate some variables.

Recall that α_{ij} is equal to 1 when a limited-stop service serves stops i and j and no other stops in between, i.e., runs express from stops i to j. Thus, every passenger whose origin or destination is between stops i and j is not served by the limited-stop service and will experience increased expected wait time of $\delta_f = \frac{1}{2}(\frac{T}{f_0 - f} - \frac{T}{f_0})$. On the other hand, those whose trips start before stop i *and* end after stop j *might* benefit from the in-vehicle time savings of c_{ij} minutes on the limited-stop service. The actual contribution to the objective function is subject to available capacity, the demand split between local and limited-stop services, and whether their origins and destinations are served by limited-stop service. By assuming that every stop before i and after j is served by limited-stop service, the *maximum possible* in-vehicle time savings from running express from stops i to j is given by $c_{ij} \min(fC, \sum_{k \in K: \Gamma^k \ni (i,j)} p^k)$. Therefore, an upper bound on the contribution of α_{ij} to the objective function value, for a given limited-stop frequency f, is given by

$$U_{ij}(f) = c_{ij} \min\left(fC, \sum_{k \in K: \Gamma^k \ni (i,j)} p^k \right) - \mu_w \sum_{\substack{\{k \in K | i < s^k < j\} \\ \cup \{k \in K | i < d^k < j\}}} p^k \delta_f. \tag{24}$$

For any pair of stops i and j, a variable α_{ij} can then be eliminated from the formulation if the upper bound $U_{ij}(f)$ is negative for a given limited-stop service frequency f. Because variables x_{ij}^k's for all $k \in K$ such that $(i, j) \in \Gamma^k$ can take positive values only when α_{ij} equals 1, the corresponding x_{ij}^k's can also be eliminated.

Additionally, we observe and prove the following property of this upper bound.

Proposition 2 *For* $0 \le f \le f' < f_0$, *if* $U_{ij}(f) < 0$, *then* $U_{ij}(f') < 0$.

Proof We first claim that the upper bound $U_{ij}(f)$ is a concave function of $f \in [0, f_0)$. From Eq. (24), the two terms inside the minimum operator are linear, and hence concave, in f. Because concavity is preserved under minimum operation, we have that the first term in Eq. (24) is concave. The second term (including the minus sign) is also concave as its second derivative is non-positive for $f \in [0, f_0)$. Lastly, because concavity is preserved under summation, we have that $U_{ij}(f)$ is a concave function of $f \in [0, f_0)$. Next, note that $U_{ij}(0) = 0$. Depending on the first derivative of $U_{ij}(\cdot)$ at 0, the value of $U_{ij}(f)$ may increase as f increases until the first derivative becomes zero. The concavity ensures that once the value of $U_{ij}(f)$ falls below 0 for some f, it remains negative for all $f' > f$. We thus establish the claim. □

In other words, once a variable α_{ij} is eliminated for some f, it will also be eliminated for any $f' > f$. More importantly, the property suggests that it is generally easier to solve the optimization model for a large limited-stop service frequency f as more variables are likely to be eliminated.

4.3 Heuristic

The underlying idea of this heuristic arises from the observation that the optimal limited-stop service route for a particular limited-stop service frequency f is almost identical to the optimal limited-stop service routes for $f - 1$. Potentially, an optimal solution for a particular limited-stop service frequency f can be valuable input to the optimization model for $f - 1$ to reduce computational complexity.

Note that for a large limited-stop service frequency f, the optimal limited-stop service route tends to skip only a few stops because of the limited capacity of the local service and the substantial increase in wait time for passengers who are not served by the limited-stop service. As a limited-stop service becomes less frequent, the local service capacity increases; the increase in wait time for local passengers diminishes; and consequently, an optimal limited-stop route tends to skip more stops. Empirically, we observe that if a certain stop is not included in the optimal limited-stop route for a limited-stop service frequency f, that stop is also not included in the optimal routes for any $f' < f$. Thus, we propose the following heuristic that can be used to solve a sequence of optimization problems for different limited-stop service frequencies.

1. Solve the optimization model for a limited-stop service frequency $f = (f_0 - 1)$. Let $\beta_i^*(f)$ denote the optimal values of β_i's for a limited-stop service frequency f.
2. For each stop i in S, if $\beta_i^*(f)$ is zero, add the constraint $\beta_i = 0$ to the optimization model.
3. Solve the optimization model with the additional constraints for a limited-stop service frequency $f = f - 1$.
4. Repeat steps 2 and 3 for less frequent limited-stop service.

We start the heuristic with a large limited-stop service frequency because it is easier to solve as suggested by the property in Proposition 2.

Finally, although the optimality of the solutions obtained from this heuristic has not been proved analytically, we have not found any instances where this heuristic leads to suboptimal solutions.

