Subline frequency setting for autonomous minibusses under demand uncertainty

K. Gkiotsalitis[a,1,*], M. Schmidt[b], E. van der Hurk[c]

[a] University of Twente, Dept. of Civil Engineering, 7500 AE Enschede, Netherlands
[b] Erasmus University Rotterdam, Dept. of Technology and Operations Management, Postbus 1738, 3000 DR Rotterdam, Netherlands
[c] Technical University of Denmark (DTU), Management Science, 2800 Kgs. Lyngby, Denmark

Abstract

Over the last years, there have been initiated several pilots with autonomous minibusses. Unlike regular bus services, autonomous minibusses serve a limited number of stops and have more flexible schedules since they do not require bus drivers. This allows the operation of a line through a flexible combination of *sublines*, where a subline serves a subset of consecutive stops in the same order as the original line. This paper studies the subline frequency setting (SFS) problem under uncertain passenger demand. We present a frequency setting model that assigns autonomous minibusses to sublines in order to exploit the available resources as much as possible and minimize the operational and passenger waiting time costs. Passenger waiting time costs may depend on the combination of several lines whose frequencies cannot be perfectly aligned for each passenger journey. We present a new estimation of the expected waiting time for passengers to improve the accuracy of the passenger waiting time costs in the case of sublines. Our SFS model is originally formulated as a MINLP and reformulated as a MILP that can be solved to global optimality. Further, we explicitly consider the uncertainty of passenger demand in the optimization process by formulating a stochastic optimization model. The performances of our stochastic and deterministic models that assign minibusses to sublines are tested under various passenger demand scenarios in the 14-stop autonomous minibus line in Eberbach, Germany and a fictional bus line with 20 bus stops. Results show potential improvements in operational costs in the range of 10-40% depending on the passenger demand profile.

Keywords: autonomous minibusses; vehicle scheduling; frequency setting; stochastic optimization; short-turning; demand uncertainty.

1. Introduction

Autonomous minibusses are gaining momentum as they are deployed in several pilots across Europe to offer last-mile solutions to travelers in urban areas. Recently,

*Corresponding author
Email addresses: k.gkiotsalitis@utwente.nl (K. Gkiotsalitis), schmidt2@rsm.nl (M. Schmidt), evdh@dtu.dk (E. van der Hurk)
[1] Tel.: +31 534 891 870

five autonomous minibus trials were launched in five European cities (Helsinki, Gjesdal, Tallinn, Lamia, and Helmond) under the EU project Fabulos (Fabulos, 2020). Autonomous minibusses have been operating in several EU trials in Frankfurt, Luxembourg, Lyon, Paris, Berlin under maximum speeds that can be up to 40 km/h (Muezner, 2018; Duss, 2018; Stein and Goebel, 2019; Modijefsky, 2019). They do not need a driver or steward on board as they are fully autonomous and they typically serve a small number of stops while providing first/last-mile services.

Tactical planning for autonomous minibusses follows to a large extent that of traditional bus lines: frequency setting, timetabling, vehicle scheduling and crew scheduling (Ceder and Wilson, 1986; Ceder, 2016). However, the last step of crew scheduling can be omitted. At the frequency settings stage, the frequency of each service line is planned considering the trade-off between the operational and the passenger-related costs (Yu et al., 2010; Szeto and Wu, 2011; Gkiotsalitis and Cats, 2018). This frequency provides also a first indication of the number of resources (vehicles) required to operate the service line (Ceder, 2011; Hassold and Ceder, 2014). The dispatching times of the assigned vehicles are determined at a subsequent step, known as timetable scheduling (Ceder, 2001; Gkiotsalitis and Alesiani, 2019).

This paper focuses on frequency setting for autonomous vehicle bus lines in the context of uncertain passenger demand and the use of *sublines*. A subline serves a specific line segment (i.e., a consecutive subset of stops of the original line), and can be obtained from the original line by performing a short-turning. Thus, sublines can provide a higher or equivalent passenger service level at lower operating costs in case of heterogeneous demand among the line. The Subline Frequency Setting problem (SFS) that is presented in this study strives to minimize the operator-related costs that include the vehicle fleet size and the vehicle running times, as well as the passenger-related costs through the assignment of optimal frequencies to all possible sublines. Our model includes a novel estimate for passenger waiting time given that multiple sublines may serve a single origin-destination pair. To evaluate the impact of uncertainty in passenger demand, we introduce a stochastic optimization SFS model and compare results under different demand profiles for a 14-stop autonomous minibus line in Eberbach, Germany and a fictional autonomous minibus line with 20 stops.

The main technical contributions of our work to the state-of-the-art are: (a) the development of a mixed-integer linear programming model for the autonomous minibus SFS problem that exploits more efficiently the available resources by placing more vehicles at line segments with higher demand, (b) the introduction of a new estimation formula for the expected passenger waiting times when several sublines serve the same stops and their frequencies cannot be perfectly aligned, and (c) the incorporation of the passenger demand uncertainties in the problem formulation with the development of a stochastic optimization model for the planning of autonomous minibusses.

The remainder of this paper is structured as follows: in section 2 we review past bus frequency setting problems that allocate the available vehicle resources to bus lines or sublines. In section 3, we introduce our SFS model. In this section, we formulate the SFS as a mixed-integer linear program (MILP) that has favorable properties when incorporating the passenger demand uncertainty in the problem formulation. This advantageous MILP formulation enables us to develop a stochastic formulation of the SFS in section 4. Our case study is detailed in section 5 where we test the performance of our deterministic and stochastic optimization solutions under different demand scenarios in a simulation study of the 14-stop autonomous minibus line in Eberbach, Germany. In section 6 we test further the performance of our deterministic and stochastic optimization solutions in a fictional, regular-sized bus line with 20 bus stops. Finally, section 7 provides the concluding remarks of our study and discusses future research directions.

2. Literature review on setting frequencies to sublines

2.1. Past studies

Frequency setting models determine the required number of trips to optimally operate a service line and the required number of vehicles to operate those trips (Ibarra-Rojas et al., 2015). Ceder (1984) proposed closed-form expressions that do not need to solve complex mathematical programs when determining the frequency of a single line. Namely, in many practical applications the frequency of a bus line is set based on *policy headways* or the *maximum loading point* (Ceder, 2016). Policy headways determine a lower bound of the line frequency and are used by operators that operate low-frequency services in suburban areas. The maximum load point method determines the frequency of a line based on the ratio of the number of passengers on board at the peak-load point to the desired passenger load of the vehicle. The maximum load point method is widely used under heavier demand scenarios and its frequency is determined based on a simple closed-form expression $f_j = \frac{\max_{s \in \mathcal{S}} P_{sj}}{\Gamma_j C}$, where f_j is the determined frequency of the examined bus line for the planning period j, P_{sj} the average number of passengers (load) observed on-board when departing from stop $s \in \mathcal{S}$ in period j, c the vehicle capacity, and $0 < \Gamma_j \leq 1$ the preferred vehicle load factor during the planning period j.

Although the maximum load point method ensures that our service supply will satisfy the maximum observed passenger load across all stops in the planning period, this crude approach can result in excessive operational costs and low productivity (see Ceder (2001)). This can be particularly seen when the average observed passenger load at the bus stop with the highest peak is several times higher than the observed bus loads at all other stops (e.g., see Fig.1 where the planned frequency should be able to accommodate almost 140 passengers at the maximum load point of stop 5, whereas in all other stops the passenger load is less than 50).

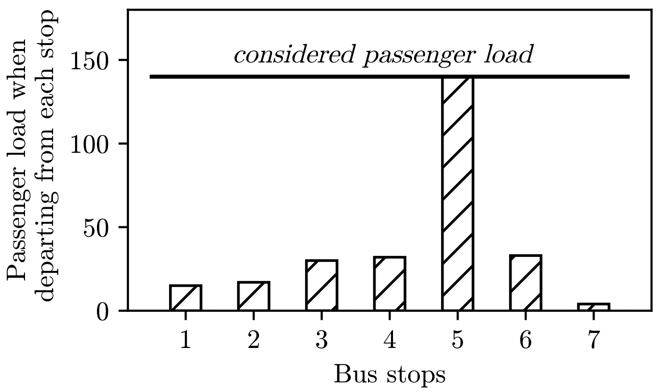

Figure 1: Example of passenger load at the maximum load point that determines the service frequency

Apart from closed-form expressions that determine the service frequency in a crude manner, there are several methods that try to find an optimal trade-off between passenger and operational-related costs (see Yu et al. (2010); dell'Olio et al. (2012); Cipriani et al. (2012); Cats and Glück (2019)). Pinto et al. (2020) proposed a joint design of multimodal transit networks and shared autonomous mobility fleet. They expanded the transit network design problem via incorporating the fleet size of shared-use autonomous vehicle mobility services as a decision variable allowing the removal of bus routes. Due to the problem's nonlinearity, they employed heuristic solution methods. Cepeda et al. (2006) proposed a frequency-based route choice model for congested transit networks which takes into account the consequences of congestion on the predicted flows as well as on the expected waiting and travel times.

Hadas and Shnaiderman (2012) used the stochastic properties of the collected data from automatic vehicle location (AVL) and automatic passenger counting (APC) systems to derive the optimal frequencies of service lines. The objective function of their optimization model aimed to minimize the empty-seat driven (unproductive cost) and the overload and unserved demand. Nikolić and Teodorović (2014) combined the network design with the frequency setting problem by determining the links and the bus frequency on each of the designed routes. To solve this problem, they employed the Bee Colony Optimization (BCO) metaheuristic. Arbex and da Cunha (2015) approached also the same problem with the use of a genetic algorithm. A new method for this problem was also proposed by Jha et al. (2019) that used multi-objective particle swarm optimization.

Verbas and Mahmassani (2013) proposed a nonlinear model for the optimal allocation of service frequencies to sublines that serve specific segments of an originally planned line. In a follow-up work, Verbas and Mahmassani (2015a) solved the vehicle allocation problem for the case of sublines that serve a subset of the stops of a line using the nonlinear solver KNITRO to find a locally optimal solution. Their objective was to assign vehicles to sublines in a more efficient way in order to improve ridership and waiting times. Later, Verbas and Mahmassani (2015b) proposed a nonlinear formulation to maximize wait time savings subject to budget, fleet, vehicle load, and policy headway constraints. The formulated program was also solved with KNITRO.

Bertsimas et al. (2020) developed nonlinear formulations for minimizing the waiting times in multimodal networks, while accounting for operator budget constraints, capacity constraints, and passenger preferences. Their proposed algorithms ran to near optimality and solved the joint frequency-setting and pricing optimization problem for public transit networks. Gkiotsalitis et al. (2019) solved the problem of allocating vehicles to sublines and interlining lines with the objective to improve the passenger waiting costs, the vehicle running costs and the depreciation costs when using more vehicles. Similar to the previous works, their nonlinear formulation did not allow the computation of a globally optimal solution resulting in the use of a genetic algorithm-based heuristic.

From the past literature, it is clear that there is an increasing number of works that address the subline frequency setting problem to utilize the available vehicles more efficiently. In Table 1 we summarize past works that consider sublines and interlining lines when setting service frequencies. It is important to note that in this study we distinguish sublines from interlining lines as follows: vehicles operating a subline serve a particular segment of a specific service line by performing a short-turning. In contrast, vehicles operating an interlining line serve segments of *more than one* service line.

Table 1: Research studies that consider sublines to allocate more vehicles to OD-pairs with higher demand

Study	Key performance indicators	Line flexibility	Demand uncertainty	Solution method
Delle Site and Filippi (1998)	Waiting times, running costs and personnel costs	Sublines: short-turning	Not considered	Locally optimal by splitting the problem into tractable subproblems
Cortés et al. (2011)	Waiting time, in-vehicle time, personnel costs and running costs	Sublines: short-turning and deadheading	Not considered	Locally optimal with applying an integrated deadheading-short-turning strategy
Verbas and Mahmassani (2015a)	Ridership and waiting time savings	Sublines: serve a subset of the entire stops of a route	Not considered	Locally optimal solution with KNITRO solver
Verbas and Mahmassani (2015b)	maximize wait time savings subject to budget, fleet, vehicle load, and policy headway constraints	Sublines: serve a subset of the entire stops of a route	Not considered	Locally optimal solution by solving an upper and a lower level problem with KNITRO
Gkiotsalitis et al. (2019)	Passenger waiting costs and vehicle running and depreciation costs	Sublines and interlining lines	Not considered	Locally optimal solution with Genetic Algorithm
This study	Waiting times, running costs and fleet size	Sublines: short-turning	Considered	Globally optimal with Gurobi solver (MILP formulation)

2.2. Contribution

One can observe from Table 1 that there is a number of works on frequency setting that consider sublines and/or interlining lines. However, none of them considers the uncertainty of passenger demand when determining the service frequencies of sublines. In addition, their nonlinear, non-convex model formulations do not allow to find globally optimal solutions resulting in the employment of heuristics that compromise the solution quality and do not offer theoretical guarantees of convergence. Given this research gap, the contributions of our work are as follows:

1. we first propose a MILP formulation for the SFS problem that can be solved to global optimality.

2. we introduce a new estimation formula for the expected waiting times of passengers

when several sublines serve the same stops and their frequencies cannot be perfectly aligned to every passenger journey.

3. we consider the passenger demand uncertainty in the planning stage by introducing a stochastic optimization model for the SFS problem.

4. we test the performance of our approach in a 14-stop minibus pilot and a fictional regular-sized bus line.

3. Problem definition and proposed Subline Frequency Setting Model

In this section we explain the assumptions we make on minibus operations and passenger behavior in order to define the SFS problem which answers the questions:

- which sublines should we establish?
- at which frequencies should the established sublines operate?

We first model the problem as a mixed-integer (non-linear) program (MINLP) and then reformulate it as a mixed-integer linear program (MILP).

3.1. Operations

We consider the frequency setting problem for one *original* line and a number of generated sublines that serve segments of the original line. We assume that the considered original line is symmetric and bi-directional, as this is currently the most typical structure of autonomous minibus lines operating in several cities (e.g., Frankfurt, Lyon, Luxembourg, Berlin, Stockholm).

The original line is characterized as a sequence of *physical stops*, which are visited in both directions. That is, a trip of the original line starts from the depot and visits all physical stops in the predefined sequence. For convenience of notation, in the remainder of this paper we associate two stops to each physical stop, one for each visiting direction. For instance, for a line with four *physical stops* (the first one denoting the depot), we refer to eight *stops* indexed from 1 to 8, with stops 1, 2, 3, and 4 referring to the four physical stops in direction from the depot, and 5, 6, 7, 8 being the stops in direction towards the depot. This is illustrated in Figure 2. We denote the ordered set of stops as $\mathcal{S}$.

Besides the original line, we consider a number of sublines. We assume that vehicles cannot park at intermediate stops between services, as these do not have the necessary parking infrastructure. Therefore, we require that all sublines start and end at one of the two terminals, where the first terminal is the depot (stop 1 in Figure 2) and the second terminal is the stop towards the opposite direction (stop 5 in Figure 2).

