CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

An Unifying Framework and Review of Semi-Flexible Transit Systems

Fausto Errico
Teodor Gabriel Crainic
Federico Malucelli
Maddalena Nonato

October 2011

CIRRELT-2011-64

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

An Unifying Framework and Review of Semi-Flexible Transit Systems

Fausto Errico[1,*], Teodor Gabriel Crainic[1], Federico Malucelli[2], Maddalena Nonato[3]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Management and Technology, Université du Québec à Montréal, P.O. Box 8888, Station Centre-Ville, Montréal, Canada H3C 3P8

[2] Dipartimento di Elettronica e Informazion, Politechnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy

[3] Dipartimento di Ingegneria, Università di Ferrara, Via Saragat 1, 44122, Ferrara, Italy

Abstract. 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 their extreme flexibility. Consequently, new transportation alternatives, combining characteristics from both the traditional transit and dial-a-ride, have been introduced. This family of transit systems is usually called semi-flexible. The twofold nature of semi flexible systems requires a complex planning activity and, consequently, a formalization of the decisions and the use of unifying models. While this has been done for traditional transit, no such effort can be found in the literature for semi flexible systems. The present paper aims at filling this gap by providing a systematic treatment of the field of semi-flexible systems. We proceed through three main steps. We first propose a general and unifying modeling framework for the class of semi-flexible transit systems taking the form of the Demand Adaptive System (DAS). We then provide a classification of the planning decisions, inspired by what has been done for the planning of traditional transit. As third and final step, we take advantage of the common framework offered by the classification to report a comprehensive and comparative literature review of the field of semi-flexible systems, including methodological contributions as well as a number of particularly significant practical experiences.

Keywords. Public transit, demand-responsive systems, dial-a-ride, semi-flexible transportation, demand-adaptive systems, planning, general modeling framework, literature review.

Acknowledgements. Funding for this project has been provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Industrial Research Chair and Discovery Grants programs, by our partners CN, Rona, Alimentation Couche-Tard, the Ministry of Transportation of Québec, and by the Fonds de recherche du Québec - Nature et technologies (FQRNT) through its Team Research grants program.

---

* Corresponding author: Fausto.Errico@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2011

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 a high degree of resource sharing and good level of service. In contrast, when the demand for transportation is low or sparse, the potential for resource sharing drops drastically, particularly because of the fixed structure of the traditional transit. For this reason, mass transit evolved towards some degree of flexibility. A first attempt in this direction extended to general customer service the well-known Dial-A-Ride systems (*DAR*), originally designed to serve people with reduced mobility. With respect to traditional transit, DAR provides a more *personalized* service by modifying itineraries, schedules and stop locations according to the transportation needs of users at a given time. At the same time, it still guarantees a certain degree of resource sharing by serving requests *collectively*.

The adaptation of DAR to servicing the 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 the 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. For similar reasons, it is difficult to integrate DAR with traditional transit services.

During the last decades, in order to address some of these issues, scientists and practitioners explored new transportation paradigms with the goal of combining regular traditional transit systems with pure on-demand services such as DAR. Efforts were aimed at efficiently coordinating traditional transit and DAR, e.g., Häll et al. (2009); Häll (2006); Hickman and Blume (2000, 2001). Proceeding toward integration, other approaches aimed to combine the different characteristics of traditional transit and DAR in the *same* transit system. Koffman (2004) and, more recently, Potts et al. (2010) report several attempts in this direction undertaken in North America. The systems that were proposed and implemented vary considerably in terms of organization, fleet managements, policies, and so on. They all present some basic common features, however. On the one hand, a set of stops with a fixed, predetermined timetable, similarly to traditional transit. On the other hand, part of the service is flexible and, as in a DAR context, users can ask for service at optional locations.. The fundamental idea behind such systems is that the regularity of the service is by itself a valuable property of public transportation because it helps users plan their trips, facilitates integration with other transportation modes, makes it possible to access the service without booking, etc. At the same time, such systems try to inject flexibility by considering "additional" time for the route of the vehicle, which can be used to possibly deviate from a basic path to operate services in a demand-responsive framework. We call such systems *semi-flexible*, to underline the fact that they are conceived as a compromise between the flexibility of DAR and the fixed structure of traditional transit. The main focus of the present paper are semi-flexible systems.

Transportation systems dedicated to service several demands with the same vehicle require complex planning activities, the resulting plan significantly determining the overall behavior and performance of the system. Semi-flexible services, combining characteristics of traditional and on-demand systems, require both a service-design phase and operational-level time and user request-dependent, adjustments of vehicles routes and schedules.

Such a high level of planning complexity requires a formalization of the decision process and the utilisation of unifying models. This has been done for traditional transit (e.g., Desaulniers and Hickman 2007). No such effort can be found in the literature for semi-flexible systems, however. As a result, even though the number of contributions to various planning aspects of semi-flexible transit system is consistent, there is a general lack of comprehensive view of the problem. Most contributions focus on specific aspects only, each assuming very particular assumptions, sometimes application specific, sometimes very simplified. This makes it hard, if not impossible, to relate, compare, and unify the efforts of different lines of research into a comprehensive view of the field.

The present paper aims at filling this gap by presenting a comprehensive literature review and a unifying modeling framework for representing and planning semi-flexible systems. We proceed through three main steps. We first propose a general and unifying modeling framework for the class of semi-flexible transit systems taking the form of the Demand Adaptive System (DAS) introduced in Malucelli et al. (1999). We show, in particular, not only that the DAS framework encompasses all the semi-flexible systems described in Koffman (2004) and Potts et al. (2010), which are the most accurate reviews on actual implementations of semi-flexible systems we are aware of, but that DAS also offers a more advanced scheduling mechanism than most.

We then provide a classification of the planning decisions for DAS, inspired by what has been done for the planning of traditional transit. It is noteworthy, however, that the more complex configuration of semi-flexible systems and operations requires richer planning processes than for traditional transit, and this is reflected in the proposed classification. The main scope of this step is to provide a common framework to help relating and comparing existing and new contributions to the field of semi-flexible systems. As third and final step, we take advantage of the common framework offered by the classification to report a comprehensive and comparative literature review of the field of semi-flexible systems, including methodological contributions as well as a number of particularly significant practical experiences.

As far as we know, no previous work addressed the above issues. The contributions of the paper therefore are the proposal of a general unifying model for semi-flexible transit system, a classification of the planning issues for these systems, and a comprehensive review of the corresponding literature and relevant practice.

The paper is organized as follows. We review the main features of traditional transit, DAR, and semi-flexible systems in Section 2. Section 3 is dedicated to, first, describing the main application classes of semi-flexible systems, second, recalling the DAS organization and characteristics and, third, showing how DAS generalizes semi-flexible systems, and thus offers an appropriate general representation for the field, and exploring the advantaged of the more complex scheduling mechanism offered by DAS. In Section 4, we discuss planning issues for semi-flexible systems and propose a classification of the associated decisions, contrasting with the cases of traditional transit and DAR. This classification is used in Section 5 to review the literature on semi-flexible systems. We conclude in Section 6.

2 Traditional transit, DAR and semi-flexible systems

Traditional transit services are particularly suited to handle situations where the demand for transportation is *strong*, i.e., when there is a consistently high demand over the territory and for the time period considered. The high degree of resource sharing by a large number of passengers makes it then possible to provide efficiently and economically high quality, i.e., frequent, services operating generally high-capacity vehicles over fixed routes and schedules. Routes and schedules may and do vary during the day, but, in almost all cases, they are not dynamically adjusted to the fluctuations of demand. In contrast, when the demand for transportation is *weak*, e.g., during out of rush-hour periods or in low-population density zones, operating a good-quality traditional transit system is very costly. The fixed structure of traditional transit services cannot economically and adequately respond to significant variations in demand. In presence of weak demand, itineraries and timetables may perfectly meet the transportation needs of the population at a specific moment, but might be completely inadequate at another time. A traditional frequent service would then be extremely expensive. On the other side, reducing service frequency, as most transit authorities do, makes the service unattractive. For this reasons, mass transit services evolved towards some degree of flexibility.

Demand Responsive Systems are a family of mass transportation services which, as the name suggests, are *responsive* to the actual demand for transportation in a specific time period. Such systems evolve towards a personalization of the transportation services, as itineraries, schedules, and stop locations may vary over time and are determined for each departure according to the particular user requests. Demand Responsive Systems were introduced under the name *Dial-a-Ride* (DAR) as door-to-door services for users with particular needs or reduced mobility, such as handicapped and elderly people (Cordeau and Laporte 2003; Wilson et al. 1971). The *flexibility* of DAR systems allows to respond to varying individual requests for transportation and provides the means to offer personalized services, while still maintaining a certain degree of resource sharing. This has lead certain transportation or city authorities to extend DAR services to more general transportation settings.

