The single-line design problem for Demand-Adaptive transit Systems: a modeling framework and decomposition approach for the stationary-demand case

Fausto Errico

Département de genie de la construction
École de technologie supérieure
and
Centre Interuniversitaire de Recherche sur les
Réseaux d'Entreprise, la Logistique et le Transport (CIRRELT)
fausto.errico@cirrelt.ca

Teodor Gabriel Crainic

Dept. management et technologie
École des sciences de la gestion
and
Centre Interuniversitaire de Recherche sur les
Réseaux d'Entreprise, la Logistique et le Transport (CIRRELT)
Université du Québec à Montréal,
teodorgabriel.crainic@cirrelt.net ,

Federico Malucelli

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)
Politecnico di Milano, federico.malucelli@polimi.it,

Maddalena Nonato

Facoltà di Ingegneria
Università degli Studi di Ferrara, maddalena.nonato@unife.it,

When demand for transportation is low or sparse, traditional transit cannot provide efficient and good-quality level service, due to their fixed structure. For this reason, mass transit is evolving towards some degree of flexibility. Although the extension of Dial-a-Ride systems to general public meets such need of adaptability, it presents several drawbacks mostly related to the their extreme flexibility. Consequently, new transportation alternatives, such as Demand Adaptive Systems (DAS), combining characteristics from both the traditional transit and Dial-a-Ride, have been introduced. For their twofold nature, DAS require careful planning. We focus on tactical aspects of the planning process by formalizing the *Single-line DAS Design Problem with Stationary Demand* and proposing two alternative hierarchical decomposition approaches for its solution. The main motivation behind this work is to provide a general methodology suitable to be used as a tool to build the tactical DAS plan in real-life conditions. We provide an experimental study where the two proposed decomposition methods are compared and the general behavior of the systems is analyzed when altering some design parameters. Furthermore, we tested the versatility of our methods on a variety of situation that may be encountered in real-life conditions.

Key words: Public Transit, Semi-Flexible Systems, Demand-Adaptive Systems, Tactical Planning, Service and Schedule Line Design

History:

1. Introduction

When the demand for transportation is consistently high during a given time period, traditional transit operates well and efficiently as it naturally allows for high degrees of resource sharing and consequent good level of service. In contrast, when the demand for transportation is low or sparse, the resource-sharing levels drastically drop, due in particular to the fixed structure of traditional transit services. For this reason, mass transit is evolving towards some degree of flexibility.

A first attempt in this direction was made by extending the well-known *Dial-A-Ride* systems (*DAR*), originally designed to serve people with reduced mobility, to general customer service (Wilson et al. 1971, Cordeau and Laporte 2003). With respect to traditional transit, DAR provides a *personalized* service by setting up itineraries, schedules and stop locations according to user expressed needs while still guaranteeing a certain degree of resource sharing by serving requests *collectively*. The adaptation of DAR to general public displays, however, a number of drawbacks, some of which follow from the extreme flexibility inherent in the system definition. Thus, for example, because the supply of transportation services changes according to needs expressed for particular time periods, neither the transit operator nor the users may predict the vehicle itineraries, stop locations, and associated schedules. As a consequence, users are obliged to book the service well in advance of the actual desired time of utilization and the actual pick up time is very much

left to the discretion of the operator. This operation mode also makes it difficult to integrate DAR with traditional transit services. Adapted DAR is therefore not compatible with the expectations and requirements of most users.

So-called *semi-flexible* systems were introduced to address these issues (Koffman 2004, Potts et al. 2010). We focus on the *Demand-Adaptive System* (*DAS*), introduced by Malucelli et al. (1999) and then treated in more general contexts in Crainic et al. (2005, 2001). DAS displays features of both traditional fixed-line bus services and purely on-demand systems such as DAR, and has been shown to generalize most semi-flexible systems (Errico et al. 2013).

A DAS bus line serves, on the one hand, a given set of *compulsory* stops according to a predefined *master schedule* specifying the time windows associated with each of these compulsory stops. This provides the traditional use of the transit line without in-advance reservations. On the other hand, similarly to DAR services, passengers may issue requests for transportation involving *optional* stops inducing detours in the vehicle routes. The fundamental idea of DAS is that the time-window mechanism introduces a degree of flexibility while providing a certain service *regularity*. This allows users to plan their trips, simplifies integration with other transportation modes, and makes the service accessible also without reservation (at compulsory stops).

Transit systems are dedicated to service several transportation requests with the same vehicle and generally require comprehensive planning activities. DAS is no exception requiring planning at the strategic, e.g., setting up the system and defining policies, tactical, e.g., designing the service and schedule, and operational where vehicle routes and schedules are adjusted depending on actual user requests. Yet, as emphasized by Errico et al. (2013), where an extensive literature review on semi-flexible system panning can be found, very few contributions address these issues. This is particularly true for the tactical aspects of the planning process, for which authors have either aimed to build closed-form analytic relations among the main service parameters based on a simplified operational framework (e.g., Fu 2002, Quadrifoglio et al. 2006, Zhao and Dessouky 2008), or addressed particular aspects of the problem only (e.g., Errico 2008, Crainic et al. 2012, Errico et al. 2016, Smith et al. 2003).

Our objective is to contribute to address this gap in knowledge and methodology. We study more particularly the tactical problem of designing the service and schedule for a single DAS line for a planning period during which the volume of demand (i.e., the number of requests per unit of time) is random but homogeneous, that is, the statistical properties of the estimated demand, its probability distribution in particular, are stable. We thus define a core tactical planning problem, the *Single-line DAS Design Problem with Stationary demand* (*S-SDDP*). The S-SDDP encompasses several interrelated decisions regarding the selection of the compulsory stops among all the potential stops in the territory, their sequencing, and the determination of the master schedule specifying the visit

4

Errico et al.: *The DAS Design Problem*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

time window for each compulsory stop. The S-SDDP represents a broad class of DAS planning cases, as well as a major building block for more complex settings, as also discussed in the paper.

We propose a mathematical formulation of the S-SDDP, which has the benefit of formally defining and capturing these different decision layers and their intricate relationships. However, the formulation is computationally intractable, even for very small problem instances. We therefore turn our attention to developing approximate solutions approach for the S-SDDP and propose two heuristic methods, called *Sequence Compulsory* and *Sequence All*. The two methods are based on different hierarchical decomposition approaches along the lines of the sets of decisions making up the problem, yielding a number of core subproblems for which efficient solution methods are known in the literature. We also performed a comprehensive computational study to compare the two decomposition strategies and to assess their versatility on a variety of different scenarios.

Summarizing, the contributions of the paper are:

• Formally define a new family of complex combinatorial problems, the S-SDDP, corresponding to the tactical service and schedule design of DAS with stationary demand;

• Introduce a mathematical formulation of the S-SDDP; to the best of our knowledge, we are the first to introduce and formally define the S-SDDP;

• Propose two hierarchical decomposition strategies to address the S-SDDP;

• Perform extensive computational experiments with two main purposes: 1) to validate and compare the solution methods, as well as the impact of variations in system parameters on its performance, and 2) to test the versatility of our methods on a variety of situations that could be encountered in real-life applications;

• Discuss the generalization and utilization of S-SDDP in broader planning contexts.

The paper is organized as follows. Section 2 provides the description of the problem setting and its main components. The main contributions of the paper are detailed in the following sections, namely, the model formulation of the S-SDDP in Section 3, and the decomposition strategies and the core subproblems in Section 4. Section 5 is dedicated to the presentation and discussion of the results of the experimental campaign we performed. We discuss the broader utilization of the S-SDDP for designing DAS services in various contexts in Section 6 and conclude in Section 7.

2. Context and problem setting

DAS, inheriting features of both traditional transit and DAR, requires a complex planning phase. To the purpose of properly contextualizing the S-SDDP, in Section 2.1 we provide some details on the DAS operating mechanism, while in Section 2.2 we generally describe the DAS planning activity, which encompasses strategic, tactical, as well as operational decision levels. The S-SDDP, which belongs to the tactical planning, is then introduced in Section 2.3.

2.1. Demand Adaptive transit Systems

In its most general form, a DAS is made up of several lines and is interconnected with the traditional transit system. Several vehicles operate on each DAS line providing service among a sequence of *compulsory* stops. Each compulsory stop is served within a predefined *time window*. The collection of time windows corresponding to the compulsory stops, including the start and end of the line, forms what we call the *master schedule* of the DAS line. This defines the traditional part of a DAS. Additional flexibility is provided by allowing customers to request service from and to *optional* stops within a given area. Such stops are visited only if a request is issued and accepted when operating the service. We denote users requesting service at an optional stop as *active users* (a-users), while users moving only between compulsory stops are called *at-compulsory-stop users* (c-users). Notice that c-users are also known in the literature as *passive* users because no reservation is needed at compulsory stops (see Frei et al. 2017, for example).
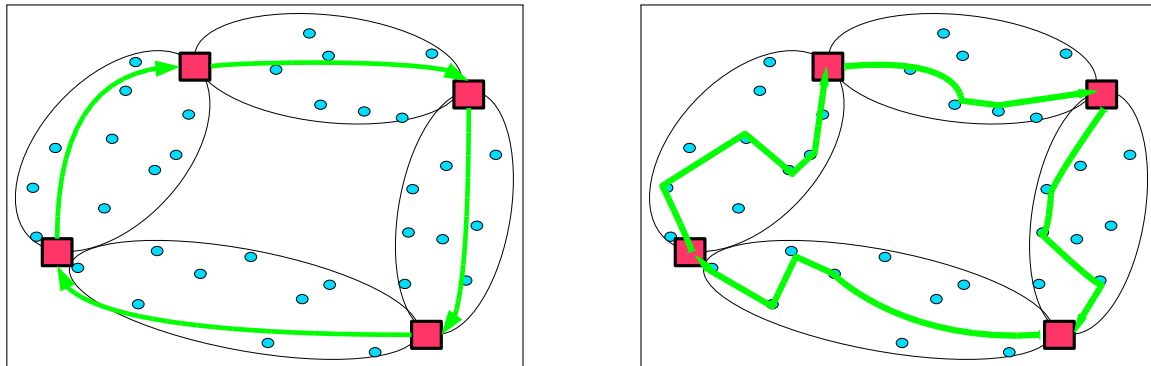
(a) A DAS vehicle serving compulsory stops and c-users only

(b) A DAS vehicle serving compulsory and optional stops, i.e., c-users as well as and a-users

Figure 1   A vehicle operating a Single-Line DAS

To serve optional stops, vehicles must generally deviate from the shortest path joining two successive compulsory stops. The region and the corresponding set of optional stops that is possible to visit between two consecutive compulsory stops is defined in advance and it is called *segment*. Figure 1a depicts a DAS line operated by a vehicle servicing only compulsory stops, while Figure 1b illustrates the same DAS line when user requests for optional stops are present.

The time window associated to a compulsory stop defines the earliest and latest vehicle departure times, *EDT* and *LDT*, respectively, for that compulsory stop. The vehicle is allowed to arrive at any moment before the LDT. If it arrives before the EDT, it has to wait at least until EDT, experiencing what we call *idle* time periods. To guaranty a uniform quality of service, time windows have a fixed *width* for all compulsory stops.

Assuming that customers have no information about the vehicle's current position, and given that the vehicle might leave at any time after the EDT, c-users need be present at the desired

Errico et al.: *The DAS Design Problem*

6   Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

compulsory stop not later than the EDT. As a consequence, c-users experience Vehicle Arrival Waiting times, which we call VAW times whenever the vehicle arrives at the compulsory stop later than the EDT (VAW times are sometimes referred to as *passive times* in the literature). It is worth observing that Global Positioning System (GPS) devices are nowadays inexpensive and can be used to provide customers with real-time information about the exact vehicle position. This information can be suitably used to reduce, if not eliminate, VAW times.

The detours induced by the activation of optional stops must be such that the time windows at compulsory stops are met. In this paper we assume the DAS1 operational policy (Malucelli et al. 1999) where service requests might be *rejected* if their acceptance leads to time window violations. If a request is accepted, users are picked up and dropped off exactly at the location they asked for. As will be clearer later, our methodology is, however, independent from the specific policy adopted at operational level. Finally notice that the time window and the segment specifications provide an *a priori* guarantee on the longest user travel time. Furthermore, time windows can be useful for users who want to plan transfers to other DAS lines, or to the traditional transit network.

2.2. DAS design planning levels

The management of mass transit systems generally requires complex planning activities. For traditional transit, the design of the system in terms of line routes is determined during the so-called strategic planning phase, timetables and vehicle schedules and routes are part of the tactical planning phase, and crew schedules are built during operational planning (Ceder and Wilson 1997). Comparatively, purely on-demand services such as DAR, need little strategic design, mainly to define service areas and the composition of the fleet (see, e.g., Diana et al. 2006, Quadrifoglio et al. 2008). The most important planning process for DAR is at the operational level, however, when routes and schedules are determined little time before each departure and are possibly dynamically modified once service has begun.