We obtain sublines by short-turning vehicles at intermediate stops. For instance, a subline in Figure 2 that starts from stop 1 and performs a short-turn at stop 3 will serve stops 1-2-3-6-7-8. Similarly, starting from the terminal at stop 5 and performing a short-turn at stop 6 will result in a subline serving stops 5-6-3-4. It becomes evident that the number of generated sublines starting at the same terminal is equal to the number of stops that can be used for short-turning. That is, in Figure 2 we have 4 sublines if we use all intermediate stops for short-turning. In the remainder of this paper, we use $\mathcal{R}$ to indicate the set of all potential lines, where 1 is the original line that serves all stops and $\langle 2, ...r, ... \rangle$ are the sublines.

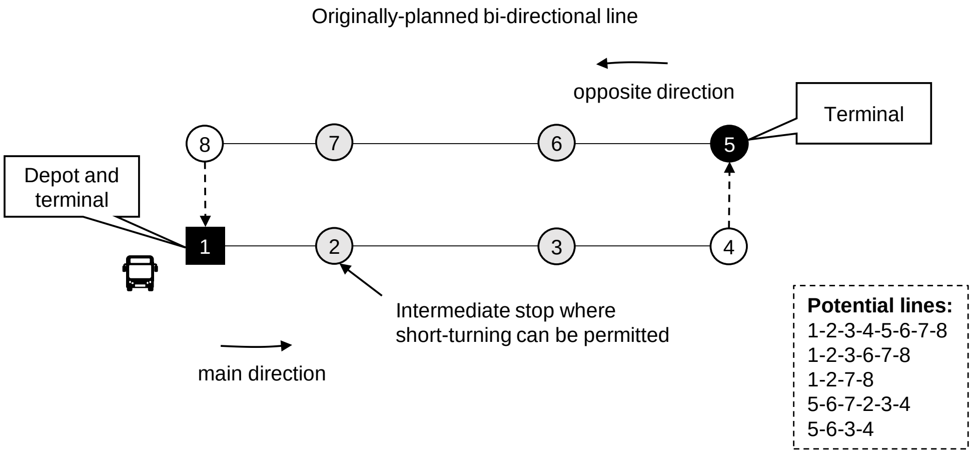

Figure 2: Generation of sublines from an originally planned bi-directional line.

Note that our SFS model will determine which sublines are deemed operational by considering their contribution to the reduction of passenger waiting times and operational costs. That is, we may not need to operate all eligible sublines but only some of them.

We first determine the *round-trip time* T_r of the original line $r = 1$ and each subline $r \in \mathcal{R} \setminus \{1\}$ assuming deterministic driving times between the stops and a fixed stopping time at each stop to let passengers board and alight. We assume that the minibusses operate according to a periodic schedule, where each potential line r has a fixed frequency, f_r, per period P. This fixed frequency f_r needs to be determined by our SFS model. For operational reasons, we impose a lower bound F on the frequency of the sublines. That is, subline $r \in \mathcal{R} \setminus \{1\}$ is either operated with a frequency of $f_r \geq F$, or it is not operated at all.

To ensure a minimum service quality for our passengers, for each OD-pair $(s, y) \in \mathcal{O}$ we require that the *service frequency for* (s, y) (that is, the number of departures from s of all possible lines that visit y during period P), f_{sy}, is equal to or higher than a *minimum allowed service frequency* Θ. I.e., if we let $\mathcal{R}_{(s,y)}$ denote the set of all potential lines that visit stops s and y, then $f_{sy} := \sum_{r \in \mathcal{R}_{(s,y)}} f_r \geq \Theta$.

We consider a limited number of available minibusses N. Not all minibusses need to be operated because there is a cost involved when deploying a minibus. Each minibus has a seating capacity of c and there is no bus driver. It is not allowed to transport standing passengers in the autonomous minibus, i.e, c is the maximum number of passengers that a minibus can transport. We also assign each vehicle exclusively to one of the possible lines. That is, a vehicle is not allowed to serve multiple sublines because it serves a specific subline under a fixed frequency. In addition, we require that at least K minibusses are assigned to the original line, $r = 1$, to ensure that the original line remains operational.

With x_r denoting the number of minibusses on a potential (sub)line r operated per period P, we consider costs related to whether a minibus is used at all, $W_1 \sum_{r \in \mathcal{R}} x_r$, and costs per time unit driven $W_2 \sum_{r \in \mathcal{R}} T_r T f_r$, where W_1 and W_2 are scaling parameters. The cost $W_1 \sum_{r \in \mathcal{R}} x_r$ is used to penalize the assignment of additional minibusses since there is a cost involved with the deployment of a minibus (for example an opportunity cost, as this minibus could have been used somewhere else in the network).

3.2. Assumptions on passenger behavior

For the base model that we formulate in this section, we assume that we are provided with the origin-destination pairs $\mathcal{O}$ and the cumulative passenger demand B_{sy} for $(s, y) \in \mathcal{O}$ for the whole planning horizon T. The planning horizon T should be selected such that the demand does not significantly change over its duration, i.e., we do not consider peak and off-peak demand within a specific planning horizon. Later, in Section 4, the deterministic passenger demand B_{sy} for $(s, y) \in \mathcal{O}$ will be replaced by stochastic demand reflecting the passenger demand uncertainty that might be observed across different days.

We also assume that passengers arrive randomly at their origin stop, as common in high-frequency services. The reason for this assumption is that recent studies have shown that passengers do not coordinate their arrivals at stops with the arrival times of buses in high-frequency services, and thus their average waiting time is half the headway (see Welding (1957); Hickman (2001); Bartholdi III and Eisenstein (2012); Cats (2014)).

Demand values represent the demand for traveling with the minibusses that serve the origin and destination stops of the passengers, i.e., we do not consider elastic demand/mode choice in our model. Finally, we assume that passengers choose the next minibus that departs from their origin stop and brings them to their destination, irrespective of the subline that this minibus might be serving. Ergo, the expected waiting time does not depend on the headways between minibusses of the same subline only, but on the headways between all *relevant* minibus departures for the passengers that can bring them to their destination. This is elaborated in section 3.3.

3.3. Estimating passenger waiting time

Different from the situation where the frequency of just the original line is determined, when operating several sublines we cannot expect that the departures relevant for a certain OD-pair will be perfectly synchronized with each other. This is illustrated in the following example: consider the situation depicted in Figure 2, where we have eight stations (four in each direction) and five potential lines (including the original line). Assume that the period length P is one hour and that the three potential lines that start from the depot operate once per hour. Then, for passengers from station 1 to station 2, it would minimize their waiting time to schedule regular departures, i.e., have a minibus depart every 20 minutes, leading to an expected waiting time of $20/2 = 10$ minutes. However, with this schedule, passengers from station 1 to station 3 would experience a gap in their schedule. This would lead to an expected waiting time of $\frac{1}{3} \cdot \frac{20}{2} + \frac{2}{3} \cdot \frac{40}{2} = 16.67$ minutes. For these passengers, it would be better if the two minibusses going until station 2 or beyond are scheduled with a headway of 30 minutes. This would lead to an expected waiting time of 15 minutes for the passengers going from station 1 to station 3. In that case though, the waiting time for the passengers from station 1 to station 2 would increase to at least $\frac{1}{4} \cdot \frac{15}{2} + \frac{1}{4} \cdot \frac{15}{2} + \frac{1}{2} \cdot \frac{30}{2} = 11.25$ minutes. This is summarized in Table 2.

	lower bound ($\frac{P}{2f_{sy}}$)	optimized for 1 to 2	optimized for 2 to 3	upper bound $\frac{P}{f_{sy}+1}$
From 1 to 2	10	10	11.25	15
From 2 to 3	15	16.67	15	20

Table 2: Expected waiting times in example

In general, if we let f_{sy} denote the service frequency for OD-pair (s, y), i.e., the number of relevant departures for OD-pair (s, y) per time period P, the expected waiting time will lie somewhere between:

- $\frac{P}{2f_{sy}}$ (if the relevant departures are perfectly synchronized)
- and $\frac{P}{2}$ (if all relevant departures take place at the same moment in time).

We refer to f_{sy} as the *service frequency* of OD-pair (s, y).

In our SFS model, we use the value $\frac{P}{f_{sy}+1}$ to estimate the waiting time of OD-pair (s, y). That is, we express the total waiting time as

$$\sum_{(s,y)\in\mathcal{O}} B_{sy}\frac{P}{f_{sy}+1}. \tag{1}$$

The value $\frac{P}{f_{sy}+1}$ is the expected waiting time between the arrival of a passenger of OD-pair (s, y) until departure of the next vehicle that serves OD-pair (s, y) under the assumption that vehicle departures are scheduled independently and randomly, assuming equal probability for each departure moment of a vehicle. In that sense, $\frac{P}{f_{sy}+1}$ constitutes a *lower* bound on the expected waiting time of a passenger with f_{sy} travel options within the period, as it can be seen in our Theorem B.1. in Appendix B. Once a set of sublines and their respective frequencies are known, these can be scheduled in a subsequent timetabling step so that the actual expected waiting times of passengers will be lower than $\frac{P}{f_{sy}+1}$.

3.4. Objective function

In the objective function of our model we strive to establish a trade-off between the reduction of (i) operational-related costs emerging from the use of additional minibusses and vehicle running times as discussed in Section 3.1, and (ii) costs related to passenger waiting times estimated as discussed in Section 3.3. We stretch again that the passenger waiting times in Eq.(2) are overestimated, given that the actual expected waiting times of passengers will be lower than $\frac{P}{f_{sy}+1}$.

Using scaling parameters W_1 and W_2 to trade-off operational costs with waiting times, we obtain objective function (2). W_1 stands for the cost per minibus, W_2 is the cost per time unit driven.

$$z(x, f) := \underbrace{W_1 \sum_{r\in\mathcal{R}} x_r}_{\text{cost of operating the minibusses}} + \underbrace{W_2 \sum_{r\in\mathcal{R}} T_r T f_r}_{\text{vehicle running times cost}} + \underbrace{\sum_{(s,y)\in\mathcal{O}} B_{sy}\frac{P}{f_{sy}+1}}_{\text{passengers' waiting time}} \tag{2}$$

3.5. Proposed SFS mathematical programming model

Our SFS formulation contains three sets of variables related to the subline frequencies. Integer variable x_r specifies how many vehicles are assigned to a potential line $r \in \mathcal{R}$. Note that a subline $r \in \mathcal{R} \setminus \{1\}$ is not deemed operational if $x_r = 0$. Next, f_r represents the selected service frequency for potential line $r \in \mathcal{R}$. This frequency needs to be integer since we assume a periodic timetable that repeats itself for every period P. Finally, a_r is a binary variable that indicates whether subline $r \in \mathcal{R} \setminus \{1\}$ is operational or not. Our initial SFS problem formulation is provided below.

Variable f_{sy} represents the realized service frequency for OD-pair $(s, y) \in \mathcal{O}$, which serves as input for the estimation of the average travel time. Furthermore, to account for passengers in our model, we use the following variables: for each potential line r and stop s, $b_{r,s}$ represents the number of passengers that board r at s, $v_{r,s}$ represents the number of passenger that alight from r at s, and $l_{r,s}$ represents the in-vehicle passenger load of potential line r at stop s, that is, the number of passengers on board of r when departing from s.

We introduce a 0-1 parameter $\Delta_{r,sy}$ which takes the value 1 if subline r is capable of serving the OD-pair $(s, y) \in \mathcal{O}$, and 0 otherwise. Our first, MINLP subline frequency setting model (Q) reads as follows:

$$(Q) \;\; \min z(x, f) := \sum_{r \in \mathcal{R}} (x_r W_1 + W_2 T_r T f_r) + \sum_{(s,y) \in \mathcal{O}} B_{sy} \frac{P}{f_{sy} + 1} \tag{3}$$

subject to:

$$f_r \leq \frac{x_r}{T_r} \qquad \forall r \in \mathcal{R} \tag{4}$$

$$f_{sy} \leq \sum_{r \in \mathcal{R}} \Delta_{r,sy} f_r \qquad \forall (s, y) \in \mathcal{O} \tag{5}$$

$$f_{sy} \geq \Theta \qquad \forall (s, y) \in \mathcal{O} \tag{6}$$

$$x_r \leq a_r M \qquad \forall r \in \mathcal{R} \setminus \{1\} \tag{7}$$

$$x_r \geq a_r T_r F \qquad \forall r \in \mathcal{R} \setminus \{1\} \tag{8}$$

$$\sum_{r \in \mathcal{R}} x_r \leq N \tag{9}$$

$$x_1 \geq K \tag{10}$$

$$x_r \in \mathbb{Z}_{\geq 0} \qquad \forall r \in \mathcal{R} \tag{11}$$

$$f_r \in \mathcal{F} \qquad \forall r \in \mathcal{R} \tag{12}$$

$$a_r \in \{0, 1\} \qquad \forall r \in \mathcal{R} \setminus \{1\} \tag{13}$$

$$b_{r,s} = \sum_{y > s} B_{sy} \frac{f_r}{f_{sy}} \Delta_{r,sy} \qquad \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \setminus \{|S|\} \tag{14}$$

$$v_{r,y} = \sum_{s < y} B_{sy} \frac{f_r}{f_{sy}} \Delta_{r,sy} \qquad \forall r \in \mathcal{R}, \forall y \in \mathcal{S} \setminus \{1\} \tag{15}$$

$$l_{r,s} = l_{r,s-1} + b_{r,s} - v_{r,s} \qquad \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \setminus \{1\} \tag{16}$$

$$l_{r,1} = b_{r,1} \qquad \forall r \in \mathcal{R} \tag{17}$$

$$l_{r,s} \leq c f_r \qquad \forall r \in \mathcal{R}, s \in \mathcal{S} \tag{18}$$

The objective function (3) is a condensed version of (2). Constraint (4) ensures that the round-trip travel time of each potential line $r \in \mathcal{R}$, T_r, together with the number of its assigned vehicles, x_r, provides an upper bound on the subline frequency f_r, namely $f_r \leq \frac{x_r}{T_r}$. Constraint (5) sets the service frequency f_{sy} of each OD-pair $(s, y) \in \mathcal{O}$ to be no larger than the total frequency assigned to all sublines r that serve OD-pair (s, y).

Note that the 0-1 parameter $\Delta_{r,sy}$ allows us to only consider the minibusses assigned to sublines $r \in \mathcal{R}$ that serve the particular OD-pair (s, y). Because the original line is always operational, $\Delta_{1,sy} = 1$ for any OD-pair (s, y). Constraint (6) ensures that each OD-pair (s, y) is served at least with minimum frequency Θ, thus guaranteeing a minimum level of service. Constraint (7) uses a very big positive number M and enforces that when subline $r \in \mathcal{R} \setminus \{1\}$ is operational, that is $x_r > 0$, then a_r should be equal to one. Otherwise, $a_r = 0$. Constraint (8) states that every subline $r \in \mathcal{R} \setminus \{1\}$ should have at least a minimum frequency of F to be deemed operational. Constraint (9) is the fleet size constraint ensuring that no more vehicles are used than the available fleet N. Constraint (10) ensures that at least K minibusses will serve all stops $s \in \mathcal{S}$ by being assigned to the original line serving all stops, line $r = 1$. Constraint (11) restricts x_r to positive integer values, and constraint (12) restricts frequency f_r to take values from a discrete set of feasible frequencies $\mathcal{F}$, thus allowing to require a minimum frequency if the subline is selected for operation. Constraint (13) defines variable a_r as binary. Constraint (14) estimates the total number of passengers that board vehicles of potential line r at stop s, by splitting the passengers of each OD-pair (s, y) equally over all relevant potential lines for (s, y). In a similar way, constraint (15) estimates the number of alighting passengers per stop and potential line. Constraints (16)-(17) keep track of the in-vehicle load per stop and per potential line. Constaint (18) ensures that the capacity restrictions are met per subline.