Table 2 Baseline values of parameters

Parameter		Baseline value
Wait time disutility weight	μ_w	1.0
Bus capacity	C	80 passengers
Travel time between adjacent stops	$t_{i,i+1}$	1.5 minutes
Dwell time at stop i	t_i^d	0.5 minutes
Travel time elasticity	e	-0.5
Total travel time from stop i to stop j on a local service	t_{ij}	$\sum_{l=i}^{j-1} t_{l,l+1} + \sum_{l=i+1}^{j-1} t_l^d$
Total travel time savings from running express between stops i and j	c_{ij}	$\sum_{l=i+1}^{j-1} t_l^d$
Rate of increase in limited-stop service demand per minute of travel time reduction for O-D pair $k = (s^k, d^k)$	a^k	$\frac{-e}{t_{s^k,d^k}}$

5 Proof of concept

Using data from a bus operator in a major city, we provide in this section numerical results obtained from our optimization model and solution approach. We also examine solution quality, computational times, and the model's sensitivity to different parameters.

5.1 Data and parameters

We obtain real-world data from a bus operator in a major city. The data set contains route information and expected O-D demands of 178 high-frequency bus routes, operating every 15 minutes or less. In this work, we focus on the weekday, morning peak schedules, from 7 a.m. to 9 a.m. ($T = 120$ minutes).

Table 2 summarizes the baseline values and expressions of the parameters we use in this work. We assume that the travel times between adjacent stops are equal for simplicity, and more importantly, to facilitate our understanding of the model behavior and solutions. The dwell times at each stop are also assumed to be equal for the same reasons. We acknowledge that instead of constants, dwell times should be a function of the expected numbers of passengers boarding and alighting at the stops, which in turn depend on the decision variables—limited-stop service route, frequency allocation, and the resulting passenger assignments. Incorporating variable dwell times however will result in an intractable model. Lastly, we simply specify the total travel time savings from running express between two stops as the sum of the dwell times at the skipped stops (i.e., every stop between the two stops). To be more precise, rigorous estimates of limited-stop service run times like the one presented in (Tétreault and El-Geneidy 2010) can be used.

In this section, our discussions focus on one particular bus service, referred to as bus service A. Bus service A consists of 35 stops along its corridor spanning almost 9 miles. With 24 bus trips, the service carries on average a total of 3,151 passengers during the two-hour morning peak period. The average trip length of passengers served by service route A is 11 stops.

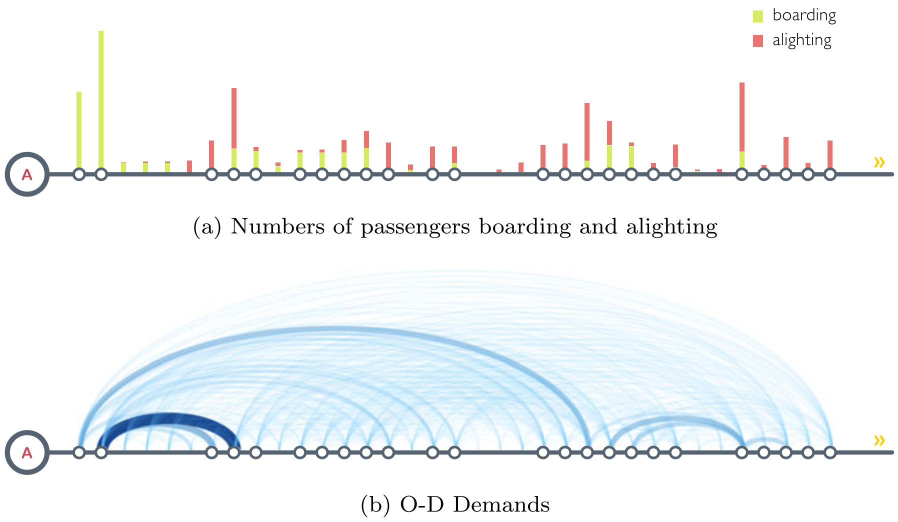


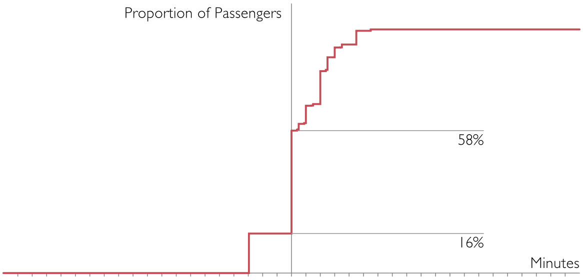
Fig. 5 Visualization of the optimal limited-stop service route together with (**a**) numbers of passengers boarding and alighting at each stop and (**b**) O-D demands

5.2 Numerical results

We implemented the solution approach proposed in Sect. 4 with Java 1.6 and IBM ILOG CPLEX 12.2. The result obtained from the optimization model shows that, by reassigning 13 out of 24 bus trips (54 %) to operate a limited-stop service, a total user welfare of 1,506 minutes can be achieved on route A. Covering 84 % of the demand, the limited-stop route skips 11 out of 35 stops (31 %) along the corridor. Figure 5 visualizes the optimal limited-stop service route together with (a) numbers of passengers boarding and alighting at each stop and (b) O-D demands. A hollow circle represents a bus stop which is served by the limited-stop route. A height of a bar at each stop in Fig. 5a indicates the number of passengers boarding/alighting at the stop. Thickness and opacity of an arc connecting two stops in Fig. 5b indicates a proportion of passengers on the O-D pair.