The extension of DAR to general public displays a number of drawbacks, however, some of which follow from the extreme flexibility inherent within the system definition. Consequently, practitioners started to experiment and implement new transportation paradigms with the scope of combining regular (traditional) transit systems with pure on demand services such as DAR. Some of the researches aimed at efficiently coordinate traditional transit and DAR, see for example Häll et al. (2009); Häll (2006); Hickman and Blume (2000, 2001). Such systems, sometimes called *Ingrated*-DAR (IDAR), allow user to transfer from a DAR vehicle to traditional transit and, possibly, again to a DAR vehicle. The main drawback of IDAR system is that the DAR component is managed at operation time (including transfers) and, consequently, inherit most of the DAR drawbacks.

A further steps toward integration has been done by combining the different nature of traditional transit and DAR in the *same* transit system. Such systems are commonly called *semi-flexible* systems. A first study reporting practical experiences of semi-flexible systems undertaken in North America and testifies the importance of semi-flexible system and their potentially very large impact in terms of cost reductions and quality of service in weak-demand scenarios can be found in Koffman (2004). A more up-to-date and complete review con be found in Potts et al. (2010). The fundamental idea behind semi-flexible systems is that the regularity of the service is by itself a valuable property of a transportation systems because, for example, helps users in planning their trips, facilitates integration with other transportation modes, and make it possible to access the service without reservation. At the same time, such systems try to inject flexibility by considering additional times that can be used to possibly deviate from a basic path to operate service in a demand responsive framework.

It should be mentioned that sometimes semi-flexible systems are called *flexible* systems. We preferred *semi-flexible* to *flexible* because the terms underlines the compromise between a rigid structure, typical of traditional transit, and the flexibility of DAR system. More details on semi-flexible systems and DAS are given in the next Section.

3 Semi-flexible systems and DAS

The main scope of the present section is to describe several types of semi-flexible systems, introduce DAS, show that it can be used as a unifying model for the whole class of semi-flexible systems, and the discuss disadvantages of this.

We first describe, in Section 3.1, several types of semi-flexible systems that have been implemented and actually operated as they are reported in Koffman (2004) and Potts et al. (2010). We then introduce the details of DAS in Section 3.2. In Section 3.3, we see how DAS can model each of the semi-flexible systems described and comment about the more general scheduling model provided by DAS. In the same section we also discuss the advantages of adopting DAS as unifying model.

3.1 Semi-Flexible Systems

In the work by Koffman (2004), an inquiry is performed to explore the state-of-the practice of semi-flexible systems undertaken in North America. Following the same lines of that work, Potts et al. (2010) performed a new and more extended inquiry, obtaining a considerable amount of new information. These two works together, form the most comprehensive and updated review of practical experience undertake in North America we are aware of. The authors also report a number of very interesting, statistically based, considerations about relations among type of services, region to serve, type of demand, operating costs, needed technologies, etc. Readers are referred to the original works for the details of the inquiries.

Both works report the same classification of semi-flexible systems according to the particular structure of the systems. In particular, the classification is as follows (in brackets, the percentage of transit agencies adopting a given type of service as reported in Potts et al. (2010)): Route Deviation (51.6%), Point Deviation (19.5%), Demand-Responsive Connector (30.5%), Requests Stops (30.9%), Flexible Route Segments (19.5%), Zone route (32.9%). These architectures are schematically sketched in Figure 1.

Summarizing, they could be described as follows:

- **Route Deviation**: Vehicles operate on a regular schedule along a well defined path and deviate to serve optional requests within a zone around the path. In some particular variant of the service, there are no, or only few, scheduled stops on the main path and users can issue what is commonly called a flag-request, i.e., the users can ask for service in any point of the main path, or at marked stops, by waving their hand when the vehicle approaches. This implies, from the operational

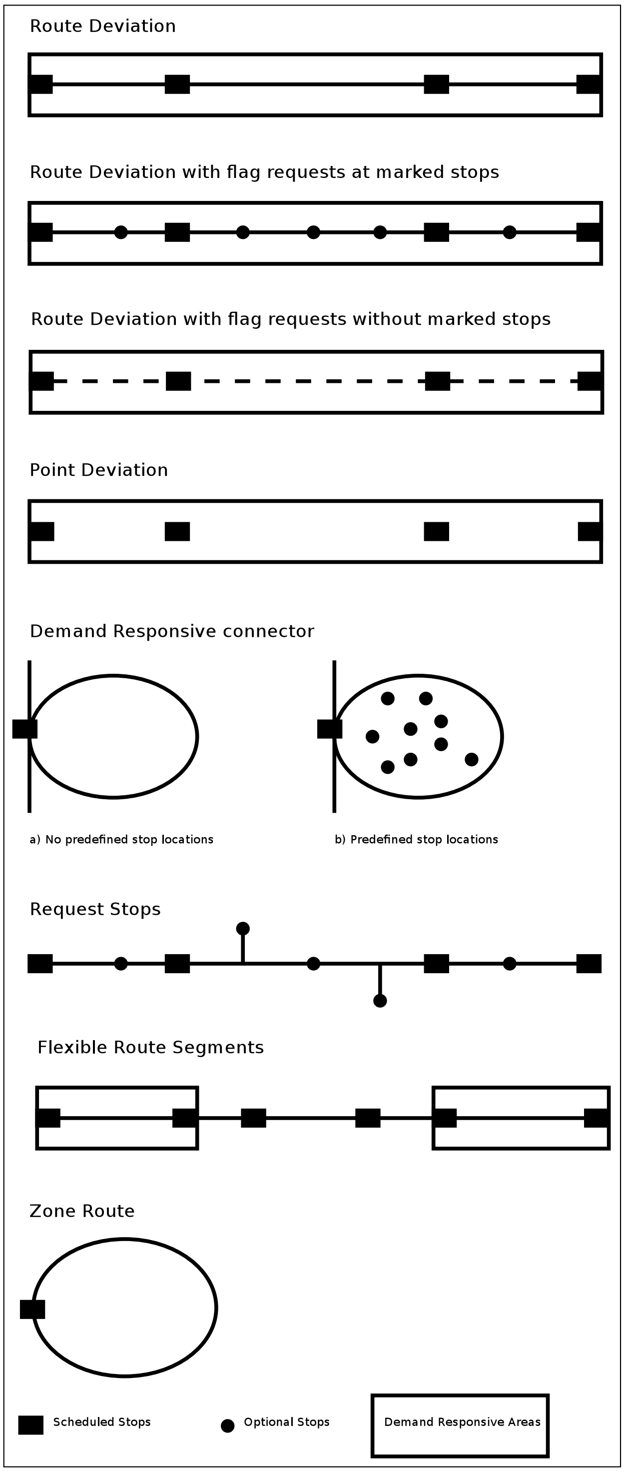


Figure 1: Types of semi-flexible systems. Picture inspired by Koffman (2004)

policy point of view, that vehicles, when deviate from the main path, are obliged to go back to the main path in the exact point where they left to accommodate the deviation. If there are marked stops on the path, the vehicle is allowed to go back to the path anywhere between the point where it left and the next marked stop. The deviations are usually constraint to a certain maximum distance from the main path.

- **Point Deviation**: The main difference with the previous architecture is that the vehicle is not constrained to follow any predefined path. Only a few stops are scheduled (the sequence of the scheduled stops is predefined and induced by the schedule istelf), while the rest of the service is demand responsive and the vehicle serves optional requests within a given zone. Given that the path between two scheduled stops might change at every departure, flag-requests are no more possible. The range of deviations and the service area are usually defined according to geographical or road network features, or the need to serve a given neighborhood.

- **Demand Responsive Connector**: Vehicles operate in a demand-responsive mode within a given zone. However, at least one scheduled transfer point is present that connects with a fixed route network. The demand-responsive service area is defined similarly to the previous case.

- **Requests Stops**: Vehicles operate in conventional fixed-route, fixed-schedule mode, but also serve a limited number of stops near the route, if requests are issued. Usually the optional stops are very few and located near the main path. Similarly to Route Deviation, in some variant of the service flag requests are allowed at any point along the main path.

- **Flexible Route Segments**: Vehicles operate in conventional fixed-route fixed-schedule mode, but switch to demand-responsive operation for a limited portion of the route.

- **Zone Routes**: Vehicles operate in demand responsive mode and cover a given zone. Usually, those systems have one or two scheduled stops, corresponding to the departure and the end of the line. Departure and end of the line can be placed at the same location. The difference with Point Deviation is in the number of scheduled stops: Point Deviation usually presents more than two scheduled stops.

The schedule mechanism is quite similar for all semi-flexible types. Scheduled stops present schedules defined by one point-in-time value. From the operational point of view, the vehicle has to leave the stops at that exact time. As a consequence, if vehicle arrives earlier, it must wait. The amount of additional time allocated between two consecutive scheduled stops is called *slack* time and its determination is usually reported being a very critical decision. In fact, short slack times might not give enough time to serve a region, while too long slack times might imply long waiting times at fixed stops.

The above semi-flexible systems differ in terms of degree of operations policies and flexibility offered. In fact, for example, the operation policy allowed in Route Deviation are much more constrained than those allowed in Point Deviation. Also, according to the particular architecture, more or less portion service is dedicated to demand-responsive part of the service. In this sense, the Request Stops system appears to be the closest to fixed-line transit while Zone Route the closest to DAR, the other architectures being located within this range.