Note that program (Q) is a mixed-integer nonlinear program (MINLP). It is mixed-integer because variables a_r are binary, variables x_r, f_r and f_{sy} are restricted to integer/discrete values. It is nonlinear because the objective function (3) as well as constraints (14)-(15) are fractional since they contain a division by one of the variables.

3.6. SFS reformulation to a MILP

Following the ideas presented in (Claessens et al., 1998) and (van der Hurk et al., 2016), we reformulate the MINLP program (Q) to a MILP.

We use again $\mathcal{F}$ as the discrete set of *acceptable* frequencies for the original line and the sublines. As sublines need to have frequencies of at least F if they are operated, and we have at most N vehicles at our disposition, it is sufficient to consider the finite set $\mathcal{F} := \left\{0, F, F+1, F+2, \dots, N \cdot \left\lfloor \frac{1}{\min_r T_r} \right\rfloor \right\}$. If certain frequencies are not desirable for design considerations, we can also further restrict this set.

Let $\zeta_{f,r}$ be a new binary variable, where $\zeta_{f,r} = 1$ if potential (sub)line r is operated with frequency $f \in \mathcal{F}$, and 0 otherwise. To ensure that exactly one line frequency per potential line is chosen, we require

$$\sum_{f \in \mathcal{F}} \zeta_{f,r} = 1 \qquad \forall r \in \mathcal{R} \tag{19}$$

Then, constraint (4) can be rewritten as:

$$\sum_{f \in \mathcal{F}} f \cdot \zeta_{f,r} \leq \frac{x_r}{T_r} \qquad \forall r \in \mathcal{R} \tag{20}$$

Similarly, let $u_{f,sy}$ be a binary decision variable, where $u_{f,sy} = 1$ if the OD-pair $(s, y) \in \mathcal{O}$ is served with frequency $f \in \tilde{\mathcal{F}}$, and 0 otherwise. That means, the number of vehicles (not necessarily of the same subline) that depart within a period from s and visit y is f. Note that if we restrict the set of subline frequencies $\mathcal{F}$ to contain only specific frequencies, our set $\tilde{\mathcal{F}}$ should allow for service frequencies between OD-pairs that arise from servicing one OD-pair with several lines. For the sake of simplicity, we use $\tilde{\mathcal{F}} := \{\max\{F,\Theta\}, \max\{F,\Theta\}+1, \max\{F,\Theta\}+2, ..., N \cdot \lceil \frac{1}{min_r T_r} \rceil\}$. To ensure that exactly one service frequency f per OD-pair is chosen, we require

$$\sum_{f\in\tilde{\mathcal{F}}} u_{f,sy} = 1 \qquad \forall (s,y) \in \mathcal{O}. \tag{21}$$

and rewrite constraints (5) as

$$\sum_{f\in\tilde{\mathcal{F}}} f \cdot u_{f,sy} \leq \sum_{r\in\mathcal{R}} \Delta_{r,sy} \sum_{f\in\mathcal{F}} f \cdot \zeta_{f,r} \qquad \forall (s,y) \in \mathcal{O} \tag{22}$$

Constraints (6) can be omitted as they are implicitly fulfilled by the definition of $\tilde{\mathcal{F}}$.

To linearize the objective function, we precompute the passenger waiting time cost $\frac{P}{f+1}$ that an OD-pair (s,y) would incur if it is served with frequency f, i.e., if $u_{f,sy} = 1$. We can then replace the third term of the objective function with

$$\sum_{(s,y)\in\mathcal{O}} B_{sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{1+f} u_{f,sy}.$$

Consequently, for any frequency $f \in \tilde{\mathcal{F}}$, we have $\frac{P}{f+1} u_{f,sy} = \frac{P}{f_{sy}+1}$ if we operate the OD-pair $(s,y) \in \mathcal{O}$ with that frequency, and $\frac{P}{f+1} u_{f,sy} = 0$ otherwise. Our objective function is reformulated as:

$$\tilde{z}(x,u,\zeta) := \sum_{r\in\mathcal{R}} \Big(x_r W_1 + W_2 T_r T \sum_{f\in\mathcal{F}} f \cdot \zeta_{f,r} \Big) + \sum_{(s,y)\in\mathcal{O}} B_{sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy} \tag{23}$$

To linearize constraints (14) and (15), we introduce a binary variable $h_{f_1,f_2,r,sy}$ which is equal to 1 when potential line $r \in \mathcal{R}$ operates with frequency $f_1 \in \mathcal{F}$ and the OD-pair $(s,y) \in \mathcal{O}$ is served by frequency $f_2 \in \tilde{\mathcal{F}}$.

We impose the constraints

$$\sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}\setminus\{0\}} h_{f_1,f_2,r,sy} = 1 \qquad \forall r \in \mathcal{R}, \forall (s,y) \in \mathcal{O} \tag{24}$$

$$2h_{f_1,f_2,r,sy} \leq \zeta_{f_1,r} + u_{f_2,sy} \qquad \forall f_1 \in \mathcal{F}, \forall f_2 \in \tilde{\mathcal{F}}, \forall r \in \mathcal{R}, \forall (s,y) \in \mathcal{O} \tag{25}$$

to ensure that for line r and OD-pair (s,y) we have a (unique) pair of frequencies f_1^*, f_2^* (constraint (24)) and to link the variables $h_{f_1,f_2,r,sy}$ to the frequency indicator variables $\zeta_{f_1,r}$ and $u_{f_2,sy}$: if, for some f_1^*, f_2^*, we have $\zeta_{f_1^*,r} = 1$ and $u_{f_2^*,sy} = 1$, then $h_{f_1^*,f_2^*,r,sy}$ is forced to be equal to 1 in order to satisfy constraint (24) given than $h_{f_1,f_2,r,sy} = 0$ for any other f_1, f_2 pair. The reason for this is that there is no other f_1, f_2 pair that results both in $\zeta_{f_1,r} = 1$ and $u_{f_2,sy} = 1$, and thus constraint (25) cannot be met if $h_{f_1,f_2,r,sy} \neq 0$.

Then, the quadratic equality constraints (14)-(15) that determined the values of $b_{r,s}$ and $v_{r,s}$ become:

$$b_{r,s} = \sum_{y>s} B_{sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{|S|\} \tag{26}$$

$$v_{r,y} = \sum_{s<y} B_{sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad \forall r\in\mathcal{R}, \forall y\in\mathcal{S}\setminus\{1\} \tag{27}$$

We summarize the changes made in the reformulated MILP $(\tilde{Q})$ that is presented below.

$$(\tilde{Q}) \quad \min \sum_{r\in\mathcal{R}} \Big(x_r W_1 + W_2 T_r T \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} \Big) + \sum_{(s,y)\in\mathcal{O}} B_{sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy} \tag{28}$$

$$\text{s.t.} \quad \sum_{f\in\mathcal{F}} \zeta_{f,r} = 1 \qquad \forall r\in\mathcal{R} \tag{29}$$

$$\sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} \le \frac{x_r}{T_r} \qquad \forall r\in\mathcal{R} \tag{30}$$

$$\sum_{f\in\tilde{\mathcal{F}}} u_{f,sy} = 1 \qquad \forall (s,y)\in\mathcal{O} \tag{31}$$

$$\sum_{f\in\tilde{\mathcal{F}}} f\cdot u_{f,sy} \le \sum_{r\in\mathcal{R}} \Delta_{r,sy} \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} \qquad \forall (s,y)\in\mathcal{O} \tag{32}$$

$$\sum_{r\in R} x_r \le N \tag{33}$$

$$x_1 \ge K \tag{34}$$

$$x_r \in \mathbb{Z}_{\ge 0} \qquad \forall r\in\mathcal{R} \tag{35}$$

$$\sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} h_{f_1,f_2,r,sy} = 1 \qquad \forall r\in\mathcal{R}, \forall (s,y)\in\mathcal{O} \tag{36}$$

$$2h_{f_1,f_2,r,sy} \le \zeta_{f_1,r} + u_{f_2,sy} \qquad \forall f_1\in\mathcal{F}, \forall f_2\in\tilde{\mathcal{F}}, \forall r\in\mathcal{R}, \forall (s,y)\in\mathcal{O} \tag{37}$$

$$l_{r,1} = b_{r,1} \qquad \forall r\in\mathcal{R} \tag{38}$$

$$l_{r,s} \le c \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} \qquad \forall r\in\mathcal{R}, s\in\mathcal{S} \tag{39}$$

$$l_{r,s} = l_{r,s-1} + b_{r,s} - v_{r,s} \qquad \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{1\} \tag{40}$$

$$v_{r,y} = \sum_{s<y} B_{sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad \forall r\in\mathcal{R}, \forall y\in\mathcal{S}\setminus\{1\} \tag{41}$$

$$b_{r,s} = \sum_{y>s} B_{sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{|S|\} \tag{42}$$

$$u_{f,sy} \in \{0,1\} \qquad \forall\ f\in\tilde{\mathcal{F}}, \forall (s,y)\in\mathcal{O} \tag{43}$$

$$\zeta_{f,r} \in \{0,1\} \qquad \forall\ f\in\mathcal{F}, \forall r\in\mathcal{R} \tag{44}$$

$$h_{f_1,f_2,r,sy} \in \{0,1\} \qquad \forall f_1\in\mathcal{F}, \forall f_2\in\tilde{\mathcal{F}}, \forall r\in\mathcal{R}, \forall (s,y)\in\mathcal{O} \tag{45}$$

Note that variable a_r and constraints (7) and (8) are not needed in this model, as we have explicitly limited the set $\mathcal{F}$ to contain only *acceptable* frequencies. For the reader's convenience, we have summarized the model's nomenclature in the Appendix (Table A.21).

This reformulation results in a MILP that guarantees global optimality and results in computational improvements over the MINLP program (Q) because its continuous relaxation is a linear program that can be solved in polynomial time by a deterministic Turing machine.

4. Assigning minibusses under passenger demand uncertainty

Most autonomous minibus pilots operate in dedicated lanes without mixed-traffic conditions and exhibit stable inter-station travel times. Nonetheless, the passenger demand might vary significantly in space and time introducing uncertainties when determining the number of vehicles assigned to potential lines. In the remainder of this section, we treat the passenger demand $B = \{B_{sy}\}$ as an uncertain parameter. We denote by $\tilde{z}(x, u, \zeta, B)$ the value of the objective function in dependence of the variables x, u, ζ and the uncertain demand B. One frequently-used approach to cope with parameter uncertainty is to search for a solution that optimizes the *expected* value of the objective function. In general, this requires knowledge of the *probability distributions* governing the uncertain parameters (in our case: the demand distribution). For our model, however, knowledge of the *expected* demand per OD-pair is sufficient to compute the solution minimizing the expectation of the objective function: due to the linearity of the expected value operator, and due to the fact that the uncertain demand variables only appear in the objective function, we have:

$$\begin{aligned}\mathbb{E}_B[\tilde{z}(x,u,\zeta,B)] :=& \mathbb{E}\left[\sum_{r\in\mathcal{R}}\left(x_r W_1 + W_2 T_r T \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r}\right) + \sum_{(s,y)\in\mathcal{O}} B_{sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy}\right] \\ =& \sum_{r\in\mathcal{R}}\left(x_r W_1 + W_2 T_r T \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r}\right) + \sum_{(s,y)\in\mathcal{O}} \mathbb{E}[B_{sy}] \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy}.\end{aligned} \tag{46}$$

In our experiments, we estimate $\mathbb{E}[B_{sy}]$ by the average observed demand $\bar{B}_{sy}$ for OD-pair (s, y) to compute the number of vehicles and the frequency assignment that minimizes the expected value of our objective function. That is, if $B_{i,sy}$ is a measurement (realization) of the passenger demand from s to y during one scenario (e.g., day of operations) $i \in \{1, 2, ..., I\}$, then

$$\sum_{(s,y)\in\mathcal{O}} \mathbb{E}[B_{sy}] \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy}$$

becomes:

$$\frac{1}{I}\sum_{i=1}^{I} \sum_{(s,y)\in\mathcal{O}} B_{i,sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy}$$

Considering the passenger demand realizations, $B_{i,sy}$, in the constraints of our optimization problem can result in infeasibilities or over-utilization of vehicles, especially when we require to serve the entirety of the passenger demand for every possible scenario (i.e., even for outlier scenarios with unexpectedly high passenger demand). For this reason, the passenger demand constraint that forces in-vehicle passenger loads to always be less than or equal to the capacity of the vehicles can be relaxed to allow a small number of unserved passengers during scenarios (days) with unexpectedly high passenger demand volumes. Considering this, our *stochastic* optimization model ($\tilde{P}$) that incorporates the realizations of the passenger demand, $B_{i,sy}$, is formulated as:

$$(\tilde{P}) \quad \min \ z(x,u,\zeta) := \sum_{r\in\mathcal{R}} \Big(x_r W_1 + W_2 T_r T \sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} \Big) + \frac{1}{I}\sum_{i=1}^{I} \sum_{(s,y)\in\mathcal{O}} B_{i,sy} \sum_{f\in\tilde{\mathcal{F}}} \frac{P}{f+1} u_{f,sy} \tag{47}$$

$$\text{s.t. Eqs. } (29) - (37) \tag{48}$$

$$b_{i,r,s} = \sum_{y>s} B_{i,sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad \forall i\in\{1,...,I\}, \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{|S|\} \tag{49}$$

$$v_{i,r,y} = \sum_{s<y} B_{i,sy}\Delta_{r,sy} \sum_{f_1\in\mathcal{F}} \sum_{f_2\in\tilde{\mathcal{F}}} \frac{f_1}{f_2} h_{f_1,f_2,r,sy} \qquad i\in\{1,...,I\}, \forall r\in\mathcal{R}, \forall y\in\mathcal{S}\setminus\{1\} \tag{50}$$

$$l_{i,r,s} = l_{i,r,s-1} + b_{i,r,s} - v_{i,r,s} \qquad \forall i\in\{1,...,I\}, \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{1\} \tag{51}$$

$$l_{i,r,1} = b_{i,r,1} \qquad \forall i\in\{1,...,I\}, \forall r\in\mathcal{R} \tag{52}$$

$$g_{i,r,s} + c\sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r} = l_{i,r,s} \qquad \forall i\in\{1,...,I\}, \forall r\in\mathcal{R}, \forall s\in\mathcal{S}\setminus\{|\mathcal{S}|\} \tag{53}$$

$$\sum_{i=1}^{I}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}-\{|\mathcal{S}|\}} \max(0, g_{i,r,s}) \leq p\sum_{i=1}^{I}\sum_{(s,y)\in\mathcal{O}} B_{i,sy} \tag{54}$$

$$u_{f,sy}\in\{0,1\} \qquad \forall f\in\tilde{\mathcal{F}},\ \forall (s,y)\in\mathcal{O} \tag{55}$$

$$\zeta_{f,r}\in\{0,1\} \qquad \forall f\in\mathcal{F},\ \forall r\in\mathcal{R} \tag{56}$$

$$h_{f_1,f_2,r,sy}\in\{0,1\} \qquad \forall f_1\in\mathcal{F}, \forall f_2\in\tilde{\mathcal{F}}, \forall r\in\mathcal{R}, \forall (s,y)\in\mathcal{O} \tag{57}$$

where constraints (49)-(54) differ from the constraints applied when solving the problem deterministically. In particular, constraints (49) and (50) determine the passenger boardings and alightings at each stop of potential line r for each passenger demand scenario i. Constraints (51) and (52) determine the in-vehicle passenger load at each stop of potential line r for each passenger demand scenario i. It is evident that if this passenger load $l_{i,r,s}$ is always lower than the vehicle capacity limit irrespective of the demand scenario i, then the provided capacity is sufficient. Because there might exist, however, some demand scenarios where the vehicle capacity is not sufficient, we introduce constraints (53)-(54) that include the newly introduced continuous variable $g_{i,r,s}$. The new variable $g_{i,r,s}$ is equal to $l_{i,r,s} - c\sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r}$ and it represents the difference between the in-vehicle load and the available capacity. If $l_{i,r,s} \geq c\sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r}$, then $g_{i,r,s}\geq 0$ and it represents the number of unserved passengers at demand scenario i for line r at stop s. When $l_{i,r,s} \leq c\sum_{f\in\mathcal{F}} f\cdot\zeta_{f,r}$, then $g_{i,r,s}\leq 0$ which represents the empty space of line r at stop s at demand scenario i. Clearly, when $g_{i,r,s}\leq 0$ there is still available space in the line and we do not have any unserved passengers.