It is evident that most bus stops with high demands for boarding and/or alighting are included in the limited-stop service route as the service can then potentially benefit a large number of passengers. However, the bus stops that are served by the limited-stop service do not necessarily have higher demands than those skipped stops. For example, stop 6, which is not served by the limited-stop service, has slightly higher demand than stop 32 (the fourth from last), which is served by the limited-stop service. This is because stopping at stop 6 affects in-vehicle travel times of a large number of passengers boarding the limited-stop service at the first and second stops. Additionally, there are very few passengers boarding at stop 6, and those who alight at stop 6 gain only little benefit from the limited-stop service. On the other hand, although there are fewer passengers alighting at stop 32, a lot of them travel longer distances, thereby contributing larger in-vehicle travel time reduction to the objective function value. Moreover, stopping at stop 32 affects in-vehicle travel times

Fig. 6 Cumulative distribution function (CDF) of travel time changes



of a smaller number of passengers on the limited-stop service who alight at the last three stops.

It is essential to understand how passenger travel times (wait and in-vehicle times) change compared to the original service. Figure 6 shows a cumulative distribution of the travel time changes. About 16 % of the passengers are not served by the limited-stop service and have to wait on average 3 minutes longer for the frequency-reduced local service. Another 42 % of the passengers are not affected by the introduction of the limited-stop service. In particular, these passengers are served by both local and limited-stop services, but the limited-stop service does not provide any travel time reductions to their trips (i.e., every stop between their origins and destinations is served by the limited-stop service). Lastly, the other 42 % of the passengers benefit from travel time reductions ranging from 0.4 to 5.5 minutes. The average travel time reduction is 2.3 minutes.

The distribution of the travel time changes suggests that passengers are not likely to make transfers between local and limited-stop service because besides the inconvenience of transferring and the possibility of additional fare costs, the potential travel time savings are offset for most passengers by additional wait times at transfer points. Specifically, if a passenger first boards the local service and connects to the limited-stop service, the additional wait time is $\frac{1}{2}(\frac{120}{13}) = 4.6$ minutes on average; and if a passenger first boards the limited-stop service and connects to the local service, the additional wait time is $\frac{1}{2}(\frac{120}{11}) = 5.5$ minutes on average. Therefore, for this particular route, our assumption that transfers between local and limited-stop services are not allowed is generally valid. Similarly for the other 177 high-frequency bus routes in our network, we also find that, in each optimal solution, the average wait time for both local and limited-stop services is larger than the average travel time reduction provided by the limited-stop service.

In terms of ridership split, 55 % and 45 % of the passengers are assigned on the local and limited-stop services, respectively. Figure 7 visualizes a bus's load profile for each service. The dashed line at the top indicates the bus capacity. The limited-stop service is generally less crowded than the local service. Moreover, it is important to note that the limited-stop buses are never full, which implies that the capacity constraints of the limited-stop service are not tight. Consequently, all the passengers who want to get on a limited-stop bus (either because it is the first bus to arrive, or because he/she *prefers* the limited-stop service) can board, and the system-optimal assignment is identical to the user-optimal assignment. This is also the case for the other bus routes in our network. Nonetheless, in general, when a limited-stop service

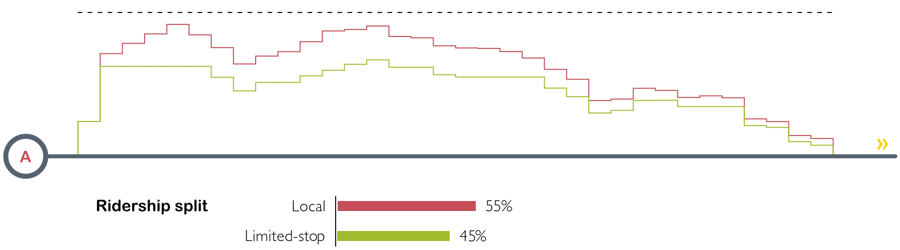


Fig. 7 Load profiles of the local and limited-stop services

capacity is reached—especially in subproblems with small limited-stop service frequencies, the optimal solution corresponds to the system-optimal assignment, and its objective function value serves as an upper bound on the total user welfare in the user equilibrium.

5.3 Computational times

One major goal of this work is to develop an optimization model together with a solution approach that can be used to solve efficiently the incremental bus service design problem for real-world bus services. We present in this section computational times and discussions on the effectiveness of the solution approach proposed earlier.

Computations are carried out on a Mac OS X machine with an Intel Core i7 2.7 GHz processor and 8 GB of RAM. We decompose the problem into 23 subproblems, one for each possible value of limited-stop service frequency f. The computational time for each subproblem is limited to 300 seconds. Table 3 summarizes the computational times for bus service A when the problem size reduction and/or the heuristic presented in Sect. 4 are applied. In order to compute the optimality gap of a solution, we obtain the optimal solution for each subproblem by applying the problem size reduction and letting the CPLEX MIP solver run without imposing a time limit.

As a baseline, we first solve each subproblem using only the MIP solver, that is, neither the problem size reduction nor the heuristic presented in Sect. 4 are applied. For f less than or equal to 11, the solver cannot obtain any feasible solution within the time allotted; and for f between 12 and 14, the solutions provided by the solver are not optimal.

