Abstract—Intelligent transportation systems based on automated data collection frameworks are widely used by the major transit companies around the globe. This paper describes the current state of the art on improving both planning and control on public road transportation companies using automatic vehicle location (AVL) data. By surveying this topic, the expectation is to help develop a better understanding of the nature, approaches, challenges, and opportunities with regard to these problems. This paper starts by presenting a brief review on improving the network definition based on historical location-based data. Second, it presents a comprehensive review on AVL-based evaluation techniques of the schedule plan (SP) reliability, discussing the existing metrics. Then, the different dimensions on improving the SP reliability are presented in detail, as well as the works addressing such problem. Finally, the automatic control strategies are also reviewed, along with the research employed over the location-based data. A comprehensive discussion on the techniques employed is provided to encourage those who are starting research on this topic. It is important to highlight that there are still gaps in AVL-based literature, such as the following: 1) long-term travel time prediction; 2) finding optimal slack time; or 3) choosing the best control strategy to apply in each situation in the event of schedule instability. Hence, this paper includes introductory model formulations, reference surveys, formal definitions, and an overview of a promising area, which is of interest to any researcher, regardless of the level of expertise.

Index Terms—Automatic passenger counting (APC), automatic vehicle location (AVL), operational control, operational planning (OP), public transportation (PT) networks.

I. INTRODUCTION

THERE currently are about 800 million vehicles running on our road networks [1]. These vehicular networks are crucial for human mobility, regardless of their type. The excessive number of vehicles running on the world’s biggest urban areas is increasingly discouraging the use of private transportation vehicles in favor of public transportation (PT).

Such large number of vehicles running worldwide is increasing the complexity of the transportation networks, particularly its operations. Therefore, it is becoming harder to maintain the efficiency of private transportation. These inefficiencies lead to road congestion, higher levels of pollution, and time and energy wastes. Moreover, the increasing price of fuel is turning private transportation into a luxury as the cars’ rising operational costs go up to levels that are unaffordable in most family budgets. For instance, the congestion indexes in the USA’s urban road networks in 2011 caused a total 5.5 billion hours in travel delays and 2.9 billion gallons of fuel wasted [2].

In the last decades, public road transportation companies have played a central role in highly populated urban areas, particularly by providing fast short-distance transportation services. Inner-city transportation networks are becoming larger to cover the increasing demand for fast and reliable transportation to and from their industrial, commercial, and residential cores. New challenges await this industry: the mentioned increase in fuel prices and its effects on ticket prices and the improved offer on railway-based services are forcing the public road transportation operations to be more reliable than before so that they can maintain their profitability. In the U.S., the savings on congestion costs caused by PT services increased nearly 131% from 1982 to 2005. However, this increase was 10% between 2005 and 2011 [2].

Consequently, monitoring their operations is now more relevant than ever. More specifically, it is important to determine if and why the bus is failing to meet the schedule. In fact, reliability problems are a major concern for both transit system passengers and operators. A service that is not on time causes an increased waiting time on stops, uncertainty on travel time (TT), bus bunching (BB) events, and, ultimately, a general dissatisfaction with the system. An unreliable service may lead to a major loss of public support as the passengers may leave these networks to find alternative transportation modes, which leads to a critical loss of revenue [3], [4].

More than relying on the traditional operational planning (OP), it is crucial to monitor and analyze what is happening with their fleets and drivers, namely, its location. There are many methods to sense vehicle position. This paper focuses on data collected through Global Positioning System (GPS). After
the Tri-Met experience kicked off in 1991 [5], many companies started to install new computer-aided bus dispatch systems. Examples of regions include New Jersey, Chicago, Minneapolis, and Seattle (USA); Ottawa and Montreal (Canada); Eindhoven and The Hague (The Netherlands) [6]; Cagliari and Genoa (Italy) [7], [8]; Melbourne (Australia) [9]; Toulouse (France) [7]; or even London (U.K.) [10]. Such dispatch systems were based both on the automatic vehicle location (AVL) and on the automatic passenger counting (APC) systems deployed on their fleets. These systems collect the location of buses usually by broadcasting the sensors’ values using an interval of 10–30 s depending on the radio capacity. Typically, AVL systems are based on GPS measurements, whereas the APC systems typically rely on estimation techniques based on door loop counts or weight sensors. These sensors are installed in every vehicle.

Initially, the service provider only wanted to monitor and control their operations (see Fig. 1 to see a possible example of a monitoring framework). Nevertheless, the advances in real-time communications and vehicle location technologies (such as WiFi, 3G, and GPS) over the last two decades have largely increased the availability of such data. There has been an increasing evolution from the old asynchronous acquisition methods, where the data acquired in each vehicle were uploaded to a main server with a large periodicity (commonly daily), to a synchronous method (i.e., real time) [6]. Therefore, it is possible to differentiate the old offline AVL systems and the real-time AVL systems. Such online technology makes it possible to produce continuous flows of data (also known as data streams). Each vehicle transmits the data with a very short (but certain) periodicity to a main server.

In the last decade, many researchers have highlighted the potential of the stored AVL data to provide insights on how to evaluate (and improve) PT reliability in mass transit companies by improving operational planning and control. The technical reports presented by James G. Strathman and his team became the backbone of state of the art on AVL-based evaluations of schedule reliability [11]–[13]. However, the real-time availability of these AVL data opened new research directions for improving PT reliability, namely, by introducing real-time decision models to support the operational control.

The aim of this survey is to review works on the improvement of PT reliability by employing operational planning and control strategies based on historical AVL data. The remainder of this section is structured as follows: first, a high-level definition of operational planning and control is presented, followed by a brief description of the survey’s scope. Finally, the structure of this paper is provided.

A. Operational Planning and Control

Often, reliability problems arise in complex PT networks with high demand. It is possible to divide the causes of reliability problems into two separate groups: internal and external. Internal causes include factors such as driver behavior, passenger boarding and alighting at stops, improper scheduling, route configuration, or interbus effects, which represent persistent problems. External causes are, by definition, more chaotic, and these include traffic congestion and accidents, weather, traffic signs, and interferences with on-street parking. The persistent problems are addressed using OP strategies, whereas the sporadic problems are mitigated by control strategies. The OP strategies are often referred to as preventive actions, which aim to avoid PT unreliability on a long-term perspective, and the control actions have a corrective purpose in a very specific and brief moment [3], [14].

