

Lee, Chungmok; Nair, Rahul

## Article

# Robust transit line planning based on demand estimates obtained from mobile phones

EURO Journal on Transportation and Logistics (EJTL)

## Provided in Cooperation with:

Association of European Operational Research Societies (EURO), Fribourg

*Suggested Citation:* Lee, Chungmok; Nair, Rahul (2021) : Robust transit line planning based on demand estimates obtained from mobile phones, EURO Journal on Transportation and Logistics (EJTL), ISSN 2192-4384, Elsevier, Amsterdam, Vol. 10, Iss. 1, pp. 1-13, <https://doi.org/10.1016/j.ejtl.2021.100034>

This Version is available at:

<https://hdl.handle.net/10419/325148>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Robust transit line planning based on demand estimates obtained from mobile phones



Chungmok Lee<sup>a</sup>, Rahul Nair<sup>b,\*</sup>

<sup>a</sup> Department of Industrial and Management Engineering, Hankuk University of Foreign Studies, South Korea

<sup>b</sup> IBM Research Europe, Ireland

## ARTICLE INFO

### Keywords:

Robust network design  
Transit line planning  
Mobile phone data  
Column generation

## ABSTRACT

The line-planning problem seeks to determine the set of fixed routes (or lines) a transit operator should run, along with associated operation frequencies. We propose an optimization algorithm for the transit-line-planning problem based on bi-level programming that exploits the problem's structure in conjunction with the range estimation of demands. The issue of conservativeness due to the range estimation is mitigated by adopting the robust optimization approach. The model was inspired by a real-world application that leverages big data available from telecommunications operators to estimate city-wide mobility patterns. Demand estimates from such sources are based on large sample sizes (often orders of magnitude larger than those used in survey-based approaches), and capture day-to-day variability in travel demand as ranges. The validity of the proposed algorithm is demonstrated by using real-world data derived from 2.5 billion call data records from Abidjan, Côte d'Ivoire.

## 1. Introduction

We present a robust bi-level program that addresses the transit-line-planning problem with capacity constraints. The line-planning problem seeks to determine a set of fixed routes on which an operator should run along with the frequencies for each route. In real life, the travelers do not just take the shortest path to their destination, as it often involves a long waiting time for public transport such as buses and/or subways. In order to reduce travelers' total travel times, the operator needs to determine, whilst taking resource limitations into account, which public transport routes to operate and how frequently they should operate, since the more frequently the transport operates, the less waiting time is expected.

Line planning is one stage in the hierarchy of the overall transit-planning process that transit operators conduct. Several recent texts provide a context for this process, (Vuchic, 2017; Ceder, 2016) gives an overview. The first stage is the design of a fixed route infrastructure (stops, streets, tracks for rail/tram services). The second stage is the transit operator's determination of the lines that should run on that infrastructure. A line is defined, in our context, as a fixed route to be operated during a particular time of day. More specifically, a line is an ordered sequence of stops. Each line has an associated frequency, which denotes the number of services the travelers see every hour. The

subsequent stages of transit planning involve timetabling and rostering of crews and vehicles. In this paper, we focus on the line-planning problem faced by operators who have a large number of routes to operate.

The travel-demand model used in this research is motivated by a real-world application for estimation of mobility demand from big data available from telecommunications operators. Demand estimations from such sources leverage large sample sizes, typically orders of magnitude larger than traditional surveys. In contrast to classical demand-estimation methods that seek a point estimate between an origin-destination (OD) pair for a specific time period, mobile-phone-based estimates provide a distribution that captures inherent day-to-day uncertainties in the travel-demand process.

Accurate estimation of travel demand can be problematic on two counts. First, classical econometric demand-estimation methods based on surveys can be prohibitively expensive for many cities in the growth markets. Second, such surveys might not be timely, since cities are growing so rapidly the changes in travel needs often outpace the frequency at which travel surveys can be conducted. African cities, for example, are growing at an annual rate of 3–5% (Kumar and Barrett, 2008).

This rapid urbanization has been accompanied by similarly rapid mobile-phone adoption. The data obtainable from mobile phones can

\* Corresponding author.

E-mail addresses: [chungmok@hufs.ac.kr](mailto:chungmok@hufs.ac.kr) (C. Lee), [rahul.nair@ie.ibm.com](mailto:rahul.nair@ie.ibm.com) (R. Nair).

offer key insights into mobility patterns; moreover, it is infrastructure-free and timely. Such data, representative of large user samples, can provide rich characterizations of travel demand. Data resolution in space and time typically enables activity-sequence reconstruction that allows for a detailed assessment of movement patterns and captures day-to-day variability inherent in the demand process. For example, the data can reveal changes in OD flows on account of special events (Dong et al., 2015) as well as people density estimation by location for security applications (Calabrese et al., 2015).

A key big-data insight from telecommunications data is that repeated observations over a period of time allow for a richer characterization of mobility-demand processes. Travel demand between two zones not only can be considered as a point estimate (as a function of time of day), but also as a random variable with a known distribution. This characterization should then be exploited in the context of transit-network design which requires a novel solution approach.

For optimization of the line-planning problem with respect to demand uncertainties, there are two dominant approaches, namely stochastic programming and robust optimization. In the stochastic programming, uncertain demand is represented by probability distributions according to certain estimated parameters. Based on these distributions, many demand scenarios are generated and considered in order to achieve solution robustness. Explicit consideration of several such discrete scenarios often results in a very large optimization problem. To address this problem, optimization algorithms such as Benders decomposition can be used, though the number of required scenarios often grows too quickly with the number of origin-destination (OD) pairs. By contrast, the other approach to the line-planning problem, robust optimization does not explicitly take each scenario into account. Rather, instead of discretized scenarios, a set of all (infinitely many) possible realizations of demand data, termed an *uncertainty set*, is defined. The goal, simply, is to protect the solution from the worst-case realization of demand over the elements of the uncertainty set. Using uncertainty sets is less computationally demanding than the traditional scenario-based approaches, if the uncertainty sets are defined carefully, which is a very advantageous property, especially in real-world applications (Ben-Tal and Nemirovski, 2007; Bertsimas and Sim, 2004).

The worst-case realization may impact not only the feasibility of constraints but also the objective function. When the coefficients of the objective function are uncertain, the so-called *objective robustness* seeks the best worst-case solution in the context of the max-min framework. Nonetheless, the objective functions with uncertain coefficients defined by the uncertainty set can be trivially reformulated to constraints by adding auxiliary decision variables.

Several alternative measures for ensuring solution robustness to input uncertainty are available. We propose a method that models the demand uncertainty as a range. There are two reasons for using range estimates and uncertainty sets to capture robustness. First, the proposed range-based characterization allows us to exploit the  $\Gamma$ -parameterization approach (Bertsimas and Sim, 2004), which enables control of the conservatism of the solutions. For realistically sized networks, competing approaches such as scenario-based approaches can result in intractably large models. Second, the apportionment of an uncertainty budget across various demand ranges generates a conservative solution that is valid for all realizations from that set. Because it is unlikely that all demands are simultaneously at their upper bound, the solution is robust against any joint-demand realizations that are encountered.

Our solution approach is based on bi-level programming, because the direct single-level formulation involves nonlinear constraints. In the bi-level formulation developed, the operator's line-planning (routes and frequencies) decisions are considered at the upper level, while the passengers' route choices are made at the lower level. For modeling of passenger route choices, we adapt the more realistic optimal strategy model originally proposed by Spiess and Florian (1989). We then extend the model to include capacity constraints, specifically, by taking into account the delays incurred from over-congested routes.

This paper's contribution to the literature is a capacity-constrained robust transit-line planning model wherein demand estimation is available as a range. We first propose a mathematical formulation for the problem by employing the robust optimization and the bi-level programming approach, which yields a so-called flow-based formulation. To address problem sizes arising in practice, the column generation method is adapted in the context of the bi-level programming by decomposing the flow-based formulation into many shortest path problems. The column generation based approach shows significantly better results in terms of computational times than the original flow-based counterpart, which enables us to tackle a real-world problem in Abidjan, Côte d'Ivoire. A Monte-Carlo simulation designed to show the value of the robust approach is also presented.

The paper proceeds as follows. Section 2 reviews the existing literature on the line-planning problem, and Section 3 introduces the proposed modeling approach and solution method. Section 4 presents the Abidjan usage case, and Section 5 discusses the relevant computational experiments. Finally, Section 6 draws conclusions.

## 2. Literature review

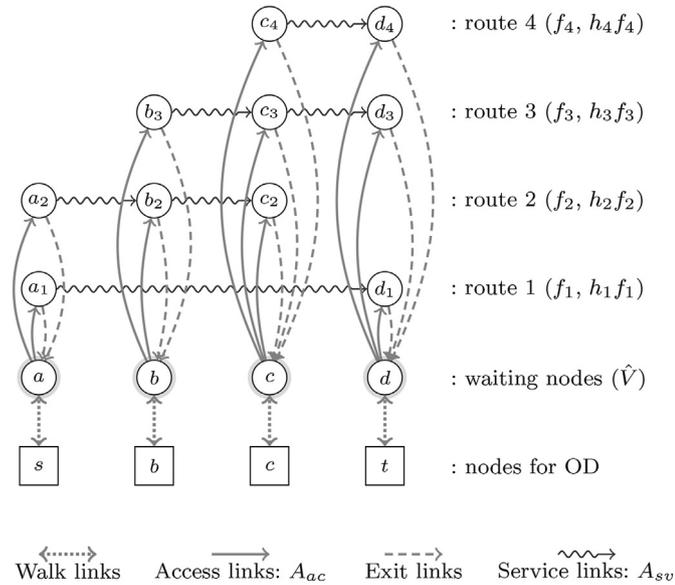
There is an extensive literature on the transit-planning process and, more specifically, the line-planning problem. The problem has been studied in several contexts and with a variety of considerations, both in the objectives and constraints. Readers are referred to the review papers on the subject (GuihaireJin-Kao, 2008; Guy and Hickman, 2007; Ibarra-Rojas et al., 2015) as well as a recent survey on deterministic methods (ZanjiraniFarahani et al., 2013). We focus herein on recent studies that have considered strategic planning and/or robustness, used similar structures of bi-level models, and exploited opportunistic sensing for transit design.

All of the previous methods address one or more of the following main challenges associated with transit planning: (a) computational challenges associated with practical-sized networks leading to special solution methods (Borndörfer et al., 2007), heuristic (Mauttone and Urquhart, 2009) or meta-heuristic approaches (AsadiBagloee and Aviceder, 2011; Cipriani et al., 2012); (b) the large feasible search space for network design (Borndörfer et al., 2007); (c) competing trade-offs between operators and users (Ceder and Wilson, 1986; Matisziw et al., 2006); (d) the balancing of design attributes such as routes, frequencies, travel directness, transfer penalties, and (e) models of user behavior and responses to designs.