The next set of computational times are obtained by using the upper bound in (24) to reduce the problem sizes before running the MIP solver. Although feasible solutions still cannot be obtained within the time limit for f less than or equal to 11, the computational times are significantly reduced for the other values of f. Moreover, the optimal solutions for f equal to 13 and 14 can now be obtained within the time limit.

For the third set of computational times, we apply the heuristic outlined in Sect. 4.3 without the problem size reduction. The heuristic enables us to obtain the optimal solutions within the time limit for all possible limited-stop service frequencies except for f equal to 1 and 2, where suboptimal feasible solutions are obtained. Note that for f equal to 1, the MIP solver stops before the time limit is reached with a nonzero

Table 3 Computational time

Limited-stop service frequency	Computational time in seconds (non-zero optimalitygap)						
	None		Problem size reduction only		Heuristic only		Both
1	–		–		63.81	(19 %)	298.93
2	–		–		299.03	(20 %)	299.07
3	–		–		298.96		299.10
4	–		–		299.10		298.88
5	–		–		225.30		137.35
6	–		–		144.66		86.10
7	–		–		113.87		64.62
8	–		–		72.03		40.58
9	–		–		38.73		28.12
10	–		–		22.83		13.86
11	–		–		14.65		5.87
12	289.38	(35 %)	297.49	(13 %)	12.72		4.29
13	289.52	(31 %)	297.79		8.85		2.20
14	289.25	(6 %)	128.12		4.49		1.09
15	266.14		40.84		3.65		0.65
16	139.66		23.16		11.78		1.94
17	94.83		8.92		23.12		2.03
18	46.83		4.36		13.54		1.73
19	36.49		1.48		20.76		0.63
20	22.00		0.60		5.49		0.20
21	7.72		0.16		5.57		0.09
22	4.29		0.05		2.36		0.04
23	1.30		0.02		1.44		0.02

optimality gap. This occurs because the heuristic adds more constraints to the original formulation according to the solution for f equal to 2, which is not optimal, and consequently leads the MIP solver to an incorrect optimal solution. The computational times are again significantly reduced from the baseline, although the problem size reduction appears to be more effective for f greater than or equal to 17.

Finally, we apply both the problem size reduction and the heuristic. The optimal solutions can now be obtained within the time limit for all possible limited-stop service frequencies, and the computational times are further reduced.

Effectiveness of problem size reduction using the upper bound in (24) is depicted in Fig. 8. In the figure, each arc connecting stops i and j represents a variable α_{ij} that has a nonnegative upper bound on the contribution to the objective function value and hence remains in the optimization model. For a limited-stop frequency of 1, only about 2 % of variables α_{ij}'s are eliminated. As implied by Proposition 2—more variables can be eliminated for higher limited-stop service frequencies, 80 % and 94 % of variables α_{ij}'s are eliminated for a limited-stop service frequencies of 14 and 23, respectively. For the latter, there are only two possible routes from the re-

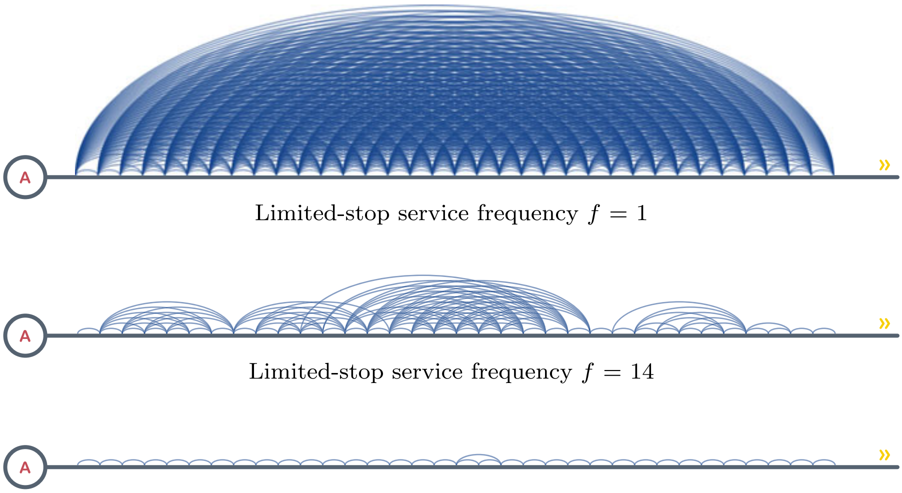


Fig. 8 Remaining variables after problem size reduction using the upper bound in (24)

maining variables—one serves every stop (like the local service) and the other serves every stop except stop 19. Therefore, the optimization model can be solved extremely quickly.

Moreover, the optimal limited-stop service routes for different limited-stop service frequencies in Fig. 9 illustrates the optimality of the heuristic for bus service A. In particular, it can be seen that when a stop is skipped in an optimal limited-stop route for a limited-stop service frequency f, the stop is also skipped in the optimal limited-stop routes for any limited-stop service frequencies smaller than f.