1) OP Strategies: A typical OP process is carried out by sequentially following four steps [15], [16].

a) Network definition: It consists of defining the lines, routes, and subsequent bus stops. Here, a route is considered as a road path between an origin and a destination, which passes by multiple bus stops. A line is defined as a set of routes (which typically consists two routes with very similar paths, but inversely ordered).

b) Schedule planning: The trips are defined by first identifying the set of bus stops for which schedule timepoints will be set (the origin–destination (O–D) stops are always part of this set). Second, timestamps are assigned to previously defined schedule timepoints. Such timestamps may be composed of an expected arrival time plus some slack time. However, in high-frequency routes, this timetabling can be also defined by setting the time between two consecutive trips in the same route (i.e., headway based) [15]. The set of resulting trips is often defined as the schedule plan (SP).

c) Definition of duties: A duty is a task that a driver and/or a bus must perform. The definition of the drivers’ duties has much more constraints than the definition of bus duties (for instance, a driver must stop regularly, governmental legislation). Commonly, the logical definition of bus duties is performed prior to the drivers’ duties.

d) Assignment of duties: It consists of physically assigning the previously defined logical duties to the companies’ drivers and buses.

The AVL-based OP strategies to improve PT reliability consist in adjusting the definitions made on such tasks using real-world data. This type of works focuses on (a) restructuring the route and adjusting the (b) existing SP. AVL-based works on this subject follow this trend. The (c and d) resource-based strategies are applied to improve the profitability rather than the company’s operations. Specifically, the (c and d) definition and assignment of duties are commonly performed by using constraint-based methods and not by analyzing AVL data. In fact, to the authors’ best knowledge, there is no work suggesting it so.

2) Control Strategies: It is reasonable to define control strategies as real-time responses to sporadic service problems [18]. The goal is to restore service normality when deviations occur (i.e., in real time) [14]. It is possible to divide these strategies into two different applications [19]: 1) maintaining...
schedule reliability using metrics such as on-time performance or headway stability (to be discussed in Section IV-A) and 2) schedule coordination at terminals/hubs to facilitate transfers [20]. This paper focuses mainly on the first type of applications.

Such strategies imply the selection of corrective actions (described in Section VI) to avoid eminent unreliable contexts, which are particularly chaotic in high-frequency routes.

B. Scope and Goals

This paper attempts to organize ideas, issues, and approaches developed by the research community for improving the reliability of public road transportation operations based on AVL data. This paper focuses on three different yet highly related approaches: 1) changing the network definition; 2) evaluating and adjusting the SP in place; and 3) automatically selecting corrective control actions in real time.

The goal is to better understand the problems and the existing solutions in the hope of providing some insights and promoting a deeper exploitation of the possibilities enabled by the data collected by automated data collection (ADC) systems. Moreover, this paper aims to provide a clear ground truth to encourage new researchers working on these areas to go further by exploring novel concepts and research lines.

C. Paper Structure

The remainder of this paper is structured as follows. Section II revises the structure, the potential, and the issues related to ADC systems, as well as historical references. Section III first describes the problem of defining a PT network by establishing its routes. Then, the works that employ data collected by ADC systems to improve this definition are presented in detail. Section IV revises the AVL-based works, which evaluate the SP reliability, whereas Section V revises works on improving the PT reliability by adjusting its SP. AVL-based control strategies and models are systematically surveyed in Section VI. Section VII explores the gaps found in the literature and proposes some challenges and opportunities to the research community. Finally, the conclusions are drawn, along with the future trends in this research area.

D. List of Acronyms

Table I contains a list of acronyms used throughout this paper.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>Actual Arrival Time</td>
</tr>
<tr>
<td>ADT</td>
<td>Actual Departure Time</td>
</tr>
<tr>
<td>ART</td>
<td>Actual Run Time</td>
</tr>
<tr>
<td>ATIS</td>
<td>Advanced Traveller Information Systems</td>
</tr>
<tr>
<td>AD</td>
<td>Arrival Delay</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>ADC</td>
<td>Automated Data Collection</td>
</tr>
<tr>
<td>APC</td>
<td>Automatic Passenger Counting</td>
</tr>
<tr>
<td>AVL</td>
<td>Automatic Vehicle Location</td>
</tr>
<tr>
<td>AR</td>
<td>AutoRegressive</td>
</tr>
<tr>
<td>BB</td>
<td>Bus Bunching</td>
</tr>
<tr>
<td>DEA</td>
<td>Data Envelopment Analysis</td>
</tr>
<tr>
<td>EWT</td>
<td>Excess Waiting Time</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HV</td>
<td>Headway Variation</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transportation Systems</td>
</tr>
<tr>
<td>KNN</td>
<td>k Nearest Neighbours</td>
</tr>
<tr>
<td>MA</td>
<td>Moving Average</td>
</tr>
<tr>
<td>OTP</td>
<td>On-Time Performance</td>
</tr>
<tr>
<td>OP</td>
<td>Operational Planning</td>
</tr>
<tr>
<td>p.d.f.</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PPR</td>
<td>Project Pursuit Regression</td>
</tr>
<tr>
<td>PT</td>
<td>Public Transportation</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forests</td>
</tr>
<tr>
<td>RTV</td>
<td>Run Time Variation</td>
</tr>
<tr>
<td>SAT</td>
<td>Scheduled Arrival Time</td>
</tr>
<tr>
<td>SDT</td>
<td>Scheduled Departure Time</td>
</tr>
<tr>
<td>SRT</td>
<td>Scheduled Run Time</td>
</tr>
<tr>
<td>SP</td>
<td>Schedule Plan</td>
</tr>
<tr>
<td>SVR</td>
<td>Support Vector Regression</td>
</tr>
<tr>
<td>TT</td>
<td>Travel Time</td>
</tr>
<tr>
<td>TTP</td>
<td>Travel Time Prediction</td>
</tr>
<tr>
<td>TTV</td>
<td>Travel Time Variability</td>
</tr>
</tbody>
</table>

II. ADC: REVIEW AND POTENTIAL

The GPS is a satellite-based navigation system developed by the U.S. Department of Defense in 1960 [21]. The system was first designed for military purposes to provide navigational fixed data on an hourly basis; however, it did not represent a reliable data source. In 1993, the number of satellites increased to 24. The system became operational and available to civilians. However, its accuracy was low (>100 m). It became available massively in 2000, and the position accuracy of a basic GPS increased to ~10 m. This technology rapidly became a standard to obtain real-time information not only in professional transportation fleets but also in individual vehicles.