Bi-level formulations have been considered in several studies. At the upper level, the design decisions are made, and at the lower level, the passengers make service choices based on service offerings (e.g., routes and operation frequencies). (Szeto and Jiang, 2014) formulated a model with upper-level decisions relating to route structure and frequencies. The lower-level problem typically models passenger route choice based on optimal strategies (Spiess and Florian, 1989), and additionally accounts for congestion and in-vehicle costs. (Constantin and Florian, 1995) proposed a bi-level program formulation based on the optimal strategies model and conducted computational experiments for three (small) real-world networks. (Goerigk and Schmidt, 2017) considers line planning when passenger route choice is integrated. This contrasts with classical approaches which impose sequence - design first and passenger assignment second. The problem is formulated as a bi-level optimization problem and solved with constraint generation. This approach is conceptually similar to the one presented here. Some works have looked to integrate all levels of transit planning (Schöbel, 2017).

To the best of our knowledge, all of the previous approaches have assumed that demand, as based on point estimation, is deterministic. Other work has considered the lower-level assignment problem in the context of variational inequalities and solved them using an agent-based simulation (Ma, 2013).

Some studies have argued that solutions to line-planning problems need to account for the uncertainty of key inputs such as travel times



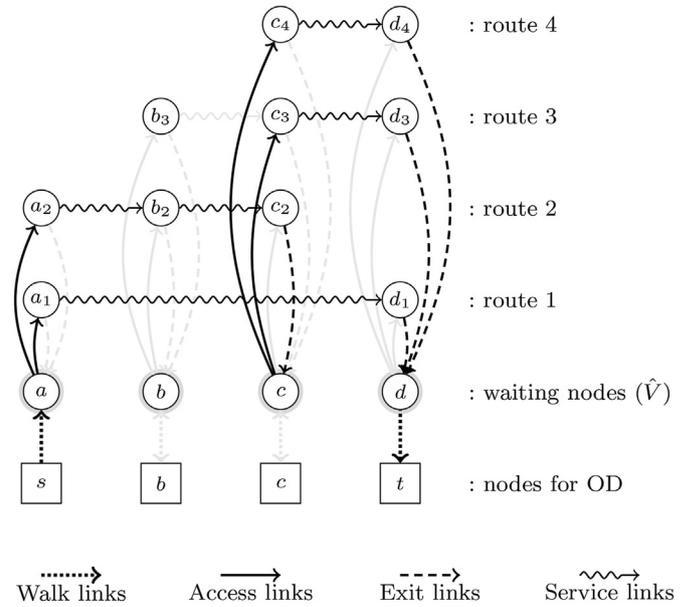
**Fig. 1.** Sample network with four OD nodes and routes. The waiting occurs at the waiting nodes. The service links incur the travel time and delays if they are over capacity. Each route is described by a (frequency, capacity) tuple. This network gives the sets defined in Section 3.1 as follows:  $R = \{1, 2, 3, 4\}$ ,  $\hat{V} = \{a, b, c, d\}$ ,  $A_a = \{(a, a_1), (a, a_2)\}$ , ...,  $A_{ac}^1 = \{(a, a_1), (d, d_1)\}$ , ...,  $A_{sv}^1 = \{(a_1, d_1)\}$  and so on.

(Yan et al., 2013) and link failures (Marín et al., 2009). Other studies, considering demand data to be uncertain, have employed a scenario-based demand-realization perspective. For example, (Amiripour et al., 2014), studying seasonal demand variations, argued that transit designs that are valid for flows during one season are likely to be sub-optimal during others. Certainly, formulations are *robust* only to the extent that the strategic design is valid for all scenarios considered.

Utilizing a bi-level model incorporating genetic algorithm based meta-heuristics, (Ukkusuri et al., 2007) approached the problem of long-term demand uncertainty by considering a discrete scenario set to minimize the expected travel costs and the standard deviation. Unfortunately, determining the scenario set is not straightforward for planning applications that arise in practice, and moreover, the number of scenarios needed to achieve a representative sample of demand uncertainty can be quite large.

From the data mining perspective, there are extensive studies on extraction of disaggregated and/or aggregate mobility patterns obtained from opportunistic sensing. Some of those researchers have sought to address transit-planning questions. For example, (Bastani et al., 2011) used GPS trajectory data from taxis to suggest minibus routes. (Chen et al., 2013) also used the same kind of data to derive OD matrices that they then used to plan nighttime services. (Calabrese et al., 2010), having conducted a detailed comparison of mobile-phone and vehicular data, showed that the trip lengths of vehicles and movements suggested by mobile phones are linearly related. (Alexander et al., 2014) compared flows derived from mobile phones, on the basis of which they suggested scaling based on census data for estimation of true demand processes. (Calabrese et al., 2011) presented methods for extraction of OD flows from mobile data and compared the respective results with gravity models. (Berlingerio et al., 2013) utilized point estimates of demand from mobile data to address the frequency-setting problem therein.

In contrast to these previous studies, the presented work aims to leverage demand distributions as opposed to point estimates. These can be estimated from big data sources like large mobile-phone data samples. The subsequently derived robust optimization model, acknowledging the inherent uncertainty associated with travel-demand uncertainty due to



**Fig. 2.** An example optimal strategy of demand  $s \rightarrow t$ . The optimal strategy is essentially a union of three distinct routes.

day-to-day variability and any potential biases associated with the data source, explicitly considers demand uncertainty as a range. To the best of our knowledge, this is the first practical-solution algorithm for the transit-network design that uses the robust optimization methodology. The proposed algorithm is implicitly able to take into account infinitely many demand scenarios (defined by an uncertainty set), and so it does not suffer from the scalability issue that often appears when the uncertain demands are considered according to a set of discrete demand scenarios.

### 3. Model

Given (a) a transit network of stops and arcs, (b) a set of current and candidate routes to operate, (c) a fleet size constraint in terms of number of vehicles, (d) estimated demand distributions for OD pairs, and (e) a model of traveler behavior, we seek to determine the set of lines to operate along with their frequencies such that the total of system-wide travel times is minimized.

#### 3.1. Notations

We consider a transit network  $G(V, A)$  that consists of a set of nodes  $V$  and a set of arcs  $A$ . The passenger (traveler) demands set  $K$  is defined by pairs of origin and destination nodes. The sets of origins and destinations are denoted by  $S \subset V$  and  $T \subset V$ , respectively. For node  $i \in V$ , let  $A_i^+$  (or  $A_i^-$ ) be a set of arcs originating from (or entering to)  $i \in V$ . The head and tail of an arc  $a$  are denoted by  $head(a)$  and  $tail(a)$ , i.e.,  $head(a) = j$  and  $tail(a) = i$  where  $a = (i, j)$ . Each arc is associated with travel time (or cost)  $c_a$ . There is a subset of nodes  $\hat{V} \subset V$  at which waiting occurs (stop nodes). Any arc emanating from waiting node  $i \in \hat{V}$  to any nodes not in  $S \cup T$  is called an *access arc* (or link). An access arc has an associated *exit arc* that is parallel but opposite in direction to the access arc. We let  $A_i \subseteq A_i^+$  be the set of access links of a waiting node  $i \in \hat{V}$ .  $R$  represents a set of all routes (lines), and each route's operating frequency is denoted by  $f_r$ . Each route  $r \in R$  passes an ordered set of *stop* nodes, while each stop node is connected by the access and exit arcs to the corresponding waiting node of the route. Any two consequent stop nodes for the same route are connected with a *service link*. Let  $A_{ac}$  and  $A_{sv}$  be the sets of access and service links, respectively. Sets of access links and service links for a route  $r \in R$  are denoted by  $A_{ac}^r$  and  $A_{sv}^r$ , respectively. Similarly, waiting nodes for a

route  $r \in R$  are denoted by  $\widehat{V}_r$ . Any access or service link  $a \in A_r$  of route  $r \in R$  has operation frequency  $f_a$  such that  $f_a = f_r$ , i.e.,  $f_a = f_r$  for all  $a \in A_r$ . A service link  $a \in A_{sv}$  also has a capacity  $h_a$  (the capacity of the vehicle operating on the arc). Let  $C$  denote the total number of vehicles available. Fig. 1 illustrates an example network with various types of arcs and nodes. Additional notations are defined in the relevant sections of the text.

### 3.2. User behavior model

We assume that travelers employ the so-called *optimal strategy* in navigating the transportation network (Spiess and Florian, 1989). The idea behind the optimal strategy is that by considering multiple routes, the waiting time can be reduced even if the routes in consideration are not the shortest ones.

In the sample network shown in Fig. 1, consider a traveler from  $s$  to  $t$  whose optimal strategy states that it is best to consider two routes (route 1 and route 2) at the same time to reduce the waiting time at node  $s$ . If the traveler took route 2 at node  $s$ , (s)he should transfer to either route 3 or route 4 at node  $c$  to get to node  $t$ . This optimal strategy is illustrated in Fig. 2, which is essentially a union of three paths between  $s$  and  $t$ . The first path is  $s \rightarrow a \rightarrow a_1 \rightarrow d_1 \rightarrow d \rightarrow t$ ; the second is  $s \rightarrow a \rightarrow a_2 \rightarrow b_2 \rightarrow c_2 \rightarrow c \rightarrow c_3 \rightarrow d_3 \rightarrow d \rightarrow t$ , and the third is  $s \rightarrow a \rightarrow a_2 \rightarrow b_2 \rightarrow c_2 \rightarrow c \rightarrow c_4 \rightarrow d_4 \rightarrow d \rightarrow t$ . It is important to note that the optimal strategy might involve multiple paths that are not necessarily the shortest in terms of travel time. Multiple paths can reduce the waiting time, because the traveler can take whatever transport arrives first at the waiting node.

Besides the waiting times, a traveler also needs to take into account the capacities of routes. The travel time of an over-congested route can be considerably longer, which results in a sub-optimal strategy. Therefore, a traveler, in estimating the possible delay, makes an implicit assumption about how many other travelers will share the same route. As it is impossible for a traveler to precisely predict the congestion levels of the routes, establishing the optimal strategy using a point estimation of the traffic level might yield a strategy that will perform poorly under the different realizations of traffic conditions. This motivates a need for a *robustly optimal strategy* that is guaranteed to be optimal against fluctuations in traffic demands. To derive the robustly optimal strategy, first we need to formally characterize the demand uncertainty.

### 3.3. Demand characterization

We assume that the demand between any two nodes is estimated by the following range.

**Definition 1.** For any  $s \in S$  and  $t \in T$ ,  $\tilde{d}_{st} \in [\widehat{d}_{st} - \bar{d}_{st}, \widehat{d}_{st} + \bar{d}_{st}]$ , where  $\widehat{d}_{st} \geq \bar{d}_{st}$

The underlying demand uncertainty and systematic biases from opportunistic sensing are likely to cause some OD pairs to show greater variance than others. For example, areas that include stadiums or large institutional buildings like hospitals are associated with greater variations in demand, which can be modeled by a wider range. Examples of this are shown in the box plot, Fig. 4, wherein changes in variability for different OD pairs can be seen.