5.4 Sensitivity analyses

The numerical results presented in Sect. 5.2 are obtained using the baseline values of parameters. In this section, we examine how changes in the values of parameters affect the total user welfare for bus service A. The results are summarized in Fig. 10. A solid circle in each plot indicates the baseline case. Additionally, the characteristics of limited-stop service in the optimal solution of each scenario are provided in Table 4.

Dwell time Dwell times depend on many factors such as traffic conditions, passenger loads, fare payment methods, busway designs, and vehicles (floor heights, door configurations). Increasing the dwell time at a stop causes increases in total in-vehicle travel times for services serving the stop and total travel time savings of limited-stop services skipping the stop. Because, in this work, the total travel time savings is simply defined as a sum of dwell times at the skipped stops, varying dwell times essentially adjusts travel time savings achievable by limited-stop services. As a result, total user welfare increases as the dwell time per stop increases. Additionally, Fig. 10a shows that these increases are more substantial as dwell times increase.

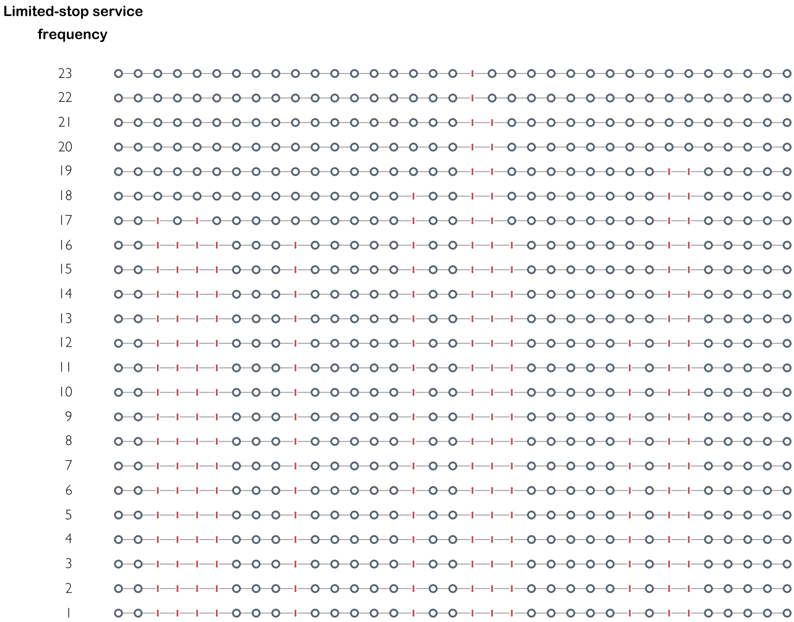


Fig. 9 Optimal limited-stop service routes for different limited-stop service frequencies

In-vehicle: wait time disutility So far, we assume that one minute of in-vehicle travel time poses the same level of disutility as one minute of wait time. In the transportation literature, wait time cost is usually assumed to be larger than in-vehicle time cost primarily because of discomfort caused by weather, safety, etc. On the other hand, transit agencies in many cities around the world now provide real-time bus arrival information, allowing passengers to schedule their arrivals at bus stops such that their wait times are minimized, regardless of the scheduled headway of the bus service. In this case, the cost of wait time amounts to the cost of schedule delay, which might be less than or equal to the in-vehicle time cost as passengers can remain productive until the next bus arrives.

According to Fig. 10b, as wait time disutility increases, the total user welfare drops at a decreasing rate. This suggests that reassigning some number of local bus trips to operate a limited-stop service is particularly beneficial to bus systems for which the wait time cost or disutility is relatively low compared to the in-vehicle time cost.

Travel time elasticity Travel time elasticity usually varies from one city to another and depends on many factors such as income, trip length and time of day. Typically, travel time elasticity ranges between −0.3 and −0.7. According to constraints (21)–(23), increasing travel time elasticity can potentially increase the proportion of passengers *preferring* limited-stop service. The increase is however subject to travel time reductions relative to travel times on the local service and the additional wait time resulting from not boarding the first arriving bus. For dwell times of 30 seconds, total user welfare decreases minimally as travel time elasticity increases (see Fig. 10c). This happens because the travel time reductions achievable by limited-stop