Various mass transit agencies now have their fleets equipped with ADC systems in order to track the vehicles’ behavior during their operation. The deployment of such systems usually consists of equipping the fleet’s vehicles with the following: 1) a GPS receiver; 2) sensors of interest; and 3) a communication device capable of communicating with a remote server. The information obtained is then uploaded to the control center. A simple implementation of this system is displayed in Fig. 2. While the GPS receiver is mainly used to track the vehicle’s spatial coordinates, other sensors may be employed to collect other types of data, such as APC. This section provides an overview of ADC systems. First, postprocessing issues on AVL data are discussed. Then, the type of information usually required by the PT companies is presented, along with the vehicle’s location.

A. AVL Data Postprocessing

The ADC systems did not provide a 100% reliable vehicle location data source. In fact, in order to be consistently used, the data often require a preprocessing unit. Commonly, it is
possible to divide this process into a two-stage process: tracking and filtering [22]. Typically, an ADC system only provides real-time vehicle location information, which may be inaccurate due to some satellite unavailability, partial/full signal blocking, or other temporary failure. The tracking stage consists of assigning AVL data to the real schedule data provided by the transit company. Therefore, it is necessary to identify and/or estimate which was the real arrival time to each one of the schedule points (i.e., a set of bus stops along a route to which an exact arrival time is assigned on a timetable). Sometimes, it is necessary to employ regression or other estimation models to fill the gaps on the AVL stream and then perform the matching stage. In the second stage (filtering), it is necessary to transform the AVL data into estimates of vehicle dynamic states (e.g., vehicle speed). Typically, such information can be inferred by employing a Kalman filter or a time series model [22] (see Table V to know more about this type of models). This stage can be disregarded if no more information is required other than the one provided by the first step, or if the AVL system installed also provides that information (e.g., trip start/end information, fuel consumptions, and maintenance status).

### B. Related Data Items

Other potentially valuable data items in addition to GPS location and APC data [6] can be acquired by ADC systems in other onboard sensors. Some of these items are outlined in the following.

1) **Stop and start moment**: With stop-level records, it is possible to obtain additional insights on dwell time and holding.
2) **Event codes**: Events include mechanical alarms and flags (e.g., overheated engine and lift in use), or driver-entered alarms.
3) **Control messages**: From the control center to the operator. Such messages report corrective actions to restore service normality (see Section VI-A to learn more about this topic).
4) **Farebox transactions**: It includes information about ticket validation, namely, its type. It makes it possible to calculate statistics about the type of passengers served by a given route.

These types of data items are already deployed in many AVL systems worldwide [6], and they are helpful in various planning and control tasks such as maintenance scheduling (2), route designing and timetabling (3), corrective control actions (1 and 3), or route designing (namely, its stops) (4).

### III. ON IMPROVING NETWORK DESIGN

The PT network design consists of defining the number of routes to build, along with their paths and bus stops. Mathematically, it is possible to formulate the problem as a graph with nodes, links, and, subsequently, routes. Let $G = (N, A)$ be a graph, where $N, A$ represents a set of nodes and links, respectively; and $R$ is a set of routes. Nodes represent road intersections and/or zones, whereas links represent a particular mode of transport between nodes. A route $r_i \in R$ is a sequence of nodes connected by the links of a single mode.

This graph is usually fed by an O-D matrix previously computed, which covers the O-D zone spatial definition, as well as the time-varying mobility needs between these areas. Bus stops can be established by setting the optimal bus spacing or by considering other demand-related constraints. To know more about this subject, refer to [23].

There are a few AVL-based works focused on improving the PT reliability by adjusting the route definition (i.e., network definition). The most common approach is changing the location of the bus stops. The work in [24] used historical GPS data to mine human mobility patterns in a major Taiwan intercity bus operation to find an optimal compromise between the location of bus stops and operational costs. The researchers did so by discovering the demand patterns using both a stochastic-demand scheduling model and heuristic-based methods to solve the models. A passenger wait cost-based model is developed in [25] to find the optimal bus stop spacing based on historical AVL data.

### IV. ON EVALUATING SP RELIABILITY

The SP reliability is a vital component for service quality. Improvements on reliability may increase the service demand and, consequently, the companies’ profitability. Low reliability levels lead to a limited growth in the number of passengers and to a decreased perceived comfort [11]. It is possible to establish three distinct axes on evaluating SP reliability [26]: 1) the unexpected increases on the waiting time on bus stops; 2) the time spent in crowded situations caused by transport overloading; and 3) delays on the passengers arrivals due to travel time variability (TTV). The first two axes are mainly related with passengers’ comfort and experience criteria. The value of such extra time consumptions varies from the passenger condition (seated or standing) [27]. However, these two aspects are mainly satisfiers: additional aspects that the passengers like to have but are not essential factors to abandon the services provided by a certain PT company. On the other hand, the last one is a fundamental issue by the disturbances that it does on the passengers’ daily activities [26]. By directly affecting the convenience and the speed of transportation, it is key to maintain the travelers’ confidence on the PT network (i.e., a dissatisfier). These priorities on the described PT quality factors are illustrated in Fig. 3. For the aforementioned reasons, this survey is focused on carrying the SP reliability evaluation by the existing TTV. Once established, it is expected that an SP meets the passengers’ demand by following their mobility needs (namely, their daily routines). Typically, service unreliability is originated by one (or many) of the following causes [14], [29]: schedule deviations at the terminals, passenger load variability, running time variability, meteorological factors, and driver behavior.