It is unrealistic to suppose that all demands are at the upper bounds simultaneously; therefore, the model makes no such assumption. Rather, a parameter  $\Gamma$  is introduced to account for the degree of demand variability. The possible demands as constrained by  $\Gamma$  eventually determine the degree to which the obtained solution should be robust; this explains why the following definition is often called the *budget of uncertainty* (Bertsimas and Sim, 2004). Note that the problem becomes a deterministic case when  $\Gamma \rightarrow \infty$  because it implies that all demands are at the upper bounds at the same time.

**Definition 2.** Budget of uncertainty: For a given  $\Gamma$ , we want the solution to be guaranteed valid for any demand realization in a set

$$D := \left\{ d \in \mathbb{R}_+^{|S||T|} \mid \begin{array}{l} d_{st} = \widehat{d}_{st} + \bar{d}_{st}\gamma_{st}, \\ \sum_{s \in S, t \in T} \gamma_{st} \leq \Gamma, \\ -1 \leq \gamma_{st} \leq 1, \forall s \in S, t \in T \end{array} \right\} \quad (1)$$

where  $\widehat{d}_{st}$  and  $\bar{d}_{st}$  are the nominal and maximum deviation demands, respectively.

### 3.4. Model development

Naturally the problem can be seen as a two-player game: operator vs. travelers. The operator decides which lines to operate for a given fleet of transports, where the number of vehicles operating and round trip time  $t_r$  on a given route  $r$  eventually determine the frequency of the route  $f_r$ . A transport can be a bus, subway, tram, or anything deployable for the public transportation. Then the feasible set of frequencies for a given set of routes  $R$  can be stated as

$$F := \left\{ f \in \mathbb{R}_+^{|R|} \mid \begin{array}{l} \sum_{r \in R} \frac{t_r}{60} f_r \leq C, \\ \underline{f}_r \leq f_r \leq \bar{f}_r, \quad \forall r \in R. \end{array} \right. \quad (2)$$

The first constraint ensures that the total number of vehicles in travel is not greater than the size of the fleet  $C$ . We assume that the frequency of a transport should be within  $[\underline{f}_r, \bar{f}_r]$  for any  $r \in R$ , which is guaranteed by the second constraints. When  $\underline{f}_r = \bar{f}_r = 0$ , it means the route  $r$  is not operated at all. Therefore, the frequency setting  $f$  determines which routes to employ and how frequently the routes in use will operate at the same time.

We assume the operator is a public sector body interested in social welfare and not profit maximization. The operator goal, in this case, is to find the optimal frequency setting  $f \in F$  that minimizes the system-wide travel time. That objective function can be stated as

$$\mathcal{F}(v, w, q) := \sum_{s \in S} \left( \underbrace{\sum_{a \in A} c_a \sum_{t \in T} \widehat{d}_{st} v_a^{st}}_{\text{journey time}} + \underbrace{\sum_{i \in V} w_i}_{\text{waiting time}} \right) + \beta \underbrace{\sum_{a \in A_{sv}} c_a q_a}_{\text{delay}} \quad (3)$$

The total travel time  $\mathcal{F}(v, w, q)$  is the sum of the total journey time, the total waiting time, and the total delay time due to over-congested service links. The parameter  $\beta$  controls the trade-off between the travel time and the congestion-incurred travel delay. Traffic volume  $v_a^{st}$  represents the fraction of flow from source node  $s$  to destination node  $t$  on arc  $a$ . Other terms, for example those related to operational costs, can also be included. However at this stage of transit planning, these costs are generally not known precisely without full roster information on crews and schedules. Note that the total journey time and total waiting time are considered with the nominal demands. This study focuses on the impact of demand uncertainty on the capacity because, without the capacity constraints, the optimal strategy of the individual traveler remains the same at higher demands. Moreover, the objective function implicitly depends on the frequency setting: the waiting time  $w$  is a function of the frequency setting because the frequency and optimal strategy uniquely determine the waiting time at the nodes.

Then, the feasible set of flow  $v$ , waiting time  $w$ , and delay time  $q$  for a given transport frequency  $f$  can be stated as

$$Y_f := \{(v, w, q) \in \mathbb{R}_+^{|A|+|\widehat{V}|+|S|+|A_{sv}|}\}$$

$$\sum_{a \in A_i^-} v_a^{st} - \sum_{a \in A_i^+} v_a^{st} = \begin{cases} 1, & \text{if } i = s, \\ -1, & \text{if } i = t, \forall i \in V, s \in S, t \in T, \\ 0, & \text{otherwise.} \end{cases}$$

$$\sum_{t \in T} \widehat{d}_{st} v_a^{st} \leq \sum_{r \in R} \delta_{ar}^s f_r w_{tail(a)}^s, \forall a \in A_{ac}, s \in S,$$

$$\sum_{s \in S} \sum_{t \in T} \widehat{d}_{st} v_a^{st} + \max_{\substack{0 \leq \gamma_{st} \leq 1, \forall s \in S, t \in T}} \left\{ \sum_{s \in S} \sum_{t \in T} \widehat{d}_{st} v_a^{st} \gamma_{st} \right\} \leq h_a \sum_{r \in R} \delta_{ar}^s f_r + q_a, \forall a \in A_{sv}$$

where  $\delta_a^r$  is 1 if  $a \in A_r$ , 0 otherwise. The above represents a traveler's set of feasible behaviors under a given operating frequency  $f$ . The first set of constraints, which is often referred to as flow-balance, ensures that all demand from the origins to destinations is served, i.e. no demand is lost. The second set of constraints relates the waiting time for a demand  $s$  at a node  $i$  to the frequencies of vehicles to ride. The last set of constraints measures the delay time on arc  $a$  according to the total sum of flows and the total capacity of the arc. Note that the capacity constraints are protected by the sum of worst-case demands defined by the budget of uncertainty. Due to this protection term, the traveler's behavior becomes robust to the variation of demands. In other words, the traveler is guaranteed to travel the optimal strategy route with a delay time less than or equal to  $q_a$ , which yields the robustly optimal strategy.

It is well-known that the last set of constraints defining  $Y_f$  can be stated as below by using the reformulation trick from [Bertsimas and Sim \(2004\)](#):

$$\sum_{s \in S} \sum_{t \in T} \widehat{d}_{st} v_a^{st} + \Gamma y_a + \sum_{s \in S} \sum_{t \in T} u_a^{st} \leq h_a \sum_{r \in R} \delta_{ar}^s f_r + q_a, \quad \forall a \in A_{sv},$$

$$y_a + u_a^{st} \geq \widehat{d}_{st} v_a^{st}, \quad \forall a \in A_{sv}, s \in S, t \in T,$$

$$y_a \geq 0, \quad \forall a \in A_{sv},$$

$$u_a^{st} \geq 0, \quad \forall a \in A_{sv}, s \in S, t \in T$$

which is a system of linear inequalities. We denote  $\tilde{Y}_f$  as  $Y_f$  with replacing the last constraints with the reformulated ones presented above. Let  $P_{Y_f}$  and  $\tilde{P}_{Y_f}$  denote the polyhedra of  $Y_f$  and  $\tilde{Y}_f$ , respectively. It is clear that the projection of  $\tilde{P}_{Y_f}$  onto the  $(v, w, q)$ -space is equivalent to  $P_{Y_f}$ .

Here, delays are assumed to grow linearly with link volumes. This approach to handling delays is close to the "line-specific overload delays" strategy proposed by [Hing-Keung Lam et al. \(1999\)](#), albeit in an equilibrium setting. Several other alternative approaches to handling the delays have been proposed in the literature, including: (a) a convex optimization by changing the link cost to a function of the flows, i.e.,  $c_a = c_a(v_a)$  (e.g., [Spiess and Florian, 1989](#)); (b) making the waiting time dependent on link flows (e.g., [De Cea and Fernández, 1993](#)) modeled route sections and used waiting links for each route section); (c) the "effective frequency" approach by which crowded buses have  $f \rightarrow 0$  (since it is harder for passengers to board them; when transports are empty (low congestion) the effective frequency becomes the same as the nominal frequency), and (d) deploying more advanced schemes, wherein capacity constraints  $\rightarrow$  higher delays  $\rightarrow$  higher boarding/dwell times  $\rightarrow$  lower frequencies.

The optimal strategy under the given frequency  $f$  minimizing the travelers' travel times can be obtained by solving

$$\min_{(v, w, q) \in Y_f} \{\mathcal{F}(v, w, q)\}. \quad (4)$$

### 3.5. Mathematical formulation

We now present two alternative formulations to solve the line-planning problem. The first one aims to jointly determine frequencies of services and flows, i.e., it makes the operator and traveler decisions jointly. The second formulation is a bi-level program that naturally separates those decisions.

#### 3.5.1. Nonlinear programming formulation

The optimal line planning problem can be stated as:

$$(P0) \min \sum_{s \in S} \left( \sum_{a \in A} c_a \sum_{t \in T} \widehat{d}_{st} v_a^{st} + \sum_{i \in \widehat{V}} w_i^s \right) + \beta \sum_{a \in A_{sv}} c_a q_a \quad (5)$$

$$\text{s.t.} \sum_{a \in A_i^-} v_a^{st} - \sum_{a \in A_i^+} v_a^{st} = \begin{cases} 1, & \text{if } i = s, \\ -1, & \text{if } i = t, \quad \forall i \in V, s \in S, t \in T, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$\sum_{t \in T} \widehat{d}_{st} v_a^{st} \leq \sum_{r \in R} \delta_{ar}^s f_r w_{tail(a)}^s, \quad \forall a \in A_{ac}, s \in S, \quad (7)$$

$$\sum_{s \in S} \sum_{t \in T} \widehat{d}_{st} v_a^{st} + \Gamma y_a + \sum_{s \in S} \sum_{t \in T} u_a^{st} \leq h_a \sum_{r \in R} \delta_{ar}^s f_r + q_a, \quad \forall a \in A_{sv}, \quad (8)$$

$$y_a + u_a^{st} \geq \widehat{d}_{st} v_a^{st}, \quad \forall a \in A_{sv}, s \in S, t \in T, \quad (9)$$

$$y_a \geq 0, \quad \forall a \in A_{sv}, \quad (10)$$

$$u_a^{st} \geq 0, \quad \forall a \in A_{sv}, s \in S, t \in T, \quad (11)$$

$$\sum_{r \in R} \frac{t_r}{60} f_r \leq C, \quad (12)$$

$$\underline{f}_r \leq f_r \leq \bar{f}_r, \forall r \in R, \quad (13)$$

$$v_a^{st} \geq 0, \forall a \in A, s \in S, t \in T, \quad (14)$$

$$w_i^s \geq 0, \forall i \in \widehat{V}, s \in S, \quad (15)$$

$$q_a \geq 0, \forall a \in A_{sv}, \quad (16)$$

where  $\delta_a^r$  is 1 if  $a \in A_r$ , 0 otherwise. The objective minimizes the total system travel time and is the same as Equation (3). Constraints (6)–(11) are from the user-behavior model  $Y_f$  with the reformulation presented in the previous section. Constraints (7) relate the frequency of vehicles  $f$  (operator decision) to the waiting time  $w$  (traveler-side decision). The nonlinearity of these constraints makes solving the problem difficult. One possible approach is using linearization techniques such as RLT ([Sherali and Shetty, 1980](#); [SheraliAmine, 1992](#)) and McCormick linearization ([McCormick, 1976](#)). However, those reformulation techniques would involve the adding of (possibly exponentially) many constraints and variables, while still not being able to guarantee the optimality of the solutions, because constraint (7) is not convex. Constraints (12) and (13) ensure the frequencies are feasible as defined in (2).