As discussed, when $g_{i,r,s} \geq 0$ the allocated capacity for line r, $c\sum_{f\in\mathcal{F}} f \cdot \zeta_{f,r}$, is lower than the in-vehicle passenger load at stop s for a demand scenario i, and we have unserved passengers at that stop. To reduce the number of unserved passengers, we only allow a small percentage p (%) of unserved passengers. Given that $\sum_{i=1}^{I} \sum_{(s,y)\in\mathcal{O}} B_{i,sy}$ are all the passengers across all demand scenarios $i = \{1, 2, ..., I\}$, we allow up to $p\sum_{i=1}^{I} \sum_{(s,y)\in\mathcal{O}} B_{i,sy}$ unserved passengers. This is achieved by constraint (54). Note that if $p = 0\%$, we would like each subline to be able to serve all passengers at all stops for every demand scenario i. However, this might result in infeasibilities for demand scenarios that are extreme outliers. Constraint (54) is nonlinear because it includes the $\max(0, g_{i,r,s})$ term which is equal to the number of unserved passengers at scenario i for line r at stop s. This constraint can be linearized by replacing it with constraints (58)-(63) where $\sigma_{i,r,s}$ is a newly introduced continuous variable representing the unserved passengers and $y_{i,r,s}$ a newly introduced binary variable which indicates whether there are unserved passengers at demand scenario i for line r at stop s.

$$\sum_{i=1}^{I} \sum_{r\in\mathcal{R}} \sum_{s\in\mathcal{S}-\{|\mathcal{S}|\}} \sigma_{i,r,s} \leq p\sum_{i=1}^{I} \sum_{(s,y)\in\mathcal{O}} B_{i,sy} \tag{58}$$

$$\sigma_{i,r,s} \geq 0 \qquad \forall i \in \{1, ..., I\}, \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \tag{59}$$

$$\sigma_{i,r,s} \geq g_{i,r,s} \qquad \forall i \in \{1, ..., I\}, \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \tag{60}$$

$$\sigma_{i,r,s} \leq My_{i,r,s} \qquad \forall i \in \{1, ..., I\}, \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \tag{61}$$

$$\sigma_{i,r,s} \leq g_{i,r,s} + M(1 - y_{i,r,s}) \qquad \forall i \in \{1, ..., I\}, \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \tag{62}$$

$$\sigma_{i,r,s} \in \mathbb{R}, \quad y_{i,r,s} \in \{0, 1\} \qquad \forall i \in \{1, ..., I\}, \forall r \in \mathcal{R}, \forall s \in \mathcal{S} \tag{63}$$

Constraints (58)-(63) linearize constraint (54) because they force $\sigma_{i,r,s}$ to be equal to $\max(0, g_{i,r,s})$ for any $i \in \{1, ..., I\}, r \in \mathcal{R}, s \in \mathcal{S}$. In more detail, when we have unserved passengers ($g_{i,r,s} \geq 0$), constraints (59)-(63) will force $y_{i,r,s}$ to be equal to 1 and $\sigma_{i,r,s}$ to be equal to $g_{i,r,s}$. When, however, the capacity of the operating vehicles of the line is sufficient ($g_{i,r,s} \leq 0$), then constraints (59)-(63) will force $y_{i,r,s}$ to be equal to 1 and $\sigma_{i,r,s}$ to be equal to 0.

Remark: We should note that constraints (58)-(63) make program $(\tilde{P})$ less compact and increase the complexity of the optimization problem because they introduce multiple variables with $I \times |\mathcal{R}| \times |\mathcal{S}|$ elements and multiple additional integrality constraints. A less complex formulation, that does not consider the total number of unserved passengers, is a formulation that does not allow the in-vehicle load of a (sub)line to exceed a pre-defined limit at any stop. This would just require to use constraints:

$$l_{i,r,s} \leq p'c\sum_{f\in F} f \cdot \zeta_{f,r} \qquad \forall i \in \{1, ..., I\}, r \in \mathcal{R}, s \in \mathcal{S} \tag{64}$$

where $p' = 1$ if we request to serve all passengers at all demand scenarios and $p' > 1$ if we allow for a small number of unserved passengers at every stop. Replacing constraints

(58)-(63) by (64) will result in a more compact and less computationally complex model, but it will not enforce an upper limit to the total number of unserved passengers.

5. Case study: 14-stop autonomous minibus line in Eberbach, Germany

5.1. Case study description

Our case study is a bi-directional autonomous minibus line operating in Eberbach, Germany. The line's length is 750 m (1500 m when performing a round trip). The minibus line has two terminals, one at the location of the depot (stop 1) and one at the location of the change of direction (stop 8). This line has 7 stops in each direction, which are indexed as $\mathcal{S} = \langle 1, 2, ..., 14 \rangle$ in a sequential order, starting from stop 1. Note that each physical stop has two indexes. One when the direction of the trip is from stop Restaurant & Hortus Ludi to stop Parkplatz, and one when the direction is from Parkplatz to Restaurant & Hortus Ludi. That is, the physical stop Restaurant & Hortus Ludi has index 1 for trips operating in the direction Restaurant & Hortus Ludi → Parkplatz and it also has index 14 for trips operating in the direction Parkplatz → Restaurant & Hortus Ludi. Figure 3 presents the topology of the 14-stop autonomous minibus line, the terminals (Parkplatz to Restaurant & Hortus Ludi and Parkplatz), and the inter-station travel times in minutes. Stops {1,2,3,4,5,6,7} correspond to trips that operate in the direction Restaurant & Hortus Ludi → Parkplatz and stops {8,9,10,11,12,13,14} correspond to trips that operate in the direction Parkplatz → Restaurant & Hortus Ludi.

In terms of size, this autonomous minibus line is a typical autonomous minibus line since most autonomous minibusses operating in European cities (e.g., Luxembourg, Lyon, Paris, Berlin, Frankfurt) serve less than 7 stops per direction.

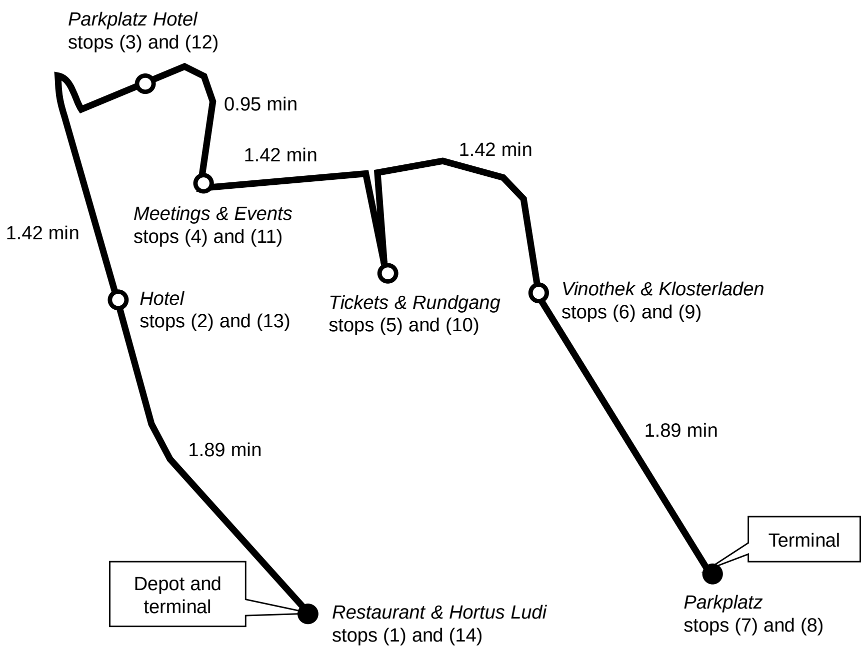

Figure 3: Topology of the 14-stop autonomous minibus line operating in Eberbach, Germany. Each one of the 7 physical stops has two indexes depending on the trip direction when visiting that stop

Given our two terminals and considering that we can use any intermediate stop to perform a short-turn, we can generate 10 sublines. That is, we have a total of 11 potential lines and we seek to find (a) which ones of them should be deemed operational, and (b) what would be the frequency for each operational line. The generated lines are provided in Table 3 together with their round-trip travel times.

Table 3: List of potential lines. Line $r = 1$ it the original line and lines 2,...,11 are sublines. Symbol $-$ indicates a change in line direction.

Line ID, r	Served stops of the line	Round-trip travel time, T_r (min)
1	1,...,7−8,...,14	17.98
2	1,...,6−9,...,14	14.2
3	1,...,5−10,...,14	11.36
4	1,...,4−11,...,14	8.52
5	1,...,3−12,...,14	6.62
6	1,2−13,14	3.78
7	8,...,13−2,...,7	14.2
8	8,...,12−3,...,7	11.36
9	8,...,11−4,...,7	9.46
10	8,...,10−5,...,7	6.62
11	8,9−6,7	3.78

Note that a line represented in Table 3 as 1,...,7−8,...,14 indicates a line that serves stops 1,2,3,4,5,6,7, changes direction, and then serves stops 8,9,10,11,12,13,14. For instance, line 6 serves stops 1,2, changes direction (short-turning), and then serves stops 13,14.

We assume that we have a total number of $N = 36$ minibusses. We choose a planning horizon of $T = 6$ h in which we assume homogeneous demand. Frequencies are expressed in vehicles per hour (that is, $P = 1$ h). We require that at least $K = 2$ minibusses are assigned to the original line. The type of the autonomous minibusses is Types Arma DL3 from Navya and their capacity is $c = 8$ passengers[2].

In this case study, a subline is deemed operational if it has a frequency of at least $F = 1$ minibus per hour. To attain periodic line schedules, we restrict the set of possible frequencies: each possible line $r \in \mathcal{R}$ can receive a frequency from the set $\mathcal{F} = \{0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 30, 60\}$, where each frequency is expressed in vehicles per hour. We assume $\Theta = 2$ trips/h as minimum allowed frequency to ensure a minimum level of service between any OD-pair $(s, y) \in \mathcal{O}$ with strictly positive non-zero demand. The scaling parameter related to the cost of operating an extra minibus is set to $W_1 = 3$, and the cost of a unit increase in the total running times $W_2 = 1.5$.

5.2. Passenger demand scenarios

The number of passengers willing to travel between any OD-pair s, y may vary significantly from day to day. In this section, we explain how we generate demand scenarios for our test cases.

We are specifically interested in the investigation of the effect of sublines and stochastic optimization in normal demand profiles and demand profiles skewed towards the center or

[2] https://www.probefahrt-zukunft.de/

the terminals of the line. Because of this, we consider the following four cases to sample from:

(1) Skewed demand profile to the left terminal (stops 1-4 and 11-14)

(2) Skewed demand profile to both terminals (stops 1-3 and 5-7 and 12-14 and 8-10)

(3) Skewed demand profile to the center (stops 3-5 and 10-12)

(4) Balanced demand, i.e., the expected demand is the same on each line segment and thus there is no peak on a particular segment of the line

The demand profiles in these four cases are presented schematically in Figure 4.

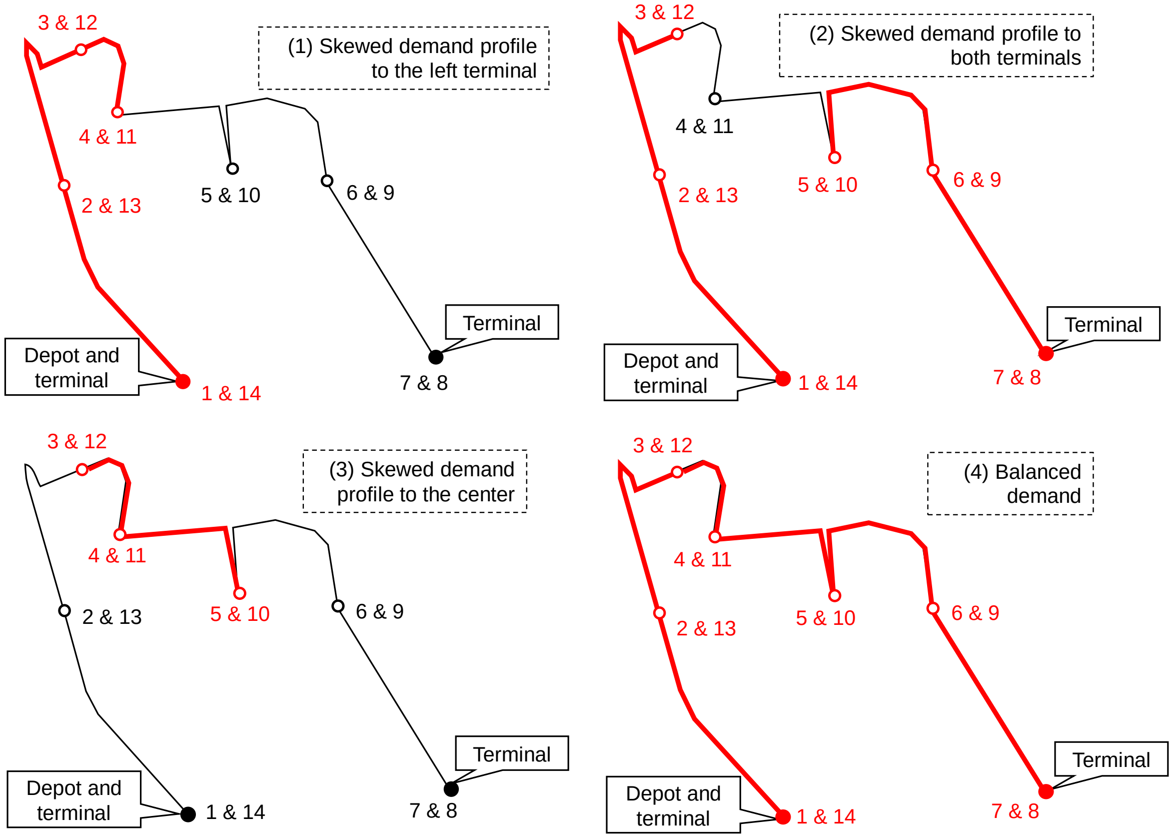

Figure 4: Passenger demand profile in each one of the four considered cases. Line segments in red have higher demand levels than segments in black. Each stop has two identification numbers, one for each direction.