Today’s urban areas are characterized by a constant evolution of road networks, services provided, and location (for instance, new commercial and/or leisure facilities). Therefore, it is highly important to automatically assess how the SP suits the needs of an urban area. An efficient evaluation can lead to important changes in an SP. These changes will lead to a reduction in operational costs (for instance, by reducing the number of daily trips in a given route) and/or a reliability improvement in the entire transportation network, which will increase the quality of the passengers’ experience and, therefore, the number of customers.
An SP consists of a set \( S = \{ S_1, S_2, \ldots, S_k \} \) of \( k \) schedules, which provide detailed information about every trip running on previously defined routes. Each schedule contains a timetable \( t_i : i \in \{1, \ldots, k\} \). Different routes may have different timetables. Nevertheless, they share the number \( k \) of schedules and the daily coverage \( C_i \) of each schedule.

A schedule planning process for a given route relies on the following three steps: the first step is defining the number \( k \) of schedules and their individual coverage \( C_i \); second, the schedule timepoints are chosen among all bus stops in the route; and finally, the third step is defining timetables \( t_i \) for each route schedule \( S_i \) containing the time the buses pass at each scheduled timepoint (per trip). This process is done for all routes. It should be guaranteed that the number \( k \) of schedules and the coverage \( C_i \) are the same for all routes so that the passengers easily memorize the SP. To learn more about this topic, the reader should refer to the survey on Urban Transit Operational Planning in [15].

From the aforementioned definition of SP, it is possible to divide the SP evaluation into two different dimensions: the suitability of the number of schedules \( k \) and of the set of their daily coverages \( C = \{ C_1, C_2, \ldots, C_k \} \) and the reliability of their timetables \( \{t_1, \ldots, t_k\} \) (to test whether the real arrival times of each vehicle at each bus stop are meeting the previously defined timetable). Although there is an obvious impact on the definition of the timetable, to the authors’ best knowledge, there is no research in the literature addressing the evaluation of the number of schedules and their daily coverage.

This section defines and reviews evaluation methodologies with regard to the reliability of timetables. Fig. 4 illustrates the service improvement cycle based on the AVL data.

A. Evaluation Metrics and Requirements

When evaluating an SP, it is important to differentiate low-frequency services and high-frequency services [30]: in low-frequency services, passengers arrive at the bus stops shortly before the bus’s scheduled services, whereas in high-frequency services, the customers tend to arrive at the stops randomly [31]–[34]. In the first scenario, punctuality is the main metric, whereas the service regularity is the most important metric in high-frequency routes. There is no exact boundary between these two scenarios. In early 2000, Fan and Machemehl conducted a data-driven experiment in Austin, Texas (USA), where they identified a 10-min threshold [35]. Recent studies have also used 10–12 min as a threshold between low- and high-frequency services [26], [36].

Polus presented in 1979 a landmark paper [37] proposing four measures of performance for evaluating SP reliability on arterial routes: overall TT, congestion index, overall travel speed, and delay. All these measures were route based and highly focused on the operational perspective. The first three are mainly variations of the remaining ones—which are based on ratios between the actual and expected run times. Delay was a more sophisticated measure defined as all the time consumed while traffic is impeded in its movement somehow—but also reported as hard to obtain by then.

The AVL data enabled the possibility to extend this analysis to other granularities than route based such as segment based or stop based. Following such advances, four main indicators were first proposed by Nakanishi [38] and followed by other similar studies [8], [11]. These indicators are outlined as follows: 1) on-time performance (OTP); 2) run time variation (RVT); 3) headway variation (HV); and 4) excess waiting time (EWT). The first two indicators are more applicable to low-frequency routes, whereas the last two focus on the high-frequency routes [11], [30], [39]. This set of indicators is the most widely known formulations of these metrics, which have been used on multiple studies in the last decade (which are detailed in Section IV-B). They are formally presented here.

OTP indicates the probability that buses will be where the schedule says they are supposed to be. It is possible to represent this metric by an arrival delay (AD) in a given trip \( i \), i.e., \( AD_i \) as function of both the scheduled arrival time, i.e., \( SAT_i \), and the actual arrival time, i.e., \( AAT_i \). Therefore, it can be defined as follows [39]:

\[
AD_i = AAT_i - SAT_i.
\] (1)

The RVT represents the variation on the run times performed by each trip. Some introductory concepts on this subject will be presented in the following. Typically, the TT reports the trip duration, from terminal to terminal, and is often referred to as round-trip time [16]. TT is often used to define the time required to go from one point of interest to the other [40]. This last definition is used in this survey. One of the factors that mostly affect the RVT is the dwell time, which is the total time the bus has to stay at a given bus stop for passenger boarding and alighting [41]. From the passenger perspective, a larger variation can mean a longer waiting time in some stops and/or missed transfers. From the operational planners’ perspective, greater RVT translates into higher costs as a result of the extra hours that must be added to accommodate passenger load [39]. This indicator is more appropriate for routes that cover long distances, facing many traffic lights and regular traffic delays [42].
Given a set of $n$ trips of interest, it is possible to compute the RTV as follows [11]:

$$ RTV = n^{-1} \times \sum_{i=1}^{n} |SAT_i - AAT_i|/AAT_i. \tag{2} $$

In high-frequency routes, where the trips start within very short headways, the OTP is not that relevant [43]. The HV represents the probability that controllers are able to maintain a regular and stable headway between each pair of vehicles running in the same routes.