#### 3.5.2. Bi-level programming formulation

The two-player interpretation of the above problem makes the following alternative, bi-level-programming-based formulation possible:

$$(P1) \min_{f \in F} \mathcal{Z}(f)$$

$$\text{s.t.} \mathcal{Z}(f) = \min_{(v, w, q) \in Y_f} \mathcal{F}(v, w, q).$$

Here, the optimal value function  $\mathcal{Z}(f)$  is defined by an optimal solution of a lower-level optimization problem from the user perspective. The upper-level denotes the transit-operator-determined frequencies. The lower-level problem asserts the given frequencies by evaluating the total travel time and penalty for over-capacity on the link. Bi-level programming problems are generally NP-hard (Colson et al., 2005; Chinchuluun et al., 2009). Hereafter, we denote the upper-level problem as (P1)<sup>U</sup> while the lower-level problem as (P1)<sup>L</sup>. The problem (P1) can be seen as a decomposition of the problem (P0) so that both problems (P1)<sup>U</sup> and (P1)<sup>L</sup> are linear. Therefore, for the lower-level problem, the frequency setting  $f \in F$  should be given from the upper-level problem, and the optimal solution of the lower-level problem would provide an evaluation of the goodness of the frequency setting to the upper-level problem.

### 3.6. Solution approach for Bi-level programming formulation

The solution approach is summarized as follows. First, we set an initial frequency setting  $f_0$ ; then, a gradient direction to reduce the total travel time is identified by solving the traveler-side problem (P1)<sup>L</sup>, and the frequency  $f$  is updated using the descent direction and step size. This procedure is repeated until the frequency solution converges. It is noteworthy that the most computationally demanding step is obtaining the descent direction, which involves solving a large-sized user-behavior problem for the robustly optimal strategy. We propose the column generation based approach, which enables us to benefit from demand-wise decomposition and the use of very fast specialized algorithms for the shortest path problem. More detailed explanations follow below.

#### Algorithm 1. Frank-Wolfe algorithm for (P1)

---

**Algorithm 1** Frank-Wolfe algorithm for (P1)

---

```

1: procedure FRANKWOLFE
2:    $f_0 \leftarrow \underline{f}$ 
3:    $l \leftarrow 0$ 
4:   repeat
5:      $l \leftarrow l + 1$ 
6:     Obtain:  $\hat{\sigma}$  ▷ Theorem 1
7:     Solve problem:  $f^* \leftarrow \arg \min_{f \in F} \{-f^T \hat{\sigma}\}$  ▷ Direction-finding step.
8:     Solve problem:  $\alpha^* \leftarrow \arg \min_{0 \leq \alpha \leq 1} \{\mathcal{G}(\alpha f^* + (1 - \alpha)f^l)\}$  ▷ Step-length-finding step.
9:      $f^{l+1} \leftarrow \alpha^* f^* + (1 - \alpha^*) f^l$ 
10:    until  $\mathcal{G}(f^{l+1}) - \mathcal{G}(f^l) < \epsilon$ 
11: end procedure

```

---

We consider Algorithm 1 to solve (P1). Let  $\hat{\sigma}$  be a descent direction of the optimal value function  $\mathcal{Z}(f)$  at a given frequency  $\hat{f}$ . In order to use Algorithm 1, the  $\hat{\sigma}$  should be known for any  $f \in F$ , which is given by Theorem 1. For a given (or fixed)  $f$ , let  $\pi$ ,  $\phi$ , and  $\mu$  be the dual variables associated with constraints (7), (8), and the remaining constraints of the problem (P1)<sup>L</sup>, respectively.

**Theorem 1.** For any  $f \in F$ , let  $\hat{w}$  and  $(\hat{\pi}, \hat{\phi})$  be the primal and dual optimal solutions of (P1)<sup>L</sup>, respectively. Consider a vector.

$$\hat{\sigma} := \left[ - \sum_{s \in S} \sum_{a \in A_{ac}^s} \hat{w}_{tail(a)}^s \hat{\pi}_a^s - \sum_{a \in A_{sv}^s} h_a \hat{\phi}_a \right]_{r \in R}.$$

Then,  $\hat{\sigma}$  is a descent direction if  $\hat{\sigma} \neq 0$ . Moreover, the descent direction  $\hat{\sigma}$  becomes the negative of the gradient, i.e.  $-\nabla \mathcal{Z}(f)$ , if the primal and dual optimal solutions are unique, where  $\nabla \mathcal{Z}(f)$  is the gradient of  $\mathcal{Z}(f)$  at  $f \in F$ .

*Proof.* We let  $\hat{Y}_f \subset Y_f$  and  $\hat{Z}_f$  denote the set of primal and dual optimal solutions, respectively, of the problem  $\min_{(v,w,q) \in Y_f} \mathcal{F}(v, w, q)$  for the given  $f \in F$ .

For a given  $f \in F$ , we consider the Lagrangian function  $\mathcal{L}(f, v, w, q, \pi, \phi, \mu)$  for the problem  $\min_{(v,w,q) \in Y_f} \mathcal{F}(v, w, q)$ . Because, for a given  $f \in F$  the

problem  $\min_{(v,w,q) \in Y_f} \mathcal{F}(v, w, q)$  is an LP satisfying the constraint qualification, as shown in Still (2018) the directional derivative  $D_{\hat{\sigma}} \mathcal{Z}(f)$  in the direction  $\hat{\sigma}$  such that  $\sum_{r \in R} \hat{\sigma}_r^2 = 1$  at the point  $f$  is given as

$$D_{\hat{\sigma}} \mathcal{Z}(f) = \min_{(v,w,q) \in \hat{Y}_f} \max_{(\pi, \phi, \mu) \in \hat{Z}_f} \sum_{r \in R} \hat{\sigma}_r \frac{\partial \mathcal{L}(f, v, w, q, \pi, \phi, \mu)}{\partial f_r}.$$

Note that  $\mathcal{L}(f, v, w, q, \pi, \phi, \mu)$  can be seen as a function of  $f$ , which gives the partial derivatives of  $\mathcal{L}(f, v, w, q, \pi, \phi, \mu)$ . For any  $r \in R$ , we have,

$$\frac{\partial \mathcal{L}(f, v, w, q, \pi, \phi, \mu)}{\partial f_r} = \sum_{s \in S} \sum_{a \in A_{ac}^s} w_{tail(a)}^s \pi_a^s + \sum_{a \in A_{sv}^s} h_a \phi_a \leq 0$$

because  $\pi_a^s \leq 0$  and  $\phi_a \leq 0$ .

Now we consider a pair of primal and dual optimal solutions  $(\hat{v}, \hat{w}, \hat{q}) \in \hat{Y}_f$  and  $(\hat{\pi}, \hat{\phi}, \hat{\mu}) \in \hat{Z}_f$ . We claim that  $\hat{\sigma} \in \mathbb{R}_+^{|R|}$  defined as below is a descent direction if  $\hat{\sigma}_r \neq 0$ :

$$\hat{\sigma}_r := - \frac{\partial \mathcal{L}(f, \hat{v}, \hat{w}, \hat{q}, \hat{\pi}, \hat{\phi}, \hat{\mu})}{\partial f_r} = - \sum_{s \in S} \sum_{a \in A_{ac}^s} \hat{w}_{tail(a)}^s \hat{\pi}_a^s - \sum_{a \in A_{sv}^s} h_a \hat{\phi}_a, \quad \forall r \in R.$$

Then, we need to show that

$$D_{\hat{\sigma}} \mathcal{Z}(f) = \min_{(v,w,q) \in \hat{Y}_f} \max_{(\pi, \phi, \mu) \in \hat{Z}_f} \sum_{r \in R} \hat{\sigma}_r \frac{\partial \mathcal{L}(f, v, w, q, \pi, \phi, \mu)}{\partial f_r} < 0,$$

which implies that  $\hat{\sigma}$  is a descent direction. Due to (B), it is clear that  $D_{\hat{\sigma}} \mathcal{Z}(f) \leq 0$ . Assume, for a contradiction, that  $D_{\hat{\sigma}} \mathcal{Z}(f) = 0$ . Then, there must be  $\tilde{r} \in R$  such that

$$\hat{\sigma}_{\tilde{r}} > 0 \text{ and } \frac{\partial \mathcal{L}(f, \hat{v}, \hat{w}, \hat{q}, \tilde{\pi}, \tilde{\phi}, \tilde{\mu})}{\partial f_{\tilde{r}}} = 0, \text{ for some } (\tilde{\pi}, \tilde{\phi}, \tilde{\mu}) \in \hat{Z}_f,$$

which implies that  $\tilde{\pi}_a^s = 0$  for all  $s \in S$  and  $a \in A_{ac}^{\tilde{r}}$  such that  $\hat{w}_{tail(a)}^s > 0$ .

Note that  $\mathcal{L}(f, \hat{v}, \hat{w}, \hat{q}, \tilde{\pi}, \tilde{\phi}, \tilde{\mu})$  has terms

$$\tilde{\pi}_a^s \left( \sum_{r \in R} \delta_a^{\tilde{r}} f_r \hat{w}_{tail(a)}^s - \sum_{t \in T} \hat{d}_{st} \hat{v}_a^{st} \right)$$

for  $s \in S$  and  $a \in A_{ac}^{\tilde{r}}$ . Because  $\tilde{\pi}_a^s = 0$  for all  $s \in S$  and  $a \in A_{ac}^{\tilde{r}}$  such that  $\hat{w}_{tail(a)}^s > 0$ , there should be  $(\tilde{\pi}, \tilde{\phi}, \tilde{\mu}) \in \hat{Y}_f$  such that  $\mathcal{L}(f, \tilde{v}, \tilde{w}, \tilde{q}, \tilde{\pi}, \tilde{\phi}, \tilde{\mu}) < \mathcal{L}(f, \hat{v}, \hat{w}, \hat{q}, \hat{\pi}, \hat{\phi}, \hat{\mu})$ , which means that  $(\tilde{v}, \tilde{w}, \tilde{q}) \notin \hat{Y}_f$ . This gives a contradiction.