For each one of these distributions, we draw two sets of samples to use in our experiments: one for the computation of the best subline network based on the stochastic models, and a second, independent, set of samples for the evaluation of the solutions proposed by the deterministic and stochastic models. Each sample contains 100 demand scenarios for each one of the four demand profiles presented in Figure 4.

5.3. *Model comparison*

We compare the solutions of the following models:

- the *deterministic no sublines model (DNS)*: this model uses the average passenger demand from the 100 sampled demand scenarios as input and computes the optimal frequency of the original line without considering sublines. This model computes optimal frequencies by solving the deterministic MILP described in $(\tilde{Q})$ after setting $x_r = 0$ for all sublines $r \in \{2, 3, ..., 11\}$

- the *deterministic sublines model (DWS):* this model also uses the average passenger demand from the 100 sampled demand scenarios and computes the optimal frequencies of all (sub)lines by solving the deterministic MILP $(\tilde{Q})$

- the *stochastic sublines model (SWS):* this model uses the 100 sampled demand scenarios as input and determines the line frequencies when requesting to satisfy at least a percentage p of the overall passenger demand when solving the model $(\tilde{P})$

We note that in the SWS we request to find a solution that results in less than at most $p = 1\%$ of unserved passengers when implemented in the 100 sampled demand scenarios. That is, the solution of the SWS is requested to satisfy at least 99% of the overall demand from the 100 sampled demand scenarios. This choice is made because, after systematic testing, we observed that for all demand profiles considered in this case study the solution of the DWS that satisfies all passenger demand on the average case is also capable of satisfying the passenger demand of more than 98% of the sampled demand scenarios. That is, the DWS offers already good solutions that perform well under passenger demand variations and the SWS explores more conservative solutions that will result in less than 1% unsatisfied passengers in the expense of using more resources (minibusses).

The deterministic and stochastic models are implemented in Python 3.8 and solved using the optimization solver Gurobi 9.1.2 that employs branch-and-bound and dual simplex as a solution method to solve MILP problems. The experiments are conducted on a cloud computing service (Microsoft Azure - F2s v2) with 2 CPUs and 4096 MB RAM. To enhance reproducibility, the demand data used in this case study and the software code are publicly released on GitHub (2021).

5.4. *Numerical experiments*

5.4.1. *Case 1: skewed demand profile to the left terminal*

We first start with the case of the skewed demand profile to the left terminal (stops 1-4 and 11-14) presented in Figure 4. Table 4 presents the number of model variables (column 2), constraints (column 3), the gap between the incumbent upper and lower bound of B&B (column 4), the number of required simplex iterations for exploring the nodes of the B&B tree (column 5), and the computation times of solving the three models for this demand profile (column 6). We note that a gap of 0% means that a globally optimal solution is found because the incumbent solution of the MILP has the same performance as the solution of the best-performing linear relaxation from all of the current leaf nodes in the B& B tree. Note that solving the SWS requires considerably more computation time because $(\tilde{P})$:

- uses all 100 sampled demand scenarios as input in the optimization process resulting in an increased number of constraints and variables.

- has a considerably higher number of integral constraints due to its additional variables $\sigma_{i,r,s}$, $y_{i,r,s}$ and $g_{i,r,s}$ resulting in an extensive exploration of the B&B tree to find the globally optimal solution.

Table 4: Convergence and computation times

	compactness indicators				
model	constraints	integer variables	simplex iterations	gap	comp. time (s)
DNS	8 460	8 834	317	0%	0.4
DWS	91 760	91 294	172 147	0%	64
SWS	206 668	106 794	15 325 632	0%	22 031

The optimal number of vehicles assigned to each service line and the corresponding frequencies, as well as total running time and objective value for the three models, are presentend in Table 5.

Table 5: Optimal number of vehicles x_r and frequencies f for (sub)line r for the three models for the 100 sampled demand scenarios that correspond to the demand profile of case 1

	DNS	DWS	SWS		DNS	DWS	SWS
x_1	18	3	3	f_1	60	10	10
x_2	0	0	2	f_2	0	0	5
x_3	0	4	3	f_3	0	20	15
x_4	0	5	5	f_4	0	30	30
x_5	0	0	0	f_5	0	0	0
x_6	0	0	0	f_6	0	0	0
x_7	0	0	0	f_7	0	0	0
x_8	0	0	0	f_8	0	0	0
x_9	0	0	0	f_9	0	0	0
x_{10}	0	0	0	f_{10}	0	0	0
x_{11}	0	0	0	f_{11}	0	0	0
Total number of vehicles:	18	12	13				
Vehicle running times (h):	107.88	66.24	67.68				
Waiting time estimate (min):	0.98	1.49	1.41				
Objective function value:	233.08	161.23	163.80				

The DNS solution will result in a frequency of 60 trips per hour at each segment of the original service line. The solutions that consider sublines though, will result in higher frequencies at the segments closer the left terminal since the demand is skewed at this part of the service line. These optimal segment-level frequencies when using sublines are presented in Figure 5.

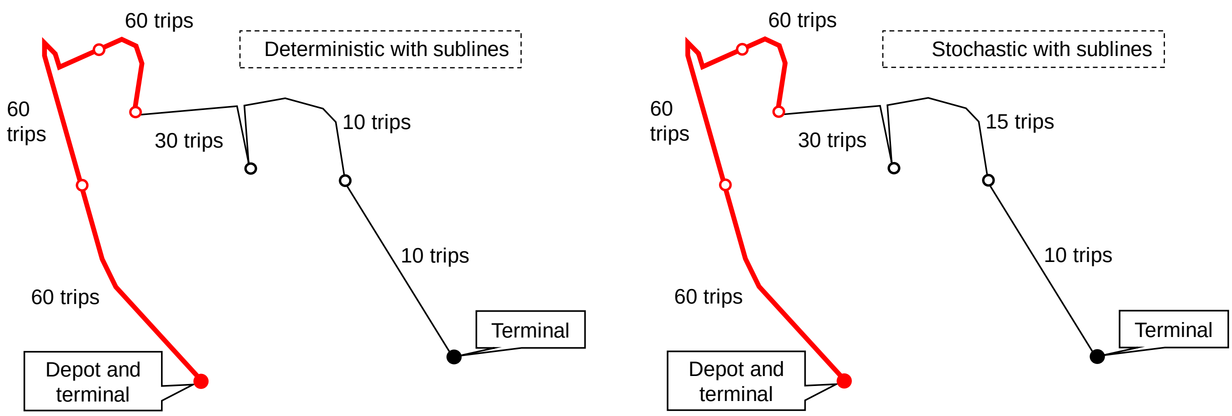

Figure 5: Frequencies in trips per hour at each line segment when implementing the DWS and SWS solutions for case 1 with skewed demand to the left terminal (depot)

From Table 5 one can note that the DNS solution performs significantly worse than the DWS and SWS solutions, which do consider sublines. In particular, the DNS solution requires to deploy 6 and 5 more vehicles, respectively. In addition, it has increased operational costs because its vehicles should run for a running time of 107.88 h within the 6-hour planning period T (the running time is calculated as $\sum_{r\in\mathcal{R}} T_r T f_r$).

As expected, the SWS solution results in slightly increased operational costs compared to the DWS solution. This is a result of the more conservative nature of the stochastic model that seeks to serve more than 99% of the overall passenger demand over all 100 sampled scenarios. Note, however, that the DWS solution already serves more than 98% of the overall passenger demand over all 100 sampled scenarios that the computation is based on. The results show that to achieve the additional 1% demand coverage of the SWS, we need to use one more vehicle, and vehicle running times slightly increase.

We now proceed to the evaluation of our three solutions. We use the same passenger demand profile in our sampling, but we generate 100 different (unseen) demand samples. We then use our already derived solutions and we perform 100 simulations to evaluate the performance of each one of the solutions in terms of unserved passenger demand. The results are presented in Table 6. In these simulations, we assign the new passenger demand from the 100 different samples to the service supply offered by the DNS, DWS and SWS solutions, respectively. If the assigned demand to vehicles exceeds the capacity, then the remaining passengers are considered to be unserved. We consider that unserved passengers leave the service line and do not wait for the next trip of this service line.

Table 6: Unserved passengers

solution	unserved passengers	% of the total demand
deterministic no sublines (DNS)	315	0.30%
deterministic with sublines (DWS)	1568	1.49%
stochastic with sublines (SWS)	1124	1.07%

In Table 7 we also present the waiting time estimate of all passengers in the 100 new demand scenarios. This waiting time estimate considers only the served passengers because we assume that the unserved passengers are leaving the system. For each one of the new 100 demand scenarios we compute the estimate of the total waiting time of all passengers. Column 1 presents the estimate of the total waiting time for the median demand scenario. Column 2 reports the standard deviation of the waiting time estimate from the 100 demand scenarios. Column 3 presents the estimate of the total waiting time of passengers for the best-case demand scenario of the 100 considered scenarios. Column 4 presents the waiting time estimate for the worst-case demand scenario.

Table 7: Estimate of the total waiting time of passengers in hours

solution	median	st dev	min	max
DNS	17.34	2.69	10.59	23.03
DWS	26.21	3.01	18.70	33.84
SWS	24.78	2.93	17.34	32.31

5.4.2. Case 2: skewed demand profile to both terminals

We now consider the case with the skewed demand profile to both terminals presented in Figure 4 (stops 1-3 and 5-7, and 12-14 and 8-10). The computation times of solving the three models for this case are presented in Table 8.

Table 8: Convergence and computation times

	compactness indicators				
model	constraints	integer variables	simplex iterations	gap	comp. time (s)
DNS	8 460	8 834	86	0%	1
DWS	91 760	91 294	132 027	0%	60
SWS	206 668	106 794	15 975 801	0%	23 029

The optimal number of vehicles assigned to each service line and the corresponding frequencies, as well as total running time and objective value for the three models, are presented in Table 9.

Table 9: Optimal number of vehicles x_r and frequencies f for (sub)line r for the three models for the 100 sampled demand scenarios that correspond to the demand profile of case 2

	DNS	DWS	SWS		DNS	DWS	SWS
x_1	9	5	9	f_1	30	10	30
x_2	0	0	0	f_2	0	0	0
x_3	0	0	0	f_3	0	0	0
x_4	0	0	0	f_4	0	0	0
x_5	0	2	4	f_5	0	15	30
x_6	0	0	0	f_6	0	0	0
x_7	0	0	0	f_7	0	0	0
x_8	0	0	0	f_8	0	0	0
x_9	0	0	0	f_9	0	0	0
x_{10}	0	2	4	f_{10}	0	15	30
x_{11}	0	0	0	f_{11}	0	0	0
Total number of vehicles:	9	9	17				
Vehicle running times (h):	53.94	46.80	93.60				
Waiting time estimate (min):	1.94	2.19	1.12				
Objective function value:	142.23	135.96	211.17				

The DNS solution results in a frequency of 30 trips per hour at each segment of the original service line. The segment-level frequencies for the DWS and SWS solutions are presented in Figure 6.

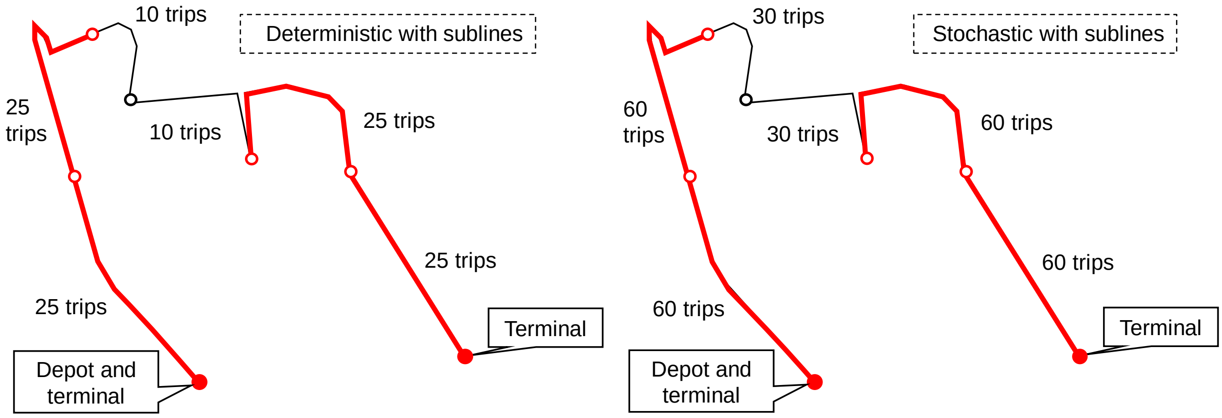

Figure 6: Frequencies in trips per hour at each line segment when implementing the DWS and SWS solutions for case 2 with skewed demand to both terminals

From Table 9 one can observe that the SWS solution uses the same number of vehicles as the SNS solution. However, four of these vehicles are used to serve only part of the network: two of them serve stations 1,2,3-12,13,14 and the other two stations 8,9,10-5,6,7. In this way, demand on the more frequented parts of the network can be covered more efficiently. This leads to a decrease of 13% in vehicle hours driven.

The SWS solution results in considerably increased operational costs compared to the solutions computed with the deterministic models. To achieve a service that serves more than 99% of the overall passenger demand across the 100 sampled scenarios, one would need to deploy 8 more minibusses and increase the vehicle running times to 93.6 h. We should note here, however, that the solution of the SWS model is very conservative in this case because, even if it allowed to satisfy only 99% of the overall passenger demand, the found solution satisfied 100% of it. Clearly the constraint of satisfying 99% of the overall demand was very restrictive in this particular case and a stochastic solution with improved running costs could have been derived if this limit was relaxed.

We now proceed to the evaluation of the three solutions. Using 100 different (unseen) demand samples, we perform 100 simulations to evaluate the performance of the solutions of the three models in terms of unserved passenger demand. The results are presented in Table 10. Notably, the solution of the SWS model is so conservative that satisfies all passenger demand even for the new demand samples. Because of the excessive supply, this solution results also in an average passenger waiting time estimate of only 1.12 min.