Let $f_{i,j}$ be the frequency (i.e., scheduled headway) established between a given pair of trips $(i, j)$, whereas $H_{i,j}$ represents the observed headway on such pair of trips at a bus stop of interest, i.e., $b$. The headway ratio on the bus stop $b$, i.e., $H_r_{i,j}$, is defined as follows [11], [39]:

$$ H_{r_{i,j}} = \left( \frac{H_{i,j}}{f_{i,j}} \right) \times 100 \tag{3} $$

where the value 100 represents a perfect SP matching. Given a set of $n$ trips of interest, it is possible to compute the standard deviation and the mean value of $H_r$ ($\mu_r$ and $\sigma_r$, respectively). We can do it so by calculating every possible $H_{r_{i,j}}$ for $i \in \{1, \ldots, n-1\}$ at a bus stop $b$. Then, it is possible to obtain the HV at bus stop $b$ throughout these $n$ trips as follows [44]:

$$ HV_b = \frac{\sigma_r}{\mu_r} \tag{4} $$

The EWT is an estimation of the excessive waiting time that passengers experience as a consequence of unreliable service. It is possible to calculate the EWT at a bus stop $b$, i.e., $EWT_b$, as a function of $HV_b$. A possible way to do so is presented as follows [45]:

$$ EWT_b = \frac{\sigma_r^2}{2 \times \mu_r^2} \tag{5} $$

The bus stop $b$ used to compute statistics on the first two indicators is the destination bus stop. For the last two indicators, any bus stop can be considered a reference if it has a frequency scheduled to it, i.e., $f_{i,j}$. Commonly, such statistics are computed by the transit companies aggregating its values to a fixed time granularity (typically, 1-h periods) [8], but they can be also computed according to the trip. The following section presents a review of the evaluation of SP reliability by measuring the previously mentioned indicators on historical AVL data.

### B. Review on SP Evaluation Studies

Many works have evaluated schedule reliability by measuring the aforementioned indicators on historical AVL data sets. Strathman et al. [11], [39] evaluated schedule reliability on the Tri-Met by measuring indicators (1–4), whereas the work by Bertini et al. [46] solely focuses on the first two ratios. Traditionally, the HV was often disregarded by the transit planners due to the intrinsic chaos assumed (as the schedule timepoints on the timetables are not the central variable to confirm service reliability). Nevertheless, recent advances have changed this reality: in [13], AVL/APC data were considered to evaluate the impact of the HV on the operational control. Another perspective of the Tri-Met data is presented in [17], where an analysis of indicators (2–4) demonstrated the feasibility of using AVL data along with other data sources to better accomplish their evaluation. Lin and Ruan [47] formulated a probability-based headway regularity metric (HV). Then, the authors tested their approach using AVL data from Chicago. In [48], relations between transit assignment, BB events, and operation models are mined from the location-based data. This study aimed to identify irregularities in HV’s distribution function caused by an inadequate SP. The reliability of an express service implemented in Montreal, Canada, is evaluated in [49] by employing the first two indicators. A large-scale evaluation was performed by Hounsell et al. [10], where the data acquired through the iBus (an AVL/APC framework installed on a bus fleet running in the city of London, U.K.) were used to evaluate all the four main indicators of schedule reliability.

Another approach to evaluate schedule reliability on a route is the segment-based one. It consists of identifying segments/parts of a route where there are greater schedule deviations, and therefore, the SP should be adjusted by changing the timetable or by introducing bus priority lanes and/or traffic signals in intersections. One of the first authors to realize such work was Horbury [50] based on the HV. In [51], it is proposed to measure indicators (1 and 2) using stop-based metrics and to identify the causes for larger deviations through an empirical framework. The work in [52] proposes a way of identifying where the schedule is unreliable by evaluating the first two indicators on the schedule timepoints.

Recently, the methodological approach to evaluate SP reliability has evolved from the key indicators to using nonparametric deterministic methods such as data envelopment analysis (DEA) [53]. The main advantage enabled by employing such a complex method is the possibility of directly comparing metrics from distinct dimensions by introducing decision-making units. Lin et al. [54] used AVL data to establish confidence intervals for the DEA scores based on the four indicators previously introduced. Despite its usefulness in identifying cost-based relationships between the resources used and the service produced, the DEA models are not addressed in this survey as they usually address a wider scope on the companies’ management than our own.

Many of the aforementioned works have often employed the four traditional transit measures to evaluate schedule reliability. Nevertheless, few works have been successful in identifying the factors behind poor performance measurements. Such measurements focus mainly on the passengers’ perspective about the service. Recently, innovative approaches have emerged...
on this research topic, such as the day-to-day variability. Mazloumi et al. [9] proposed to determine the nature and shape of TT distributions for different departure time windows at different times of the day, using data from a route running in Melbourne, Australia. Factors causing TTV in public transport are also explored using regression methods. A method for finding interesting contexts to justify RTV is proposed in [55]: Distribution rules\(^1\) are employed to identify particular conditions that lead to systematic bus delays. The HV is explored using a sequence mining approach in [57]. The goal is to highlight sequences of bus stops where a failure to meet the schedule systematically leads to BB situations further stops ahead in the route. A comparative overview of the aforementioned studies on evaluating SP reliability is presented in Table II. The Granularity column represents the maximum detail level used when calculating the indicators’ values. Recently, an innovative study was presented by Chen et al. [58], where three novel metrics were proposed to address three distinct granularities, namely, stop, route, and network levels. This approach seems promising. However, it also fails to deliver a unique indicator on the SP reliability.

The contribution of this ongoing generator of historical trip data to evaluate SP reliability is that it replaces the old estimations on TTV with real values [46]. The findings of the evaluations previously described consisted of identifying unreliable schedule timepoints [52], [55], [59] or badly designed bus priority lanes [10]. Some evaluations also build dwell time models that help to understand how this variable changes from trip to trip and throughout the day [10].

The four metrics are well established in the literature. However, they focus mainly on the passengers’ perception of service quality, particularly the EWT. The OTP can help the planners identify the exact schedule timepoints to be changed, whereas the RTV shows a more general perspective on network service, which can lead to more profound studies on the drivers’ behavior, terminal dispatching policies, or on the current schedule’s slack. The HV is the most used metric. Even so, it is possible to observe that the company’s perspective on such RTV is not addressed as a primary goal of these evaluation studies.

Nevertheless, even if it is possible to identify what is happening and where changes must be performed to improve SP reliability, it is not easy to identify how it is possible to improve it. The next section focuses on the use of AVL data to develop and/or improve schedule planning.