Note that (A) becomes

$$D_{\hat{\sigma}} \mathcal{Z}(f) = \sum_{r \in R} \hat{\sigma}_r \frac{\partial \mathcal{L}(f, \hat{v}, \hat{w}, \hat{q}, \hat{\pi}, \hat{\phi}, \hat{\mu})}{\partial f_r}$$

when the optimal solution  $(\hat{v}, \hat{w}, \hat{q}, \hat{\pi}, \hat{\phi}, \hat{\mu})$  is unique, i.e.,  $|\hat{Y}_f| = |\hat{Z}_f| = 1$ , which implies that  $\hat{\sigma} = -\nabla \mathcal{L}(f)$ .

**Theorem 1** implies that  $\hat{\sigma}$  at any  $f$  can be obtained easily from the primal and dual optimal solutions of  $(P1)^L$ . Note that the descent direction may depend on the choice of a pair of primal and dual optimal solutions  $(\hat{v}, \hat{w}, \hat{q}) \in \hat{Y}_f$  and  $(\hat{\pi}, \hat{\phi}, \hat{\mu}) \in \hat{Z}_f$ . For this study, we take the primal and dual optimal solutions of the problem  $(P1)^L$  solved by the dual Simplex method. Then, the direction finding problem at  $f^l$  can be stated as follows.

$$(DF) \min - \sum_{r \in R} \hat{\sigma}_r f_r \quad (17)$$

$$\text{s.t. } \sum_{r \in R} \frac{f_r}{60} \leq C, \quad (18)$$

$$\underline{f}_r \leq f_r \leq \bar{f}_r, \forall r \in R. \quad (19)$$

The above is an LP problem that can be solved easily by LP solvers. Moreover, the problem is also a variant of the fractional Knapsack problem which can be solved in  $O(|R|\log|R|)$  as shown in the following theorem.

**Theorem 2.** (DF) can be solved in  $O(|R|\log|R|)$ .

*Proof.* Since (DF) is a fractional Knapsack problem, an optimal solution is obtained by selecting the maximum possible  $f_r$  for each  $r \in R$  in decreasing order of  $\frac{\hat{\sigma}_r}{60}$ , where the sorting is done in  $O(|R|\log|R|)$ .

For the step length finding problem, using discretized values between 0 and 1 for  $\alpha$ , we solved the problem

$$\min_{\alpha \in \{0, 0.1, 0.2, \dots, 1\}} \{ \mathcal{L}(\alpha f^* + (1-\alpha)f^l) \}, \quad (20)$$

whose optimal solution is denoted as  $\alpha^*$ , where  $f^*$  is the optimizer of (DF). This implies that we need to solve  $(P1)^L$  many times. Recall that problem  $(P1)^L$  is to find an optimal strategy with the capacity constraints under demand uncertainty, which minimizes the total travel time, for a given frequency  $f$ . If there are no capacity constraints, the problem  $(P1)^L$  is separable by the source nodes  $s \in S$ , which results in  $|S|$ -independent problems for the optimal strategies. For these problems, we may use a dynamic programming algorithm developed by Spiess and Florian (1989), which is very fast. Unfortunately, the problem is no longer separable with capacity constraints, so their approach cannot be used.

On the other hand, the problem  $(P1)^L$  can be stated by removing (12)-(13) and fixing  $f$  in (P0). According to the linear programming duality, the inner maximization problem in constraints (8) can be reformulated (Bertsimas and Sim, 2004). Then, we have the following formulation, called the ‘‘flow-based’’ formulation.

$$(G) \min (5)$$

$$\text{s.t. } (6), (7), (14), (15), (16),$$

$$\sum_{s \in S} \sum_{t \in T} \hat{d}_{st} v_a^{st} + \Gamma y_a + \sum_{s \in S} \sum_{t \in T} u_a^{st} \leq h_a \sum_{r \in R} \delta_{ar}^r f_r + q_a, \quad \forall a \in A_{sv}, \quad (21)$$

$$y_a + u_a^{st} \geq \bar{d}_{st} v_a^{st}, \quad \forall a \in A_{sv}, s \in S, t \in T, \quad (22)$$

$$y_a \geq 0, \quad \forall a \in A_{sv}, \quad (23)$$

$$u_a^{st} \geq 0, \quad \forall a \in A_{sv}, s \in S, t \in T. \quad (24)$$

Note that (G) is a linear programming problem with a fixed  $f$ . However, the size of the problem grows rapidly with the size of the network and especially

with the number of demands. Specifically, each demand pair  $(s, t)$  increases the size of the problem by  $|A| + |A_{sv}|$ . A typical transit network should consider a large number of demands; consequently the size of the problem may be too large to solve as it is. To overcome this, we propose a column generation based approach in the next section.

### 3.6.1. Column generation for solving (G)

For a given source and sink pair  $s - t$ , any feasible solution of (G) can be represented as a union of paths between  $s$  and  $t$  due to the flow balance constraints (6). Let  $P_{st}$  be a set of all paths between  $s$  and  $t$ . Then it is easily seen that the following relation

$$v_a^{st} = \sum_{p \in P_{st}} \sigma_a^p z_p^{st},$$

holds, where  $\sigma_a^p = 1$  if path  $p \in P_{st}$  uses arc  $a$ , 0 otherwise. We introduce new extended variables  $0 \leq z_p^{st} \leq 1$  representing the fraction of demand between  $s$  and  $t$  using path  $p$ . By using the above relation, we consider the extended formulation.

$$(EG) \min \sum_{s \in S} \left( \sum_{a \in A} c_a \sum_{t \in T} d_{st} \sum_{p \in P_{st}} \sigma_a^p z_p^{st} + \sum_{i \in \hat{V}} w_i^s \right) + \beta \sum_{a \in A_{sv}} c_a q_a \quad (25)$$

$$\text{s.t. } \sum_{p \in P_{st}} z_p^{st} = 1, \quad \forall s \in S, t \in T, \quad (26)$$

$$\sum_{t \in T} d_{st} \sum_{p \in P_{st}} \sigma_a^p z_p^{st} \leq \sum_{r \in R} \delta_{ar}^r f_r w_{tail(a)}^s, \quad \forall a \in A_{ac}, s \in S, \quad (27)$$

$$\sum_{s \in S} \sum_{t \in T} \hat{d}_{st} \sum_{p \in P_{st}} \sigma_a^p z_p^{st} + \Gamma y_a + \sum_{s \in S} \sum_{t \in T} u_a^{st} \leq h_a \sum_{r \in R} \delta_{ar}^r f_r + q_a, \quad \forall a \in A_{sv}, \quad (28)$$

$$y_a + u_a^{st} \geq \bar{d}_{st} \sum_{p \in P_{st}} \sigma_a^p z_p^{st},$$

$$\forall a \in A_{sv}, s \in S, t \in T, \quad (29)$$

$$y_a \geq 0, \quad \forall a \in A_{sv}, \quad (30)$$

$$u_a^{st} \geq 0, \quad \forall a \in A_{sv}, s \in S, t \in T, \quad (31)$$

$$z_p^{st} \geq 0, \quad \forall a \in A, s \in S, t \in T, \quad (32)$$

$$w_i^s \geq 0, \quad \forall i \in \hat{V}, s \in S, \quad (33)$$

$$q_a \geq 0, \quad \forall a \in A_{sv}. \quad (34)$$

Since there can be exponentially many paths, we incorporate the column generation based approach. Consider a column generation master problem with a limited set of paths. Let  $(\hat{\theta}, \hat{\pi}, \hat{\phi}, \hat{\lambda})$  be the dual optimal solution of the column generation master problem, where  $\theta$ ,  $\pi$ ,  $\phi$ , and  $\lambda$  are the dual variables associated with (26), (27), (28), and (29), respectively. The column generation master problem is optimal if the condition

$$\hat{d}_{st} \left( \sum_{a \in A(p)} c_a - \sum_{a \in A(p) \cap A_{ac}} \hat{\pi}_a^s - \sum_{a \in A(p) \cap A_{sv}} \hat{\phi}_a + \frac{\bar{d}_{st}}{\hat{d}_{st}} \sum_{a \in A(p) \cap A_{sv}} \hat{\lambda}_a^{st} \right) - \hat{\theta}_{st} \geq 0, \quad \forall p \in P_{st}, s \in S, t \in T,$$

holds, where  $A(p) \subseteq A$  is a set of arcs in path  $p$ .

Then, the column generation sub-problem for demand  $s - t$  is given as follows:

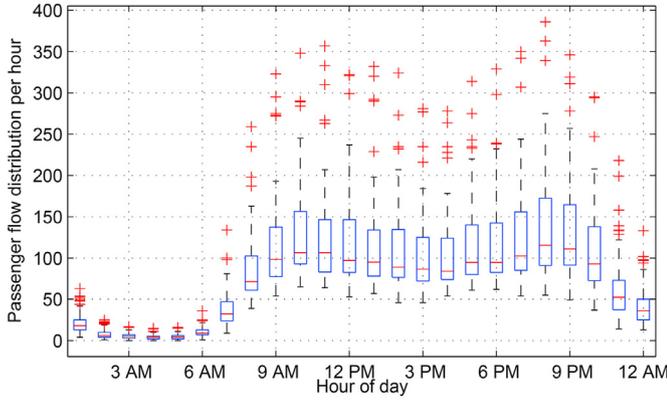


Fig. 3. Flow distribution for top 100 OD flows by hour of day.

$$(\text{SP})_{\xi_{st}} = \min \sum_{a \in A(p)} \tilde{c}_a \quad (35)$$

$$\text{s.t. } p \in P_{st}, \quad (36)$$

where

$$\tilde{c}_a = \begin{cases} c_a - \hat{\pi}_a^s, & \text{if } a \in A_{ac}, \\ c_a - \hat{\phi}_a + \frac{\bar{d}_{st} \hat{\lambda}_a^{st}}{\bar{d}_{st}}, & \text{if } a \in A_{sv}, \\ c_a, & \text{otherwise.} \end{cases}$$

Note that  $\tilde{c}_a$  is nonnegative for any  $a \in A$  provided that  $c_a \geq 0$  since  $\hat{\pi} \leq 0$  and  $\hat{\phi} \leq 0$ . A new path  $p^*$  is added to the master problem if  $d_{st} \xi_{st} < \hat{\theta}_{st}$ , where  $p^*$  is an optimizer of  $\xi_{st}$ .

Since every cost of the graph for the column generation problem is nonnegative, very fast shortest path algorithms such as Dijkstra's can be used. In practice, it is sufficient to solve the column generation subproblem  $|S|$  times, because the Dijkstra's algorithm will provide all shortest paths from a source to all destinations.