DAS, inheriting features of both traditional transit and DAR, requires a careful planning phase. Errico et al. (2013) reported a classification of the planning decisions in the standard strategic, tactical and operational hierarchy, together with a detailed review of the related literature.

Summarizing, at the strategic level, the region to be served is identified and partitioned into subregions, each corresponding to the area a single DAS line will service. An initial set of compulsory stops (possibly empty) to be used as transfer points is also determined. The desired quality of service and frequency of service is established for each service area. Service quality measures might include expected travel time, as well as transfer time from one line to another.

At the tactical level, the DAS line is built for every subregion identified at the strategic level, this process including decisions about the location of additional compulsory stops and their sequence,

time windows, and segments. This *backbone* of the line plays the same role for the transit authority and the users of a DAS as the schedule in traditional transit systems. It defines only partial itineraries and schedules, however. The backbone may be completed at operational level, when the actual requests for transportation become known. Thus, for each departure time, the actual itinerary and schedule is built to incorporate the additional optional stops corresponding to the accepted a-user requests, while respecting the constraints imposed by the line backbone.

2.3. The S-SDDP: problem setting

In this subsection we formally define the S-SDDP and delimit its scope.

General assumptions

The S-SDDP belongs to the tactical planning and in particular focuses on the design of a single DAS line. It assumes that strategic choices were previously made and are available. This includes the service area to be served by the DAS line with the set of potential stops, some of which will become compulsory. Given the particular role of some of the compulsory stops, e.g., major intermodal terminals or locations of interest in the city, a number of them could be decided during strategical planning. For the sake of generality, however, we assume the initial set of compulsory stops empty and let the tactical planing provide the actual selection. Decisions regarding the "structural" definition of the service, namely the shape to give to the line, e.g., circular, single leg to be traveled back and forth or more complex shapes, are generally taken at the strategic level. In the following, we address the complex case of a circular line serviced in clockwise and counter-clockwise directions.

Temporal assumptions and the stationary-demand hypothesis

The general transportation-demand itineraries are assumed determined at the strategic level of planning and are considered as given. While demand volumes may vary during a typical operating day, it is often the case that the operating horizon is made of several homogeneous periods characterized by *stationary demand*, that is, periods during which the variations in the number of requests per unit of time and origin-destination pair display constant statistical properties. Two inputs to our problem follow from this. First, an *estimation of the demand* for transportation between pairs of potential stops, built, e.g., on a combination of historical data and customer-preference inquiry, give the probability laws regulating the number of customer requests for each origin-destination pair. Second, the *service frequencies* for the line, established at strategic level according to capacities of the vehicles that will operate it and the desired level of service, indirectly give the number of departures for the planning period considered. The stationary-demand planning period may then be divided into a number of identical and consecutive time *slices*, defined as the time length corresponding to two consecutive occurrences of the line. The S-SDDP is thus a time-independent

8

Errico et al.: *The DAS Design Problem*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

problem over the planning period and may be solved for a single departure only. We show in Section 6 how to use the S-SDDP to address the general SDDP, and how the methodology developed for the stationary case may be adapted to address more general time-dependent cases.

Spatial assumptions

We follow the common practice to discretize the service area in *demand points* or stops, thus implying a discretization of functions of the service area, such as demand distributions. Different discretization levels correspond to different data aggregations and might imply a different number of demand points for a given area. Some other authors (e.g., Fu 2002, Quadrifoglio et al. 2006) assume a continuous service area instead. We observe that, while a discretization process generally implies a loss of information, to the best of our knowledge, work assuming continuous service areas generally require uniform spatial distributions. Given that we aim at modeling real-life contexts where uniformity assumption do not usually hold, discretization appeared more suitable.

The S-SDDP

Given the previous hypothesis, the S-SDDP integrates decisions on the selection of the compulsory stops, the definition of the main structure of the line by sequencing the compulsory stops, the partition of the service area into segments, i.e., the assignment of optional stops to segments, and the definition of the time windows at compulsory stops, yielding the schedule of the line. The objective is to identify the best design and schedule that services the estimated demand while optimizing a generalized measure of system performance accounting for both operational costs and the quality of service offered to customers. As an indirect measure of the operational costs, and potential fare prices, we adopt the widely accepted practice of considering those as proportional to the expected vehicle travel times or distances. On the other hand, we measure the level of service by accounting for two elements. The first one is what we call the Generalized Users' Travel Time (GUTT) which, beyond the vehicle-moving times, also accounts for the two DAS-specific time components due to the time windows mechanism, namely idle and VAW times. Notice that the GUTT is also called *latency* in the literature. The second element accounts for the number of customers boarding or alighting at compulsory stops. Observe that, similarly to most systems addressing costs and quality measures, the objectives indicated above are conflicting, as to a higher degree of efficiency usually corresponds a lower quality of service.

3. Modeling the S-SDDP

The S-SDDP is a very complex problem involving several interrelated decision layers. To the best of our knowledge it has never been introduced nor formalized before. In this section we propose a mathematical formulation whose principal purpose is to formally define the problem in all of its aspects. Section 3.1 introduces some basic notation, Sections 3.2 and 3.3 develop the

topological design and timing part of the formulation, respectively, while Section 3.5 is dedicated to the objective function. Section 3.6 provides a brief summary.

3.1.   Basic notation

Let us represent the service area as a directed graph $G = (N, A)$ where the node set N corresponds to the demand points or potential stops and the arc set A corresponds to possible direct connections between pairs of potential stops. Node 1 denotes the terminal and it is the first compulsory stop. We associate a travel time t_{ij} with each arc $(i, j) \in A$.

Recall that, in our context, decisions have to be taken when the actual requests are not known yet. Thus, let $f_{hk}(d)$ be the probability mass function of the transportation requests, where $f_{hk}(d) = \text{Prob}\{\hat{d}_{hk} = d\}$ and $\hat{d}_{hk}$ stands for the random number of requests between each origin h and destination k. The demand may then be described by an origin-destination matrix D with probability mass functions as entries, i.e., $D = [f_{hk}(\cdot)]$. In our case, given the stationary-demand assumption, the matrix D does not explicitly depend on time. Let $\bar{d}_{hk}$ be the expected number of requests between h and k and the related matrix $\bar{D} = [\bar{d}_{hk}]$. Also let $\bar{\pi}_{hk} = f_{hk}(0)$ be the no-service-request probability for a given origin h and destination k. It is worth noticing that our solution method only requires the knowledge of the demand matrix $\bar{D} = [\bar{d}_{hk}]$ and the probabilities $\bar{\pi}_{hk}$, which is a less demanding assumption than requiring the estimate of the full probability distributions $f_{hk}(d)$.

We associate a time window with each compulsory stop and let δ be the fixed width of the time windows of all compulsory stops. Finally, recall that a *segment* defines the set of optional stops that can be visited between consecutive compulsory stops.

3.2.   The topological design

Let us introduce compulsory stop selection variables

$$y_i = \begin{cases} 1 \text{ if compulsory stop } i \text{ is selected,} \\ 0 \text{ otherwise,} \end{cases}$$

where y_1 is set to 1.

Moreover, let us introduce one variable for every pair of possible compulsory stops (hence for every arc of the graph)

$$x_{ij} = \begin{cases} 1 \text{ if compulsory stops } i \text{ and } j \text{ are consecutive,} \\ 0 \text{ otherwise.} \end{cases}$$

Clearly, the following relation linking vectors x and y holds:

$$y_i = \sum_{(i,j) \in A} x_{ij} \quad \forall i \in N \tag{1}$$

In order to specify the sequence constraints for the compulsory stops, we can introduce suitable cycle constraints:

$$\sum_{(j,i)\in A} x_{ji} - \sum_{(i,j)\in A} x_{ij} = 0 \quad \forall i \in N \tag{2}$$

Possible subtours will be avoided by the time window constraints that we discuss later in this section.

Turning to the segment definition, we introduce the following variables that assign stops to segments:

$$z_{ij}^h = \begin{cases} 1 \text{ if stop } h \text{ is assigned to segment } (i,j). \\ 0 \text{ otherwise.} \end{cases}$$

Variables z_{ij}^h implicitly define the set N_{ij} of stops which may be visited between compulsory stops i and j. By convention, if $z_{ij}^h = 1$ for some h, then we set $z_{ij}^i = 1$ and $z_{ij}^j = 0$. The following constraints link x, y and z variables:

$$z_{ij}^h \leq x_{ij} \quad \forall h \in N, \forall (i,j) \in A, \tag{3}$$

$$z_{ij}^i = x_{ij} \quad \forall (i,j) \in A, \tag{4}$$

$$z_{ij}^j = 0 \quad \forall (i,j) \in A, \tag{5}$$

and

$$\sum_{(i,j)\in A} z_{ij}^h = 1 \quad \forall h \in N, \tag{6}$$

where relations (4) and (5) imply that compulsory stops are assigned to the segment they initiate, while (6) prescribes that every stop must be assigned to a segment.

3.3. The time windows

We now focus on the set of relations regulating the choice of the time window at each compulsory stop and for each direction of the line. The time required to travel a segment plays a crucial role. We find however two levels of challenges here: first, the actual transportation requests are not known at the moment of determining the S-SDDP, thus the problem is inherently stochastic; second, once the transportation requests are known, a new optimization problem needs to be solved at each departure. In fact, the actual vehicle itinerary must be established according to the specific operational policy adopted by the transit agency, which could allow, for example, to reject or delay some requests, or to propose alternative itineraries. In this paper we specialize our formulation for the DAS1 policy treaded in Malucelli et al. (1999), Crainic et al. (2005). For the sake of simplicity, we do not explicitly introduce decision variables for the operational policy. We rather assume that an optimal itinerary is available once the transportation requests are revealed, and we introduce variables associated with the time required to travel segments.

More formally, given two consecutive compulsory stops, i and j, the set of *active* optional stops $\hat{N}_{ij} \subseteq N_{ij}$, i.e., stops for which a request is issued, will normally vary at each departure and is considered a random set. Consequently, the optimal path to service all stops in $\hat{N}_{ij}$ is determined at each departure and its duration $\hat{T}_{ij}$ is also a random variable. Let us focus, for the moment, on the first segment in one of the two possible directions of the line, say $(1, j)$. Let $\hat{T}_j$ be the random variable describing the arrival time at compulsory stop j. Assuming that at compulsory stop 1 the departure time is $t_1 = 0$, we have $\hat{T}_{1j} = \hat{T}_j$. For some ϵ close to 1, but not greater than 1, we define the ϵ-quantile of $\hat{T}_j$ as $T_j^\epsilon = Q(\epsilon, \mathcal{F}^c(\hat{T}_j))$, where $Q(\cdot, \cdot)$ is the quantile function, and $\mathcal{F}^c(\cdot)$ the cumulative distribution function. T_j^ϵ can be interpreted as the maximal trip duration between 1 and j guaranteeing that all requests are satisfied with probability at least ϵ. We also define $\hat{T}_j^\epsilon$, the random variable obtained from $\hat{T}_j$ by discarding the values greater than T_j^ϵ. The random variable $\hat{T}_j^\epsilon$ accounts for the fact that the service is guaranteed only for the ϵ-portion of the active sets with the shortest travel time. The value of ϵ is a design parameter and, by assigning different values to it, it is possible to obtain different levels of robustness. For example, the setting $\epsilon = 1$ allows to travel through all the stops in $N_{1,j}$, if required. In our application, we require that the vehicle never arrives at j later than T_j^ϵ. We thus introduce a variable b_j, which is only defined if $y_j = 1$, representing the latest arrival time at compulsory stop j, and impose that $T_j^\epsilon \leq b_j$. We define the corresponding time window as $[a_j = b_j - \delta, b_j]$.

A similar mechanism also holds for the generic segment (i, j), except that the departure time $\hat{T}_i^p$ in i is a random variable that can be obtained from the arrival time $\hat{T}_i^\epsilon$ by discarding the values smaller than a_i and setting $\mathcal{F}^c(\hat{T}_i^p) = \mathcal{F}^c(\hat{T}_i^\epsilon)$ for all values in $[a_i, b_i]$. $\hat{T}_i^p$ accounts for the fact that if the vehicle arrives in i before a_i it cannot leave until a_i. Observe that, to simplify the notation, we make the assumption that the service time is zero (the extension to non-zero service times is trivial). In consequence, we can write $\hat{T}_j = \hat{T}_i^p + \hat{T}_{ij}$. The probability distribution of $\hat{T}_j$ can be derived via convolution as $\mathcal{F}^d(\hat{T}_j) = \mathcal{F}^d(\hat{T}_i^d) * \mathcal{F}^d(T_{ij})$, where $\mathcal{F}^d(\cdot)$ denotes the probability density function and $*$ the convolution operator. Finally, T_j^ϵ can be obtained as for the single segment case.