V. On Improving Schedule Planning

Schedule planning strategies aim at reducing the likelihood of schedule deviations responding to persistent and predictable problems [14]. The questions brought about by the researchers when regarding schedule planning address both the evaluation and improvement of company timetabling. The timetable adjustments can be proposed in three perspectives: 1) slack-time-based perspective; 2) travel-time-based perspective; and 3) headway-based perspective.

In the real world, there is no such thing as a perfect SP. The PT operations will certainly experience some TTV, which will lead to some unreliability comparatively to the previously defined timetables. The aforementioned techniques try to reduce this SP unreliability as much as possible. Typically, the travel-time-based strategies to improve the SP consist of changing the scheduled round-trip times. For that, these strategies use some inference method in order to predict such variables. However, any prediction produced has an associated likelihood. Consequently, such prediction values need to be tuned before going to the public schedule.\(^2\) One of the most common tuning strategies consists of adding slack times to these predictions (particularly in low-frequency routes) based on such variability, as suggested by [55]. A distinct strategy is the headway-based ones: they try to establish optimal bus frequencies by computing a balanced relationship between the expected demand and the available resources. This section presents a systematic revision of these optimizing strategies for improving the schedule timetables.

A. Tuning Up the Schedule Using Slack Times

Operational planners may add slack times to the timetable when they are building a schedule for a low-frequency route. The slack time is the difference between the scheduled and actual expected arrival times [60], [61]. It may be seen as an admissible RTV on a schedule timepoint. The amount of slack introduced can produce large-scale effects on the SP reliability. Insufficient slack times reduce the likelihood of a bus catching up when it falls behind. On the other hand, excessive slack times will reduce the service frequency or increase the amount of resources required, namely, buses and drivers. The definition of an optimal slack time is a problem that can be expressed as a tradeoff function between the service frequency and its reliability. This section addresses the problem of improving SP reliability by determining an optimal slack time based on AVL data. It is possible to define the optimal slack time as follows [60]–[62]: let \(SRT\) be the scheduled run time for a scheduled trip of interest \(i\) between the schedule timepoints \(b_s\) and \(b_d\), respectively; whereas \(ART\) is the actual run time. It is possible to define both using

\[
SRT_{i,b_s,b_d} = SAT_{i,b_d} - SDT_{i,b_s}, \quad ART_{i,b_s,b_d} = AAT_{i,b_d} - ADT_{i,b_s}
\]

where \(SAT\) and \(AAT\) represent the scheduled arrival time and the actual arrival time, respectively; whereas \(SDT\) and \(ADT\) represent the scheduled departure time and the actual departure time between the schedule timepoints \(b_s\) and \(b_d\), respectively. Then, it is possible to define the expected run time in the same context, i.e., \(E(RT)\) as

\[
E(RT)_{i,b_s,b_d} = n^{-1} \sum_{j=1}^{n} ART_{j,b_s,b_d}
\]

where \(n\) represents the number of previous occurrences of the scheduled trip \(i\) (i.e., the same service, day type, and scheduled departure time in previous days) considered to compute \(E(RT)\). Finally, it is possible to define the optimal slack time to be added to the schedule point \(b_d\) in the trip \(i\), i.e., \(s_i^{b_d}\) as follows [60], [61]:

\[
st_i^{b_d} = SRT_{i,b_s,b_d} - \left( \frac{SRT_{i,b_s,b_d}^2}{E(RT)_{i,b_s,b_d}} \right).
\]

\(^1\)Distribution rules were first proposed in [56]. They are related to quantitative association rules but can be seen as a more fundamental concept, useful for learning distributions.

\(^2\)The operational timetable may differ from the one distributed to the general public to improve the passengers’ perception on the quality of service.
Although it is common to add slack to the schedule, research focusing on setting appropriate slack times based on historical AVL data is scarce. It is particularly surprising if we consider that the slack time is defined according to the mean TT, which can be easily computed using AVL data.

To the authors’ best knowledge, there are only four works employing AVL data to optimize slack times in timetables: Dessouky et al. [62] found an optimal slack ratio of 0.25 to be added to the SP in place on a transit agency operating in Los Angeles, USA. In [63], two heuristics are proposed to solve the timetabling problem as an optimization problem by employing two heuristic procedures, namely, ant colony [64] and genetic algorithms [65]. Both the travel and slack times were considered output variables. Such methodology was calibrated using location-based data from a bus route in Melbourne, Australia. Yan et al. [61] proposed a novel optimization model to schedule design by taking into account the bus TT uncertainty and the bus drivers’ schedule recovery efforts. The goal with this model was to find the optimal slack time to add to the schedule running in the city of Suzhou, China. Finally, a distribution rule-based methodology is proposed in [55], and the goal here is to find particular conditions that lead to schedule unreliability. Then, an optimal slack time equation is formulated based on the probability density function (pdf) found in each unreliable context in the city of Porto, Portugal.

By adding slack time to their schedules, the operational planners expect to increase not only the passengers’ satisfaction on service reliability but also the flexibility of both the operators and controllers to take actions for the vehicles to recover their scheduled times. Often, the slack also addresses regulatory questions with regard to the maximum operator driving time, as well as other terminal bus dispatching issues. Due to its importance for the operations and perception on the service quality, further research should be conducted on this topic based on historical AVL data. Nevertheless, this tuning strategy highly depends on the scheduled arrival times. The next section presents a comprehensive review of AVL-based research techniques to improve these times.

B. TTP

One of the most common transportation problems is the travel time prediction (TTP). The literature on this topic is extensive. TTP problems can be used in several contexts such as fleet management, logistics, individual navigation or mass transit, planning, monitoring, and control. This section provides a review of the research focusing on TTP based on AVL data to improve public road transportation planning and monitoring.

TTP consists of predicting the TT for a given trip (or segment). The TT function is formally defined in the following. Let \( TT_{i,j} \) be the run time between two bus stops of interest \( b_i, b_j : j > i \). It is possible to compute TT as

\[
TT_{i,j} = \sum_{k=i}^{j-1} dwT_k + RT_{(k,k+1)}
\]  

(9)

where \( RT_{(k,k+1)} \) is the nonstop running time in the road segment between two consecutive bus stops \( b_k, b_{k+1} \), and \( dwT_k \) is the dwell time on the bus stop \( b_k \).