### 3.6.2. Early termination of column generation

Algorithm 1 entails solving (EG) many times. Specifically, at each iteration, we solve problem (EG) twelve times by changing  $f$  (one for obtaining  $\nabla \mathcal{Z}(f^l)$  and eleven for the step-length search, since the step-length problem is solved for discrete values in the unit interval). Therefore, we keep all generated columns so that the subsequent column generation will require only a small number of new columns. Moreover, the column generation can be terminated early if the condition

$$\tilde{Z}_{EG} + \sum_{s \in S, t \in T} (d_{st} \xi_{st} - \hat{\theta}_{st}) \geq \hat{Z}, \quad (37)$$

holds, where  $\tilde{Z}_{EG}$  and  $\hat{Z}$  are the objective value of the current column generation master problem and the best objective value (i.e., incumbent solution), respectively. The validity of the above condition is easily seen from the so-called Lagrangian bound (i.e., the optimal value of a column generation master  $\leq$  the current objective value of the column generation master + the minimum of reduced costs).

## 4. The case of Abidjan, Côte d'Ivoire

The mobile-phone data, made available by Orange (2014), consisted of 2.5 billion call-detail records (CDRs) for Côte d'Ivoire. A CDR is generated each time a call or a text-message transaction occurs. Around the city of Abidjan, this dataset consisted of anonymized CDRs for half a million users over an arbitrary two-week period within a five month

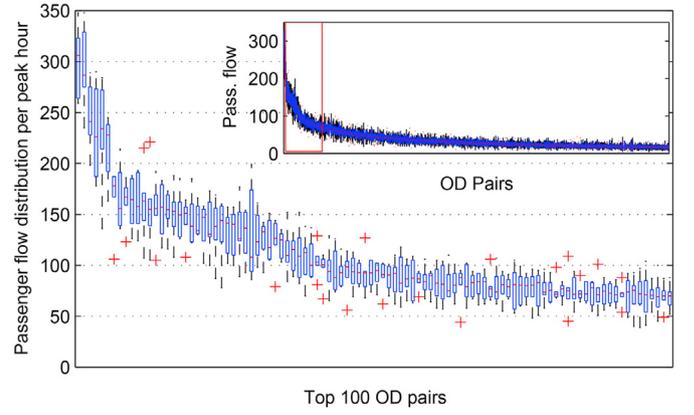


Fig. 4. Box plot of passenger flow distribution for top 100 OD flows (inset shows top 1000 OD flows).

interval.

Each CDR contains a time and location when and where a call/text transaction occurred along with an associated ID. The spatial component is encoded as the mobile-phone antenna that handled the transaction: typically, but not necessarily, the closest antenna to the user. In the city of Abidjan, a set of 407 antennas are considered. These towers, which are analogous to Traffic Analysis Zones (TAZ), serve as the reference for production and attraction zones for travel. Each individual in the dataset therefore represents a sequence of time-space events between these zones.

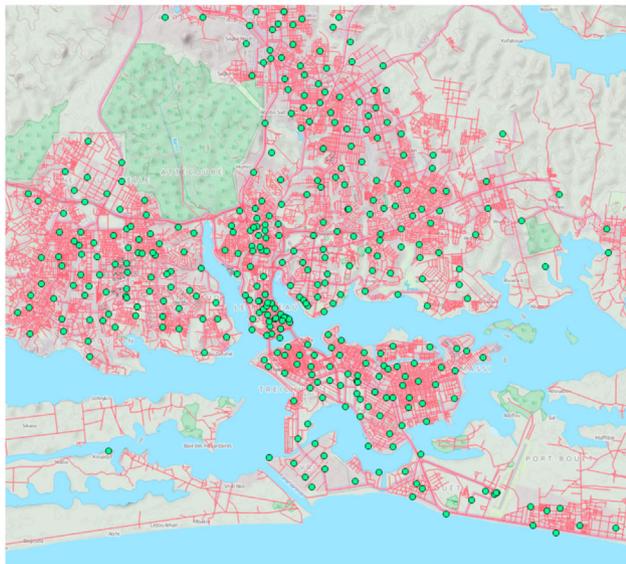
CDRs by themselves are mostly infrequent snapshots of disaggregated behavior. (Berlingerio et al., 2013) demonstrated, for the same dataset, how such infrequent snapshots can be transformed into (partial) activity sequences. The main intuition presented therein, is that repeat observations at the disaggregated level allow for activity-sequence reconstruction, even if the constituent daily activity patterns are partial.

At the aggregate level, this variability detected in individual behaviors along with their likelihood, translates to demand, as a distribution, between any two points on the network. The same OD matrix derived by Berlingerio et al. (2013) is used here, with the exception that the range values are specified. Readers are referred to their paper for the data-extraction algorithms used. Fig. 4 shows the variability of the top 100 OD pairs during peak periods along with a box plot. As shown in Fig. 3, the data during the peak periods have a higher spread.

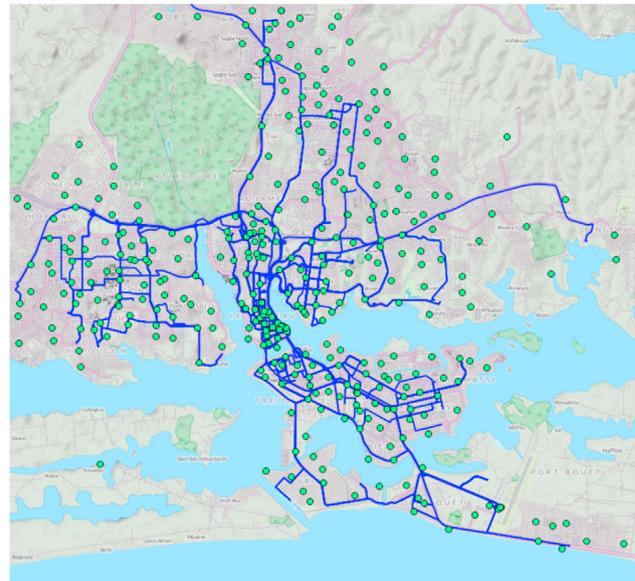
In order to specify the random demand variable  $\tilde{d}_k$ , as a range, for each of the 107,761 OD pairs derived from the CDR data, the median of the CDR-based flows is considered as  $\hat{d}_k$  and one standard deviation as  $\bar{d}_k$ . In practice, some scale up of demand as a function of mobile operator market share as well as population statistics such as those suggested in the literature can be considered (Alexander et al., 2014).

From the supply-side, Sub-Saharan cities across Africa, with the exception of Lagos, have seen organized public transport deteriorate over the past few decades (Gwilliam, 2011). Due to cost inefficiencies and higher fares, public operators have ceded ground to informal operators who have the majority of the transport mode share. In Abidjan, a city with 4.5 million citizens, the public operator SOTRA has a fleet of 439 buses. This is complemented by roughly 5,000 mini-buses and roughly 11,000 shared taxis (*woro-woro*).

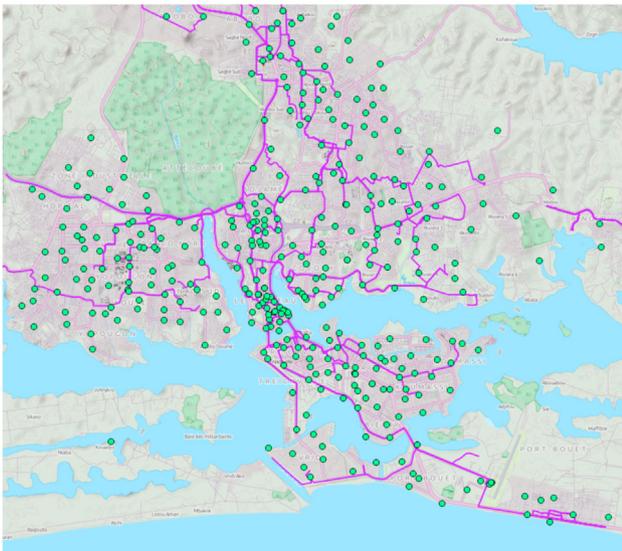
Data on the 92 SOTRA routes were considered as the baseline network shown in Fig. 5b. The frequencies of these services were extracted from their public website. Additionally, a set of future potential routes, called the candidate route set, were considered, as shown in Fig. 5c. Roughly half of those routes were generated based on high-density corridors as observed in the CDR data (Berlingerio et al., 2013); the other half of the candidates were generated by a shortest path algorithm between zones with high volume of passenger flows.



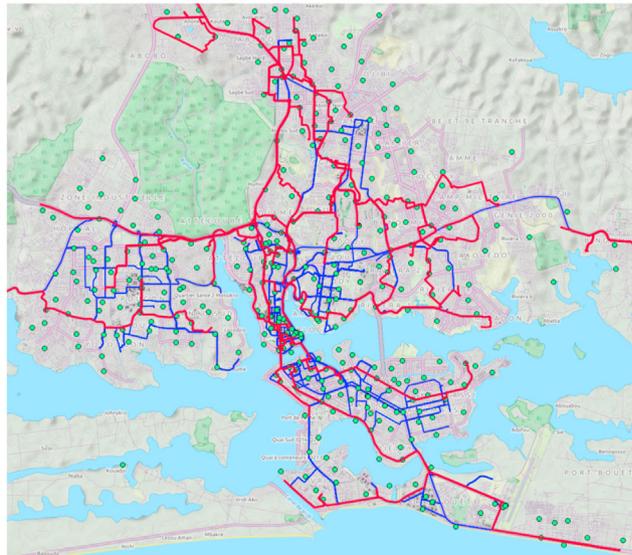
(a) Cell tower locations



(b) Existing routes



(c) Candidate routes



(d) Optimized routes with  $\Gamma = 200, C = 300$ , see Section 5 for details

Fig. 5. Abidjan mobile-phone towers and public transport network.

Table 1

Comparison of computational times of the proposed approach with those of two alternative methods for finding the gradient direction. ( $C = 300, \beta = 1, h = 100$ ) it\* = iterations.