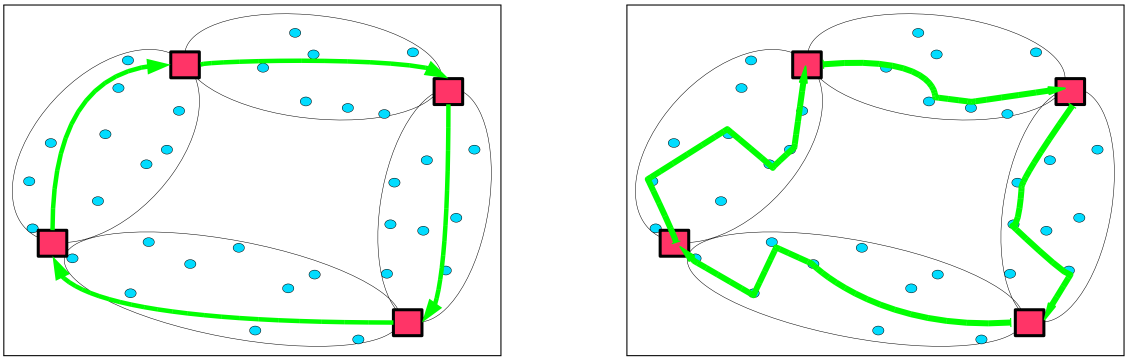
Despite the differences, it is possible to identify some common features: 1) there is a set of the stops displaying some form of fixed, predetermined timetable, similarly to traditional transit; 2) part of the service is flexible and, as in a typical demand responsive context, users can ask for service at optional locations in some predefined service area; 3) in order to allow flexibility, the predetermined timetables need to account for some additional time needed to make deviations possible. Moreover, some of the architectures described above are closely related. For example, Request Stops can be seen as a Route Deviation where only a few deviations are possible. Zone Route can be seen as a particular case of Point Deviation where only one or two scheduled stops are present. Also, Zone Route is similar to a Demand Responsive Collector, in case the scheduled stops are transfer points to traditional transit. Route Deviation can be seen as a Point Deviation with a more restricted operation policy, and so on.

The above interrelations actually suggest that a unifying modeling framework is desirable. In Section 3.3 we will see how DAS, introduced in the next section, can undertake such a role.

3.2 Demand Adaptive System

DAS was introduced in Malucelli et al. (1999) and can be described as follows. In its most general form, a DAS is made up of several lines and is interconnected to 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*. This makes up the traditional part of a DAS. Additional service and flexibility is provided by allowing customers to request service from and to *optional* locations in a given service area. Such locations are served only if a request is issued and it is accepted in the system. The set of optional stops may, or may not, be predefined and, in the latter case, door-to-door service is allowed. We identify users requesting service at an optional stop as *active*, while users moving exclusively between compulsory stops are identified as *passive*.

To serve optional stops, the vehicle must generally deviate from the shortest path joining two successive compulsory stops. The region and consequently the set of optional locations that is possible to visit between two consecutive compulsory stops, is defined



(a) DAS serving compulsory stops only. Passive users only

(b) DAS serving compulsory and optional stops. Passive and active users

Figure 2: A Single-Line DAS

in advance and it is denoted *segment*. Figure 2a depicts the basic DAS service of the compulsory stops, while Figure 2b illustrates the same DAS line when user requests for optional stops are present. The set of compulsory stops together with the segments definition, is denoted *Topological Design* of a DAS line. The collection of time windows corresponding to the compulsory stops, including the start and end of the line, is denoted *Master Schedule* of the DAS line. Notice that a DAS line does not need to be circular, i.e. the last compulsory stop does not need to be physically located in the same location of the first one.

The mixture of compulsory stops and time windows constitutes the "backbone" of a DAS line and accomplishes important functions: it guarantees a certain regularity of the service and at the same time allows user to access the service without in-advance reservation. Also, transfers among DAS lines and among these and regular transit lines take place at compulsory stops. Time windows play an important role in this context because they establish time relations among different DAS and traditional lines which share some compulsory stops. On the other hand, time windows influence the amount of flexibility available for demand-responsive service. In any case however, the detours associated to optional locations must be consistent with the time windows at compulsory stops.

To this regard, a specification must be done regarding the scheduling mechanism of DAS. A time window defines the earliest and the latest departure at a given compulsory stop. If vehicles arrive within the time window, they are allowed to leave at any time. Vehicles are never allowed to leave after the latest departure time. However, if vehicle arrives before the earliest departure time, it must wait.

As a final comment, observe that definition of DAS does not rely on specific assumptions about the service area, such as its geographical features and shape, or on the road network configuration. Similarly, except for some mild assumptions on the demand configuration (some locations are supposed to be more attractive than others) no particular hypothesis is done on the distribution of the demand over space and time. Finally, and perhaps most importantly, no explicit assumption is done on the operation policy adopted. In the next section, we will see that the above points make DAS a very general and powerful modeling tool for semi-flexible transit.

3.3 DAS: a general model for Semi-Flexible systems

The main correspondences between semi-flexible systems and DAS elements are schematically depicted in Table 1. Generally speaking, for what concerns topological aspects,

	Semi-flexible	DAS
Topology	Scheduled Stops	Compulsory Stops
	Demand responsive areas	Segments
Schedules	Departure times at scheduled stops	Time windows on compulsory stops

Table 1: Correspondence between semi-flexible systems and DAS elements

scheduled stops in semi-flexible systems are represented by compulsory stops in DAS, while zones served in demand responsive mode for semi-flexible systems can be represented by segments in DAS. Regarding the schedules, which, for semi-flexible systems, are defined by a single point-in-time value for each scheduled stop, in DAS become a more generalized scheduling mechanism represented by time windows, whose operational rules has been specified in the previous section. Notice however that, by fixing time windows widths to zero for all the compulsory stops in DAS, the meta-schedule reduces to the usual schedule of semi-flexible systems.

Given that the definition of DAS does not assume any particular geometric shape of the service area, or specific policy at operation time, to detail the the correspondences summarized in Table 1 for each of the semi-flexible system, is relatively easy:

- Route Deviation. In addition to the correspondences in Table 1, an operation policy must be specified obliging the vehicle to follow the main path. In case flag requests are allowed, the operational policy must oblige the vehicle to return on the path in the exact point where the path was left, otherwise a less strict operational policy can be implemented where the vehicle can return on the basic path wherever between the point of deviation and the next marked stop.

- Point Deviation. Correspondences in Table 1 are sufficient. Differently from the previous case, no specific operational policy has to be assumed.

- Demand Responsive Connector. Transfer points correspond to compulsory stops. Time windows assure that transfer times are as scheduled. Segments correspond to the demand responsive areas. The case of one single transfer point can be represented by a DAS line with only one segments, the two compulsory stop representing the same physical transfer point. Given that DAS allows both situations where the set of optional locations is made up of every potential physical address in the service area, or a predefined subset of stops, the two variant of Demand Responsive Connector depicted in Picture 1 are equally modeled.

- Requests Stop. This case is very similar to Route Deviation, but more constraint. In this case segments might contain only one optional location.

- Flexible Route Segment. Stops of the traditional part of the line correspond to compulsory stops. The fact that no optional stops are served in this part of the service, is represented by empty segments between compulsory stop. Demand-responsive portions of the service corresponds to non-empty segments.

- Zone Routes. Given the similarities with the Demand Responsive Connector, DAS will model this case in the same way.

The master schedule mechanism of DAS presents several significant advantages for the way flexibility distribution is handled, with respect to the single point-in-time schedule usually adopted in semi-flexible systems. The operation policy of semi-flexible systems, in fact, is such that, whenever the vehicle arrives at a compulsory stop earlier than the scheduled time, it must wait. This, in practice, prevents the possibility to use for one segment a portion of time allocated for another segment, hence builds a sort of flexibility barrier between consecutive pairs of segments. The meta-schedule can drastically reduce this effect. In fact, similarly to the usual slack time mechanism, when a vehicle arrives before the earliest departure time, the vehicle must wait. On the other hand, whenever the vehicle arrives at the compulsory stop within its time window, the vehicle is allowed to leave at any time. By this mechanism, flexibility can commute from one segment to the next according to the actual requests.

We conclude the present section by recalling the relevant advantages of adopting a general model such as DAS. Solution frameworks developed to address planning issues arising in DAS are generally suitable to be applied to the whole class of semi-flexible systems. Such solution frameworks can be then tailored to tackle and take advantage of the specific features of a given actual application, such as the specific road network, or the operation policy. In such a context, it is also easier to modularize the solution process and relate and merge together different line of research.