Table 10: Unserved passengers

solution	unserved passengers	% of the total demand
deterministic no sublines (DNS)	5035	4.70%
deterministic with sublines (DWS)	6775	6.32%
stochastic with sublines (SWS)	0	0.00%

The total passenger waiting times at each scenario are also computed. Table 11 provides the median, the standard deviation, the min and the max values of the total passenger waiting time estimates. Note that the excessive supply provided by the solution of the SWS model results in considerably lower waiting times compared to the DNS and DWS.

Table 11: Estimate of the total waiting time of passengers in hours

solution	median	st dev	min	max
DNS	34.59	4.27	26.06	48.32
DWS	38.98	4.57	26.58	52.86
SWS	19.71	2.24	15.71	26.94

5.4.3. Case 3: skewed demand profile to the center

We now consider the case with the skewed demand profile to the center (stops 3-5 and 10-12) presented in Figure 4. The convergence and computation times of solving the three models for this case are presented in Table 12.

Table 12: Convergence and computation times

	compactness indicators				
model	constraints	integer variables	simplex iterations	gap	comp. time (s)
DNS	8 460	8 834	86	0%	2
DWS	91 760	91 294	369 076	0%	137
SWS	206 668	106 794	15 525 524	0%	22 072

The optimal number of vehicles assigned to each service line and are presented in Table 13.

Table 13: Optimal number of vehicles x_r and frequencies f for (sub)line r for the three models for the 100 sampled demand scenarios that correspond to the demand profile of case 3

	DNS	DWS	SWS		DNS	DWS	SWS
x_1	9	6	2	f_1	30	20	6
x_2	0	0	0	f_2	0	0	0
x_3	0	0	6	f_3	0	0	30
x_4	0	0	0	f_4	0	0	0
x_5	0	0	0	f_5	0	0	0
x_6	0	0	0	f_6	0	0	0
x_7	0	0	1	f_7	0	0	4
x_8	0	2	4	f_8	0	10	20
x_9	0	0	0	f_9	0	0	0
x_{10}	0	0	0	f_{10}	0	0	0
x_{11}	0	0	0	f_{11}	0	0	0
Total number of vehicles:	9	8	13				
Vehicle running times (h):	54.00	47.34	73.19				
Waiting time estimate (min):	1.94	2.09	1.55				
Objective function value:	128.55	117.21	165.40				

The DNS solution results in a frequency of 30 trips per hour at each segment of the original service line. The segment-level frequencies for the DWS and SWS solutions are presented in Figure 7.

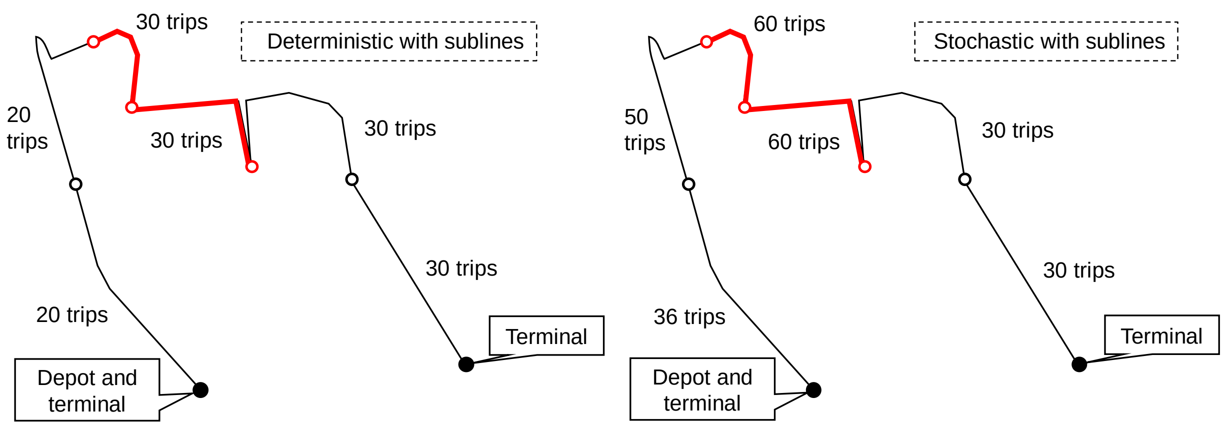

Figure 7: Frequencies in trips per hour at each line segment when implementing the DWS and SWS solutions for case 3 with skewed demand to the center

From Table 13 one can note that when the demand is skewed towards the center, there is still a slight benefit when using sublines. This benefit is not as significant as in the cases where the demand is skewed towards the terminals, but it still results in using one minibus less and reducing the vehicle running times by more than 6 hours (12% improvement).

The SWS solution that is designed to serve more than 99% of the overall passenger demand increases again the operational costs, both in terms of required vehicles and vehicle running times.

Using 100 different (unseen) demand samples, we perform 100 simulations to evaluate the performance of the solutions of the three models in terms of unserved passenger demand. The results are presented in Table 14.

Table 14: Unserved passengers

solution	unserved passengers	% of the total demand
deterministic no sublines (DNS)	4524	7.06%
deterministic with sublines (DWS)	5210	8.13%
stochastic with sublines (SWS)	132	0.21%

The total passenger waiting times at each scenario are also computed. Table 11 provides the median, the standard deviation, the min and the max values of the total passenger waiting time estimates. Note that the excessive supply provided by the solution of the SWS model results in lower waiting times compared to DNS and DWS.

Table 15: Estimate of the total waiting time of passengers in hours

solution	median	st dev	min	max
DNS	20.84	3.24	12.81	28.16
DWS	22.55	3.28	14.20	30.16
SWS	16.64	2.09	10.92	21.77

5.4.4. Case 4: balanced demand

We now consider the fourth and final case where the demand does not have a peak at specific line segments (see Figure 4). The computation times of solving the three models for this case are presented in Table 16.

Table 16: Convergence and computation times

	compactness indicators				
model	constraints	integer variables	simplex iterations	gap	comp. time (s)
DNS	8 460	8 834	3	0%	0.8
DWS	91 760	91 294	79 921	0%	41
SWS	206 668	106 794	10 432 412	0%	14 041

The optimal number of vehicles assigned to each service line are presented in Table 17.

Table 17: Optimal number of vehicles x_r and frequencies f for (sub)line r for the three models for the 100 sampled demand scenarios that correspond to the demand profile of case 4

	DNS	DWS	SWS		DNS	DWS	SWS
x_1	6	6	6	f_1	20	20	20
x_2	0	0	0	f_2	0	0	0
x_3	0	0	0	f_3	0	0	0
x_4	0	0	0	f_4	0	0	0
x_5	0	0	0	f_5	0	0	0
x_6	0	0	0	f_6	0	0	0
x_7	0	0	0	f_7	0	0	0
x_8	0	0	0	f_8	0	0	0
x_9	0	0	0	f_9	0	0	0
x_{10}	0	0	0	f_{10}	0	0	0
x_{11}	0	0	0	f_{11}	0	0	0
Total number of vehicles:	6	6	6				
Vehicle running times (h):	35.96	35.96	35.96				
Objective function value:	121.89	121.89	121.89				

As one might have expected, the models that consider sublines result in the same optimal solution as the DNS model that does not consider sublines since there is no demand peak at a specific line segment. All models assign 6 vehicles to the original line and no vehicles to sublines since the constant demand levels across all line segments do not require the use of sublines.

5.4.5. Conclusions on the case study

Based on the experiments undertaken in this section, we conclude that establishing sublines is particularly useful when demand is skewed. In the first case discussed in section 5.4.1, where passenger demand is skewed towards one terminal, we achieve up to 39% vehicle running time reductions when using sublines compared to using just one long line. This reduction in vehicle running times in our experiments decreases to 13% for the case where demand is skewed towards both terminals (section 5.4.2) and to 12% when it is skewed towards the center (section 5.4.3). Finally, when the demand is balanced (section 5.4.4), considering sublines does not bring any extra benefits.

Furthermore, it is interesting to see that solutions based on average demand already provided coverage for 98% of the passengers over the sampled demand scenarios. However, when we test the solutions on different (unseen) scenarios, we see that in cases 2 and 3 the percentage of unserved passengers is significantly higher for the deterministic DWS solution compared to the SWS solution (see Tables 10 and 14). There, we can see that the SWS solution leads to less unserved passengers - although the threshold of 99% is not met on the test scenarios either. More sample scenarios may be required in the computation of the solution to be able to fully meet this threshold. This is investigated in our next section where we use 500 sampled demand scenarios instead of 100 when solving $(\tilde{P})$.

The additional robustness against demand variations that we gain when using the SWS model, comes at the price of considerably more vehicles and increases significantly the vehicle running times. This can be taken into consideration by policy makers that might need to decide about the trade-off between offering sufficient capacity even at worst-case demand scenarios and reducing the operational costs.

6. Case study: 20-stop fictional line

6.1. Description

The longer a minibus line is, the more benefits may be gained by short-turning. Autonomous minibus lines are currently operating on relatively short lines: e.g., the line with 7 physical stops studied in Section 5 is the longest autonomous minibus pilot in Germany[3]. Notwithstanding this, there might be longer autonomous minibus lines with more stops in the near future. For this reason, in this section we conduct experiments using a fictional 20-stop line (10 stops per direction) that is presented in Figure 8. We consider a fixed inter-station travel time of 3 minutes between any pair of successive stops.

In Section 5 we observed that the biggest benefit of operating sublines occurs when the demand profile is skewed towards one of the two terminals. For this reason, we focus on such a demand profile in this fictional case study.

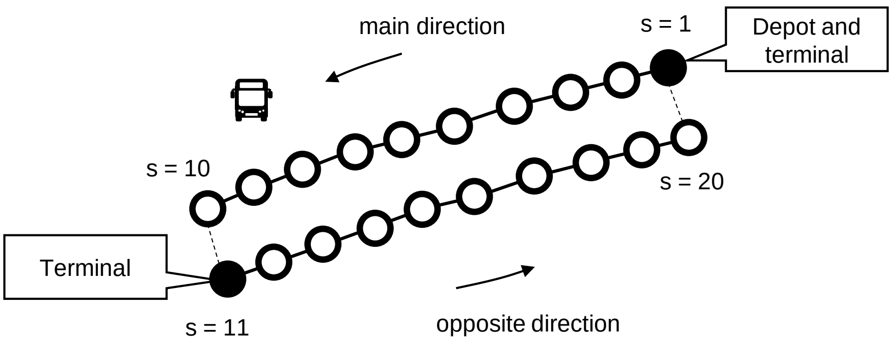

Figure 8: Line topology

To model demand, we draw 500 samples of passenger demand, which we model using independent uniform distributions for each OD-pair. Using Tukey's boxplot convention, we display the mean, interquartile range, minimum/maximum points, and the outliers of the sampled demand data in Figure 10.

The demand profile is schematically presented in Figure 9, where segments with higher demand are highlighted in red.

[3]https://www.probefahrt-zukunft.de/

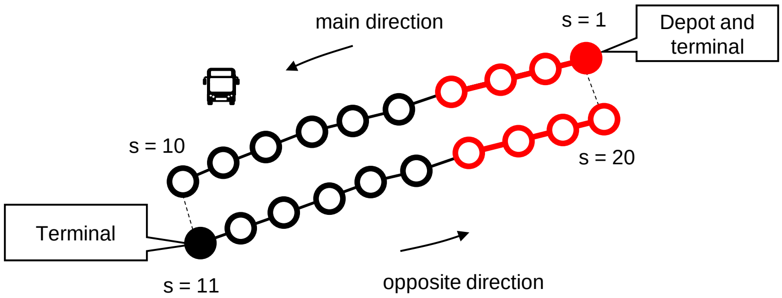

Figure 9: Considered passenger demand profile from which we draw the 500 passenger demand samples. Segments in red have higher demand levels

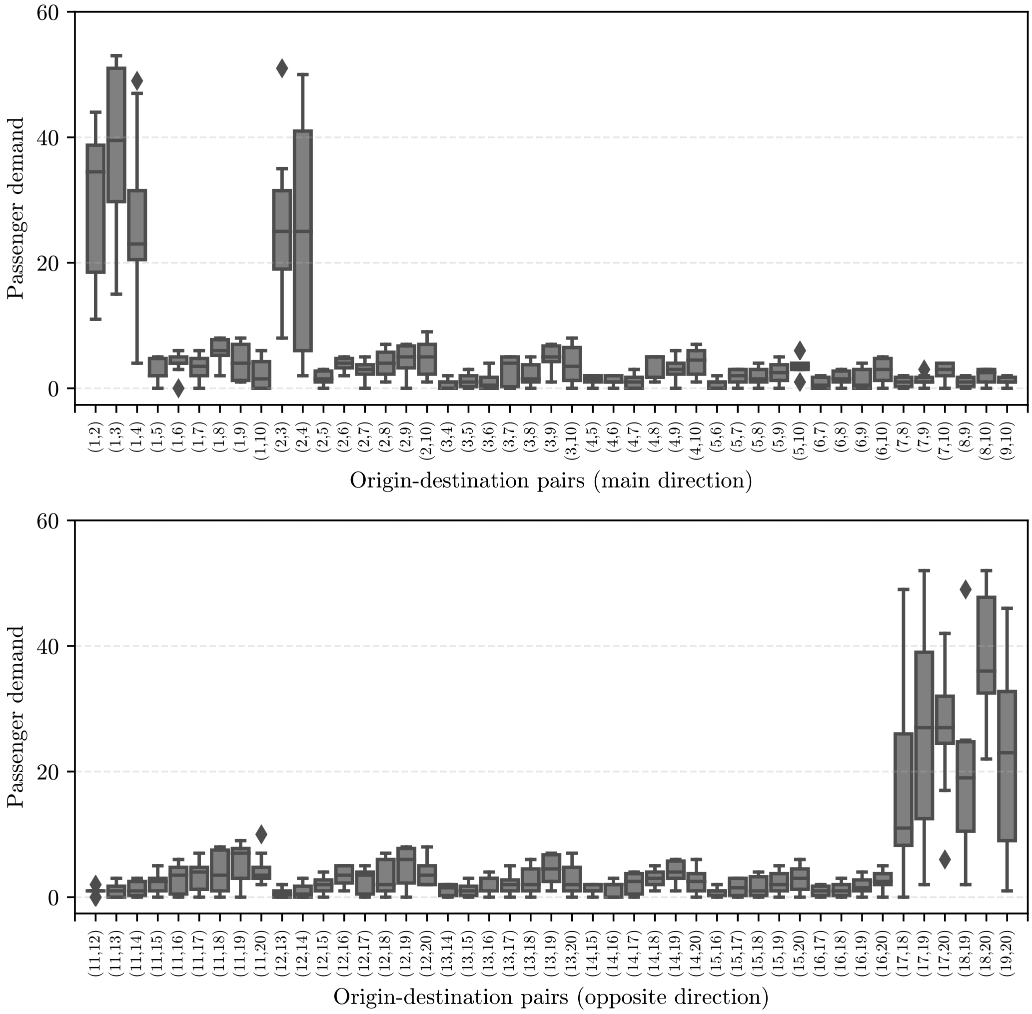

Figure 10: Tukey boxplot of the considered passenger demand data

Due to the skewed demand profile, we only consider sublines that are generated from the depot (sublines which start at stop 1 and end at stop 20) since it will not be beneficial to operate sublines starting from stop 11 and ending at stop 10. As a result, this bus line has 8 sublines (one for each intermediate stop in the direction from the depot). The original line serves stops $1, ..., 10 - 11, ..., 20$. Subline $r = 2$ serves stops $1, 2, ..., 9 - 12, 13, ..., 20$. Subline $r = 3$ serves stops $1, 2, ..., 8 - 13, 14, ..., 20$. This continues in a similar fashion until subline $r = 9$ that serves stops four stops: $1, 2 - 19, 20$. The average round-trip travel times of the potential lines are $(T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9) = (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)$ in hours.