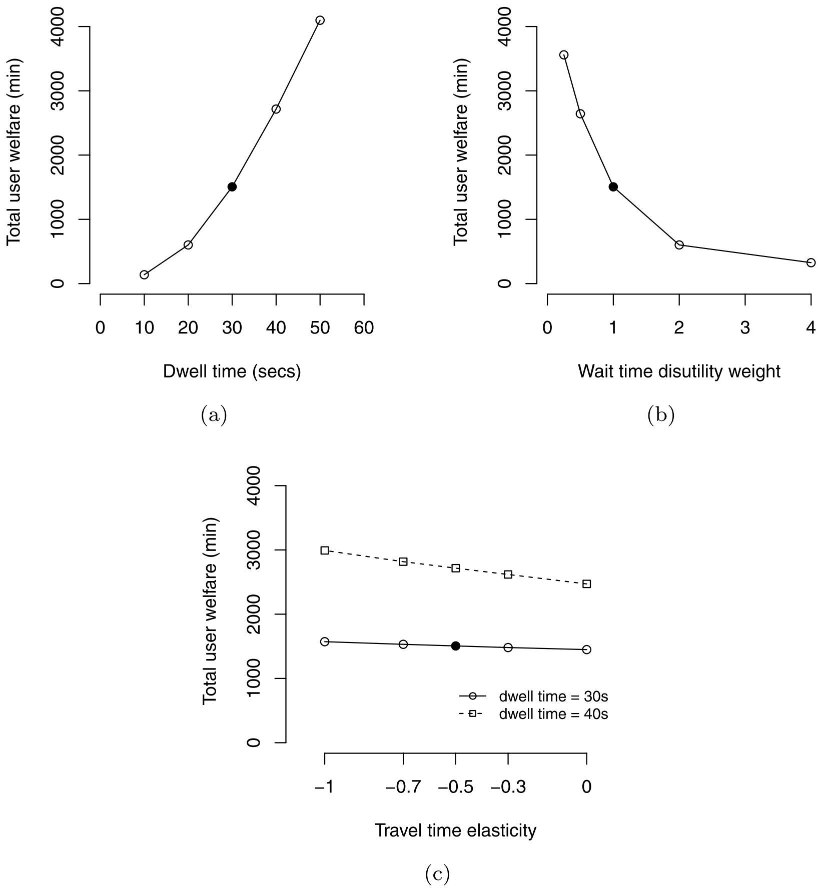


Fig. 10 Sensitivity analyses

services are small relative to the travel times on the local service. Specifically, only 1.4 % of passengers *prefer* the limited-stop service in the baseline case. We also perform a sensitivity analysis for dwell times of 40 seconds to see the effect of travel time elasticity when the travel time reductions achievable by limited-stop services are higher. In this case, the changes in total user welfare are more significant. Lastly, note that because we assume a fixed O-D matrix, travel time elasticity only affects total user welfare through the ridership split between the local and limited-stop services. In reality, it also affects the number of new riders attracted to the limited-stop service.

6 Insights into limited-stop bus service design

The tractability of the optimization model together with our efficient solution approach allows us to solve the incremental bus service design problem for all 178

Table 4 Optimal solutions from sensitivity analyses

	Total user welfare (in-vehicle minutes)	Limited-stop service		
		Frequency	No. of stops skipped	Ridership (%)
Dwell time (s)				
10	138	14	2	43
20	602	10	10	33
30	1,506	13	11	45
40	2,716	14	12	52
50	4,098	15	12	56
Wait time disutility weight				
0.25	3,560	17	13	58
0.5	2,643	16	12	57
1	1,506	13	11	45
2	602	12	5	37
4	326	12	2	39
Travel time elasticity (dwell time = 30 s)				
−1.0	1,572	14	11	55
−0.7	1,531	13	11	45
−0.5	1,506	13	11	45
−0.3	1,481	13	11	42
0.0	1,449	12	11	39
Travel time elasticity (dwell time = 40 s)				
−1.0	2,992	15	12	60
−0.7	2,818	15	12	57
−0.5	2,716	14	12	52
−0.3	2,618	14	12	50
0.0	2,471	14	11	44

high-frequency bus routes in our data set, in which the longest bus route consists of 104 stops along its corridor spanning 23.4 miles. Given the optimal solutions for these bus services with different characteristics, we examine the impacts of some key attributes discussed in Scorcia (2010), Leiva et al. (2010), and Larrain et al. (2010) on the potential benefits of limited-stop services.

The attributes of bus services we consider in this work are demand volume (passengers), service frequency (buses), route length (stops), average trip length (stops), and demand variability (dimensionless). For a loop service, because once a bus completes its service in one direction, it continues the service in the reverse direction right away, we treat the service in both directions as a continuous service on one long corridor, and hence the route length is given by the total number bus stops along both directions. Additionally, we measure demand variability using the coefficient of variation of the total demand (boarding and alighting) at each bus stop. In order to fairly compare the benefits of limited-stop services among different bus services, we