4 Issues in DAS planning and evaluation

The problem faced by transit agencies to offer good level of transit services in an efficient and, at the same time, attractive way is extremely challenging because of the complexity and number of interrelated decisions. DAS makes no exception from this point of view. On the contrary, because it displays features of different transit systems types, i.e. traditional transit on the one hand, and demand responsive systems on the other, it sums up the complexity of the problems, pose new challenges, and requires new development. It is common practice to decompose very difficult planning problems in several hierarchically related planning decisions levels and classify decisions according to the well known distinction in *strategic*, *tactical*, *operations* levels. This process is usually followed by an *evaluation* phase where the behavior and the performance of the system are analyzed under different conditions and planning policies. It is important to notice that such classifications are in many case not rigid because some decisions are on the border between decision levels and the final attribution to one or the other is sometimes arbitrary. Despite this, such classifications are important and useful because they contribute to create a common framework and standards where different lines of research can more easily compare and contribute one to the other.

A classification of planning decision for DAS is described in Section 4.1. In Section 4.2 we briefly examine the specific requirements of the evaluation phase for DAS.

4.1 Planning DAS

It is beyond the scope of the present paper to give, for traditional transit and DAR, a detailed definition of the planning decision problem and their attribution to planning levels. We will only summarize and give references for details. Table 2 schematically depicts the situation. For traditional transit, (see, for e.g. Desaulniers and Hickman 2007; Guihaire and Hao 2008), the design of the system in terms of line routes, size of the fleet and approximated frequencies are determined during the strategic planning phase. At this level, also approximate assignment of passengers to routes is done as a mean to obtain a first evaluation of the solution quality. Adjustments of frequencies to demand variations, timetables and vehicle schedules are determined in the tactical planning phase. Crew schedules are built during operational planning. Comparatively, purely demand-responsive systems need little strategic and tactical design mainly to define service areas and the composition of the fleet (see, for e.g. Diana et al. 2006; Quadrifoglio et al. 2008b). The most important planning process for DAR is at the operational level, however, when routes and schedules are determined little time before actual operations and are possibly dynamically modified once service has begun.

The case of DAS is different. Its twofold nature implies a more complex planning

Decision Level	Traditional Transit	DAS	DAR
Strategic	Definition of the service area	Definition of the service area	Definition of the service area
	Set of line routes with transfer points	A set of subregions that should be covered each by one line and potential compulsory stops	-
	Frequencies and Fleet size	Frequencies and Fleet size	Fleet size
Tactical	-	For each line additional compulsory stops	-
	-	Sequencing of compulsory stops and segment definition	-
	Definition of timetables	Master Schedule	-
	Adjustment of frequencies and timetable to demand fluctuation	Adjustment of frequency and Master Schedule to demand fluctuation	-
Operations	-	For each departure of the a line, the actual bus itinerary	For each departure of the a line, the actual bus itinerary
	Vehicle Scheduling	Vehicle scheduling	Vehicle scheduling
	Driver Scheduling	Driver scheduling	Driver scheduling
	-	Real time modification of the itineraries	Real-time modification of the itineraries

Table 2: Planning Decision levels for traditional transit, DAR and DAS

process and consequently requires revisiting the decision assignments to levels of the well known planning hierarchy. A possible attribution of decisions to planning decisions for DAS goes according the following scheme. The identification of the region to be served by the DAS systems and its partition into several subregions, each region corresponding to a single DAS line, are determined at strategic level. At this level candidate transfer points, represented as a subset of potential compulsory stops, might be defined. Similarly to traditional transit, also a certain quality standard, mainly reflected into size of the fleet and service frequencies for each line, is established. Input data at this decision level are the road network of the potential zone to be covered and o/d demand matrices built on suitable geographical clusters (demand points) and defined over suitable time intervals. Objective functions should minimize the travel times or, as frequently done in traditional transit, some generalized cost function where additional factors are taken in account.

The tactical level defines the *backbone* of the DAS lines, i.e. the fixed part if it. For each line, the location of the additional compulsory stops, their sequence, the partition of the service area into segments, and the time windows are established. Observe that the investigation of the relationship between dimension of the service area and masters schedule is specific to this planning phase. Beyond the output of the strategic level (mainly, the zone to be covered by a single DAS line, possibly some transfer points and frequencies), the inputs include o/d demand matrices (possible more refined than for strategic level), road network, and some quality standard regarding riding times, as well as some quality requirements such as user waiting times, probability to serve future requests etc.

At the operations planning level, once the design of the lines is given by the tactical level, and the requests become known, the actual schedules are built. The new schedules must be compatible with the master-schedule and are recomputed at every departure of the line. The objectives are typically to serve as many requests as possible while trying reduce the operating costs. At operations level crew schedule also should be determined.

4.2 Evaluating DAS

In public transit, the evaluation process is usually undertaken either for a particular transit line or group of lines as stand-alone systems, or for the overall transportation system of a given city or region. The evaluation of transit lines as stand-alone systems aims to tune operating parameters impacting the system performance, to draw contingency plans, and to derive performance measures under various scenarios for cost-benefit analyzes and integration into system-wide evaluation methods. Given that the number and the complexity of the planning decision is higher for DAS than for traditional transit or DAR, higher is the number of evaluation steps that must be done. This generally results in a more complex evaluation process. For example, given that the system-design phase is almost absent in DAR, there is no need to evaluate that step, while for DAS is necessary. Similarly, given that operations of traditional transit of much more simple than for DAS, the relative evaluation procedures will considerably easier.

Transit lines are also part of comprehensive models of the transportation system of a city or region. These models integrate demand and mode-choice modeling, as well as a representation of the transportation supply of the region, i.e., the multimodal transportation infrastructure and services. The latter generally includes private (e.g., automobile, bicycles, and pedestrian) and public (e.g., bus lines, light rail, etc.) transportation means, as well as, sometimes, an approximation of the freight-vehicle flows. The assignment of demand to the transportation network supply, according to the behavior of the various classes of users considered, provides the means to simulate the behavior and performance of the transportation system. Several methods and software instruments are available to perform such studies. So-called static methods simulate the transportation system for an average demand, often during peak-hour periods. Traditional transit lines are represented through their fixed lines and headways, and particular assignment algorithms are used to compute the passenger itineraries using these lines (and private transportation means, eventually). DAR lines are not usually represented, due to the non-regularity of their lines and operations. Similar issues are also partially affecting the representation of DAS lines. Yet, due to its regular service component, a DAS line might be specified by introducing the basic route and compulsory stops, together with average travel times derived from a stand-alone evaluation of the line. Methods based on Dynamic Traffic Assignment principles are increasingly used to analyze the time-dependent behavior of transportation systems. Traditional transit lines start to be represented in such systems, vehicles being followed according to their planned schedule (headway and average speed). A representation of DAS lines similar to the one described for static models could also be used in this case. Yet, given the dynamic nature of such simulations, a challenging perspective is available for semi-flexible systems. lines.

5 Literature review: practice and methods

The literature on DAS and semi-flexible systems is very heterogeneous not only for the variety of investigated topics, but also for the variety approaches and methodologies. Generally speaking, we can identify two main types of contributions: one describing practical experiences, with no or little methodological contribution, the other developing methodologies to address several aspects of the planning process.

A summary of the works reported in this review is schematically depicted in Table 3, where references are classified by decision level and approach. Given the high number of works on practical application, in this review we presented what we considered the most significant according to the following criteria: either they are pioneering application, or have a clear vision of the planning process and need for tools, or make steps toward building planning tools or served as base for later methodological development. This family of works includes Ambrosino et al. (2003); Bruun and Marx (2006); Durvasula et al. (1998, 1999); Farwell (1998); Farwell and Marx (1996); Koffman (2004); Potts et al. (2010); Rosenbloom (1996); Welch et al. (1991). We avoided works merely describing specific application, even if successful. Rather, we preferred to cite works reviewing these kind of experiences, such as Potts et al. (2010).

We tried to be as exhaustive as possible about the second type of works, i.e. methodological contribution to semi-flexible and DAS planning. Strategic planning and evaluation are among the less studied topics in the filed of semi-flexible transit. The reason for this is perhaps related to the fact that in order to carry out such planning activities some approximation is needed about lower planning level phases and consequent system behavior. It is then natural that studies on tactical and operations planning levels developed more. In fact, much higher is the research effort that has been put in tactical planning issues. Works focusing on these aspects can be roughly classified in two main families: one is based on continuous approximation, the other on combinatorial optimization. The first set of works, which includes Alshalalfah (2009); Fu (2002); Quadrifoglio et al. (2006); Zhao and Dessouky (2008) usually consider a simplified operating environment and continuous approximation is used to find some close form analytic expression describing relation among several systems parameters and design feature. The other set of works, which includes Crainic et al. (2010); Errico (2008); Errico et al. (2011a,b); Smith et al. (2003), usually do not make particularly restrictive assumption on the operating framework nor on the type of input (demand distributions or geometric requirements of the service area), but usually require a considerable computing effort. The last group of works focuses on operational aspects of semi-flexible systems, in particular, vehicle routing and scheduling aspects. This set of works includes Crainic et al. (2005); Malucelli et al. (2001, 1999); Quadrifoglio et al. (2008a, 2007).