Summarizing, the constraints determining the time windows then are:

$$T_j^\epsilon \leq b_j \quad \text{if } y_j = 1, \forall j \in N \tag{7}$$

$$T_j^\epsilon = Q(\epsilon, \mathcal{F}^c(\hat{T}_j)) \quad \text{if } y_j = 1, \forall j \in N \tag{8}$$

$$\mathcal{F}^d(\hat{T}_j) = \mathcal{F}^d(\hat{T}_{1j}) \quad \text{if } x_{1j} = 1, \forall j \in N \tag{9}$$

$$\mathcal{F}^d(\hat{T}_j) = \mathcal{F}^d(\hat{T}_i^p) * \mathcal{F}^d(\hat{T}_{ij}) \quad \text{if } x_{ij} = 1, \forall\, (i, j) \in A \tag{10}$$

It should be noticed that random variables $\hat{T}_j$, $\hat{T}_j^\epsilon$, and consequently $\hat{T}_j^p$, are all functions of the segment-defining variables z_{ij}^h, but for simplicity we did not make this dependency explicit in

the notation. In the same spirit, we did not report in the above constraints the relations linking probability distribution functions with their cumulative versions.

The definition of the time windows in the opposite direction follows the same principles and, in particular, the following holds:

$$T_j'^\epsilon \leq b_j' \quad \text{if } y_j = 1, \forall j \in N \tag{11}$$

$$T_j'^\epsilon = Q(\epsilon, \mathcal{F}^c(\hat{T}_j')) \quad \text{if } y_j = 1, \forall j \in N \tag{12}$$

$$\mathcal{F}^d(\hat{T}_j') = \mathcal{F}^d(\hat{T}_{1j}') \quad \text{if } x_{j1} = 1, \forall j \in N \tag{13}$$

$$\mathcal{F}^d(\hat{T}_j') = \mathcal{F}^d(\hat{T}_i'^p) * \mathcal{F}^d(\hat{T}_{ij}') \quad \text{if } x_{ji} = 1, \forall (j, i) \in A \tag{14}$$

where the *prime* symbol has been used to specify values for the opposite direction. Given two consecutive compulsory stops i and j, we refer to the difference $b_j - b_i$ as the *distance* between the time windows of i and j, for the given direction of the line.

We notice that, while in this paper we assume that the operational policy only services the ϵ-portion of the active sets with the shortest travel time, implying the rejection of some requests, our model and solution method can be easily extended to other operational policies, provided that the relation between $\hat{T}_i$ and b_i is suitably adapted to the operational policy. Furthermore, observe that ϵ is a design parameter and the decision maker can modify its value to obtain the desired level of service. Finally notice that constraints (10) and (14) will forbid subtours and can be seen as a generalization of the classic Miller-Tucker-Zemlin subtour-elimination constraints.

3.4. The generalized users' travel time

We remind that the GUTT is a generalised measure accounting for three components, namely the traditional vehicle-moving times, as well as the idle and VAW times at compulsory stops (for users who are on the vehicle and waiting for the vehicle, respectively). Thus, on the one hand, we need to consider the passenger flow on each arc including those among optional stops. Another issue arising when considering the GUTT is that, since we are in the presence of a circular line, each passenger can board at a stop preceding the terminal and alight at a stop following the terminal, thus using two consecutive trips of the same line. Moreover, since we are planning both directions at once, we should consider for each request the minimum traveling time between the two options. Finally, for the sake of generality, we also consider VAW times as a component of the travel time for c-users.

By denoting g_{hk} the generalized travel time of a single user, the GUTT can be expressed as:

$$GUTT = \sum_{h \in N} \sum_{k \in N, k \neq h} \bar{d}_{hk} g_{hk} \tag{15}$$

In the following we show how to model g_{hk}. To this scope, we recall that the arrival time at a compulsory stops might differ from the departure time. For this reason, and also to enable the modeling of the VAW times, we differentiate between arrival and departure times. For each stop h, we define the random variable representing the departure time as:

$$\hat{t}_h^p = \begin{cases} \hat{T}_i^p + \hat{\tau}_{ih} & \text{if } z_{ij}^h = 1 \text{ and } h \neq i \\ a_i & \text{if } z_{ij}^h = 1 \text{ and } h = i, \end{cases} \tag{16}$$

where $\hat{\tau}_{ih}$ represents the random traveling time from i to h. In the case $h = i$, we account for the VAW time at i. In the opposite direction:

$$\hat{t}_h^{\prime p} = \begin{cases} \hat{T}_j^{\prime p} + \hat{\tau}_{jh}^{\prime} & \text{if } z_{ij}^h = 1 \text{ and } h \neq i \\ a_i^{\prime} & \text{if } z_{ij}^h = 1 \text{ and } h = i. \end{cases} \tag{17}$$

For each stop k, we define the random variable representing the arrival time:

$$\hat{t}_k^a = \begin{cases} \hat{T}_i^p + \hat{\tau}_{ik} & \text{if } z_{ij}^k = 1 \text{ and } k \neq i \\ \hat{T}_i^\epsilon & \text{if } z_{ij}^k = 1 \text{ and } k = i, \end{cases} \tag{18}$$

and for the opposite direction:

$$\hat{t}_k^{\prime a} = \begin{cases} \hat{T}_j^{\prime p} + \hat{\tau}_{jk}^{\prime} & \text{if } z_{ij}^h = 1 \text{ and } k \neq i \\ \hat{T}_i^{\prime \epsilon} & \text{if } z_{ij}^k = 1 \text{ and } k = i. \end{cases} \tag{19}$$

Finally, by defining $T = \max_i b_i$ and $T' = \max_i b_i'$ the overall upper bound on traveling time in the clockwise and counter-clockwise directions, we can write:

$$g_{hk} = \min\{\mathbb{E}[(T + \hat{t}_k^a - \hat{t}_h^p) \mod T], \mathbb{E}[(T' + \hat{t}_k^{\prime a} - \hat{t}_h^{\prime p}) \mod T']\}. \tag{20}$$

In the above expression, the expectation operator $\mathbb{E}$ is taken over all possible realizations of the customer requests. The components in the minimization operator correspond to the expected travel time in the clockwise and the counter-clockwise direction, respectively. The modulo operator ensures that the travel time for user itineraries traversing the terminal is suitably computed.

3.5. The objective function

Turning to the objective function, we remind that is it made of three main components: the operating costs, the GUTT, and the number of customers boarding or alighting at compulsory stops.

The operating cost component is given by the following term:

$$\sum_{(i,j) \in A} c_{ij} x_{ij}, \tag{21}$$

where c_{ij} accounts for the cost of operating segment (i, j). The exact definition of this cost is application dependent and it could be set, e.g., to be proportional to the expected segment travel time $\hat{T}_{ij}$ or distance, and could incorporate fixed costs.

The GUTT component can be expressed by combining equations (15) and (20):

$$\sum_{h \in N} \sum_{k \in N, k \neq h} \bar{d}_{hk} \min\{\mathbb{E}[(T + \hat{t}_k^a - \hat{t}_h^p) \mod T], \mathbb{E}[(T' + \hat{t}_k'^a - \hat{t}_h'^p) \mod T']\}.$$

Finally, the third component can be expressed as:

$$\sum_{h \in N} \sum_{k \in N, k \neq h} (\bar{d}_{kh} + \bar{d}_{hk}) y_h. \tag{22}$$

Maximizing this expression encourages to select compulsory stops in such a way that the highest number of requests has at least the origin or the destination corresponding to a compulsory stop.

3.6. Summary

To summarize:

- Decision variables:

$$x_{ij} = \{0, 1\} \qquad\qquad\qquad \forall \, (i, j) \in A$$
$$y_i = \{0, 1\} \qquad\qquad\qquad \forall \, i \in N$$
$$z_{ij}^h = \{0, 1\} \qquad\qquad\qquad \forall \, h \in N, \forall \, (i, j) \in A$$
$$b_i, b_i' \geq 0 \qquad\qquad\qquad \forall \, i \in N$$
$$T_i^\epsilon, T_i'^\epsilon \geq 0 \qquad\qquad\qquad \forall \, i \in N$$
$$T, T' \geq 0$$

- Objective function:

$$\min \left\{ \sum_{(i,j) \in A} c_{ij} x_{ij} - \sum_{h \in N} \sum_{k \in N, k \neq h} (\bar{d}_{kh} + \bar{d}_{hk}) y_h + \right.$$
$$\left. \sum_{h \in N} \sum_{k \in N, k \neq h} \bar{d}_{hk} \min \left\{ \mathbb{E}[(T + \hat{t}_k^a - \hat{t}_h^p) \mod T], \mathbb{E}[(T' + \hat{t}_k'^a - \hat{t}_h'^p) \mod T'] \right\} \right\}$$

- Constraints

$$y_i = \sum_{(i,j) \in A} x_{ij} \qquad\qquad\qquad \forall i \in N$$
$$\sum_{(j,i) \in A} x_{ji} - \sum_{(i,j) \in A} x_{ij} = 0 \qquad\qquad\qquad \forall i \in N$$
$$z_{ij}^h \leq x_{ij} \qquad\qquad\qquad \forall h \in N, \forall (i,j) \in A$$
$$z_{ij}^i = x_{ij} \qquad\qquad\qquad \forall (i,j) \in A$$

$$z_{ij}^j = 0 \qquad\qquad \forall (i,j) \in A$$

$$\sum_{(i,j) \in A} z_{ij}^h = 1 \qquad\qquad \forall h \in N$$

$$T_j^\epsilon \leq b_j \quad \text{if } y_j = 1 \qquad\qquad \forall j \in N$$

$$T_j^\epsilon = Q(\epsilon, \mathcal{F}^c(\hat{T}_j)) \quad \text{if } y_j = 1 \qquad\qquad \forall j \in N$$

$$\mathcal{F}^d(\hat{T}_j) = \mathcal{F}^d(\hat{T}_{1j}) \quad \text{if } x_{1j} = 1 \qquad\qquad \forall j \in N$$

$$\mathcal{F}^d(\hat{T}_j) = \mathcal{F}^d(\hat{T}_i^p) * \mathcal{F}^d(\hat{T}_{ij}) \quad \text{if } x_{ij} = 1 \qquad\qquad \forall \, (i,j) \in A$$

$$T_j'^\epsilon \leq b_j' \quad \text{if } y_j = 1 \qquad\qquad \forall j \in N$$

$$T_j'^\epsilon = Q(\epsilon, \mathcal{F}^c(\hat{T}_j')) \quad \text{if } y_j = 1 \qquad\qquad \forall j \in N$$

$$\mathcal{F}^d(\hat{T}_j') = \mathcal{F}^d(\hat{T}_{1j}') \quad \text{if } x_{j1} = 1 \qquad\qquad \forall j \in N$$

$$\mathcal{F}^d(\hat{T}_j') = \mathcal{F}^d(\hat{T}_i'^p) * \mathcal{F}^d(\hat{T}_{ij}') \quad \text{if } x_{ji} = 1 \qquad\qquad \forall (j,i) \in A$$

$$\hat{t}_h^p = \begin{cases} \hat{T}_i^p + \hat{\tau}_{ih} & \text{if } h \neq i \\ a_i & \text{if } h = i \end{cases} \text{ if } z_{ij}^h = 1 \qquad\qquad \forall h \in N, \forall (i,j) \in A$$

$$\hat{t}_h'^p = \begin{cases} \hat{T}_j'^p + \hat{\tau}_{jh}' & \text{if } h \neq i \\ a_i' & \text{if } h = i \end{cases} \text{ if } z_{ij}^h = 1 \qquad\qquad \forall h \in N, \forall (i,j) \in A$$

$$\hat{t}_k^a = \begin{cases} \hat{T}_i^p + \hat{\tau}_{ik} & \text{if } k \neq i \\ \hat{T}_i^\epsilon & \text{if } k = i, \end{cases} \text{ if } z_{ij}^k = 1 \qquad\qquad \forall k \in N, \forall (i,j) \in A$$

$$\hat{t}_k'^a = \begin{cases} \hat{T}_j'^p + \hat{\tau}_{jk}' & \text{if } k \neq i \\ \hat{T}_i'^\epsilon & \text{if } k = i \end{cases} \text{ if } z_{ij}^k = 1 \qquad\qquad \forall k \in N, \forall (i,j) \in A$$

$$T = \max_i b_i$$

$$T' = \max_i b_i'$$

4. Addressing the S-SDDP

The formulation proposed in the previous section describes a very challenging decision problem. This is easily seen by noticing, beyond the large number of binary and continuous variables, the non-linearities and the expectation operator in the objective function, as well as the non-linearity implicit in many of the time-related constrains where quantile functions and convolution operators are involved. For this reason, rather than trying to directly implement the formulation in a computer code, we decided to turn our attention to build a framework to develop heuristic solution methods.

In this section we propose a solution method based on the hierarchical decomposition of the S-SDDP according to its several decision layers. As it can be seen by inspecting the formulation in the previous section, the S-SDDP is made of three main groups of decisions, namely *group 1*: y establishing the compulsory stops; *group 2*: z and x for the segment definition and sequencing; and *group 3*: b and the time-related variables T and T^ϵ to define the time windows at the compulsory stops.