Although some approaches seem quite simple, various methodologies are employed to TTP problems from different research areas. It is possible to divide such approaches into four distinct categories [66], [67]: 1) machine learning and regression methods; 2) state-based and time series models; 3) traffic theory-based models; and 4) historical database models. This last family of naive approaches consists of simple averages and other types of time-varying Poisson processes whose average TT or speed is achieved by its historical values depending on the day type and/or on the period of the day [68]. Its simplicity is commonly reported as an important drawback when representing the complex relationships between the TT and other variables usually established in urban PT networks. Consequently, they present a poor approach to TTP [67], [69], which is not addressed in this survey. To facilitate the reader’s understanding of this review, Table III presents a detailed description of some complex regression models employed in TTP works.

It is possible to differentiate short- and long-term TTP problems according to the prediction horizon considered. It is common to define such threshold between 60 and 180 min [67], [69]. The long-term TTP is most commonly used for the SP definition—which is the function addressed by this survey. This is an interesting problem due to the existing amount of historical AVL data in the agency databases used today. To accomplish such goal, the prediction should be valid for a long period (for instance, TTP for Monday trips at 8 A.M. should be as most accurate as possible for the entire forecasting horizon). However, the state of the art on long-term TTP is nearly nonexistent. To the authors’ best knowledge, there are only two works on it: Klunder et al. [83] used \( k \) nearest neighbors (kNN) with the input variable departure time, weekday, and date; a comparison of state-of-the-art regression algorithms [support vector regression (SVR), project pursuit regression (PPR), and random forests (RF)] for long-term TTP is presented in [88].

Almost every TTP approaches consider a short-term horizon. The short-term TTP is commonly related to the real-time information on arrival time provided to the clients by the Advanced Traveler Information Systems (ATIS) in place. It is very useful to passengers as it improves both their traveling experience and their transfers [69]. The techniques employed to solve this kind of problems are necessarily different from the long-term TTP and are not directly applicable to the planning stage. However, there are many synergies and commonalities between these two problems and, consequently, between the approaches used to solve each one of them.

1) Regression is the most common approach for both.
2) Some of the regression algorithms employed are applicable in both cases or easily adaptable for that purpose.
3) Information provided by the ATIS on the short-term TT may reduce some passenger-centered TTV, namely, excessive passenger loading at some bus stops and/or major hub stations. This effect will cause a chain reaction by reducing first the \( ADT_i \) and, consequently, the \( SDT_i \) and the TT associated with such stops \( SRT_i \) [see (6)]. As further discussed in Section VII-B2, the authors believe that the rich literature on short-term TTP can present useful lessons to improve the current studies on long-term TTP. For these reasons, the short-term TTP was included in this survey’s scope. From now on, we will refer to TTP using a short-term horizon.

3In addition to the obvious differences in time horizons, the seasonality detected and the importance of the decision variables are completely different from one problem to the other. The differences have a relevant impact when it comes to finding the relationships between these variables and the target variable.
The remainder of this section reviews the works using the approaches to TTP from categories (1–3), followed by methods to evaluate the reliability of such numerical predictions.

1) Machine Learning and Regression Methods: These methods are proposed to infer the arrival times (i.e., a dependent variable) using a mathematical function based on a set of independent variables (i.e., decision variables). Over the last two decades, regression models have been the state of the art on this kind of approach. Works using such type of approaches are summarized in Table III. In addition, providing accurate TTP, the regression models are also commonly capable of estimating the impact that each input variable has on the target variable (i.e., TT). A dwell-time-based simple linear regression model is employed by [92] and [94]–[96]. However, complex models such as SVR, kNN, PPR, and artificial neural network (ANN) are the most popular approaches to this problem due to their ability to find complex nonlinear relationships between the target variable and the independent ones.

ANN is the most successful regression method employed on TTP problems: [66], [72], [74], and [75] used it over location-based data, whereas the works in [71] and [73] used APC data. However, it presents four main drawbacks compared with other regression methods: 1) a time-consuming training procedure [92]; 2) the input–output function is unknown; 3) a reasonable knowledge of the problem is usually required to perform an optimal feature selection, hidden layers, and learning rate [92]; and 4) overfitting is highly possible [71].

Approaches promising to mitigate three of these four limitations have recently been presented: (2) in [78], a method to perform analysis of variance [97] to determine feature selection in order to perform ANN-based TTP; (3) in [77], a genetic algorithm [65] is proposed to find the optimal values for the ANN parameters in TTP context; (4) [77] and [78] proposed to find prediction intervals rather than optimal values for TT to handle the uncertainty within the ANN predictive models. Such prediction intervals reduce the possibilities of overfitting and can be used to optimize the schedule’s slack times. However, such additions to the basic ANN model decrease even more the (1) traditionally slow training process by adding complex preprocessing stages.

The SVR has the advantage of being able to incorporate different types of kernels to find the optimal boundary [92], whereas kNN and kernel-based regression models deal more adequately with missing data or with outliers. The AVL-based kNN models for TTP emerged recently [81], [84]–[86]. Some works report that they can outperform ANN [81], [98]. In addition to the aforementioned characteristics, the kNN is an approach which, conversely to the ANN or the SVR, does not require any assumption about the functional form of the relationship between the dependent variables or the statistical distribution of data (i.e., nonparametric). However, similarly to ANN/SVR methods, its reliability depends on the availability of a sufficiently large quantity of data [98].

Although they are useful, most methods reported do not provide a clear input–output function as linear regression models do. Surprisingly, there are not many works comparing more than two regression methods for TTP [81], [92], and there is simply one focusing on ensemble models [88]. Table IV presents a comparison between the aforementioned regression methods on this specific context. This comparison follows [99, Sec. 10.7]. Although the regression models have been emphasized, there are other types of multivariate prediction models employed to TTP. A novel approach was presented by [100], where they employed clustering algorithms, such as the K-means, to identify a set of past trips with trajectories identical to our own. Then, the data were used to perform a segment-based TTP using AVL data collected in Taipei. A good example of this is the work by [101]: the historical GPS data measurements from a bus network are clustered into three distinct speed profiles by using a fuzzy C-means clustering algorithm. A pdf-based method to improve timetables is proposed by [103]. It aimed to maximize the on-time density area. Such maximization occurs

4It is a black box-type function that only provides an output and not a relationship between the independent variables and the target variable.