The scaling parameter related to the cost of operating an extra minibus is set to $W_1 = 3$, and the cost of a marginal increase in the total running times $W_2 = 1.5$. There is a maximum fleet of $N = 36$ vehicles available. The planning period is $T = 6$ h. A subline is deemed operational if it has a frequency of at least $F = 1$ minibus per hour. The minimum allowed frequency to ensure a minimum level of service between any OD-pair $(s, y) \in \mathcal{O}$ with strictly positive non-zero demand is $\Theta = 2$ trips/h. To attain periodic line schedules, each line $r \in \mathcal{R}$ can receive a frequency from the set $\mathcal{F} = \{0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 30, 60\}$ where each frequency is expressed in vehicles per hour. The minimum number of minibusses that need to be assigned to the original line serving all stops stops is $K = 2$.

Using the above data, we consider the mean passenger demand values per OD-pair to compute the DNS and the DWS solutions. We compute two solutions with the SWS model, one that aims to satisfy 98% of the passenger demand across all 500 sampled scenarios and one that aims to satisfy 99% of it. The computation times of solving these models are presented in Table 18.

Table 18: Computation times

model	simplex iterations until convergence	computation time (s)
deterministic no sublines (DNS)	92	1.41
deterministic with sublines (DWS)	258 330	285.22
stochastic with sublines (SWS - 98%)	10 312 562	21 796.82
stochastic with sublines (SWS - 99%)	10 064 762	19 776.47

The solutions of the three models are presentend in Table 19.

Table 19: Assigned vehicles per subline for each solution for the 20-stop case

solution	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	Total number of vehicles	Vehicle running times (h)
DNS	27	0	0	0	0	0	0	0	0	27	146
DWS	9	4	0	0	0	2	3	0	0	18	101
98% SWS[a]	18	0	0	0	0	0	3	0	0	21	115
99% SWS[b]	18	0	0	0	5	0	0	0	0	23	127

[a]this solution satisfies $\geq$ 98% of the passenger demand in the 500 samples
[b]this solution satisfies $\geq$ 99% of the passenger demand in the 500 samples

A first interesting finding is that the DWS solution requires 9 less vehicles and 45 less vehicle running hours compared to the DNS solution. This improvement is even greater than the operational cost improvement in the 14-stop line in Eberbach where we had a reduction of six vehicles and 41 vehicle running hours. This quantifies the improvement potential in lengthier lines when using sublines to reduce operational costs.

It is interesting to note that the DWS solution satisfies more than 95% of the overall demand in the 500 sampled scenarios. If higher certainty is required, i.e.., 98% or 99%, more or longer lines are required, as we can see when we compare the two SWS solutions with the DWS solution.

Interestingly, in this larger scenario, even the conservative SWS solution (b - 99%) has lower vehicle running times than the DNS solution. This underlines the important role of using sublines to allocate resources more efficiently when the passenger demand is skewed towards one terminal.

In Table 20 we present the number of unserved passengers when implementing each solution in 500 new passenger demand scenarios that are sampled from the same probability distribution as the scenarios presented in Figure 10. The DNS solution results in slightly more unserved passengers compared to the conservative SWS solution ($b - 99\%$), 19 vs 4 unserved passengers, even if it uses four more vehicles and 19 more vehicle running hours. Note also that the average passenger waiting time remains below 3 minutes regardless of the implemented solution.

Table 20: Unserved passengers

solution	unserved passengers	% of the total demand
deterministic no sublines (DNS)	19	0.048%
deterministic with sublines (DWS)	152	0.380%
98% stochastic with sublines (SWS) [a]	38	0.095%
99% stochastic with sublines (SWS)[b]	4	0.010%

[a]this solution satisfies $\geq$ 98% of the passenger demand in the 500 samples

[b]this solution satisfies $\geq$ 99% of the passenger demand in the 500 samples

7. Concluding Remarks

In this work, we introduced a novel frequency setting model that assigns autonomous minibusses to sublines. This model, originally formulated as a MINLP, is reformulated as a MILP that can be solved to global optimality. Based on that model, we explicitly consider the uncertainty of passenger demand in the optimization process by formulating a stochastic optimization model. Notably, the stochastic model is based on the sample average approximation method and maintains a MILP formulation.

Our deterministic and stochastic models that assign autonomous minibusses to sublines were tested against a baseline model that assigns vehicles only to the original line. In our first case study, we considered various demand profiles, such as higher demand levels at the line segments close to the terminals, at the center of the line, and constant demand across all links. Our experiments confirmed that the potential of savings by using sublines is higher when demand is skewed and quantified this benefit.

When comparing the SWS solutions with the DWS solutions, we found that the decrease in unserved passengers when using SWS requires the deployment of more vehicles and a significant increase in the running time. This can be taken into consideration by policy makers that might need to decide about the trade-off between offering sufficient capacity even at worst-case demand scenarios and reducing the operational costs.

To fully explore the benefit of sublines, we also performed experiments on a 20-stop line considering a skewed demand profile towards one terminal. Results from this larger case showed that we can significantly reduce the number of assigned vehicles and the vehicle running times while the average passenger waiting time is barely affected. Similarly to the first case study, the SWS solutions offered marginal benefits in terms of being able to serve more passengers at worst-case demand scenarios while resulting in increased operational costs.

7.1. Limitations

Our assumption that passengers arrive at stops randomly, i.e., not based on the timetable, is most applicable in systems with high frequencies, i.e., where an OD-pair is served at least four times an hour. Waiting time estimates in our model are based on the assumption that each passenger can actually board the next minibus serving his/her OD-pair - which may not be possible in practice. This can lead to an underestimation of waiting times, in particular in crowded systems. Another limitation is that we assume that a passenger will use the next minibus that serves his/her origin-destination pair. We do not allow passengers to use the next minibus to travel a few stops and then wait for another minibus that will transfer the passenger to his/her final destination. Although rare, in practice there might be passengers who are willing to split their trip into more stages even if the travel time until reaching their final destination will be exactly the same.

7.2. Future research

In terms of future research directions, in this work the generated sublines serve segments of the originally planned lines and are a product of short-turning. In future research, this can be expanded by considering interlining lines where the same vehicle can be used by more than one line as an additional option. In this case, it may be beneficial to step away from the assumption that each line is operated periodically, towards a system where vehicle runs occur on demand. Furthermore, it would be interesting to investigate passenger waiting times more closely. Finally, experiments can be expanded beyond autonomous minibusses to consider the implications of potential electric minibusses that have specific requirements in terms of vehicle charging.

REFERENCES

Arbex, R. O., da Cunha, C. B., 2015. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. Transportation Research Part B: Methodological 81, 355–376.

Bartholdi III, J. J., Eisenstein, D. D., 2012. A self-coördinating bus route to resist bus bunching. Transportation Research Part B: Methodological 46 (4), 481–491.

Bertsimas, D., Sian Ng, Y., Yan, J., 2020. Joint frequency-setting and pricing optimization on multimodal transit networks at scale. Transportation Science 54 (3), 839–853.

Cats, O., 2014. Regularity-driven bus operation: Principles, implementation and business models. Transport Policy 36, 223–230.

Cats, O., Glück, S., 2019. Frequency and vehicle capacity determination using a dynamic transit assignment model. Transportation Research Record 2673 (3), 574–585.

Ceder, A., 1984. Bus frequency determination using passenger count data. Transportation Research Part A: General 18 (5-6), 439–453.

Ceder, A., 2001. Public transport scheduling. Handbook of Transport Systems and Traffic Control 3, 539–558.

Ceder, A., 2016. Public transit planning and operation: Modeling, practice and behavior. CRC press.

Ceder, A., Wilson, N. H., 1986. Bus network design. Transportation Research Part B: Methodological 20 (4), 331–344.

Ceder, A. A., 2011. Public-transport vehicle scheduling with multi vehicle type. Transportation Research Part C: Emerging Technologies 19 (3), 485–497.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transportation research part B: Methodological 40 (6), 437–459.

Cipriani, E., Gori, S., Petrelli, M., 2012. Transit network design: A procedure and an application to a large urban area. Transportation Research Part C: Emerging Technologies 20 (1), 3–14.

Claessens, M., van Dijk, N. M., Zwaneveld, P. J., 1998. Cost optimal allocation of rail passenger lines. European Journal of Operational Research 110 (3), 474–489.

Cortés, C. E., Jara-Díaz, S., Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. Transportation Research Part A: Policy and Practice 45 (5), 419–434.

Delle Site, P., Filippi, F., 1998. Service optimization for bus corridors with short-turn strategies and variable vehicle size. Transportation Research Part A: Policy and Practice 32 (1), 19–38.

dell'Olio, L., Ibeas, A., Ruisánchez, F., 2012. Optimizing bus-size and headway in transit networks. Transportation 39 (2), 449–464.

Duss, A.-C., 2018. Contern to host luxembourg's first driverless bus.
URL https://h2020-avenue.eu/contern-to-host-luxembourgs-first-driverless-bus/

Fabulos, 2020. Robot buses as part of urban public transportation.
URL https://fabulos.eu/

GitHub, 2021. Subline frequency setting for autonomous minibusses under demand uncertainty: data and software code.
URL https://github.com/KGkiotsalitis/subline-frequency-setting.git

Gkiotsalitis, K., Alesiani, F., 2019. Robust timetable optimization for bus lines subject to resource and regulatory constraints. Transportation Research Part E: Logistics and Transportation Review 128, 30–51.

Gkiotsalitis, K., Cats, O., 2018. Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. Transportation Research Part C: Emerging Technologies 88, 187–207.

Gkiotsalitis, K., Wu, Z., Cats, O., 2019. A cost-minimization model for bus fleet allocation featuring the tactical generation of short-turning and interlining options. Transportation Research Part C: Emerging Technologies 98, 14–36.

Hadas, Y., Shnaiderman, M., 2012. Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. Transportation Research Part B: Methodological 46 (8), 1068–1084.

Hassold, S., Ceder, A. A., 2014. Public transport vehicle scheduling featuring multiple vehicle types. Transportation Research Part B: Methodological 67, 129–143.

Hickman, M. D., 2001. An analytic stochastic model for the transit vehicle holding problem. Transportation Science 35 (3), 215–237.

Ibarra-Rojas, O., Delgado, F., Giesen, R., Muñoz, J., 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological 77, 38–75.

Jha, S. B., Jha, J. K., Tiwari, M. K., 2019. A multi-objective meta-heuristic approach for transit network design and frequency setting problem in a bus transit system. Computers & Industrial Engineering 130, 166–186.

Modijefsky, M., 2019. Lyon launches new autonomous public transport service.
URL https://www.eltis.org/in-brief/news/lyon-launches-new-autonomous-public-transport-service

Muezner, P., 2018. First self-driving bus service ioki operating in germany.
URL https://ecourbanhub.com/ioki-germany-deutsche-bahn-self-driving-autonomous-bus/

Nikolić, M., Teodorović, D., 2014. A simultaneous transit network design and frequency setting: Computing with bees. Expert Systems with Applications 41 (16), 7200–7209.

Pinto, H. K., Hyland, M. F., Mahmassani, H. S., Verbas, I. Ö., 2020. Joint design of multimodal transit networks and shared autonomous mobility fleets. Transportation Research Part C: Emerging Technologies 113, 2–20.

Stein, J., Goebel, N., 2019. Berlin tests driverless buses.
URL https://www.dw.com/en/germany-berlin-tests-driverless-buses/a-50055426

Szeto, W. Y., Wu, Y., 2011. A simultaneous bus route design and frequency setting problem for tin shui wai, hong kong. European Journal of Operational Research 209 (2), 141–155.

van der Hurk, E., Koutsopoulos, H. N., Wilson, N., Kroon, L. G., Maróti, G., 2016. Shuttle planning for link closures in urban public transport networks. Transportation Science 50 (3), 947–965.

Verbas, İ. Ö., Mahmassani, H. S., 2013. Optimal allocation of service frequencies over transit network routes and time periods: formulation, solution, and implementation using bus route patterns. Transportation research record 2334 (1), 50–59.

Verbas, İ. Ö., Mahmassani, H. S., 2015a. Exploring trade-offs in frequency allocation in a transit network using bus route patterns: Methodology and application to large-scale urban systems. Transportation Research Part B: Methodological 81, 577–595.

Verbas, İ. Ö., Mahmassani, H. S., 2015b. Integrated frequency allocation and user assignment in multimodal transit networks: Methodology and application to large-scale urban systems. Transportation Research Record 2498 (1), 37–45.

Welding, P., 1957. The instability of a close-interval service. Journal of the operational research society 8 (3), 133–142.

Yu, B., Yang, Z., Yao, J., 2010. Genetic algorithm for bus frequency optimization. Journal of Transportation Engineering 136 (6), 576–583.

Appendix A. Nomenclature

Table A.21: Nomenclature

Sets	
$\mathcal{S}$	ordered set of stops of the minibus line in both directions, $\mathcal{S} = \langle 1, 2, ..., s, ... \rangle$
$\mathcal{R}$	set of all potential lines $\mathcal{R} = \langle 1, ..., r, ... \rangle$, where line 1 is the original line that serves all stops $s \in S$ and $\langle 2, ...r, ... \rangle$ are the generated sublines.
$\mathcal{O}$	set of OD-pairs with passenger demand. Note that if there is no passenger demand between stops $s \in \mathcal{S}$ and $y \in \mathcal{S}$, then $(s, y) \notin \mathcal{O}$
$\mathcal{F}$	discrete set of frequencies for each potential line
$\tilde{\mathcal{F}}$	discrete set of frequencies for each OD-pair
Parameters	
T	planning horizon
P	period length of the periodic schedule
B_{sy}	passengers willing to travel from stop s to y in our demand-homogeneous planning horizon, where $(s, y) \in \mathcal{O}$.
$\Delta_{r,sy}$	$\Delta_{r,sy} = 1$ if subline r serves the OD-pair $(s, y) \in \mathcal{O}$.
T_r	round-trip travel time of line $r \in \mathcal{R}$
N	number of available minibusses
Θ	minimum allowed service frequency, $\Theta > 0$, to ensure a minimum level of service for any passenger traveling from stop s to stop y, where $(s, y) \in \mathcal{O}$
K	minimum number of minibusses that should be assigned to the original line, where $K \leq N$
W_1	the cost of operating an extra minibus
W_2	the cost of a marginal increase in the total vehicle running times
F	minimum frequency of a subline to be deemed operational
M	a very large positive number
c	minibus capacity
Variables	
x_r	number of minibusses assigned to potential line $r \in \mathcal{R}$
f_{sy}	the service frequency of OD-pair $(s, y) \in \mathcal{O}$ in vehicles per time period P.
f_r	the service frequency of line $r \in \mathcal{R}$ in vehicles per time period P
a_r	binary variable, where $a_r = 1$ if subline r is deemed operational and 0 otherwise
$b_{r,s}$	number of passengers that board line r at stop s
$v_{r,s}$	number of passenger that alight from line r at stop s
$l_{r,s}$	in-vehicle passenger load of line r at stop s

Appendix B. Expected passenger waiting time

Theorem Appendix B.1. *Consider f vehicles (possibly operating on different sublines) that cover an OD-pair (s, y), that is, both stations s and y lie on the vehicle route in this order. Under the assumptions that*

- *the vehicles operate according to a periodic schedule that has a length of P minutes,*
- *that the departures of different vehicles at s are scheduled independently of each other,*
- *the probability of each vehicle to be scheduled to depart in minute τ is uniform over the period, and*
- *for each passenger, the probability to arrive at origin s in minute is uniform over the period, in particular passengers do not time their arrivals based on the schedule, and passenger arrivals are independent of each other*
- *each passenger can board the next vehicle serving his OD-pair*

the expected waiting time of a passenger, i.e., the time between their arrival at the origin station and the departure of the next vehicle towards the destination is $\frac{P}{f+1}$.