We thus propose to decompose the S-SDDP along these three decision groups according to a hierarchical order where we first determine the compulsory stops (group 1), we then define segment definition and sequencing (group 2) and we finally establish the time windows (group 3). In particular, we propose two variants of this hierarchical decomposition differing in the way the decision variables in group 2 are fixed. The resulting variants are called *Sequence Compulsory* (SC) and *Sequence All* (SA) and are schematically represented in Figure 2, which is more detailedly explained in the next subsection.

Three common *core* problems arise form these decompositions: the selection of compulsory stops (core problem 1 in Section 4.2), a stop sequencing problem (core problem 2 in Section 4.3) and the time window definition (core problem 3 in Section 4.4).

4.1. Hierarchical decompositions

In the SC strategy, the first step determines the set of compulsory stops (Figure 2 Step 1) by solving the core problem 1 and setting the values of variables y. The second step fixes vectors x and z by a two-phase procedure: a) the sequence of compulsory stops is determined by solving core problem 2 and thus fixing the value of variables x (Figure 2, SC-Step 2a), and b) determines the values of variables z by assigning optional stops h to segments (i, j) according to a geometrical proximity criterion:

$$z_{ij}^h = \begin{cases} 1 \text{ if } (i,j) = \arg\min\{t_{i'h} + t_{hj'} + t_{hi'} + t_{j'h}, \ (i', j') \mid x_{i'j'} = 1\}, \\ 0 \text{ otherwise,} \end{cases} \tag{23}$$

where the four terms of the argmin operator account for the two traveling directions (Figure 2, SC-Step 2b). At this stage of the procedure, variables y, x and z have been fixed, and we name the obtained partial S-SDDP the *topological* design. Finally, time windows are determined by solving core problem 3 (Figure 2, Step 3). As a result, variables b as well as the remaining time-related variables T and T^ϵ are fixed.

The SA strategy follows the main lines of SC, with the exception that the second step, where the values of vectors x and z are determined, differs. This step is addressed by a different two-phase procedure: a) a sequence among all stops, regardless if they are optional or compulsory, is obtained by solving core problem 2 (Figure 2, SA-Step 2a) and b) the obtained ordering among stops is used to determine the sequence of compulsory stops, and to define the segment composition (Figure 2, SA-Step 2b). In particular, for any sequence found at step a) if two compulsory stops i and j are linked by a path only made of optional stops, the corresponding segment is formed and $x_{ij} = 1$. Furthermore, any optional stop h visited in the sequence between i and j computed in a) is assigned to the formed segment and $z_{ij}^h = 1$. In other words, phase a) limits the set of feasible solutions while phase b) fixes the values of x and z accordingly. At this point, having determined

the values of vectors y, x and z, we obtained a feasible topological design which is used as input for the last step to determine the time windows, which is in common with the SA procedure. It is worth mentioning that the sequence determined at Step 2a of the SA strategy has the only purpose of determining the segment assignments for the tactical level, while the actual vehicle itinerary is recomputed at operations time according to the observed service requests and consistently with the tactical design.

The overall solution approach is depicted in Figure 3, where both decompositions can be encompassed. The leftmost block is composed of two sub-blocks corresponding to steps 1) and 2) of the hierarchical decomposition. Different implementations of the "Sequencing and segment definition" sub-block yield either the SC or the SA variant. In both cases, the output of this first design phase is given by the *topological design*, which is made of the set of compulsory stops, their sequence, and the segment definition. The last step of the design is the definition of the time windows (*master schedule*), which is common to both decompositions.

4.2. Core problem 1: the selection of compulsory stops

Regarding the role of the compulsory stops, the philosophy of DAS is to use them to naturally capture high demand patterns without the need of explicitly issuing requests. Thus, typical candidate locations for compulsory stops will be demand-attractor poles, such as hospitals, schools, libraries, railway stations, and others. Furthermore, transportation companies will generally be interested in servicing all the transportation requests, and only occasionally be in the position of rejecting some. From a mathematical point of view, identifying high demand patterns may be quantified through the service-request probability at a given location. Thus, we adopted the simple approach to select compulsory stops in locations where the service-request probabilities are "close" to 1. How close can be modulated by the transportation company according to its risk attitude and it is a design parameter. The proposed method has the advantage that service-request probabilities can be easily retrieved from the demand data. Furthermore, this proposed approach can easily integrate compulsory stops established with other methods, e.g., the ones selected as transfer points at the strategic level, or the ones needed because of other possible managerial or political reasons.

4.3. Core problem 2: stop sequencing

Both hierarchical decompositions perform an ordering of the stops. Among different models addressing ordering issues, the Traveling Salesman Problem with Generalized Latency (TSP-GL) is the most suitable since the objective function accounts for both operating cost and GUTT. The TSP-GL has been introduced in Errico et al. (2016), where a branch-and-cut approach based on the use of a Benders reformulation and a set of valid inequalities was developed.
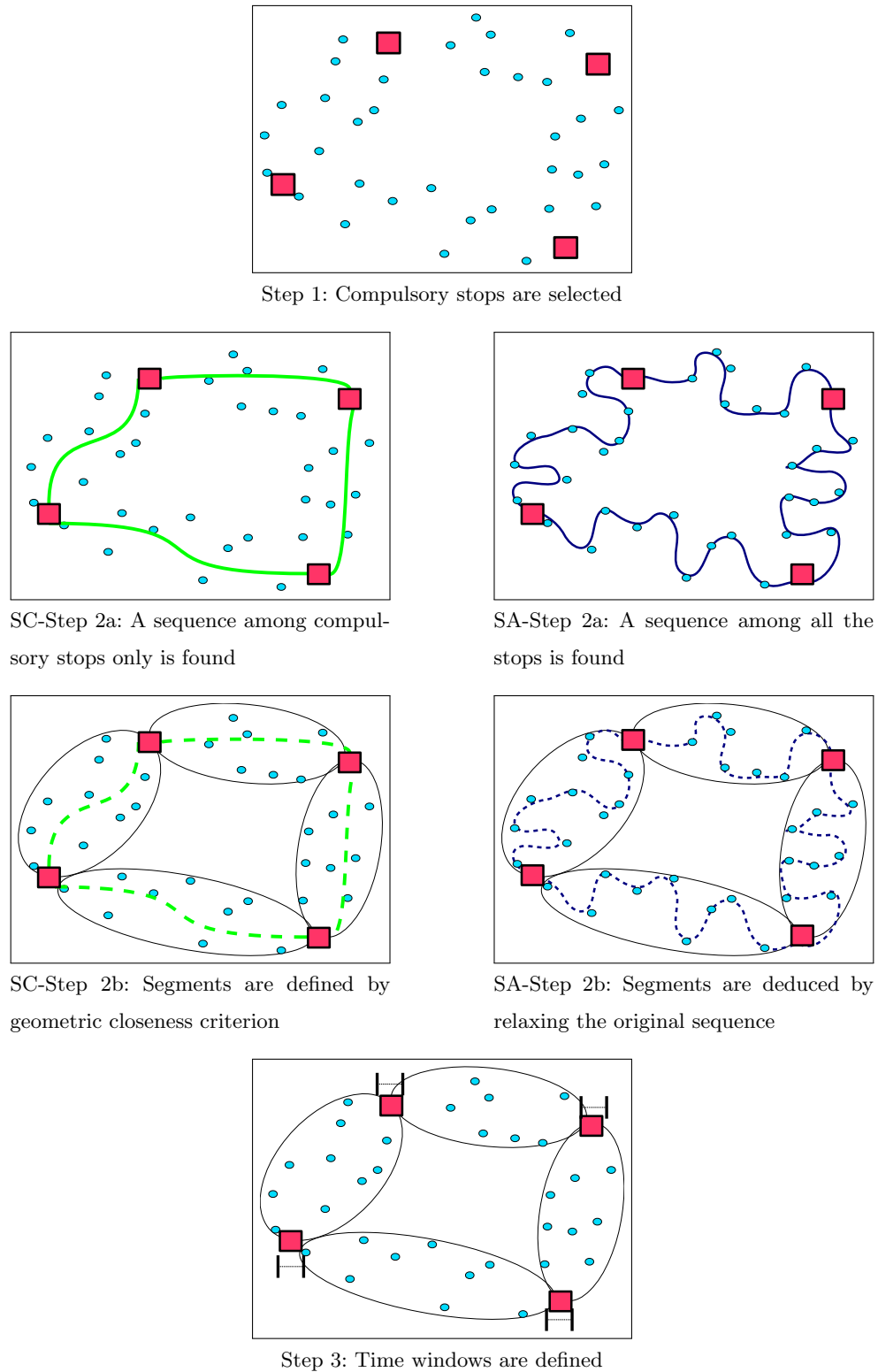
Step 1: Compulsory stops are selected

SC-Step 2a: A sequence among compulsory stops only is found

SA-Step 2a: A sequence among all the stops is found

SC-Step 2b: Segments are defined by geometric closeness criterion

SA-Step 2b: Segments are deduced by relaxing the original sequence

Step 3: Time windows are defined

Figure 2 **The two hierarchical decomposition variants: SC and SA**
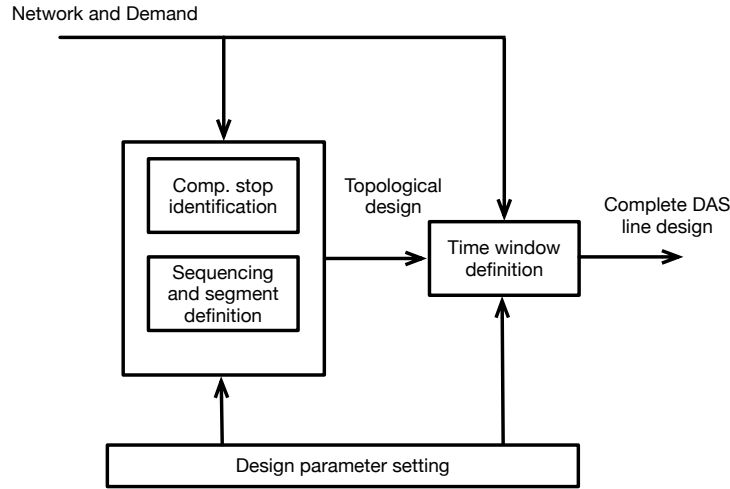
Figure 3   Structure of the Design Process

Formally, the TSP-GL can be described as follows. Consider a complete and undirected graph $G' = (N', E')$ where N' is the node set and E' the edge set. A non-negative cost c'_{ij} is associated with each edge $[i, j]$. For each pair of nodes h and k we are given the matrix of the expected transportation demand $\bar{D}' = [\bar{d}'_{hk}]$ as well as a traveling time t'_{hk}. Similarly to the Traveling Salesman Problem (TSP), the TSP-GL finds a Hamiltonian circuit $\mathcal{H}$ among all the nodes. Differently from the TSP, the objective function of the TSP-GL has two components: the cost of the circuit $\mathcal{H}$, and the average user travel time. The main design parameter related to the TSP-GL is the relative weight of the two terms of the objective function. The TSP-GL expresses the objective function in the following form:

$$\min \quad (1 - \alpha) \sum_{[i,j] \in \mathcal{H}} c'_{ij} + \frac{\alpha}{\sum_{h,k \in N'} \bar{d}'_{hk}} \sum_{h,k \in N'} \bar{d}'_{hk} \min \left\{ \sum_{(i,j) \in \overrightarrow{\mathcal{H}}_{hk}} t'_{ij}, \sum_{(i,j) \in \overleftarrow{\mathcal{H}}_{hk}} t'_{ij} \right\} \tag{24}$$

where $\overrightarrow{\mathcal{H}}_{hk}$ and $\overleftarrow{\mathcal{H}}_{hk}$ denote the portion of Hamiltonian circuit $\mathcal{H}$ from h to k in one or the opposite direction.

When $\alpha < 0.5$, the optimization process emphasizes efficiency aspects and the resulting sequences are expected to be less expensive. On the contrary, when $\alpha > 0.5$, the optimization process emphasizes the user level of service and the resulting sequences are expected to be longer, but with shorter average user travel times.

When decomposing the S-SDDP according to the SA strategy, the set of items to be ordered corresponds to the whole set of stops, hence $N' = N$ and $E' = \{[i, j] \in N', i < j\}$. The GUTT accounts for the demand among all stops, and consequently $\bar{D}' = \bar{D}$. Finally $c'_{ij} = c_{ij}$ and $t'_{ij} = t_{ij}$.

When adopting the SC strategy, only compulsory stops have to be ordered. Let us call N_c such a set. Thus $N' = N_c$ and $E' = \{[i, j] \in N', i < j\}$. Differently from the previous case, the GUTT only

accounts for the demand among compulsory stops, hence $\bar{D}' = [\bar{d}_{hk}]$, with $h, k \in N'$. This choice is justified by the fact that assignments of optional stops to segments has not been determined at this stage of the decomposition strategy. Finally, as for the SA strategy, $c'_{ij} = c_{ij}$ and $t'_{ij} = t_{ij}$.