Experiences			Ambrosino et al. (2003); Bruun and Marx (2006); Durvasula et al. (1998, 1999); Farwell (1998); Farwell and Marx (1996); Koffman (2004); Potts et al. (2010); Rosenbloom (1996); Welch et al. (1991)
Methods	Strategic & Evaluation		Crainic et al. (2009); Daganzo (1984a); Errico (2007); Errico et al. (2006); Pratelli and Shoen (2001)
	Tactical	Continuos Appr.	Alshalalfah (2009); Fu (2002); Quadrifoglio et al. (2006); Zhao and Dessouky (2008)
		Combinatorial Opt.	Crainic et al. (2010); Errico (2008); Errico et al. (2011a,b); Smith et al. (2003)
	Operations		Crainic et al. (2005); Malucelli et al. (2001, 1999); Quadrifoglio et al. (2008a, 2007)

Table 3: Decision levels for DAS and references

The rest of the section is organized as follows: practical experiences are reviewed in Section 5.1, strategic planning and evaluation in Section 5.2, tactical planning in Section 5.3, operations planning in Section 5.4. Some concluding remarks are reported in Section 5.5.

5.1 Significant practical experiences

Along time, many transportation agencies have been committing to consultants or themselves actuated, inquiries, overview of already existent systems, and preliminary studies on feasibility of the semi-flexible system, profitability, technological requirements. Consequently, there is a very vast literature reporting practical implementations of semi-flexible In this section we report, according to the forementioned criteria, the most representative works. In the presentation of the selected material, we chose to stress on the planning requirements the case studies underlined, instead of reporting the details of the application. For details, readers are referred to the original works.

To the best of our knowledge, the work by Welch et al. (1991) is the first documented attempt of reducing the number of stops in traditional transit lines in those cases where long and sinuous route designs implied deterioration of the service quality. The work describes a practical problem faced by the San Diego Metropolitan Transit Development Board. Because of the geographical configuration of the region (canyons and mesas) many commercial activity and neighborhoods developed far form the major road network arterials where bus would normally operate. As a consequence, the Board received and accorded many requests of deviating the bus route to serve some additional areas, called by authors, Out Of Direction (OOD) zones. However, with the time, they realized that this was implying a deterioration of the quality of the service for *through* passengers, i.e., passengers whose origin or destination was not within the limit of the OOD zone. For this reasons, they considered to eliminate some of those OOD. Welch et al. (1991) developed a three step methodology for the evaluation of OOD zones which can be used as a tool to decide what OOD might be worth to eliminate. The most important step was the definition of a index, called OOD Impact Index, which gives an estimation of the relative inconvenience of through passengers with respect to the number of OOD passenger, i.e. passenger whose alight or boarding stops were located within the limit of the considered OOD. The OOD Impact Index was defined as follows: (through ridership x OOD travel time)/OOD ridership.

Even though the work of Welch et al. (1991) does not specifically focus on semi-flexible transit, it is relevant to semi-flexible systems. In fact, the work by Durvasula et al. (1998) uses the idea in Welch et al. (1991) as a basis to construct a semi-flexible line. The purpose of their study was to demonstrate the technical feasibility of operating a route deviation bus service with the help of "off-the-shelf" GIS software products. The Peninsula Transportation District Commission provided the case study setting. Out of 13 transit fixed lines they select two lines as a candidate to be converted in Route Deviation lines, chosen for the low ridership in the majority of the stops. Following Welch et al. (1991), the authors identify several "out of direction" zones (OOD) and to be considered for demand-responsive service. In this way they obtained basic paths for the two Route Deviation lines. At this point, they faced the problem of defining the service zone and the slack times. As for the service zone, they considered corridors with three candidate maximum deviation from the main route: 400, 800 and 1200 meters. Furthermore, they partitioned the overall service area in segments. Regarding the slack times, they considered that the total amount of slack available for the whole line would be the difference between a fixed and predefined maximum line time ride (which they arbitrarily fixed to 60 minutes) and the time needed to travel along the shortest path among the compulsory stops. Given the total available slack time, they proposed two alternative ways to distribute this among segments: 1) proportionally to the shortest path joining consecutive compulsory stops; 2) proportionally to the number of candidate optional stops (deduced as the number of Americans with Disabilities Acr users within the segment). They implemented a decision supporting tool by using GIS features together with software packages such as ArcView, Avenue (programming language to interface ArcView) and Network Analyst (a module able to compute shortest and Hamiltonian paths on the street network), etc. Such decision supporting tool simply checks the feasibility of the insertion of a given request in a current vehicle schedule and accommodate it if possible. Using such tool they simulate the behavior of the two Route Deviation and perform a sensitivity analysis by testing the different design settings of the width of the serve area and slack time distribution method.

The series of papers by Farwell and Marx (1996), Farwell (1998) and Bruun and Marx (2006) provide another well documented and very interesting example of planning challenges involved in the actual implementation of three route deviation services in Northern Virginia (the company named the service "OmniLink"). The papers describe different implementation phases of the service for a total time range of over ten years. The service was conceived as an economically affordable way of fulfill the Americans with Disabilities Act and, at the same, time provide service to a wider range of customers. The fist implementation of the services had the structure of a Route Deviation system with a service corridor of approximatively 2.4km wide where the final layout was the result of a quite involved planning process which had first to identify and analyze the demand structure and then, in a dial and try fashion, with the help of the feedback from the population they refined it. The author report the determination of the running times and dispatching procedure as the most elusive aspects of the service planning: "Without real information on when and where deviation requests or ridership demand would be manifest, determining where to build in slack, and how much, became a critical operational puzzle". This first implementation of the service was characterized mostly by the absence of ITS technologies, while in the second implementation such functionality became fully operative and authors report great practical advantages both for the ease of communication between dispatch center and drivers and for the possibility to correct and adjust plans based on real time information. The authors make positive comments on the applicability of such systems and show how ridership almost continuously increased over the years. Finally the authors comment on the capability of such systems of probing new demand. They say: “it is interesting to note that productivity actually increased when service hours were expanded – and indication of how much latent demand existed”.

In the same period, the interest for demand responsive and semi-flexible systems was growing fast also in Europe where, starting from year 1996, the European Commission funded the project SAMPO (Systems for the Advanced Management of Public Transport Operations), and its follow-up SAMPLUS in 2000. The main scope of the projects was to “assess the potential and the effectiveness of the introduction of telematic technologies in the provision of Demand Responsive Transport Services”. SAMPO considered five potential sites for the implementation of several demand responsive services. In three sites were considered DAS systems: Florence (Italy), Hasselt (Belgium), and Kilkenny region (Ireland). The nature of the project was very general and took into account multiple aspects of the sites and included the analysis of the user need, definition and classification of different demand responsive systems (among which a route deviation service), the definition of a formal platform for the use of advanced technology, a verification and demonstration phase, etc. The whole set of *deliverables* each describing the single steps of the project can be found at the website indicated in the references SAMPLUS (2000); SAMPO (1996). A good summary of the both projects can be found in Ambrosino et al. (2003). Despite such a huge amount of research and publications, the projects itself did not contribute to the methodological challenges the implementation of demand responsive and semi-flexible systems requires. In fact, while the importance of optimization technique in the operations and real time dispatch of the service was generally recognized, the details of those topic were not treated and reported as “managed by proprietary software packages”. Furthermore, the study does not generally stress on the different nature of semi-flexible systems, with respect to purely demand responsive systems and the specific design requirements the former have with respect to the latter.

During the last decade, an increasing number of transit agencies considered the possibility of implementing some form of DAS and this resulted in new technical reports and reviews on the state of the practice. Most of them follow the main lines of the above described works and their interest is limited to the fact that they report an increasing interest in those systems (see for example Scott (2010)). On the other hand, we would like to underline the important contribution to the field of semi-flexible systems given in Koffman (2004) and, more recently, in Potts et al. (2010) which, as shown in Section 3.1, classified different types of semi-flexible systems and put the basis for a wider discussion on semi-flexible systems.

As a final consideration of this section, we notice that in almost all the practical implementations of semi-flexible systems there is no or very limited use of optimization techniques in the planning activities. With the result that most decisions are taken according to some common sense criteria. This fact is underlined in both Potts et al. (2010) and Scott (2010). The former pushes the concept to the point of affirming that those techniques are not useful in most cases. This an evident sign of the fact that a good portion of practitioners still not have clear enough the potential benefit of the optimization tools and efforts should be done in that direction.

5.2 Strategic Planning and Evaluation

An important question arising at strategic level is to decide what type of transit system is better suited to serve a given geographical area, given some social/economical conditions and demand features. The work by Daganzo (1984a) addresses such a problem. The author considers three possible transit alternatives: fixed route traditional transit, door-to-door dial a ride and Checkpoint dial-a-ride. The main difference between door-to-door DAR and checkpoint DAR is that possible origin/destination locations are finite and predefined (but arbitrarily dense) in the latter, continuously distributed in the former. The author's scope is to find the range of demand volume in which Checkpoint DAR becomes the most convenient option. Based on a very simplified transit model, the author analytically derives the optimal design of fixed route service and Checkpoint DAR (given by the optimal number of vehicles and the per unit area number of stops). The case of Checkpoint is more involved than the fixed route as a routing strategy has to be supposed first. To the scope, the author first defines a tour among all the checkpoints. Then, once demand is known, checkpoints with no requests are skipped and the vehicle serves the next in the sequence. The initial sequence is obtained as an approximate solution of the Traveling Salesman Problem. Finally, Door-to-Door DAR is seen as a limit case of Checkpoint DAR where the distance among checkpoints tends to zero.. The author compare a generalized cost function for the three cases under different demand volume levels and finds that there is only a small range of demand volume which makes the checkpoint DAR more convenient than the others. Although the checkpoint system is quite different from DASs, as it considers no predefined schedule and a specific operational policy, it is interesting and related to DASs because it represent a first step towards the introduction of fixed structure in a pure demand responsive system.