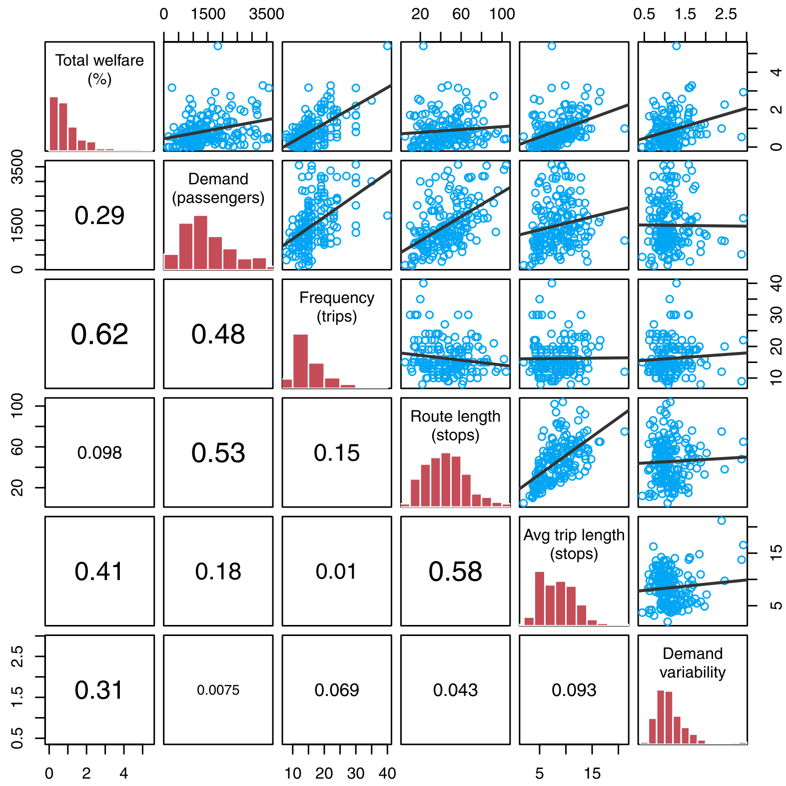


Fig. 11 Statistics and correlations of different bus service attributes and total user welfare

calculate the total user welfare attained by an optimal solution as a percentage of the total travel time, including both wait and in-vehicle travel time, of all passengers.

Figure 11 summarizes the statistics and correlations of different attributes and total user welfare. Specifically, the diagonal panels show the distributions of each attribute in our data set. The upper diagonal panels are scatter plots for each pair of attributes, and their associated correlation coefficients are provided in the lower diagonal panels.

Service frequency has the highest correlation with total user welfare. This is reasonable because if an original service frequency is low, reassigning some bus trips to operate a limited-stop service will drastically increase wait times for those who are only served by the reduced frequency local service. Although it is commonly known that limited-stop services are promising for high-demand corridors, the correlation between demand and total user welfare is not particularly high. One possible reason is that, despite the high demand, many passengers may only make short trips, thereby not gaining large benefits from limited-stop service. Nevertheless, because high demand generally implies a large number of passengers who can potentially benefit from a limited-stop service, the correlation coefficient between the demand and the total user welfare *in minutes*, as opposed to the relative percentage, is as high as 0.56.

With the second highest correlation with total user welfare, average trip length is another key attribute that determines the benefits of limited-stop services. For every

O-D pair, it takes into account both the number of passengers and how much they can potentially benefit from limited-stop services. Even though route length is highly correlated with average trip length, it is barely correlated with total user welfare, and hence not an accurate indicator of a successful limited-stop service. Again, this is simply because passengers do not necessarily travel along the entire long route. However, because a long bus route typically serves more passengers, the correlation coefficient between the route length and the *absolute* total user welfare (in minutes) is as high as 0.34.

Lastly, demand variability exhibits correlation with total user welfare to some extent. High demand variability corresponds to a concentration of demands, for both boarding and alighting, at certain stops. Consequently, high demand variability allows a limited-stop service to serve a number of passengers without making frequent stops and therefore increases the potential benefits of limited-stop services. Recall that we measure demand variability using the coefficient of variation of the total demand at each stop. As a result, it does not capture how high-demand stops are distributed along the corridor. In particular, if the high-demand stops are close together, having a limited-stop service serving all the stops will result in minimal travel time reduction, while having a limited-stop service skipping some of the stops will increase wait times of many passengers. In short, high demand variability can only partially indicate the demand patterns or load profiles that allow for successful limited-stop services. Additionally, Scorcia (2010) points out that it is important to have some minimal level of demand at low-demand stops. Otherwise, local service may spend short amounts of time at the stops or even skip them often, and hence, limited-stops services provide little additional travel time savings. Because in this work, we assume a constant dwell time at each stop, our results cannot demonstrate this observation.

7 Conclusions and future work

This paper addresses the incremental bus service design problem, in which we seek to modify a given bus service by optimally reassigning some number of bus trips to operate a limited-stop service without incurring extra operating costs. We formulate a mixed integer nonlinear model to determine the limited-stop service to be operated in parallel with the local service, and to optimize its associated frequency to maximize total user welfare. Exploiting some theoretical properties of the model, the proposed solution approach consists of three parts: decomposition, problem size reduction, and a heuristic. Although the optimality of the heuristic has not been proved analytically, we have not found any instances where the heuristic leads to suboptimal solutions.

Using real-world data from a bus operator in a major city, as a proof of concept, we provide numerical results together with detailed discussion regarding solution quality, including the distribution of travel time changes and ridership split between local and limited-stop services. The reported computational times demonstrate the tractability of the model and effectiveness of the solution approach. The sensitivity analyses shows that

- as travel time savings achievable by limited-stop services increases, the benefits of limited-stop services, measured by total user welfare, increase at an increasing rate;

– reassigning some number of local bus trips to operate a limited-stop service is particularly beneficial to the bus systems for which the wait time cost is relatively low compared to the in-vehicle time cost; and
– the impact of travel time elasticity is noticeable only when the travel time reductions achievable by limited-stop services are large relative to the travel times on the local service.