K	$\Gamma$	Column Generation			LP-based			Flow-based
		FW it*	CG it*	Columns	LP Time (s)	SP Time (s)	Tot. time (s)	Tot. time (s)
100	0	12	469	1526	13.5	25.9	111.3	276.7
	20	14	823	2800	105.6	45.1	270.0	205.5
500	0	8	933	9232	127.0	250.2	647.5	2371.3
	100	9	1142	14961	1045.8	307.7	1875.9	7597.2
1000	0	14	1267	19128	289.7	659.3	1638.5	11841.2
	200	11	1066	25747	2079.4	564.1	3666.9	29361.5

### 5. Computational results

Experiments were designed to demonstrate the solvability of the

model presented in Section 3 and the computational aspects of the solution procedure. A Monte-Carlo simulation was employed to demonstrate the benefit of the robust solution over the deterministic

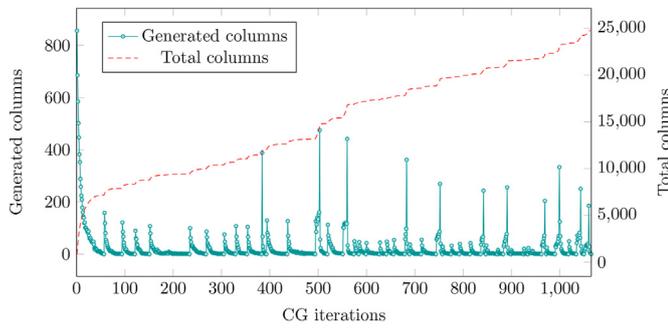


Fig. 6. Numbers of generated columns during the column generation iterations. ( $|K| = 1000, C = 300, \Gamma = 200, h = 100, \beta = 1$ ).

alternatives. All of the computational experiments were conducted on a MacPro machine with a 3.5 GHz 6-Core Intel Xeon E5 and 32GB RAM. The algorithm was implemented in Python, and CPLEX 12.6 was used as the linear programming solver. The total number of ODs estimated from the data was 107,761. Of these, roughly 15,000 ODs had non-negligible flows; the rest were discarded. This large number of ODs rendered the problem virtually impossible to solve. To address this, a clustering procedure was employed to yield aggregate demands of 100, 500 and 1000 OD pairs. The clustering was performed by merging geographically proximate OD flows.

Table 1 compares the computation times of the proposed algorithm with two alternative methods for finding the gradient direction (i.e., column generation, (EG), and the flow-based formulation, (G)). The number of Frank-Wolfe iterations and column generation iterations are reported in the columns “FW iterations” and “CG iterations”, respectively. Recall that we need to solve (EG) many times by trying different  $f$  (i.e., Frank-Wolfe iterations). It is noteworthy that “CG iterations,” represents the total sum of column generation iterations (not a single run of (EG)). The column “time (LP)” represents the time spent in solving the column generation master problem, while the column “time (SP)” indicates the time spent solving the column generation subproblems. It is clearly shown that the proposed column generation based approach is significantly faster than the flow-based formulation, and that the performance difference becomes more significant with the number of demands ( $|K|$ ). Another interesting thing to note is that with  $\Gamma > 0$ , the time for solving the column generation master problem (“time (LP)”) becomes dominant, which is understandable if the increased size of the column generation master problem is noted. We found that the CPLEX’s Python library is not so efficient in retrieving the dual solution values of the column generation master problem, which results in a large difference

Table 2

Reductions of total travel times from the baseline assignments. ( $|K| = 1000, \Gamma = 200, \beta = 1, h = 100$ ).

C	Total travel times		Reduction (%)
	Baseline	Optimized	
50	154,531,965.2	149,878,986.5	3.0
100	153,487,059.4	146,647,194.9	4.5
150	152,893,349.5	143,550,639.9	6.1
200	152,466,981.9	141,683,777.5	7.1
250	152,128,481.0	134,940,946.6	11.3
300	151,770,179.3	132,783,104.7	12.5

between the “total time” and the sum of “time (LP)” and “time (SP).” This implies that a more efficient implementation with a native CPLEX interface such as C or C++ would make the performance of the column generation even more dominant.

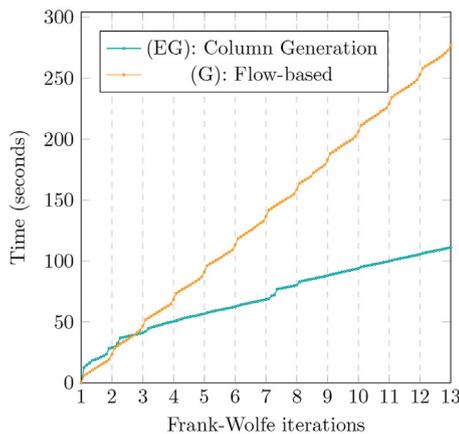
Fig. 6 illustrates the numbers of generated columns during the column generation iterations. It is clearly seen that in the early iterations, many columns are generated. The progress of the column generation becomes faster after the early iterations, which makes subsequent solving of (EG) even faster. The “peaks” shown in the figure appear when the frequency setting  $f$  was changed significantly. Since, during the Frank-Wolfe iterations, we solve (EG) repeatedly by changing  $f$ , a considerably different  $f$  would result in significantly different optimal strategies for travelers, which would require many new columns.

Fig. 7 plots the computational time increments over the Frank-Wolfe iterations. The left of the figure is obtained from a small-size problem ( $|K| = 100$ ), while the right is obtained from a large-sized problem ( $|K| = 1000$ ). The stair-like behavior is due to the significant changes of  $f$  (i.e., step 9 of Algorithm 1). It is clearly seen that the column generation outperforms the flow-based formulation, and that the benefit of column generation increases with the progress of the Frank-Wolfe iterations. This is due to (i) the fact that only a small number of new columns needs be generated after the initial generation, and (ii) the frequent early terminations of column generation.

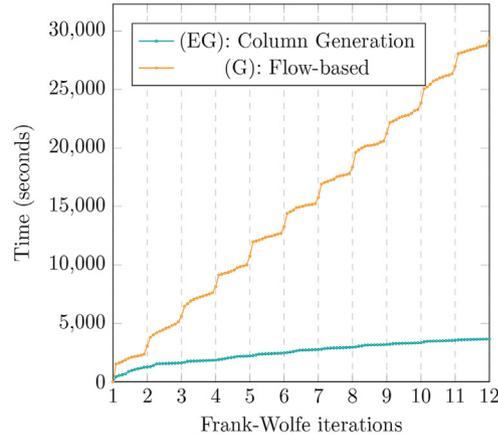
Table 2 summarizes the reductions of total travel times obtained from the proposed approach. The reduction is calculated from the baseline assignment

$$f_r = \frac{60 \times C}{|\hat{R}| \times t_r}, \text{ for all } r \in \hat{R},$$

where  $\hat{R}$  is a set of new candidate routes. Note that the baseline assignment means that we distribute the vehicles evenly to all new routes. The results show that, as expected, the reduction becomes larger as the



(a)  $|K| = 100, \Gamma = 0, C = 300, h = 100, \beta = 1$



(b)  $|K| = 1000, \Gamma = 200, C = 300, h = 100, \beta = 1$

Fig. 7. Increments of computational time over Frank-Wolfe iterations.

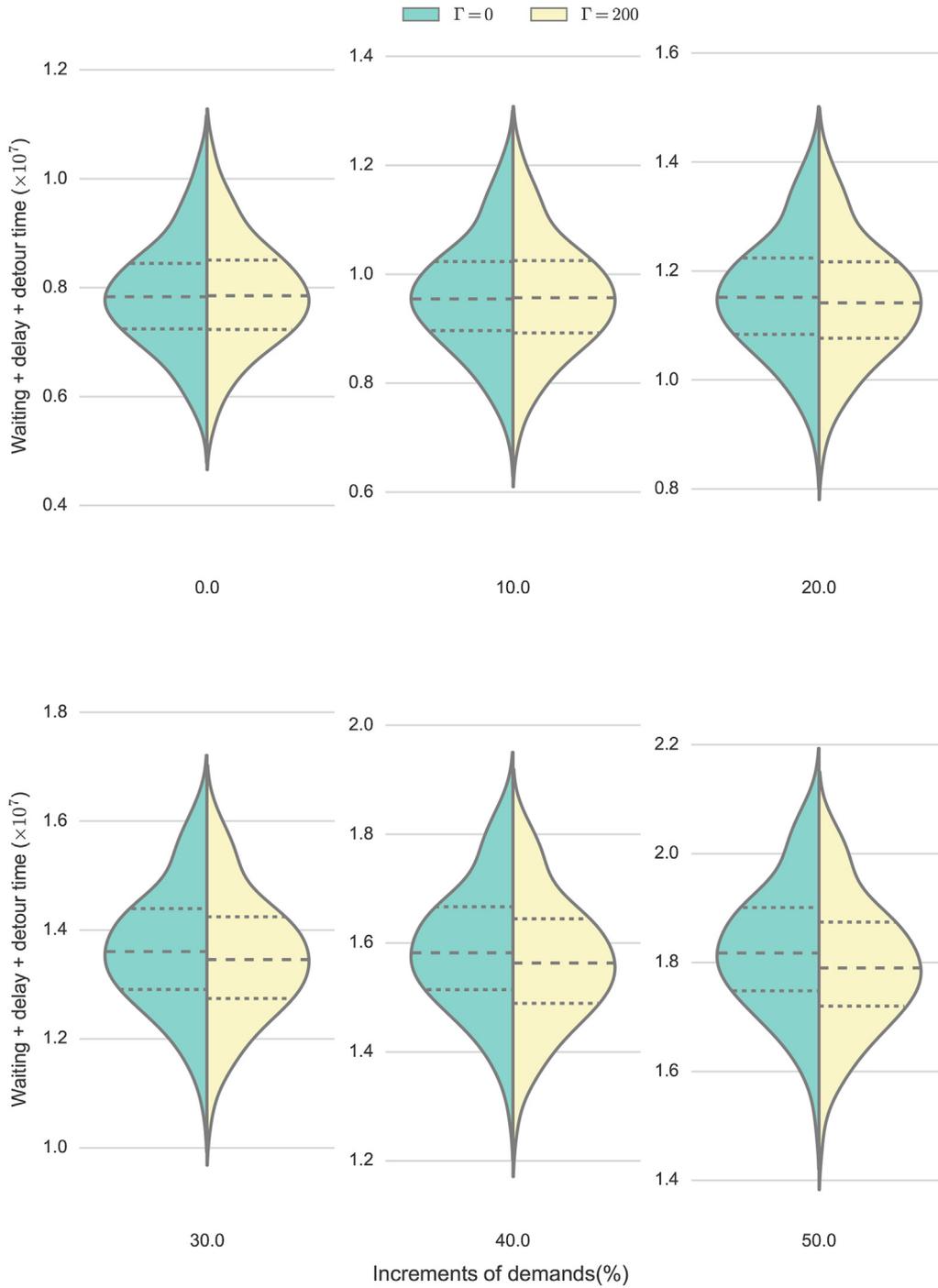


Fig. 8. Monte-Carlo simulations with varying demands. ( $|K| = 1000, C = 300, h = 100, \beta = 1$ ).

number of new vehicles deployed increases.

**Robustness of Solutions:** To measure the robustness of solutions obtained from the proposed approach, we conducted a Monte-Carlo simulation test. From the nominal demands  $\hat{d}_k$  and maximum deviation  $\bar{d}_k$ , we generated 100 random demand scenarios following a normal distribution  $\mathcal{N}(\hat{d}_k, \bar{d}_k)$  for each  $k \in K$ . Then, we evaluated the deterministic solution ( $\Gamma = 0$ ) and the robust solution ( $\Gamma = 200$ ) with all demand scenarios populated.