Similar strategic issues are also addressed in Errico et al. (2006) and Errico (2007). The works consider the problem of a comparative analysis of the performance of pure DAR and DAS under different demand levels and types. The papers put the basis for a simulation framework and preliminary analysis are performed. Advantages of DAS over DAR in presence of *polarized* demand (i.e. demand showing a few attractive *poles*) are conjectured.

The work by Pratelli and Shoen (2001) also addresses the problem of defining the area that should be served by a semi-flexible line. The semi-flexible line analyzed can be seen as a Request Stops, where a sequence of compulsory stops have been predefined.

Between each pair of compulsory stops they consider a single potential optional location. For example, if there are n compulsory stops, they consider $n-1$ potential optional stops. The problem analyzed by the authors is that of choosing, among the set of optional locations, a subset that will be included in the design of the line. If a location is chosen, users are allowed to make a request for it, otherwise they have to walk to the closest compulsory stop in order to access the service. The objective function takes into account three components: total increment of travel time, walking times, increment of waiting time at compulsory stops. The authors propose a mathematical programming formulation and discuss the possible application to a real case in Imola (Italy).

Regarding the evaluation process, the only work we are aware of facing such a topic is by Crainic et al. (2009). In this work the authors briefly review the evaluation precess for traditional transit and identify two main uses: as a stand-alone system when planning policies and design parameters need to be tuned or in a more global context (e.g. at city or regional transit planning level) where the demand and mode-choice or dynamic traffic assignment models are employed in order to simulate effectiveness and impact of specific transit mode on the overall transportation system. The authors focus on the stand-alone case, which is considered the first necessary building block, and discuss an evaluation framework addressing the specific requirements of DAS.[1]

5.3 Tactical Planning

Tactical planning is the most studied topic in semi-flexible literature, and, in particular, relations among service area, and amount of slack-time or time window design, received particular attention. Most of the works can be classified in two main categories, according to the solution method: one adopts Continuous Approximation, the other, Combinatorial Optimization. We review the first category in Section 5.3.1, the second in Section 5.3.2. Some general considerations are reported in Section 5.3.3.

5.3.1 Works based on Continuous Approximation

All the woks reviewed in the present section consider a simplified operational framework with the scope of building closed-form analytic relations between the main design parameter of the line. Although the notation used in those works slightly differs, the assumptions are mostly the same. The service area is represented as a rectangle of given width and length, where the length is considerably higher than the width. The service is performed along the horizontal direction (the length) and compulsory stops are located in the middle of the two vertical edges of the rectangle. The demand is modeled as a set

[1] Some of the material at the end of section 4 can be moved here. See also footnote there.

of locations, either pickups or drop-offs, uniformly and continuously distributed on the service area, with a specific, given, per unit density. The vehicle is considered to have constant speed, and move on an infinitely dense grid road network according to linear paths running parallel or orthogonal to the edges of the service area.

The first paper of this kind we are aware of, is by Fu (2002) and addresses the problem of defining the optimal amount of slack time needed to service a single rectangle as described above, while optimizing an objective function made up of three components: the operator cost, the service benefit and user costs. The resulting model is a linear program in one variable with a feasible region bounded by three constraints, which can be trivially solved analytically. The author completed the study by simulating the operations by a tool called SimParatransit (Fu 2001) originally devised for simulation of paratransit operation and adapted (the author does not explain how) to the scope. The simulation was performed in the simplified operational framework considered when building the analytical model, and focused on estimating the effects of slack time changes on idle times, and number of feasible deviations. Even if the model was able to reflect some general tendencies, it substantially failed in capturing the details of the system behavior.

The paper by Quadrifoglio et al. (2006) also considers the same simplified environment, with the addition of considering vehicles with infinite capacity. The scope of the paper is to derive upper and lower bounds of the expected velocity along the main direction (the length) of a vehicle, under the given design parameters and demand density, while serving optional demand. As a first step, using arguments very similar to Fu (2002), and assuming the so called no-backtracking policy introduced in Daganzo (1984b) (the vehicle never moves backward with respect to the main direction), the authors derive a lower bound on the expected longitudinal velocity. As a second step, it is shown that, under certain conditions, the no-backtracking policy is optimal. The authors use this fact to derive the first upper bound on the expected longitudinal velocity by considering subsets of requested points satisfying such conditions and characterized by the fact that the Hamiltonian path trough such subset is certainly shorter than the one on all the requested stops. To compute expected values, the authors need to compute the probability distribution of the number of points belonging to such subsets and, as this calculation is very complex, the authors resort to a continuous approximation. The second upper bound is obtained by considering the total travel time as the summation of the time to travel from each requested stop to its closest neighbor. This is equivalent to considering a relaxation of the Hamiltonian path and thus provides an upper bound of the longitudinal velocity. The authors consider a simulation study where the operational policy adopted is an insertion heuristic algorithm as described in Quadrifoglio et al. (2007) and results are compared with the approximated upper and lower bounds. The authors claim that, with some exception, the approximated values fit the simulated ones sufficiently well. Finally, based on the previous results, the authors give an estimate relation between longitudinal velocity and service *capacity*, defined as number of optional locations the vehicle is able to service in a given time.

Similar aspects, but different methods, are investigated in Zhao and Dessouky (2008) where the system capacity (in the meaning specified above) is treated. The authors analyze the relationship between service cycle time, and the length and width of the service area. They consider the same simplified setting as in Quadrifoglio et al. (2006). Considering the programmed time duration duration of the line, T, and the distribution probability of the actual arrival time T_R, and under the further assumption that $(E(T_R) < T)$, the authors call on queueing-theory results and derive a Wiener-Hopr integral equation represent the delay distribution. From this, and by approximating to the exponential distribution the more complex travel-time distribution, they derive approximated analytical relations among the length and the width of the square and L.. Finally the authors test the correctness of their analytical model by simulating a non back-tracking nearest insertion operational strategy. The authors claim that the experimental results obtained are in line with the analytical approximations derived.

Finally, the works by Alshalalfah and Shalaby (2008, 2009, 2010) basically follow the lines of Fu (2002); Quadrifoglio et al. (2006); Zhao and Dessouky (2008) and only slightly change the perspective as they compare with a service operating on the same territory as a traditional line.

5.3.2 Works based on Combinatorial Optimization

Relations between the dimension of the service area and the amount of slack time allocated to provide service flexibility has been addressed in the work by Smith et al. (2003) which is based, in turn, on the work by Durvasula et al. (1998), where two Route Deviation were studied (see Section 5.1 for more details). In particular, the authors tried to optimally determine the dimension of the service area and the maximum allowable deviation from the basic route. The authors consider three possible maximum deviation values (possibly varying for each segment) and two slack time policies (all the segments have the same policy). The authors build a multi-objective nonlinear choice model where the contrasting objectives are the maximization of the feasible deviations and the minimization of slack time. The problem is solved by a gradient method where the evaluation of the objective function at each iteration is performed by a (non-optimal) GIS based tool, described in Durvasula et al. (1998). Given the very few variables and possible values considered, the authors, using out-of-the-shelf solvers, are able to provide heuristic solutions for specific case at hand. No study on computational efficiency, nor bounds on the solution quality were reported.

Several aspects of the DAS planning activity are treated, with a particular focus to the tactical level, in Errico (2008) and Errico et al. (2011b). In these works the authors introduce and formalize what they call the Single-line DAS Design Problem (SDDP). The SDDP assumes that the territory to be covered by the DAS line has been determined, the travel times between any pair of potential demand point in the territory is known, and that a measure of the transportation demand among the potential is available. For a given time horizon, the SDDP has to select what locations in the service area are more suitable to be compulsory stops, their sequence, the partitioning of the service area in segments, and the determination of the master schedule. To address the SDDP, the authors examine several hierarchical decompositions and focus on two, mainly differing for the sequence of decisions. In both decompositions, a few core problems can be identified: the selection of compulsory stops, a particular sequencing problem called the General Minimum Latency Problem (GMLP) and the problem of the definition of the time windows called the Master Schedule Problem (MSP). After a description of the core problems and the parameters governing their outputs, the authors provide a computational study to simulate the behavior of the system under several design parameter settings and demand volumes. The results prove that the design obtained is sound and that the master schedule mechanism has several advantages over the single points in time schedule. For example, results prove that, for a given probability to serve all requests, the total time required by a vehicle to complete the service, is much higher for small time windows than for large. Moreover, most of the performance measures worsen in the former case.