Moreover, we solve the optimization model for 178 bus routes with different characteristics in order to examine the impacts of some key attributes on the potential benefits of limited-stop services. We find that service frequency and average trip length are highly correlated with the total user welfare attained by the optimal solutions, while demand volume and route length show reasonable correlation with total user welfare only in an absolute sense. Lastly, demand variability exhibits correlation with total user welfare only to some extent as it still depends on the underlying demand profile.

As in most mathematical models in the public transit network design literature, we make certain assumptions to simplify the problem and ensure tractability of the model. In terms of future research, the following are particularly interesting directions:

– *Relaxing the fixed O-D matrix assumption.* This assumption prohibits passengers whose origins or destinations are not served by a given limited-stop service from walking to nearby stops that are served by both local and limited-stop services. The assumption can be restrictive if the travel time savings from a limited-stop service are considerable and sufficiently justify walking to other bus stops. Additionally, this ignores potential ridership increase in the long run.
– *Considering multiple bus routes that share a segment along their corridors simultaneously.* This is important because introducing a limited-stop service for one bus route will not only affect the ridership split between the local and limited-stop services, but might also lead to ridership shift from one bus route to another.
– *Taking into account shorter running times of limited-stop service.* As a result of shorter running times of limited-stop service, an operator might be able to operate more bus trips over a period under consideration using the same number of vehicles. This however will increase the operating costs, and the objective function can no longer omit the operator cost to ensure profitability.
– *Incorporating limited-stop service frequency into the demand model.* In the presented model, the portion of passengers preferring a limited-stop service only depends on the travel time savings, not the frequency of limited-stop service. This can be addressed by (1) redefining a^k as a function of limited-stop frequency f, instead of a constant; and/or (2) subtracting the expected additional wait time for limited-stop service (in equivalent in-vehicle minutes) from the total travel time savings in the demand model. Although the resulting demand model will become nonlinear in f, the presented solution approach still works as it decomposes the problem into subproblems with fixed values of f.

Acknowledgements The authors thank Professor Nigel H.M. Wilson and the anonymous reviewers for their useful comments.

References

Ceder A (1989) Optimal design of transit short-turn trips. Transp Res Rec 1221:8–22. http://trid.trb.org/view.aspx?id=308479

Ceder A (2007) Public transit planning and operation: theory, modelling and practice. Elsevier, Butterworth-Heinemann

Ceder A, Stern H (1981) Deficit function bus scheduling with deadheading trip insertions for fleet size reduction. Transp Sci 15(4):338–363. doi:10.1287/trsc.15.4.338. http://transci.journal.informs.org/cgi/content/abstract/15/4/338

Ceder A, Wilson N (1986) Bus network design. Transp Res, Part B, Methodol 20(4):331–344. doi:10.1016/0191-2615(86)90047-0

Desaulniers G, Hickman M (2007) Public transit. In: Barnhart C, Laporte G (eds) Transportation, handbooks in operations research and management science, vol 14. Elsevier, Amsterdam, pp 69–127. Chap. 2. doi:10.1016/S0927-0507(06)14002-5

Furth P (1985) Alternating deadheading in bus route operations. Transp Sci 19(1):13–28. doi:10.1287/trsc.19.1.13

Furth P (1986) Zonal route design for transit corridors. Transp Sci 20(1):1–12. doi:10.1287/trsc.20.1.1. http://transci.journal.informs.org/cgi/content/abstract/20/1/1

Furth P (1987) Short turning on transit routes. Transp Res Rec 1108:42–52. http://trid.trb.org/view.aspx?id=288816

Guihaire V, Hao J (2008) Transit network design and scheduling: a global review. Transp Res, Part A, Policy Pract 42(10):1251–1273. doi:10.1016/j.tra.2008.03.011

Jordan W, Turnquist M (1979) Zone scheduling of bus routes to improve service reliability. Transp Sci 13(3):242–268. doi:10.1287/trsc.13.3.242

Larrain H, Giesen R, Muñoz J (2010) Choosing the right express services for bus corridor with capacity restrictions. Trans Res Rec: J Trans Res Board 2197:63–70. doi:10.3141/2197-08

Leiva C, Muñoz J, Giesen R, Larrain H (2010) Design of limited-stop services for an urban bus corridor with capacity constraints. Transp Res, Part B, Methodol 44(10):1186–1201. doi:10.1016/j.trb.2010.01.003

Schwarcz S (2004) Service design for heavy demand corridors: limited-stop bus service. Master thesis, Massachusetts Institute of Technology

Scorcia H (2010) Design and evaluation of BRT and limited-stop services. Master thesis, Massachusetts Institute of Technology

Tétreault P, El-Geneidy A (2010) Estimating bus run times for new limited-stop service using archived AVL and APC data. Transp Res, Part A, Policy Pract 44(6):390–402. doi:10.1016/j.tra.2010.03.009

Transportation Research Board (2002) Estimating the benefits and costs of public transit projects: a guidebook for practitioners. TCRP Report 78

Transportation Research Board (2007) Bus rapid transit practitioner's guide. TCRP report 118

Ulusoy Y, Chien S, Wei C (2010) Optimal all-stop, short-turn, and express transit services under heterogeneous demand. Trans Res Rec: J Trans Res Board 2197:8–18. doi:10.3141/2197-02