It should be noticed that in both decomposition strategies the sequencing step is performed before the time windows have been determined. This lack of information might imply sequences with a certain degree of suboptimality, especially concerning time window-related components such as idle and VAW times. We report detailed information about these performance measures when analysing the computational results is Section 5. We will see that, in spite of this lack of information, the obtained solutions perform relatively well.

4.4.   Core problem 3: the time windows definition

Both decomposition strategies require, as a final step, to determine the time windows associated to compulsory stops. This is obtained by solving the *Master Schedule Problem* (MSP) introduced in Crainic et al. (2012). Summarizing, the MSP considers two kind of inputs. On the one hand, the *topological* design, i.e., the ordered set of compulsory stops, and the assignment of optional stops to segments. On the other hand, for each stop i, a service-request probability π_i is supposed to be known. For a given line direction, and for each segment, the MSP estimates the probability distributions of the random variables $\hat{T}_j^\epsilon$, $\hat{T}_j^p$ and the values T_j^ϵ, and determine the time windows $[a_j, b_j]$ for each compulsory stop accordingly. The objective is to minimize the latest arrival time at the terminal in such a way that the probability to service *all* the transportation requests is at least $\epsilon \approx 1$. As previously specified, we consider time windows having the same width δ. Consequently, the parameters regulating the MSP are ϵ and δ.

To properly use the MSP solution method, we must associate the demand from h to k with one of the two line directions. Once again, the critical aspect of such an assignment is that time windows are not yet defined, and we can only estimate the travel times between h and k. Among the several possible methods to determine these assignments, the one that experimentally gave the better results is the following. Let us denote $\overrightarrow{\mathcal{P}}_{hk}$ and $\overleftarrow{\mathcal{P}}_{hk}$ the set of segments (i, j) between h and k in one or the opposite direction. We assign the demand from h to k to the line direction corresponding to the argument giving the minimum of the following expression

$$\min \left\{ \sum_{(i,j) \in \overrightarrow{\mathcal{P}}_{hk}} \mid N_{ij} \mid, \sum_{(i,j) \in \overleftarrow{\mathcal{P}}_{hk}} \mid N_{ij} \mid \right\} \tag{25}$$

To define the values of π_i, by recalling the notation introduced in Section 3, we can express the no-service-request probability from h to k as $\bar{\pi}_{hk} = f_{hk}(0)$. Hence

$$\pi_i = 1 - \prod_{k \in N, k \neq i} \bar{\pi}_{ik} \prod_{k \in N, k \neq i} \bar{\pi}_{ki} \tag{26}$$

It is worth noticing that the *distance* between two consecutive time windows is proportional to the flexibility available to service the segment included between them, while the time window width is related to how much of this flexibility can be transferred from one segment to the following one.

When designing the master schedule, one should take into account the following conflicting factors. On the one hand, allocating long distances between consecutive time windows allows to service more optional stops if needed, but it also increases the possibility to experience idle times. On the other hand, designing wider time windows allows to transfer the flexibility between adjacent segments, but the VAW times might considerably increase.

From the solution method point of view, the most challenging part is the estimation of the probability distribution of the time needed to travel a given segment. Crainic et al. (2012) propose an efficient sampling mechanism where probability distributions and time windows are built according to equations (7) - (14). In particular, a confidence interval on the values of T_j^ϵ is iteratively built for each compulsory stop. The size of the samples is increased until a suitable convergence criterion is satisfied.

It is worth mentioning that, while Crainic et al. (2012) assume an operational policy which guarantees service to the ϵ-portion of the active sets with the shortest travel time, the proposed method is based on sampling and can be easily extended to more complex operations policies not implying, for example, the rejection of a request.

5.   Computational experience

The proposed experimental study has two main purposes: 1) to compare the hierarchical approaches described in Section 4 (SC and SA) and evaluate how the provided tactical designs respond to changes in the design parameters (Section 5.1) and 2) to test the flexibility of our methods on a variety of situations that could be encountered in real-life applications, such as the presence of a road network, transportation demand distributed unevenly over the service area, and high demand volumes (Section 5.2).

5.1.   Comparisons and sensitivity analyses

To carry out the comparisons between the SA and the SC approaches and to perform sensitivity analyses with respect to changes in the design parameters, we proceeded according the following methodology: 1) we randomly generated the service area, as well as the demand distributions; 2) we then considered several design parameter settings; we applied the SC and SA solution approaches to obtain two single-line DAS designs for each parameter setting considered 3) for each parameter setting and design type (SC or SA), we simulated the operations of the system on a set of service-request scenarios generated according to the given demand distributions. The tactical designs are
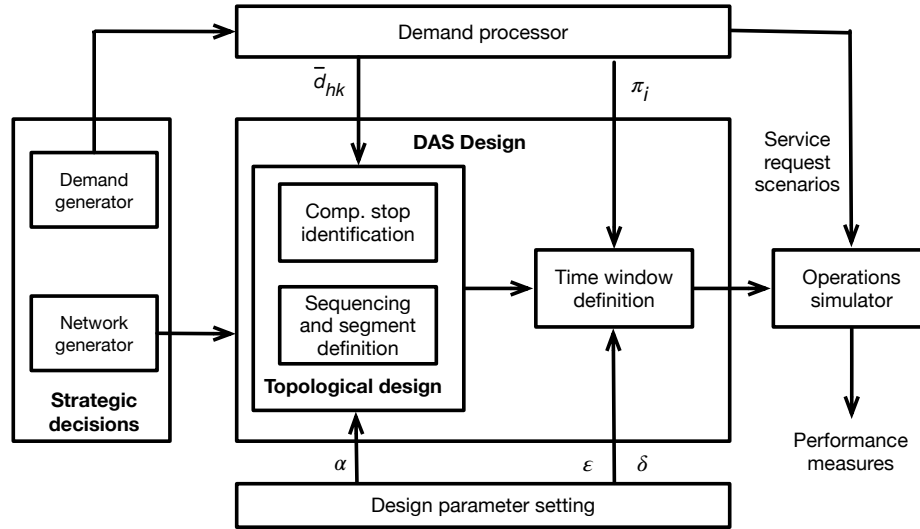
Figure 4   **Structure of the experimental method**

compared according to several performance measures such as idle time, VAW times, maximum service time, GUTT, etc.

Figure 4 illustrates the general scheme of the experimental method. The leftmost block encloses the two components of the instance generator, one providing the service area and the locations of the demand points, the other generating demand probability distribution matrices. The block in the center of Figure 4 represents the design process as described earlier. The block on the top, named *Demand processor*, is in charge of transforming the initial demand distribution in expected demand matrices for the sequencing problem and in probabilities π_i for the selection of compulsory stops and the MSP. The lowermost block represents the parameter settings used in the experimentation, namely the relative weight α of the GUTT vs. operational costs in the TSP-GL, the time-window width δ, and the minimum probability of serving all requests ϵ for the MSP. The demand processor is also in charge of sampling transportation-requests scenarios from the given demand distribution. These request scenarios form the input of the simulator, which is represented by the rightmost block in Figure 4.

The service area considered in this first part of the experimental campaign is a square with 40 demand points uniformly distributed in space. In order to set up realistic costs and distances, we scaled the edge of the square and speed values such that the duration of the Hamiltonian tour visiting the whole set of points is approximatively 6500 seconds. Such a value is important because it represents the time needed by a hypothetical traditional transit line to service the whole set of demand points (in the following we indicate such a value as Traditional Service Time, TST). Consequently it represents an upper bound of the *service* time of the DAS line (i.e., the time the vehicle needs to service the line, including riding times, idle times, time to serve customers) and

can be used to measure how much it is gained in service time when operating a DAS line instead of a traditional line.

Given that we consider the stationary-demand case, demand probability distributions are assumed constant. As previously mentioned, in real-life contexts, only a small percentage of the demand points is significantly more attractive (more requested) than others. To represent this fact, demand distributions are generated in such a way that approximatively 10% of the nodes are more requested than the others. This resulted in a service area with 4 attractive demand points. For the other demand points, the resulting average value of π_i (probability of at least one service-request for demand point i, see Section 4.4) is approximatively 30% (in Section 5.2 we also consider the case case of an almost saturated DAS line, with probabilities π_i approximatively 70%).

We considered the following design parameter values: $\alpha = \{0.1, 0.9\}$ to evaluate the effects of opposite emphasis in the objective function of the sequence problem, $\delta = \{300, 500\}$ seconds to evaluate the effects of the time-window width on idle and VAW times, and $\epsilon = \{0.85, 0.95\}$ to evaluate different policies regarding the reliability of the system (the higher the ϵ, the more reliable the system, as the probability to serve all the requests becomes higher).

The simulator operates according to the DAS1 policy (Malucelli et al. 1999), which can be described as follows: requests for transportation might be *rejected* if their acceptance causes infeasibility with respect to the master schedule. If a request is accepted, users are picked up and dropped off exactly at the location they asked for. The solution methods used to solve the operational problem are reported in Malucelli et al. (2001) and Crainic et al. (2005). In our experiments, for each topological design type, parameter setting, and input scenario, we simulated operations over 100 instances with about 25 requests on average.

We considered several measures when evaluating the performance of a particular line design: the LDT at the last compulsory stop of the line, the per-user GUTT (including VAW times), the per-user GUTT without VAW times, average per-segment idle times, average per-segment VAW times, percentage of rejected requests, maximum occupancy of the vehicle, defined as the maximum number of passengers on the vehicle at the same time. The LDT at the last compulsory stop is important because it represents the Upper bound on the Service Time of the DAS line (UST). The UST is not only important by itself, providing an indirect measure of the operating costs of the line, but also because it quantifies, when compared with the TST, how much it is gained in terms of service time compared with a traditional service. Consequently, in our tables, we usually report the UST and two related values, the Lower bound on the Service Time Improvement (LSTI = TST - UST), and the Actual Service Time Improvement (ASTI = TST-AST), where AST is the Actual Service Time (the time the vehicle actually arrives at the last compulsory stop). The per-user GUTT gives us the average time spent by users in the vehicle, including idle and VAW

Experiment	Parameters		
	α	ϵ	$\delta(s)$
Basic setting	0.1	0.95	500
α	0.9	0.95	500
ϵ	0.1	0.85	500
δ	0.1	0.95	300

Table 1   **Overview of the experimentation in Section 5.1.** $\pi_i \approx 30\%$

times. Given that VAW times might be avoided with the use of GPS devices, we also report GUTT values without them. As previously mentioned, our decomposition approach requires to perform the sequencing step before the time windows are determined, possibly implying that the GUTT component in the objective function does not correctly capture idle and VAW times. We thus also show in our tables the average per-segment idle and VAW times. Regarding the maximum occupancy of the vehicle, we observe that the tactical design, and so the S-SDDP, does not take into account capacity explicitly because this issue, related with the frequency and quality level of the service, belongs to the strategic planning. It can be interesting, however, to monitor the maximum occupancy as a way to verify that the considered time-slice intervals are reasonable and, consequently, the demand aggregation is sound.

For each input scenario and design parameter change, we analyze the resulting effects in terms of the performance measures of the system and normally we report results averaging on the number of instances. To avoid redundancies and for ease of presentation, when the variation of a certain design parameter value has different effects on the two design approaches, we underline this fact and we quantify it in tables and figures. On the contrary, when the effects are similar for the two approaches, we only report results for one of them. The experimentation is done by first defining a *basic* parameter setting and then comparing it with several alternative parameter settings differing from the basic one in the value of exactly one parameter. The parameter values in the basic setting are $\alpha = 0.1$, $\epsilon = 0.95$, $\delta = 500$ seconds. The other tested parameter and instance configurations are summarized in Table 1.