Except for the selection of the compulsory stops, which is addressed directly in Errico et al. (2011b), the other two core problems are addressed in separate papers. The authors define the set of compulsory stops by simply selecting those locations which are almost surely requested for service. This does not prevent to add any number of compulsory stops possibly indicated by external entities or as a result of the strategic planning. The second core problem, the GMLP, is addressed in Errico (2008) and Errico et al. (2011a). The GMLP is a variant of the Traveling Salesman Problem with a more complex objective function taking into account, beyond the traditional costs related to the routing, a measure of the amount of time the users pass on the vehicle (*latency*). The authors propose multicommodity-flow based formulation and investigate the polyhedral relations of the GMLP and its relations with the TSP polyhedron. A Branch and Cut approach based on Benders decomposition is proposed for the solution of the GMLP. In addition, feasibility cuts of the classical Benders decomposition are substituted by several class of valid inequality from the TSP polyhedron, considerably speeding up computations. Instances up to 70 nodes are optimally solved in the time limit imposed.

The last core problem, the MLP, is addressed in Crainic et al. (2008) and Crainic et al. (2010). The MPL can be defined as follows: given the topological design of a DAS line (i.e. the compulsory stops, their sequence, and the partition of the service area into segments) and the probability for each location (or demand point) to be requested for service, find a set of time windows such that the probability of serving all issued requests is within a given threshold and the total maximum time needed to travel the line is minimized. The authors first address the problem of determining the probability distribution of the arrival time at the last compulsory stop when one single segment is considered. To such a scope, authors propose a method to efficiently sample request scenarios and computing the corresponding travel time values. Consequently, authors consider the general case of a line with more segments and show that, based on the previous results, together with mechanisms involving the time windows limits, convolution can be used to iteratively compute the probability distribution of the arrival times at each compulsory stop and the corresponding time windows.

5.3.3 Some considerations on tactical planning

Even though both families of works presented in the previous sections address what we identified as tactical planning aspects, different approaches and findings probably make for different applications..

Given the number of simplifying hypothesis, the works in Section 5.3.1 appear more as fast and approximated evaluation tools of tactical plans, suitable to be used, for example, at the strategic level, rather than tools to build the design of actual semi-flexible lines where the complexity of a actual application is considered, where the demand is not uniformly and continuously distributed on the service area, the service areas do not have nice geometric shapes, and vehicles do not move along infinitely dense grid road network.

A different perspective is assumed by most of the works in Section 5.3.2 where hypothesis made on the service area, demand distributions, operational policy, etc, are much milder. Such works generally propose application independent methodologies and provide tools to obtain the actual design of a generic semi-flexible system. As a consequence, such works have a wider range a potential applications and seem more suitable to address the complexity of real-life application. On the other hand, such generality of the approach has a cost in terms of computational complexity and, contrary to the set of works in Section 5.3.1, their running times possibly make them less suitable to be used as fast evaluation tools of tactical plans.

5.4 Operations Planning

The DAS has been described and formally introduced for the first time in Malucelli et al. (1999), even though some concepts had been anticipated in Malucelli et al. (1997, 1998). In Malucelli et al. (1999) the authors identify three alternative operational policy:

- DAS1. Requests may be rejected if their acceptance causes infeasibility of the deviation with time windows. If a request is accepted, users must be picked up or drop off exactly where they asked.
- DAS2. Users are always picked up in the requested location but they may be

dropped off in the vicinity of the requested alighting stop if the requested drop off location implies infeasibility.

- DAS3. Users are always served, but they cab be picked up and dropped off at locations in the vicinity of the requested ones, and in this case the service pays a penalty.

Assuming the tactical plan of the DAS line given (i.e. compulsory stops and their sequence, segments and time windows), they propose linear integer formulations for the three policies above. The authors address the case of a single vehicle and provide generalizations to the multi-tour case. The authors discuss in detail about the fact that some model were more suitable to be treated by Lagrangean approaches, and other by Column Generation. They also consider hybrid approaches where heuristics are used in combination with Lagrangean approach and Column Generation. The basic idea is that both the methods provide with a set of "promising" good vehicle tours. The heuristic would then start with one of those tours and iteratively try to swap *segment-paths*, i.e, paths within a segment, from the set of good tours. They finally make some considerations about the online case. Even thought the details of the algorithms are presented, no computational result is reported.

The work by Malucelli et al. (2001) and the related paper Crainic et al. (2005) address the solution of the DAS1 policy, as previously defined, for the single vehicle and single tour case. Authors use some of the algorithmic ideas in Malucelli et al. (1999), such as that good paths obtained by Column Generation or Lagrangean Relaxation can be further improved by heuristic methods. Given a vehicle tour, a general framework where segment-paths an be swapped with candidate others, is discusses in terms of feasibility verification and alternative path selection criteria. The authors first develop a memory-enhanced constructive heuristic inspired by the class of GRASP algorithms where paths are chosen according to a suitable score function and randomization is obtained by perturbing some of the parameters affecting the score function. As an alternative solution approach, a Tabu Search heuristic is also developed, where the neighbor of a current tour is defined as another tour where at least one segment-path has been swapped with a candidate other, and several *move* evaluation procedures are proposed. Finally, the authors consider several hybrid algorithms combining the two heuristic approaches described above. An extensive computational study proved that a Tabu Search variant called PPKS outperformed the others on relatively simple instances, while a hybrid approach called TS_Gdiv was the best algorithm on harder instances. The quality of the solutions was excellent when compared with the best solutions found by a commercial exact solver after very long computing times.

In the work by Quadrifoglio et al. (2007), a quite simple insertion heuristic is proposed to manage the operations of a semi-flexible system considering a dynamic perspective., i.e., requests are issued during the planning horizon. The semi-flexible system considered have characteristics similar to what described in Quadrifoglio et al. (2006) and, in particular, the service area is rectangularly shaped, with a sequence of compulsory stops located along the longitudinal axis. Slack times between compulsory stops are supposed to be known. The procedure tries to sequentially insert requests as they arrive, and as soon as possible. In case the insertion is not feasible for the current service, the algorithm tries to insert it in the next service, until one feasible service is found. The insertion is evaluated according to a cost function accounting for the extra ride time for all passengers, the extra waiting time at the already inserted stops, and the slack time portion used by the insertion. In order to balance the greedy behavior of trying to insert requests as soon as possible, the authors consider two thresholds: 1) maximum slack time used by a single insertion 2) maximum backtrack implied by an insertion. An analysis of the performance of the system is done by considering several rectangular service area with the same area but different edge ratios.

The same problem addressed in the paper by Quadrifoglio et al. (2007), has been studied in Quadrifoglio et al. (2008a), but considering the static case, i.e., all requests are supposed to be known in advance. The authors propose a MIP formulation for the multi-tour case where requests might be postponed to a later occurrences of the line if service would imply infeasibility in the desired occurrence. If no feasible occurrence is found, the problem is considered infeasible. The authors develop several set of *non-valid* inequalities, to the scope of cutting away feasible solutions which are provably not optimal. The main idea behind such inequalities is that the users will always choose to be picked up and dropped off during the same occurrence of the line. They perform some computational tests on small instances (up to 30 nodes) and results show that among the three family of inequalities proposed, one is more effective.

5.5 Concluding remarks

The literature review reported above highlights the existence of a consistent number of contribution to the field of semi-flexible transit systems. However, it also emphasizes a high fragmentation of problems and adopted approaches, making it hard, for example, to compare and relate different research lines. Moreover, there are many aspects of the planning process that still deserve a significant research effort. Generally speaking, the less studied planning aspects are the strategic level and the evaluation phase. But also for tactical planning and operations planning, a significant amount of work is still needed.

Regarding tactical planning, for example, we observe that, even if the problems are widely affected by uncertainty, and uncertainty is somewhat taken into account more or less explicitly in the existing works, no stochastic programming approach has ever been implemented. For what the operations planning is concerned, we observe that only a very limited number of operation policies have been addressed. Moreover, most the operations policies have, at least, a static and a dynamic version, and this multiplies the research needs. A step that, in the authors' opinion, is important and should be done in the next future, is to build, and make available for the community, data, such as service area, road network, demand, etc, related to candidate zones to be served by semi-flexible systems and use these cases as testbed to compare and evaluate different approaches. The whole field of semi-flexible transit would greatly benefit of this.

Finally, regarding the implementations of semi-flexible systems in actual practice, we observe a constantly increasing interest of transit agencies for these kind of transit systems. However, no or very little use of planning tools based on optimization techniques can be observed. In our opinion, steps should be done to help practitioners understanding the advantages of optimization tools for the planning process.

6 Final conclusions and perspectives

The main scope of the present paper was to provide a systematic treatment of the field of semi-flexible systems. After underlining the importance of such a step, which was lacking in literature, we believe we reached our scope by accomplishing three main tasks: 1) We proposed and showed that the Demand Adaptive System is a general and unifying modeling framework for the whole class of semi-flexible transit systems. 2) We provided a classification of the planning decisions for DAS according to the well known distinction in strategic and evaluation, tactical, and operation planning level. 3) We reported a comprehensive review of both practical and methodological works in literature. The main advantage of adopting DAS as a general model for semi-flexible is that solution frameworks developed for DAS can be easily tailored to capture specific requirements of a given application. The classification of planning decisions, together with the comprehensive literature review, helps in relating one to another existing works, and possibly will avoid fragmentation of approaches for future works.