Proof. We divide the derivation of the expected waiting time for a passenger into two steps: In the first step, we compute the expected waiting time of a passenger based of *given* departure times. In the second step, we use the result of the first step to compute the expected waiting time for uniformly and independently distributed departures, as specified in the second and third assumption of the theorem.

Without loss of generality, we assume that the first vehicle departs at time $t_1 = 0$ from s (we choose an arbitrary vehicle and define the period start to be the point in time at which the vehicle starts).

Known departure times. Assume now that we already have a given schedule which specifies the departure times at station s to be $t_1 = 0, t_2, \ldots, t_f$ with $t_i < P$. Then we can sort the departures times of the vehicles $i = 2, \ldots, f$ in increasing order to obtain the sequence $t_{(2)}, t_{(3)}, \ldots, t_{(f)}$ and split the period into intervals $(t_{(i-1)}, t_{(i)}]$ for $i = 2, \ldots, f+1$ with $t_{(f+1)} := P$. (Note that some intervals may have a width of 0 if vehicles depart at the same time.)

The probability for a passenger to arrive in interval $(t_{(i-1)}, t_{(i)}]$ is $\frac{1}{P}(t_{(i)} - t_{(i-1)})$ (based on the fourth assumption) and the expected waiting time for a passenger arriving at a random moment in this interval is $\frac{1}{2}(t_{(i)} - t_{(i-1)})$.

That is, we can compute the expected length of the passenger's waiting time w (assuming that each moment of passenger arrival at the station is equally likely) for given vehicle departures $t_1 = 0, t_2, \ldots, t_f$ as

$$E[w|t_1, \ldots, t_f] = \frac{1}{2} \cdot \frac{1}{P} \left[\sum_{i=1}^{f} (t_{(i+1)} - t_{(i)})^2 \right] = \frac{1}{2} \cdot \frac{1}{P} \left[\sum_{i=2}^{f} [2t_{(i)}^2 - 2t_{(i)}t_{(i+1)}] + P^2 \right] \tag{B.1}$$

where we use that $t_{(1)} = t_1 = 0$ and $t_{(f+1)} = t_{(1)} + P = P$ based on the first assumption.

Unknown departure times. We now proceed with the result from the first step, to compute the expected value of the waiting time under the second and third assumption. As we do not know the departure times of the vehicles in this step (except for the one departing at $t_1 = 0$ by definition), we compute the expected value over all combinations of departure times. Assuming equal and independent probabilities for the departure times, the probability of a certain combination of departure times in a period of P minutes is $P(\tau_2 = t_2, \tau_3 = t_3, \ldots, \tau_f = t_f) = \frac{1}{P^{f-1}}$.

We therefore have

$$\begin{aligned} & E[w] \\ &= \frac{1}{2} \cdot \frac{1}{P} \int_{(t_2,\ldots,t_f)\in[0,P]^{f-1}} P(\tau_2 = t_2, \tau_3 = t_3, \ldots, \tau_f = t_f) \cdot \left[\sum_{i=2}^{f} [2t_{(i)}^2 - 2t_{(i)}t_{(i+1)}] + P^2 \right] d(t_2, \ldots, t_f) \\ &= \frac{1}{P^f} \int_{(t_2,\ldots,t_f)\in[0,P]^{f-1}} \left[\sum_{i=2}^{f} [t_{(i)}^2 - t_{(i)}t_{(i+1)}] + \frac{1}{2}P^2 \right] d(t_2, \ldots, t_f) \end{aligned}$$

Let $V = \{v_1, v_2, \ldots, v_f\}$ (with v_1 being the vehicle that defines the period start) denote the set of considered vehicles. We consider $M := \{v_{(1)} = v_1, v_{(2)}, v_{(3)}, \ldots, v_{(f)}\}$, the set of all ordered vehicle sequences. Note that $|M| = (f-1)!$.

For each sequence $m \in M$, let S_m denote the set of tupels of arrival times that correspond to the order of sequence m, i.e., $S_m := \{(t_{(1)}, t_{(2)}, t_{(3)}, \ldots, t_{(f)}) : 0 = t_{(1)} \leq t_{(2)} \leq t_{(3)} \leq \ldots, t_{(f)}\}$, with $t_{(i)}$ denoting the departure time of vehicle $v_{(i)}$.

We rewrite the multiple integral (B.2) as a sum of integrals over the sets S_m, and rewrite these as iterated integrals.

$$E[w] \tag{B.2}$$

$$= \frac{1}{P^f} \sum_{m \in M} \int_{(t_{(1)}, t_{(2)}, t_{(3)}, \ldots, t_{(f)}) \in S_m} \left[\sum_{i=2}^{f} [t_{(i)}^2 - t_{(i)} t_{(i+1)}] + \frac{1}{2} P^2 \right] d(t_{(2)}, \ldots, t_{(f)}) \tag{B.3}$$

$$= \frac{1}{P^f} \left(\sum_{m \in M} \left(\int_{t_{(f)}=0}^{t_{(f)}=P} \cdots \int_{t_{(3)}=0}^{t_{(3)}=t_{(4)}} \int_{t_{(2)}=0}^{t_{(2)}=t_{(3)}} \left[\sum_{i=2}^{f} [t_{(i)}^2 - t_{(i)} t_{(i+1)}] + \frac{1}{2} P^2 \right] dt_{(2)} dt_{(3)} \ldots dt_{(f)} \right) \right) \tag{B.4}$$

$$= \frac{1}{P^f} \cdot (f-1)! \cdot \left(\int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \left[\sum_{i=2}^{f} \underbrace{\tau_i^2}_{A_i} - \sum_{i=2}^{f} \underbrace{\tau_i \tau_{i+1}}_{B_i} + \underbrace{\frac{1}{2} P^2}_{C} \right] d\tau_2 d\tau_3 \ldots d\tau_f \right)$$

The last step follows as all $(f-1!)$ summands are identical.

We now integrate A_i, B_i and C separately.

Integration of A_i. for $i = 2, \ldots, f$

$$\int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} A_i d\tau_2 d\tau_3 \ldots d\tau_f$$

$$= \int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \tau_i^2 d\tau_2 d\tau_3 \ldots d\tau_f$$

$$= \int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_i=0}^{\tau_i=\tau_{i+1}} \tau_i^2 \cdot \left(\frac{1}{(i-2)!} \tau_i^{i-2} \right) d\tau_i d\tau_{i+1} \ldots d\tau_f$$

$$= \int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_i=0}^{\tau_i=\tau_{i+1}} \frac{1}{(i-2)!} \tau_i^i d\tau_i d\tau_{i+1} \ldots d\tau_f$$

$$= \int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_{i+1}=0}^{\tau_{i+1}=\tau_{i+1}} \frac{1}{(i-2)!} \left[\frac{1}{i+1} \tau_i^{i+1} \right]_{\tau_i=0}^{\tau_i=\tau_{i+1}} d\tau_{i+1} \ldots d\tau_f$$

$$= \int_{\tau_f=0}^{\tau_f=P} \cdots \int_{\tau_{i+1}=0}^{\tau_{i+1}=\tau_{i+2}} \frac{(i-1) \cdot i}{(i+1)!} \tau_{i+1}^{i+1} d\tau_{i+1} \ldots d\tau_f$$

$$= \frac{(i-1) \cdot i}{(f+1)!} \left[\tau_f^{f+1} \right]_{\tau_f=0}^{\tau_f=P} = \frac{(i-1) \cdot i}{(f+1)!} P^{f+1}$$

Integration of B_i. For $i = 2, \ldots, f-1$

$$\begin{aligned}
&\int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} B_i d\tau_2 d\tau_3 \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \tau_i \tau_{i+1} d\tau_2 d\tau_3 \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_i=0}^{\tau_i=\tau_{i+1}} \tau_i \tau_{i+1} \cdot \left(\frac{1}{(i-2)!} \tau_i^{i-2} \right) d\tau_i d\tau_{i+1} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_i=0}^{\tau_i=\tau_{i+1}} \frac{1}{(i-2)!} \tau_{i+1} \tau_i^{i-1} d\tau_i d\tau_{i+1} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_{i+1}=0}^{\tau_{i+1}=\tau_{i+2}} \frac{1}{(i-2)!} \tau_{i+1} \left[\frac{1}{i} \tau_i^i \right]_{\tau_i=0}^{\tau_i=\tau_{i+1}} d\tau_{i+1} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_{i+1}=0}^{\tau_{i+1}=\tau_{i+2}} \frac{i-1}{i!} \tau_{i+1}^{i+1} d\tau_{i+1} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_{i+2}=0}^{\tau_{i+2}=\tau_{i+3}} \frac{(i-1)}{i!} \left[\frac{1}{i+2} \tau_{i+1}^{i+2} \right]_{\tau_{i+1}=0}^{\tau_{i+1}=\tau_{i+2}} d\tau_{i+2} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_{i+2}=0}^{\tau_{i+2}=\tau_{i+3}} \frac{(i-1)\cdot(i+1)}{(i+2)!} \tau_{i+2}^{i+2} d\tau_{i+2} \ldots d\tau_f \\
&= \int_{\tau_f=0}^{\tau_f=P} \frac{(i-1)\cdot(i+1)}{f!} \tau_f^f d\tau_f \\
&= \frac{(i-1)\cdot(i+1)}{(f+1)!} P^{f+1}
\end{aligned}$$

For $i = f$ we have $\tau_{i+1} = \tau_{f+1} = P$ and

$$\int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \tau_f P d\tau_2 d\tau_3 \ldots d\tau_f \tag{B.5}$$

$$= \int_{\tau_f=0}^{\tau_f=P} \tau_f P \frac{1}{(f-2)!} \tau_f^{f-2} d\tau_f \tag{B.6}$$

$$= \int_{\tau_f=0}^{\tau_f=P} P \frac{1}{(f-2)!} \tau_f^{f-1} d\tau_f \tag{B.7}$$

$$= P \frac{f-1}{f!} P^f \tag{B.8}$$

$$= \frac{f-1}{f!} P^{f+1} \tag{B.9}$$

Integration of C.

$$\int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} C d\tau_2 d\tau_3 \ldots d\tau_f \tag{B.10}$$

$$= \int_{\tau_f=0}^{\tau_f=P} \ldots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \frac{1}{2} \cdot P^2 d\tau_2 d\tau_3 \ldots d\tau_f \tag{B.11}$$

$$= \frac{1}{2} \frac{1}{(f-1)!} P^{f+1} \tag{B.12}$$

Continue computation of the expected value. By summing up the terms obtained in previous steps we obtain

$$
\begin{aligned}
&\mathcal{E}[w] \\
&= \frac{1}{P^f} \cdot (f-1)! \cdot \left(\int_{\tau_f=0}^{\tau_f=P} \dots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \left[\sum_{i=2}^{f} \underbrace{\tau_i^2}_{A_i} - \sum_{i=2}^{f} \underbrace{\tau_i \tau_{i+1}}_{B_i} + \underbrace{\frac{1}{2} P^2}_{C} \right] d\tau_2 d\tau_3 \dots d\tau_f \right) \\
&= \frac{1}{P^f} \cdot (f-1)! \cdot \left(\int_{\tau_f=0}^{\tau_f=P} \dots \int_{\tau_3=0}^{\tau_3=\tau_4} \int_{\tau_2=0}^{\tau_2=\tau_3} \left[\sum_{i=2}^{f} \underbrace{\tau_i^2}_{A_i} - \sum_{i=2}^{f-1} \underbrace{\tau_i \tau_{i+1}}_{B_i} + \underbrace{\tau_f P}_{B_f} + \underbrace{\frac{1}{2} P^2}_{C} \right] d\tau_2 d\tau_3 \dots d\tau_f \right) \\
&= \frac{1}{P^f} \cdot (f-1)! \cdot \left(\sum_{i=2}^{f} \frac{(i-1) \cdot i}{(f+1)!} P^{f+1} - \sum_{i=2}^{f-1} \frac{(i-1) \cdot (i+1)}{(f+1)!} P^{f+1} - \frac{f-1}{f!} P^{f+1} + \frac{1}{2} \frac{1}{(f-1)!} P^{f+1} \right) \\
&= P \cdot \left(\sum_{i=2}^{f} \frac{(i-1) \cdot i}{f(f+1)} - \sum_{i=2}^{f-1} \frac{(i-1) \cdot (i+1)}{f(f+1)} - \frac{f-1}{f} + \frac{1}{2} \right) \\
&= P \cdot \left(\frac{1}{f(f+1)} \left[\sum_{i=2}^{f-1} ((i-1) \cdot i - (i-1) \cdot (i+1)) + (f-1) \cdot f \right] - \frac{f-1}{f} + \frac{1}{2} \right) \\
&= P \cdot \left(\frac{1}{f(f+1)} \left[- \sum_{i=2}^{f-1} (i-1) + (f-1) \cdot f \right] - \frac{f-1}{f} + \frac{1}{2} \right) \\
&= P \cdot \left(\frac{1}{f(f+1)} \left[- \frac{(f-2) \cdot (f-1)}{2} + (f-1) \cdot f \right] - \frac{f-1}{f} + \frac{1}{2} \right) \\
&= \frac{1}{2} P \cdot \frac{-(f-2)(f-1) + 2(f-1)f - 2(f-1)(f+1) + f(f+1)}{f(f+1)} \\
&= \frac{1}{2} P \cdot \frac{[-(f-2) + 2f](f-1) + [-2(f-1) + f](f+1)}{f(f+1)} \\
&= \frac{1}{2} P \cdot \frac{[f+2](f-1) + [-f+2](f+1)}{f(f+1)} \\
&= \frac{1}{2} P \cdot \frac{(f^2+f-2) + (-f^2+f+2)}{f(f+1)} \\
&= \frac{1}{2} P \cdot \frac{2f}{f(f+1)} \\
&= \frac{P}{f+1}
\end{aligned}
$$

■