5.1.1.   Effects of the sequencing parameter α. By varying parameter α, we expect to obtain different sequences and segments, since small α values (i.e, $\alpha = 0.1$) give more emphasis to the efficiency of operations, resulting in short itineraries, while values of α close to 1 (i.e., $\alpha = 0.9$) give more emphasis to GUTT, thus producing possibly longer itineraries but with overall shorter travel times for passengers. However, due to the fact that stops are treated differently, SA and SC react quite differently to changes of α. Figure 5 represents the DAS lines output by SA when $\alpha = 0.1$ (on the left) and when $\alpha = 0.9$ (on the right). Numbers close to compulsory stops give the sequencing and dots with the same color define the stops belonging to the same segment.
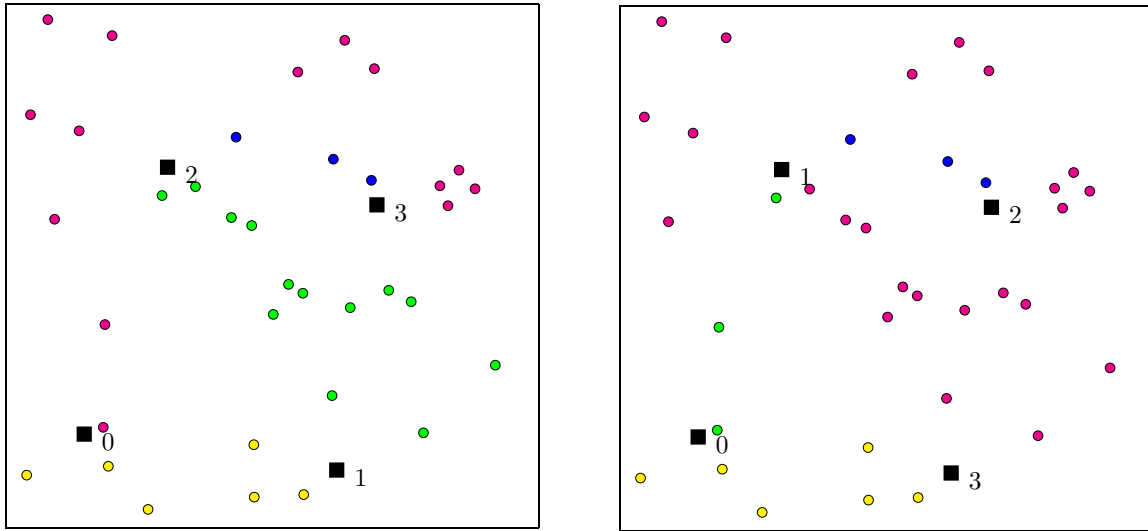
Figure 5   **SA Design approach:** $\alpha = 0.1$ **(left) and** $\alpha = 0.9$ **(right)**

The compulsory stops are the same in both cases since their selection is independent from the value of α. We observe, however, a change in the sequence of compulsory stops and the partition of demand points into segments. As expected, $\alpha = 0.1$ with emphasis on the efficiency of operations, generally results in a more linear sequence on the entire set of stops, while $\alpha = 0.9$ with emphasis on the GUTT, yields a more convoluted sequence. This has a direct consequence on the segment shapes which, for high α, tend to have a more complex configuration than for low α and this is particularly evident for the segment whose stops are colored in magenta. One could also observe that some segments remain unmodified in the two settings, e.g., the segment between compulsory stops 0 and 1 for $\alpha = 0.1$ is the same as the segment between 0 and 3 for $\alpha = 0.9$. Finally notice that, by considering compulsory stops only, when α is large, the sequence is not necessarily more involved, as in our example, though the shape of the segments is clearly more convoluted.

A similar study has been performed for the SC approach, where the sequencing step only involves compulsory stops. In this case, the relatively small number of compulsory stops does not allow α to have an influence, indeed in our example the two settings of α yield the same topological design, as illustrated in Figure 6.

We used the obtained topological designs for a numerical study and, by considering demand distributions with average $\pi_i \approx 30\%$, built the set of time windows by solving the MLP with design parameters $\delta = 500$ and $\epsilon = 0.95$. We simulated the operations over 100 service-request scenarios for each value of α and for each direction. The values of the performance measures, averaged on the number of service-request scenarios, are reported in Tables 2 and 3. The first two columns on Table
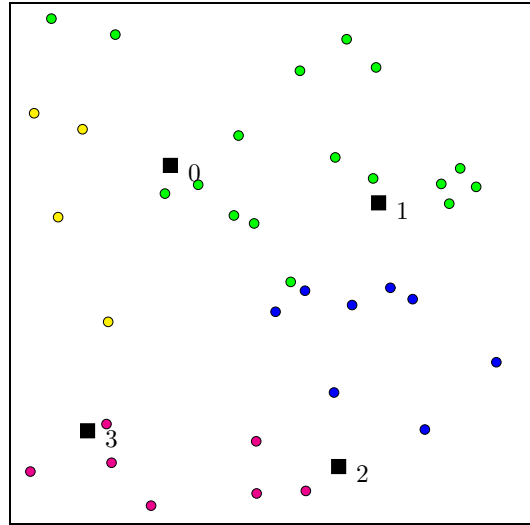
Figure 6 **SC Design approach:** $\alpha = 0.1$ **and** $\alpha = 0.9$

Design		UST (s)	%LSTI	%ASTI		%Rej	maxCap
Hier.	α			Avg	σ		
SA	0.1	5095.0	21.6	27.2	2.4	0.40	11.23
SA	0.9	5310.0	18.3	23.7	2.5	0.54	12.27
SC	0.1	5230.0	19.5	26.0	2.1	0.42	11.23
SC	0.9	5230.0	19.5	26.0	1.9	0.44	11.23

Table 2 **Different design approaches (1). Values:** $\pi_i \approx 30\%$**,** $\epsilon = 0.95$**,** $\delta = 500$

2 report the design approach as determined by the hierarchical decomposition and the value of α. From the third to the sixth column we report the UST, the percentage of the relative LSTI and ASTI (average and standard deviation). The last two columns report the average ratio of rejected requests over the total number of requests (approximatively 25 requests in each scenario) and the average maximal occupancy of the vehicle, respectively. For the same experiment, Table 3 presents some details about the timing performance of the systems. The first two columns display the design approach. The third and the fourth columns report the average value and standard deviation of the per-user GUTT (includes VAW times), respectively. The fifth and sixth columns report the average and standard deviation of the per-user GUTT without the VAW time, respectively. The next two columns report average and standard deviation of the percentage ratio between the idle time experienced by a user and the GUTT, while the last two columns display the same ratio for the VAW times.

Focusing first on SA, we observe that, as expected, the time needed on average to operate the line for $\alpha = 0.1$ is lower then for $\alpha = 0.9$, and this is confirmed by the variations of UST, LSTI, and ASTI values. A worse UST when $\alpha = 0.9$ is balanced in terms of GUTT, which is better than for the $\alpha = 0.1$ case. This tendency is best observed for the GUTT without VAW times. This is perhaps due to the fact that the considered hierarchical approach requires to optimize the sequencing step

| Design | | GUTT | | GUTT2 | | %Idle | | %VAW | |
Hier.	α	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
SA	0.1	1476.35	149.0	1396.86	123.5	3.50	4.8	10.41	7.5
SA	0.9	1409.83	167.6	1295.20	135.7	5.28	7.2	15.81	7.9
SC	0.1	1413.36	131.5	1361.63	117.3	11.90	8.2	7.12	6.0
SC	0.9	1411.17	127.7	1360.79	117.0	11.97	8.2	6.99	5.6

Table 3   **Different design approaches (2). Values:** $\pi_i \approx 30\%$, $\epsilon = 0.95$, $\delta = 500$

before the actual time windows are determined, preventing in this way the possibility to explicitly account for idle and VAW times. We finally notice that, for $\alpha = 0.9$, the idle and VAW times are slightly worse and we observe that this is related to the higher value of the UST. We also observe that idle and VAW time display a high value of the standard deviation. A possible reason is that idle and VAW times at compulsory stops are mutually exclusive, as either the vehicle arrives before the EDT (idle times) or after (VAW times). The simulation also confirms that the SC approach is weakly sensitive to the parameter α.

It is interesting to compare the two methods. The first evident difference is that SA is much more sensitive to variations of α then SC. Another difference is that the configuration of the segments for the SA case is more complex than for the SC case. Notice that, despite this complexity, the best average value of the GUTT is obtained with SA and $\alpha = 0.9$ (once again, this tendency is more evident for the GUTT without VAW times) and that the best value for UST is obtained again with SA but with $\alpha = 0.1$. However, to the best value of the GUTT also corresponds the worse value of UST and, analogously, to the best values of UST corresponds also the worst value of GUTT. This suggests that, according to the particular applications, for SA, the values of α can be modulated to obtain different efficiency/GUTT tradeoffs.

For general comments on DAS behavior, the values of LSTI are quite interesting as they emphasize, in our opinion, the potential for cost reductions of DAS in the presence of relatively low demand ($\pi \approx 30\%$ in this case). Observe that, to a consistent reduction of the service times with respect to a traditional line (LSTI around 20% for both approaches and both values of α), corresponds a very low percentage of rejected requests (less than 0.6% of the total demand in all cases). This is even more evident when we compare LSTI with ASTI. The actual service time is considerably shorter than its upper bound and this might imply advantages both on the vehicle availability and the operation costs (e.g., fuel consumption).

5.1.2. Effects of the Master Schedule Parameters, ϵ and δ. Parameters ϵ and δ concern the definition of the time windows in the MSP. To test the effects of ϵ on the system behavior, we performed simulations on two designs of a DAS line obtained from two parameter settings differing only in the value of ϵ. The rest of the parameter setting was $\alpha = 0.1$ and $\delta = 500$, and the demand scenario considered was such that $\pi_i \approx 30\%$. As in the previous case, the simulation

| ϵ | UST (s) | %LSTI | %ASTI | | %Rej | maxCap |
			Avg	σ		
0.95	5230.0	19.5	26.0	2.1	0.42	11.23
0.85	4812.5	26.0	31.7	2.5	1.27	11.11

Table 4   Master Schedule ϵ **(1). Settings:** $\pi_i \approx 30\%$, $\alpha = 0.1$, $\delta = 500$, **DesignType=SC**

| ϵ | GUTT | | GUTT2 | | %Idle | | %VAW | |
	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
0.95	1413.36	131.5	1361.63	117.3	11.90	8.2	7.12	6.0
0.85	1344.40	142.1	1269.29	111.5	7.34	7.1	10.71	7.9

Table 5   Master Schedule ϵ **(2). Settings:** $\pi_i \approx 30\%$, $\alpha = 0.1$, $\delta = 500$, **DesignType=SC**

was performed on 100 service-request scenarios for each direction. We report in Tables 4 and 5 the results averaged over the set of requests scenarios and the meaning of the column is the same of the previous table. Given that we did not observe significant differences in how SA or SC lines responded to the parameter variation, we report only the values for the SC approach.

We observe that, as expected, for $\epsilon = 0.95$ the percentage of rejected requests is less than for $\epsilon = 0.85$, but this is paid in terms of UST, LSTI, ASTI and per-user GUTT. It is interesting to observe that while the idle times increase, the VAW times decrease for higher setting of ϵ. To understand this, remember that the closer to 1 is the value of ϵ, the higher is the probability the system will be able to serve all the requests. To accomplish this, the design has to properly adjust the time windows. In particular, it seems that to an increased ϵ, the design process responds with increased *distances* between time windows. This explains the changes for idle and VAW times. In fact, if the distance between time windows increases but the demand volumes remain the same, the probability the vehicle arrives earlier than the EDT increases and consequently longer idle times are expected. Conversely, because the increased probability that the vehicle departure time is exactly at or close to the EDT, the expected VAW time decreases.

We performed for parameter δ a study similar to the one done for ϵ and performed simulations on two designs of a DAS line obtained from two parameter settings differing in the only value of δ. The rest of the parameter setting was $\alpha = 0.1$ and $\epsilon = 0.95$, and the demand scenario considered was such that $\pi_i \approx 30$. As in the previous case, the simulation was performed on 100 requests scenarios for each direction. We report in Tables 6 and 7 the results averaged over the set of requests scenarios. The meaning of the columns is the same as for the previous experimentation. Because we did not observe significant variations in the way the design produced by SA or SC responded to the parameter variation, we report only the values for the SC approach.

We recall from the discussion in Section 4.4, that time windows width is the tool to commute flexibility among adjacent segments and that the main obstacle in increasing δ is that c-users have to wait, in the worst case, for the whole length of the time window. By inspecting the values

δ	UST (s)	%LSTI	%ASTI		%Rej	maxCap
			Avg	σ		
500	5230.0	19.5	26.0	2.1	0.42	11.23
300	5400.0	16.9	20.8	1.1	0.50	11.20

Table 6 Master Schedule: δ **(1)**. Settings: $\pi_i \approx 30\%$, $\alpha = 0.1$, $\epsilon = 0.95$, **Design Type=SC**

δ	GUTT		GUTT2		%Idle		%VAW	
	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
500	1413.36	131.5	1361.63	117.3	11.90	8.2	7.12	6.0
300	1470.30	124.2	1448.73	120.2	17.10	8.4	2.89	2.7

Table 7 Master Schedule: δ **(2)**. Settings: $\pi_i \approx 30\%$, $\alpha = 0.1$, $\epsilon = 0.95$, **Design Type=SC**

corresponding to the VAW times in Table 7, we observe, as expected, a higher value for $\delta = 500$. Notice, however, that the ratio of VAW time over GUTT is still quite low. We also observe that wider time windows allow for a better use of flexibility. In fact, even though the probability that the system is able to serve all the stops is the same, both efficiency and GUTT improve when higher δ are considered, as is it possible to deduce by inspecting the values of UST, ASTI, GUTT and GUTT2. The idle times also improved because, for the mechanism described earlier, the probability the vehicle arrives at compulsory stops before the earliest departure time is lower and consequently lower are the expected idle times. Generally speaking, it seems that, when increasing δ, to a small deterioration of the VAW times corresponds a much better performance underlined by the improvements all the other performance measures. Also, wider time windows seem to find a further justification in those cases where the adoption of GPS devices to reduce the effect of VAW time is possible.