Regarding the current state of the literature on methods and practice of semi-flexible systems, several remarks were reported in Section 5.5 and we do not repeat them here. We only recall that many aspects of the planning activity still deserve significant research efforts, and this is particularly true for the strategic planning and the evaluation phase. We also underlined the importance, from the methodological point of view, to build several public available cases possibly derived from practical experiences, serving as test-beds for different solution methods. From practice point of view, we noticed that no or very little use of planning tools based on optimization techniques can be observed and that it would be very important for the Transportation Science community to to make transit agencies aware of the great potential advantages of optimization tools in the planning process.

Acknowledgments

While working on this project, T.G. Crainic was the NSERC Industrial Research Chair in Logistics Management, ESG UQAM, and Adjunct Professor with the Department of Computer Science and Operations Research, Université de Montréal, and the Department of Economics and Business Administration, Molde University College, Norway, and F. Errico was postdoctoral researcher with the Chair.

Funding for this project was provided by the Natural Sciences and Engineering Council of Canada (NSERC), through its Industrial Research Chair and Discovery Grant programs, by our partners CN, Rona, Alimentation Couche-Tard, the Ministry of Transportation of Québec, and by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT) through its Team Research Project program.

References

Alshalalfah, B. W. (2009). *Planning, Design and Scheduling of Flex-Route Transit Service.* PhD Thesis, Department of civil engineering, Univesity of Toronto.

Alshalalfah, B. W. and A. S. Shalaby (2008). Sensitivity of flex-route transit service to design and schedule-building characteristics. In *TRB 87th Annual Meeting Compendium of Papers DVD* . Transportation Research Board.

Alshalalfah, B. W. and A. S. Shalaby (2009). Feasibility and Performance of Flex-Route Transit Service in Suburban Areas. In *TRB 88th Annual Meeting Compendium of Papers DVD* . Transportation Research Board.

Alshalalfah, B. W. and A. S. Shalaby (2010). Development of Important Relationships for the Planning of Flex-Route Transit Services. In *TRB 89th Annual Meeting Compendium of Papers DVD*. Transportation Research Board.

Ambrosino, G., J. Nelson, and R. M. (Eds.) (2003). *Demand Responsive Transport Services: Towards the Flexible Mobility Agency.* ENEA, Italian National Agency for New Technologies, Energy and the Environment.

Bruun, E. and E. Marx (2006). OmniLink: Case Study of Successful Flex Route-Capable Intelligent Transportation System Implementation. *Transportation Research Record: Journal of the Transportation Research Board 1971*, 91–98.

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

Crainic, T. G., F. Errico, F. Malucelli, and M. Nonato (2008). Designing the master schedule for demand-adaptive transit systems. In *Proceedengs of the 7th International*

Conference on the Practice and Theory of Automated Timetabling, CD-ROM. CIRRELT, Montréal, Canada.

Crainic, T. G., F. Errico, F. Malucelli, and M. Nonato (2009). Proposal for Evaluation of Demand-Adaptive Transit Systems. In *TRB 88th Annual Meeting Compendium of Papers DVD* . Transportation Research Board.

Crainic, T. G., F. Errico, F. Malucelli, and M. Nonato (2010). Designing the master schedule for demand-adaptive transit systems. *Annals of Operations Research*, 1–16. 10.1007/s10479-010-0710-5.

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

Daganzo, C. F. (1984a). Checkpoint dial-a-ride systems. *Transportation Research Part B 18*(4-5), 315–327.

Daganzo, C. F. (1984b). The length of tours in zones of different shapes. *Transportation Research B 18*(2), 135–145.

Desaulniers, G. and M. D. Hickman (2007). Public Transit. In Barnhart, C. and Laporte, G. (Eds.), *Transportation*, Volume 14 of *Handbooks in Operations Research and Management Science*, pp. 69–127. North-Holland.

Diana, M., M. M. Dessouky, and 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.

Durvasula, P., B. Smith, R. Turochy, S. Birch, and M. Demetsky (1998). Peninsula transportation district commission route deviation feasibility study, final report. Technical Report VTRC 99-R11, Transportation Research Council, Virginia, USA.

Durvasula, P., B. Smith, R. Turochy, S. Brich, and M. Demetsky (1999). Investigation of route deviation transit service design. In *78th Annual Meeting of the Transportation Research Board*.

Errico, F. (2007). A Flexible Transit System. INGENIO Project, Final Report. Regional Council of Lombardia, Italy.

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, and M. Nonato (2011a). A Benders Decomposition approach for the Minimum Symmetric Hamiltonian Circuit with Generalized Latency. CIRRELT publication, Centre interuniversitaire de recherche sur les réseaux dentreprise, la logistique et le transport, Université de Montréal.

Errico, F., T. G. Crainic, F. Malucelli, and M. Nonato (2011b). The Design Problem for Single-line Demand Adaptive transit Systems. Forthcoming CIRRELT publication, Centre interuniversitaire de recherche sur les réseaux dentreprise, la logistique et le transport, Université de Montréal.

Errico, F., Malucelli, M. F. Nonato, and R. Wolfler (2006). An experimental comparison of two flexible transit systems. In *Proceedings of the 10th International Conference on Computer-Aided Scheduling of Public Transport, CASPT*.

Farwell, R. G. (1998). Demand-driven transit operations: Flex-route services. *Transportation Quarterly 52*(1), 31–43.

Farwell, R. G. and E. Marx (1996). Planning, implementation, and evaluation of omniride demand-driven transit operations: Feeder and flex-route services. *Transportation Research Board 1557*, 1–9.

Fu, L. (2001). Simulation Model for Evaluating Intelligent Paratransit Systems. *Transportation Research Record 1760*, 93–99.

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

Guihaire, V. and J.-K. Hao (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice 42*(10), 1251 – 1273.

Häll, C., H. Andersson, J. Lundgren, and P. Vrbrand (2009). The integrated dial-a-ride problem. *Public Transport 1*, 39–54. 10.1007/s12469-008-0006-1.

Häll, C. H. (2006). *A framework for evaluation and design of an integrated public transport system.* Master Thesis, Linkopings Universitet, Sweden.

Hickman, M. and K. Blume (2000). A method for scheduling integrated transit service. Working paper (see downloads).

Hickman, M. and K. Blume (2001). Modeling Cost and Passenger Level of Service for Integrated Transit Service. In Voß, S. and Daduna, J.R. (Eds.), *Computer-Aided Scheduling of Public Transport*, Volume 505 of *Lecture Notes in Economics and Mathematical Systems*, pp. 233–251. Springer, Berlin.

Koffman, D. (2004). Operational Experiences with Flexible Transit Services. Volume 53 of *TCRP Synthesis Report*. Transportation Research Board.

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

Malucelli, F., M. Nonato, and S. Pallottino (1997). Modelli di Trasporto Collettivo in Condizioni di Domanda Debole. In *Atti del III Convegno del Progetto Finalizzato Trasporti 2*. Taormina (Italy).

Malucelli, F., M. Nonato, and S. Pallottino (1998). Models and Heuristic Algorithms for Low demand Transit Systems. In *Proceedings of the Triennial Symposium on Transportation Analysis*. TRISTAN III, Puerto Rico.

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

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

Pratelli, A. and F. Shoen (2001). A mathematical programming model for the bus deviation route problem. *Journal of the Operational Research Society 52*, 494–502.

Quadrifoglio, L., M. M. Dessouky, and F. Ordóñez (2008a). Mobility allowance shuttle transit (MAST) services: MIP formulation and strengthening with logic constraints. *European Journal of Operational Research 185*(2), 481 – 494.

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

Quadrifoglio, L., M. M. Dessouky, and K. Palmer (2007). An insertion heuristic for scheduling Mobility Allowance Shuttle Transit (MAST) services. *J. of Scheduling 10*(1), 25–40.

Quadrifoglio, L., R. W. Hall, and 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.

Rosenbloom, S. (1996). Service routes, route deviation, and general public paratransit in urban, suburban, and rural transit systems. Project Report AZ-26-7000. U.S. Department of Transportation, Federal Transit Administration.

SAMPLUS (2000). Systems for Advanced Management of Public Transport Operations Plus, SAMPLUS TR 4032. European Commission. `http://cordis.europa.eu/telematics/tap_transport/research/projects/samplus.html`.

SAMPO (1996). Systems for Advanced Management of Public Transport Operations, SAMPO TR 1046. European Commission. `http://cordis.europa.eu/telematics/tap_transport/research/projects/sampo.html`.

Scott, R. (2010). Demand responsive passenger transport in low-demand situations. Technical Report 425, New Zealand Transport Agency.

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

Welch, W., R. Chisholm, D. Schumacher, and S. R. Mundle (1991). Methodology for evaluating out-of-direction bus route segments. *Transportation Research Record* (1308), 43–50.

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

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