It is noteworthy that each demand scenario has a theoretical low bound of the total travel time that can be calculated by adopting the assumption of infinitely many vehicles. Having infinitely many vehicles means that there is no waiting time and no delays, which implies that the

total travel time is the sum of the shortest of all demands. For a given demand scenario, we calculate this theoretical low bound and subtract it from the optimal objective value, which yields the “inefficiency” measure of the solution (i.e., the extent to which the objective value of the solution is worse than the theoretical best outcome for the given demand scenario). The inefficient measure consists of the waiting time, delay time, and additional detour time of travelers to avoid waiting and/or delays.

Fig. 8 plots the results of the Monte-Carlo simulation, comparing the deterministic solution ( $\Gamma = 0$ ) and the robust solution ( $\Gamma = 200$ ) for different cases of increasing demand. When the demand is not increased, the robust solution results in a slightly worse performance than the deterministic solution. However, with increasing demand, the robust solution performs better than the deterministic solution.

## 6. Conclusion

A robust transit-network design model, designed for cases in which travel demands are opportunistically sensed and estimated as ranges, is presented. Using demand data obtained from mobile-phone CDRs, the model is demonstrated for the city of Abidjan. The main results show an improvement of system-wide journey times as functions of the uncertainty budget and random realizations of demand.

The study has the following limitations. We make the assumption that all demand sensed using such methods can potentially be served by transit. The absence of a mode preference model (due to the absence of socio-demographic information associated with the mobile phone trajectories) can be assumed to increase the range uncertainty. This was not specifically considered. A second limitation is the lack of demand elasticity. The measured demand estimates are based on current supply conditions and not on future offerings. The elastic demand case is not considered here, since estimation of elasticity from such data is an open problem. The budget of uncertainty  $\Gamma$  should be carefully tuned, since this impacts the range of demands for which the robust solution is valid. The model currently assumes no change in waiting time and travel time on account of demand changes, and only considers the impact on capacity constraints. Such considerations can be relaxed to include other effects, but will make the model size even larger. Other sources of uncertainties also need to be addressed. For example, the travel routes of travelers might become dramatically different if the uncertainties of travel time and/or the operation frequencies of vehicles are considered. Any of these calls would serve as valuable future research directions. Concerning the algorithmic aspects, one should note that the computational study conducted in this paper is heavily focused on analyzing the real-world use case. This means the algorithm has been developed and tuned to tackle a specific problem instance, so the results might not be directly translated into other cases without further modifications and improvements.

There is huge potential for big-data utilization in the field of urban mobility. A more detailed characterization of demand processes under this regime is necessary if these sources of information are to impact mobility-resource decisions. Supply-side planning and the operational models could then fully leverage such information.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2018R1A1A1A05077775). Chungmok Lee was additionally supported by Hankuk University of Foreign Studies Research Fund. The authors thank Orange for sharing data for Côte d'Ivoire. We thank Anne-Marie Cromack for editing the manuscript.

## References

Alexander, Lauren, Jiang, Shan, Murga, Mikel, González, Marta C., 2014. Validation of Origin-Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data submitted for publication.

Amiripour, SM Mahdi, Avi Ceder, Avishai, Shariat Mohaymany, Afshin, 2014. Designing large-scale bus network with seasonal variations of demand. *Transport. Res. C Emerg. Technol.* 48, 322–338.

Asadi Bagloee, Saeed, Avi Ceder, Avishai, 2011. Transit-network design methodology for actual-size road networks. *Transp. Res. Part B Methodol.* 45 (10), 1787–1804.

Bastani, Favien, Huang, Yan, Xie, Xing, Jason, W., Powell, 2011. A greener transportation mode: flexible routes discovery from gps trajectory data. In: *Proceedings of the 19th*

ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '11. ACM, New York, NY, USA, pp. 405–408.

Ben-Tal, Aharon, Nemirovski, Arkadi, February 2007. Selected topics in robust convex optimization. *Math. Program.* 112 (1), 125–158.

Berlingerio, Michele, Calabrese, Francesco, Di Lorenzo, Giusy, Nair, Rahul, Pinelli, Fabio, Luca Sbodio, Marco, 2013. Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In: *Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 663–666.

Bertsimas, Dimitris, Sim, Melvyn, 2004. The price of robustness. *Oper. Res.* 52 (1), 35–53.

Borndörfer, Ralf, Grötschel, Martin, Pfetsch, Marc E., 2007. A column-generation approach to line planning in public transport. *Transport. Sci.* 41 (1), 123–132.

Calabrese, F., Pereira, F., DiLorenzo, G., Liu, L., Ratti, C., 2010. The geography of taste: analyzing cell-phone mobility and social events. In: *International Conference on Pervasive Computing*, pp. 22–37.

Calabrese, F., Di Lorenzo, G., Liu, Liang, Ratti, C., april 2011. Estimating origin-destination flows using mobile phone location data. *Pervasive Comput., IEEE 10 (4)*, 36–44.

Calabrese, Francesco, Di Lorenzo, Giusy, McArdle, Gavin, Pinelli, Fabio, Van Lierde, Erik, 2015. Real-time social event analytics. *Netmob'* 15.

Ceder, Avishai, 2016. *Public Transit Planning and Operation: Modeling, Practice and Behavior*. CRC press.

Ceder, Avishai, Wilson, Nigel HM., 1986. Bus network design. *Transp. Res. Part B Methodol.* 20 (4), 331–344.

Chen, Chao, Zhang, Daqing, Zhou, Zhi-Hua, Li, Nan, Atmaca, Tülin, Li, Shijian, 2013. B-planner: night bus route planning using large-scale taxi gps traces. In: *Pervasive Computing and Communications (PerCom), 2013 IEEE International Conference*. IEEE, pp. 225–233.

Chinchuluun, Altannar, Pardalos, Panos M., Huang, Hong-Xuan, 2009. Multilevel (hierarchical) optimization: complexity issues, optimality conditions, algorithms. In: *Advances in Applied Mathematics and Global Optimization*. Springer, pp. 197–221.

Cipriani, Ernesto, Gori, Stefano, Petrelli, Marco, 2012. Transit network design: a procedure and an application to a large urban area. *Transport. Res. C Emerg. Technol.* 20 (1), 3–14.

Colson, Benoît, Marcotte, Patrice, Savard, Gilles, 2005. Bilevel programming: a survey. *4OR 3 (2)*, 87–107.

Constantin, Isabelle, Florian, Michael, 1995. Optimizing frequencies in a transit network: a nonlinear bi-level programming approach. *Int. Trans. Oper. Res.* 2 (2), 149–164.

De Cea, Joaquin, Fernández, Enrique, 1993. Transit assignment for congested public transport systems: an equilibrium model. *Transport. Sci.* 27 (2), 133–147.

Dong, Yuxiao, Pinelli, Fabio, Gkoufas, Yiannis, Nabi, Zubair, Calabrese, Francesco, Chawla, Nitesh V., 2015. Inferring unusual crowd events from mobile phone call detail records. In: *Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 474–492.

Goerigk, Marc, Schmidt, Marie, 2017. Line planning with user-optimal route choice. *Eur. J. Oper. Res.* 259 (2), 424–436.

Guihaire, Valérie, Jin-Kao, Hao, 2008. Transit network design and scheduling: a global review. *Transport. Res. Pol. Pract.* 42 (10), 1251–1273.

Guy, Desaulniers, Hickman, M., 2007. Public transit. In: *Handbooks in operations research and management science*, 14, pp. 69–128.

Gwilliam, K.M., 2011. Africa's Transport Infrastructure. World Bank Publications.

Hing-Keung Lam, William, Gao, Z.Y., Chan, K.S., Yang, Hai, 1999. A stochastic user equilibrium assignment model for congested transit networks. *Transp. Res. Part B Methodol.* 33 (5), 351–368.

Ibarra-Rojas, Omar J., Delgado, Felipe, Giesen, Ricardo, Muñoz, Juan Carlos, 2015. Planning, operation, and control of bus transport systems: a literature review. *Transp. Res. Part B Methodol.* 77, 38–75.

Kumar, A., Barrett, F., 2008. Stuck in Traffic: Urban Transport in Africa. AICD, Background Paper. World Bank, Washington, DC.

Ma, Tai-Yu, 2013. A hybrid cross entropy algorithm for solving dynamic transit network design problem. *J. Inf. Sci. Eng.* 29 (4), 631–645.

Marín, Angel, Mesa, Juan A., Perea, Federico, 2009. Integrating robust railway network design and line planning under failures. In: *Robust and Online Large-Scale Optimization*. Springer, pp. 273–292.

Matisziw, Timothy C., Murray, Alan T., Kim, Changjoo, 2006. Strategic route extension in transit networks. *Eur. J. Oper. Res.* 171 (2), 661–673.

Mauttone, Antonio, Urquhart, Mara E., 2009. A route set construction algorithm for the transit network design problem. *Comput. Oper. Res.* 36 (8), 2440–2449.

McCormick, Garth P., 1976. Computability of global solutions to factorable nonconvex programs: Part i-convex underestimating problems. *Math. Program.* 10 (1), 147–175. Orange, 2014. Orange D4d Challenge.

Schöbel, Anita, 2017. An eigenmodel for iterative line planning, timetable and vehicle scheduling in public transportation. *Transport. Res. C Emerg. Technol.* 74, 348–365.

Sherali, Hanif D., Shetty, C.M., 1980. A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts. *Math. Program.* 19 (1), 14–31.

Sherali, Hanif D., Amine, Alameddine, 1992. A new reformulation-linearization technique for bilinear programming problems. *J. Global Optim.* 2 (4), 379–410.

Spies, Heinz, Florian, Michael, 1989. Optimal strategies: a new assignment model for transit networks. *Transp. Res. Part B Methodol.* 23 (2), 83–102.

Still, Georg, 2018. *Lectures on Parametric Optimization: an Introduction*. Optimization Online.

Szeto, W.Y., Jiang, Y., 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transp. Res. Part B Methodol.* 67, 235–263.

Ukkusuri, Satish V., Mathew, Tom V., Waller, S Travis, 2007. Robust transportation network design under demand uncertainty. *Comput. Aided Civ. Infrastruct. Eng.* 22 (1), 6–18.

Vuchic, Vukan R., 2017. *Urban Transit: Operations, Planning, and Economics*. John Wiley & Sons.

Yan, Yadan, Liu, Zhiyuan, Meng, Qiang, Jiang, Yu, 2013. Robust optimization model of bus transit network design with stochastic travel time. *J. Transport. Eng.* 139 (6), 625–634.

Zanjirani Farahani, Reza, Miandoabchi, Elnaz, Szeto, W.Y., Rashidi, Hannaneh, 2013. A review of urban transportation network design problems. *Eur. J. Oper. Res.* 229 (2), 281–302.