5.1.3. Considerations Summarizing the results of this first part of the experimentation, we observe that the decomposition methods SA and SC mostly differ in the provided topological design, while responding quite similarly to variations of parameters affecting the time windows. In particular, we observed that SA usually produced segments with more complex configurations than SC, but that those "counter-intuitive" configurations were able, under specific settings of parameter α (the relative weight in the objective function between GUTT and operating costs), to provide the design with lower UST or GUTT. We also observed that SC, when the number of compulsory stops is relatively low, is weakly sensitive to variations of α. Regarding the effects of the MSP parameters, we observed that to higher values of ϵ (the probability to be able to serve all the customers) usually correspond longer distances among time windows, with the effect of worsening the performances of the system. The effects of variation of the time windows width δ are very interesting: while to higher values of δ statically correspond increasingly worse values of VAW times, the experimentation shows that the actual increment of VAW times is quite small. The advantages derived from the possibility of commuting flexibility between consecutive segments

actually allow for shorter distances between time windows and this considerably improves the performance.

A last observation is related to the computational efficiency. Both SA and SC in the first phase have to solve the TSP-GL, in the second the MLP. The running times required by those algorithms are reported in Errico et al. (2016) and Crainic et al. (2012). The main difference between the two decomposition strategies is that SA must run the TSP-GL on the whole set of demand points while SC only on compulsory stops and this makes SA more challenging from the computational point of view than SC.

5.2.   Versatility of the proposed solution approaches

This section is primarily aimed at demonstrating the ability of the proposed solution approaches to adapt to situations that may be encountered in real-life contexts. As a secondary outcome, this section also provides further support to the conclusions drawn in the previous section. We analyse the behaviour of the proposed solution methods and tactical designs in the presence of a common type of road network, such as a Manhattan network (Section 5.2.1) in the presence of customers non-uniformly distributed over the service area (Section 5.2.2) and in presence of a high demand (Section 5.2.3). In order to avoid repetitions, only one solution approach is analysed in each of the following sections.

5.2.1.   Effects of a street network
The experimental campaign in Section 5.1 was entirely conducted on an Euclidean network, while in urban and sub-urban settings, vehicle movements are generally constrained to a road network. Therefore, it is important to explore the impact of the presence of a road network on the results obtained in the previous section. We thus performed a new experiment where we considered the same service area as in Section 5.1, but we assumed a road network in the shape of a grid, with the region divided in 25 squares. The choice of this network topology is motivated by the fact that many North American cities are built on a grid-like network. To make the experiment coherent with the considered road network, we moved the demand points to the closest point on the grid. We then applied the two hierarchical decompositions and analysed the obtained results for $\alpha = \{0.1, \ 0.9\}$, while the other parameters were set to $\delta = 500$ and $\epsilon = 0.95$. The general methodology of the experimentation follows the lines of Section 5.1. It is worth observing that in order to perform this experiment no modification was needed in the proposed algorithms because the only formal change with respect to Section 5.1 were in the values of the input distance matrix.

For brevity, in this section we report the results obtained for the SA approach only. Figure 7 and Tables 8 and 9 summarize the results, where the color-coding of the figure and the meaning of the columns in the tables are the same as in the previous section.
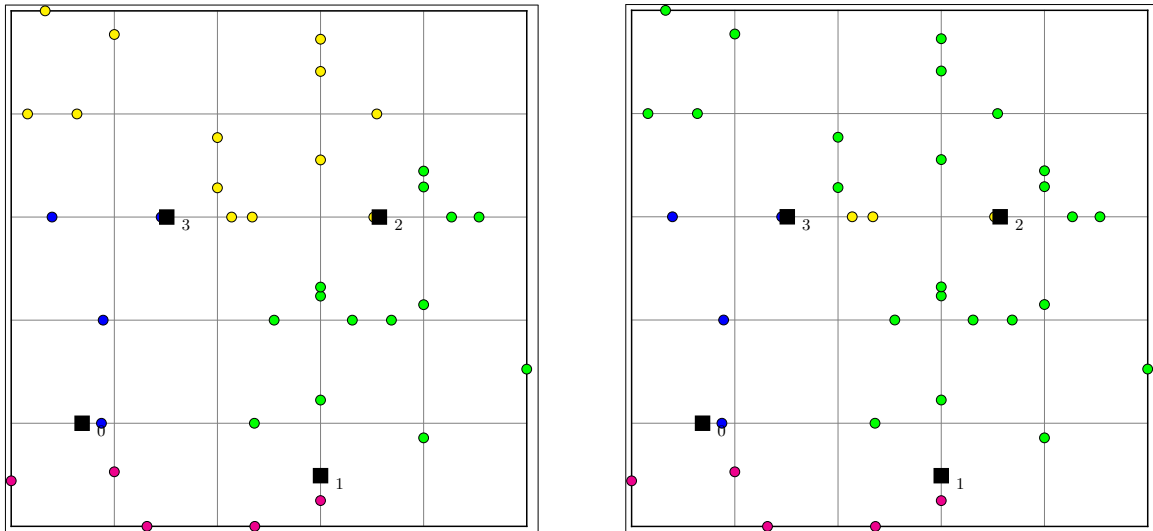
Figure 7   SA approach on a grid-like network: $\alpha = 0.1$ **(left)** and $\alpha = 0.9$ **(right)**

Design		UST (s)	%LSTI	%ASTI		%Rej	maxCap
Hier.	α			Avg	σ		
SA	0.1	6137.5	20.9	25.7	1.9	0.51	12.02
SA	0.9	6577.5	15.2	20.4	1.8	0.24	12.52

Table 8   **SA on a grid-like network(1). Values:** $\pi_i \approx 30\%$, $\epsilon = 0.95$, $\delta = 500$

Design		GUTT		GUTT2		%Idle		%VAW	
Hier.	α	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
SA	0.1	1719.75	154.4	1655.54	138.7	9.50	8.0	7.33	4.1
SA	0.9	1617.68	184.1	1545.11	169.3	9.56	10.7	8.80	5.6

Table 9   **SA on a grid-like network (2). Values:** $\pi_i \approx 30\%$, $\epsilon = 0.95$, $\delta = 500$

Examining Figures 5 and 7, we observe that the presence of a road network can have a significant impact on the way optional stops are assigned to segments, as well as on the sequence of compulsory stops. These figures also confirm the tendency of the SA approach to designing intricate segments. By comparing Tables 2 and 3 to 8 and 9, we notice a significant increase in the UST, as well as in the GUTT values. This is not surprising because the network topology makes the routing choices more constrained. However, by comparing the performance of the DAS line with the traditional line, namely the values LSTI and ASTI, we observe that they are similar to those in the previous experiment. This is due to the fact that the network restrictions also impact the traditional TST by increasing it from 6500 to 7758 seconds. The rest of the performance measures are comparable with the previous experiment. We also observe that by setting the parameter $\alpha = 0.9$ and adopting the SA approach, it is possible to reduce the values of the GUTT, even thought the corresponding segment configuration appears to be more convoluted, thus confirming the analyses in Section 5.1.1.

5.2.2.   Effects of clustered customers The demand points considered in Section 5.1 were generated according to the uniform distribution on the service area. However, real-life contexts

frequently present non-uniform distribution of the customers' location. For this reason, in the present section we investigate how our methods and tactical designs behave in presence of non-uniformly distributed customers. We thus generated a new service area displaying clustered demand points. In particular, we generated 60% of the total number of demand points according to the uniform distribution. For the remaining 40% we proceeded as follows. We divided the region in 16 squares of equal size, and we picked two such squares randomly. The remaining demand point were generated randomly, but constrained to be located in the picked squares. Regarding the demand volumes, this instance turned out to have one more attractive demand point, resulting in a total of 5 poles. For the other demand points, the average probability of at least one service-request for stop i is approximatively $\pi \approx 30$, as in Section 5.1. We then applied the two hierarchical decompositions and analysed the obtained results for $\alpha = \{0.1, 0.9\}$, while the other parameters were set to $\delta = 500$ and $\epsilon = 0.95$. It is worth noticing that no modification was required in the proposed algorithms in this case also, given that the only changes were in the input distance and average demand matrices. For brevity, in this section we report the results obtained for the SC approach only. Figure 8 and Tables 10 and 11 summarize the results.
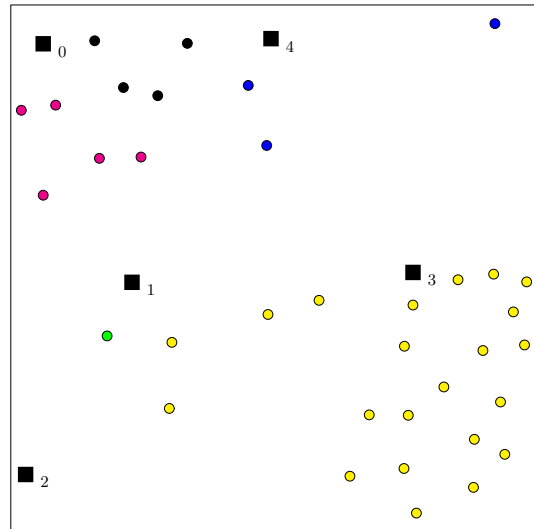

Figure 8   **SC approach on clustered customers:** $\alpha = 0.1$ and $\alpha = 0.9$

Design		UST (s)	%LSTI	%ASTI		%Rej	maxCap
Hier.	α			Avg	σ		
SC	0.1	5547.5	16.0	22.0	1.9	0.29	14.72
SC	0.9	5547.5	16.0	22.0	1.9	0.29	14.72

Table 10   **SC on clustered customers. Values:** $\pi_i \approx 30\%$, $\epsilon = 0.95$, $\delta = 500$

| Design | | GUTT | | GUTT2 | | %Idle | | % VAW | |
Hier.	α	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
SC	0.1	1520.57	135.0	1439.99	116.7	4.92	3.8	10.42	4.8
SC	0.9	1520.13	133.9	1439.55	116.6	4.93	3.8	10.43	4.7

Table 11 **SC on clustered (2). Values:** $\pi_i \approx 30\%$, $\epsilon = 0.95$, $\delta = 500$

As clear in Figure 8 and from the similarity in the values of the first and second rows in Tables 10 and 11, this experiment confirms the weak sensitivity of the SC approach to the parameter α. Moreover, we observe that the SC approach confirms the tendency showed in Section 5.1 to designing relatively simple segments. This experiment cannot be directly compared with the basic case in Section 5.1.1 because the location of the demand points and the number of attractive poles differ. However, it is still interesting to make some considerations when confronting Tables 2 and 3 with Tables 10 and 11. We notice a weak increase in the UST, as well as the GUTT values. By comparing the performance of the DAS line with respect to the traditional line and by inspecting the values LSTI and ASTI, we observe that there is a slight deterioration. This might be due to the fact that in this experiment we have five segments as opposed to four in Section 5.1, thus requiring the vehicle to always pass by an additional demand point. The maximum occupation also slightly increased given that the overall number of requests turned out to be marginally higher, due to the new attractive pole.

5.2.3. High demand. DAS systems are known to be most suitable for low-to-medium demand volumes (see Potts et al. 2010, Errico et al. 2013, for example). However, demand could considerably vary during the course of a day, as we discuss more in depth in Section 6. It is then important to observe how the system performs under different levels of demand volumes. We therefore considered the same service area and demand points as in Section 5.1, but increased the demand volumes to an average value of $\pi_i \approx 70\%$ (probability of at least one service-request for the optional demand point i). Observe that $\pi_i = 70\%$ implies demand volume close to levels typical of traditional transit. We then applied the two hierarchical decomposition methods, and analysed the results obtained with design parameters set as for the basic case, namely $\alpha = 0.1$ and $\delta = 500$, $\epsilon = 0.95$. For brevity, in this section we comment the results obtained for the SC approach only. Furthermore, because the values of π_i have no influence on the topological design, we focus on the results related to the master schedule. We report a comparison in Tables 12 and 13. The meaning of the columns is as previously, but we inserted a new column in Table 12, position sixth, to report also the average total amount of requests.

We observe that to assure the value of $\epsilon = 0.95\%$ (the probability that the system is able to service all requests) with high demand ($avg(\pi_i) \approx 70\%$), UST, LSTI, ASTI and GUTT considerably worsen, in particular the UST. This is due to the fact that the obtained Master Schedule increased the

34

Errico et al.: *The DAS Design Problem*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

$avg(\pi_i)$	UST (s)	%LSTI	%ASTI		TotReq	%Rej	maxCap
			Avg	σ			
0.3	5230.0	19.5	26.0	2.1	25.8	0.42	11.23
0.7	6097.6	6.2	12.1	2.1	49.7	0.22	20.57

Table 12   Effects of demand volumes (1). Settings: $\alpha = 0.1$, $\epsilon = 0.95$, $\delta = 500$, **DesignType= SC**

$avg(\pi_i)$	GUTT		GUTT2		%Idle		%VAW	
	Avg (s)	σ	Avg (s)	σ	Avg	σ	Avg	σ
0.3	1413.36	131.5	1361.63	117.3	11.90	8.2	7.12	6.0
0.7	1658.02	117.0	1585.39	98.5	6.96	7.1	8.56	5.9

Table 13   Effects of demand volumes (2). Settings: $\alpha = 0.1$, $\epsilon = 0.95$, $\delta = 500$, **DesignType= SC**

distance between consecutive time windows to allocate more time to service requests. Concerning idle and VAW times we observe that for $avg(\pi_i) = 70\%$ idle times slightly decrease while VAW times increase. The reasons for this are linked to the time windows mechanism as analysed in Section 5.1.2: because the number of requests are likely higher, the resulting probability that the vehicle arrives earlier than the earliest departure time at compulsory stop decreases. Finally, while the performances of the DAS generally deteriorate when $avg(\pi_i) = 70\%$, by inspecting columns reporting the LSTI and ASTI, we also notice that to a reduction of the service time close to 6% in the worse case and close to 12% on average, corresponds a very low percentage of unserviced requests (less than 0.3%).

5.2.4.   Considerations We tested the proposed solution methods, as well as the obtained tactical designs on a number of cases that could be encountered when designing a DAS tactical plan in real-life applications, in particular we studied the impact of a grid-like network, of non-uniformly distributed customers, and of high demand volumes. Overall, the main conclusions of Section 5.1 were confirmed, in particular the tendency of the SA approach to produce topological design with convoluted segments, and the low sensitivity of the SC approach to parameter α. The presence of a grid-like network did not change the performance ratios with respect to a traditional line, but the overall line duration increased, as expected, due the network restrictions. The presence of non-uniformly distributed customers does not seem to have a significant impact on the system performances, while the presence of an additional compulsory stop is likely to be the source of performance deterioration. Finally, even though guaranteeing high percentage of accepted request under high demand volumes generally requires an increased time duration of a DAS line (UST), the results showed the there is a significant gain with respect to a traditional line, even though reduced. We conclude by noticing that the experimentation in the present section did not require any modification to the proposed algorithms, as the only required changes were in the input data, thus showing once again the versatility of the proposed approach.

6.   Applicability of the S-SDDP in real-life contexts

The S-SDDP and the solution methods presented in this paper stand on the stationary-demand assumption. In this section we will discuss about the applicability of the S-SDDP for the general non-stationary case.

In transit contexts it is common to deploy seasonal tactical plans (for example, for summer or fall). Within each season, it is also common to produce several tactical plans, each accommodating some specific time period (week days, weekends, or nigh hours, for example). These plans repeat during the given season (the week-day plan is repeated at each day of the week in the season, for example), but are typically modified at each season. However, even when restricting the tactical plan to a specific time period such as a week-day, demand fluctuations may be observed, thus violating the stationary-demand hypothesis made in this paper.

The question of how a transit systems should accommodate demand fluctuations is a general issue and does not concern DAS lines only. Most planning techniques for traditional transit build the general transit network and approximated service frequencies in a first phase, based on aggregated data. Frequencies and timetables are adjusted in a second phase, according to the expected demand in a given time period. In practice, the transit system only marginally adapts to the demand fluctuations, while the general service structure stays the same. Nowadays technologies, however, open the door to more advanced management strategies of the transit systems, potentially accommodating a higher degree of flexibility.

When defining the Single Line DAS Design Problem for the general non-stationary case (SDDP), we need to establish what type of action the system is allowed perform in order to accommodate demand variations. The range of possible strategies is wide, the choice depending on several factors including planning objectives, type of users, available technologies, transit agency policies, etc. A complete investigation of all the possible alternatives is beyond the scope of this paper. However we describe here three representative settings. One possibility is to consider the design responding to every demand variation. At the other extreme of the spectrum, one can consider the design of the system fixed and not responding to any demand variations. In between, there is a range of solutions where part of the design remains fixed and part changes according to demand fluctuations. Let us now see how the solution of the S-SDDP can be used to obtain the solution of the SDDP for the three case considered.

In the first scenario considered, the design responds to every demand fluctuation. This means that the design might change at every stationary time period present in the time horizon. To see how the solution of the S-SDDP can be used, it is sufficient to consider the stationary time periods in the time horizon and partition it in time slices according to the frequency requirements established at strategic planning level. Then consider, for each stationary period, demand data

related to a single slice and solve the S-SDDP. The solution of the SDDP simply consists of the repetition of the S-SDDP solutions for each stationary period.

The second scenario considers the design never responding to demand fluctuation. This means that the design needs to account for different demand volumes at the same time. In this case it is possible to aggregate data over the whole time horizon and make the approximation that the demand is stationary. One can than deduce demand data relative to a single time slice interval. Solve the S-SDDP and repeat the same solution for every slice in the considered time horizon.

The last scenario considers the design partly fixed and partly responding to demand variations. There are several possible ways to choose what part of the design has to be fixed. We consider here a case inspired from a common practice in traditional transit. In such a context, the stops and the sequence of a bus line are usually predefined and fixed, independently of demand volumes. However, schedules might vary in correspondence to demand variations (for examples, schedules differ for rush and night hours). A similar approach for the SDDP would consider a design where the topological part is fixed while the master schedule is adjusted according to demand variations. To obtain the topological design, similarly to the second scenario, one should aggregate the data over the whole time horizon, deduce demand data relative to a single slice, find the topological design and keep it fixed along the time horizon. To obtain the master schedules, similarly to the first scenario, one should consider, for each homogeneous period, demand data related to a single slice, and solve the MSP. Then repeat the master schedule relative to it for every time slice in a given homogeneous period,

The brief analysis above explains how the methods developed in the present paper to build the S-SDDP can be easily used to build the solution of the SDDP for the general non-stationary case.

7. Conclusions

The present paper addressed tactical planning issues for the Demand Adaptive System, which is a general model for a wide class of *semi-flexible* transit systems. In particular, we introduced and mathematically formulated the single-line DAS design problem with stationary demand. For this difficult decision problem, we proposed two alternative hierarchical decomposition strategies, the Sequence All and the Sequence Compulsory (SA and SC, respectively). SA and SC differ in the decision-taking order, but they share the resulting core sub-problems.

We performed an extensive computational study which was made of two main steps, in the first we studied the effects of the main design parameters on the system performances, while in the second we tested the versatility of the proposed approaches on a number of situations that could be encountered in real-life applications.

The first part of the experimental campaign showed that, under low-to-medium demand volumes, the DAS performances are attractive when compared to traditional transit, as to a significant

reduction in terms of service times corresponded very small percentages of requests that could not be serviced within the computed time windows. The comparison between SA and SC revealed that, although SA usually produces more complex configuration of the segments, it is more sensitive to variations of the relative weight of the GUTT with respect to operating costs in the objective function. In fact, under specific settings of this weight, SA produced the best designs from the GUTT or operational-cost points of view. On the other hand, SA is computationally more demanding than SC. Results also showed that SA and SC respond similarly to variations in the values of design parameters related to the master schedule. In particular, to higher probability values to serve all the customers corresponded a general deterioration of the system performances. Results also showed that it is generally more advantageous to set relatively high values of the time window widths because the actual observed VAW times only slightly increased. Furthermore, wider time windows allowed for a better use of flexibility and this considerably improved performances.

This general behaviour was confirmed in the second part of the experimental campaign, where we tested the proposed approaches on a service area with a grid-like network, as well as with non-uniformly distributed customers. The number of compulsory stops seemed to have an impact on the service time duration of the DAS line. When tested under high demand volumes, the performance measures generally deteriorated, although results showed a significant residual advantage with respect to a traditional line. The main outcome of this second part of the experimentation was that the proposed methodology could be easily applied to a number of situations that can be encountered in real-life conditions.

We finally showed how the methodology developed for the S-SDDP can be applied to solve the general SDDP, illustrating three representative particular cases.

As for future work, we believe that it would be a valuable contribution to perform a case study on a real existing area. The major challenges will likely be in determining the limits of the service area and obtaining correct estimations of the demand volumes and, consequently, in defining the time-slices. From the methodological point of view, we observed how both hierarchical decompositions require to optimize the sequencing step before the time windows are determined, and that this makes it difficult to provide sequences explicitly accounting for idle and VAW times. A possible generalization of the present paper would explicitly embed the stochasticity of the customer presence in the sequencing problem, adopting, for example, a two-stage stochastic optimization model with recourse. Another interesting research avenue is to extended our model and methodology to the case of multiple DAS lines. The major research challenge will be to manage possible transfers among DAS lines.

Acknowledgments

38

Errico et al.: *The DAS Design Problem*
Article submitted to *Transportation Science*; manuscript no. (Please, provide the mansucript number!)

References

Ceder, A., H. M. Wilson. 1997. Public Transport Operations Planning. Charles ReVelle, Arthur E. McGarity, eds., *Design and Operation of Civil and Environmental Engineering systems*. John Wiley & Sons, Inc., New York, 395–434.

Cordeau, J.-F., G. Laporte. 2003. The Dial-a-Ride Problem (DARP): Variants, Modeling Issues and Algorithms. *4OR* **1**(2) 89–101.

Crainic, T. G., F. Malucelli, M. Nonato. 2001. Flexible Many-to-few + Few-to-many = An Almost Personalized Transit System. *Preprints TRISTAN IV - Triennial Symposium on Transportation Analysis*, vol. 2. Faculdade de Ciências da Universidade de Lisboa and Universidade dos Açores, São Miguel, Açores, Portugal, 435–440.

Crainic, T. G., F. Malucelli, M. Nonato, F. Guertin. 2005. Meta-Heuristics for a Class of Demand-Responsive Transit Systems. *INFORMS Journal on Computing* **17**(1) 10–24.

Crainic, T.G, F. Errico, F. Malucelli, M. Nonato. 2012. Designing the master schedule for demand-adaptive transit systems. *Annals of Operations Research* **194** 151–166.

Diana, M., M. M. Dessouky, N. Xia. 2006. A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research Part B: Methodological* **40**(8) 651 – 666.

Errico, F. 2008. The design of flexible transit systems: models and solution methods. Ph.D. thesis, Politecnico di Milano, Italy.

Errico, F., T. G. Crainic, F. Malucelli, M. Nonato. 2013. A survey on planning semi-flexible transit systems: Methodological issues and a unifying framework. *Transportation Research Part C: Emerging Technologies* **36** 324–338.

Errico, F., T. G. Crainic, F. Malucelli, M. Nonato. 2016. A Benders Decomposition Approach for the Symmetric TSP with Generalized Latency arising in the design of Semi-flexible Transit Systems. *Transportation Science* **51**(2) 706–722.

Frei, C., M. Hyland, H. S. Mahmassani. 2017. Flexing service schedules: Assessing the potential for demand-adaptive hybrid transit via a stated preference approach. *Transportation Research Part C: Emerging Technologies* **76** 71 – 89.

Fu, L. 2002. Planning and Design of Flex-Route Transit Services. *Transportation Research Record* **1791** 59–66.

Koffman, D. 2004. Operational Experiences with Flexible Transit Services. Vol. 53 of TCRP Synthesis Report. Transportation Research Board.

Malucelli, F., M. Nonato, T. G. Crainic, F. Guertin. 2001. Adaptive Memory Programming for a Class of Demand-Responsive Transit Systems. Voß, S., Daduna, J.R., eds., *Computer-Aided Scheduling of Public Transport*, *Lecture Notes in Economics and Mathematical Systems*, vol. 505. Springer, Berlin, 253–273.

Malucelli, F., M. Nonato, S. Pallottino. 1999. Some Proposals on Flexible Transit. Ciriani, T.A., Johnson, E.L., Tadei, R., eds., *Operations Research in Industry*. McMillian, 157–182.

Potts, J. F., M. A. Marshall, E. C. Crockett, J. Washington. 2010. A Guide for Planning and Operating Flexible Public Transportation Services. Vol. 140 of TCRP Synthesis Report. Transportation Research Board.

Quadrifoglio, L., M. M. Dessouky, F. Ordóñez. 2008. A simulation study of demand responsive transit system design. *Transportation Research Part A: Policy and Practice* **42**(4) 718 – 737.

Quadrifoglio, L., R. W. Hall, M.M. Dessouky. 2006. Performance and design of mobility allowance shuttle transit services: Bounds on the maximum longitudinal velocity. *Transportation Science* **40**(3) 351–363.

Smith, B. L., M. J. Demetsky, P. K. Durvasula. 2003. A Multiobjective Optimization Model for Flexroute Transit Service Design. *Journal of Public Transportation* **6**(1) 89–100.

Wilson, H., I. Sussman, H. Wang, B Higonnet. 1971. Scheduling algorithms for dial-a-ride systems. Tech. Rep. USL TR-70-13, Urban Systems Laboratory, MIT, Cambridge, MA.

Zhao, J, M. M. Dessouky. 2008. Service capacity design problems for mobility allowance shuttle transit systems. *Transportation Research Part B: Methodological* **42**(2) 135 – 146.