5In this context, outliers may be trips with TT largely higher than expected due to some random event or other technical reason.

6Fuzzy clustering denotes that each sample may belong to more than one cluster with a certain likelihood to each one of the considered partitions [102].
when the mean schedule adherence data are at the center of the on-time range.

Recently, promising trajectory-based models employing machine learning techniques are being proposed to address TTP in this context. Reference [104] presented a nearest neighbor trajectory (i.e., based on a kNN model) technique that identifies the historical trajectory that is most similar to the current partial trajectory of a vehicle. A TTP is provided by inferring the future trajectory of a vehicle. Similar trajectory-based approaches are proposed in [86] and [100]. However, such approaches are not applicable to long-term TTP because they mainly provide techniques that depend highly on the information available on the route segment already cruised by the vehicle (i.e., while the trip is in progress).

2) State-Based and Time Series Models: These types of approaches only rely on the most recent data samples, disregarding the remaining historical data. The time series models assume that the TT is a linear/nonlinear combination of its historical values [105]. The state-based approaches usually assume that the future state of the dependent variables only relies on the most recent states. When compared with the other data-driven methods previously described, the present methods do not depend as much on the quantity of data, and they do not require a large training period, since they mainly represent online learning algorithms. These algorithms are well known for their ability to react to unexpected events that may affect the expected traffic flow, such as heavy rains, traffic jams, sports matches, and car accidents. Consequently, they are powerful short-term predictors due to their ability to learn and update in real time, which does not occur with batch learning methods such as the ANN or the kNN [67], [69], [106]. Nevertheless, the performance of these reactive models deteriorates when facing longer forecasting horizons (e.g., [107]). An overview of the most commonly used state-based/time series models is presented in Table V. Time series models assume that the future TT on a given route depends only on its historical values [72].

The strength of these models is their high computational speed. However, they are commonly said to be unable to be built over online data, but only on historical data [69]. Nevertheless, recent studies have demonstrated the opposite [120]. Despite being widely used for traffic flow prediction [121]–[123], time series models are not so common in bus TTP. One of the explanations may be their high sensitivity to changes in the relationship between historical and real-time data, particularly when a stationary data distribution is assumed [105]. Rajbhandari [111] proposed an autoregressive model to capture the temporal variations of bus TT, whereas [117] proposed moving averages in the same context. A self-adaptive exponential smoothing-based algorithm was proposed for interzone link TTP [119].

State-based models are widely reported in TTP literature because they are capable of handling congested traffic situations [67]. The most commonly used state-based model is the Kalman filter [22], [71], [113]–[115], [124]. Its main advantage compared with Markovian approaches is its ability to filter noise in the data, which is extremely relevant in online learning tasks for short-term prediction problems. Cathey and Dailey [22] turned a sequence of AVL measurements into a sequence of vehicle state estimations (i.e., vehicle speed) to predict the arrival time by employing a Kalman filter. A similar approach was followed in [115] and tested using data from buses running in Chennai, India. A model based on two Kalman filter algorithms was developed by [114] to predict running and dwell times alternately in an integrated framework. Such filters use real-time AVL and APC data, respectively.

Lin and Bertini [109] employed a simple Markov chain to predict trip arrival times at each bus stop by formulating a probabilistic transition model between “on schedule” and “behind schedule” states (which will represent the probability of the bus getting back on schedule during the remaining trip). A finite state machine was employed by [110] based on the very same concepts. A Markov model to predict the propagation of bus delays to downstream stops is proposed by [111] for TTP.

In fact, these types of online models are not capable of dealing with long-term TTP alone. However, some works suggest that these models can be used as a complement to regression models [71], [79], [113], [124]. Such approaches are promising. The regression models can handle complex relationships...
between multiple dependent variables by analyzing historical data. The online learning models are capable of using the stream of GPS data to refine these predictions. Commonly, this type of hybrid models employs the Kalman filter as an online building block. To the author’s best knowledge, Wall [124] was the first to suggest this in 1998. The author employed a linear regression model to handle the TTP, whereas the Kalman filter was simply used to track the exact vehicle position based on the real-time stream of AVL data, which was not very accurate at the time (~100 m; see Section II to read more about such issue). In [71] and [79], the Kalman filter is proposed to fine-tune the TTP prediction produced by an ANN model based on APC data, whereas the work in [113] does the same with an SVR model.

3) Flow Conservation Equations and Traffic Dynamics Models: This class of techniques applies relations between traffic variables, obtained from the traffic flow theory, to perform TTP from flow data [67]. Such relations aim to formulate the variables, obtained from the traffic flow theory, to perform the prediction produced by an ANN model based on APC data, whereas the work in [113] does the same with an SVR model.

4) On Evaluating TTP Models: One of the most commonly used metric on TTP problems for the model evaluating task is the root-mean-square error (RMSE) [66], [71], [72], [90], [92], which is now formally defined. Let \( y_i \) be the modeled value (i.e., prediction) for the data point \( y \) at instant \( t \) of the traffic flow speed, on the road segment TT, and on dwell time. It is possible to compute RMSE as

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2},
\]

Although it is relevant, discussing which are the most adequate metrics to evaluate the prediction error is not addressed in this paper. To learn more about this topic, the reader should refer to [128, Sec. 5.8].

C. Setting Optimal Frequencies

In high-frequency routes, the arrival times are not that relevant to the passengers’ perception of quality service and even of operational planning and control. Instead, optimal frequencies are set for such routes, and the reliability studies on these routes usually try to find whether the headway is stable [43]. However, research on setting optimal frequencies in bus timetables based on historical AVL data is scarce. First, a formal definition of the problem is presented based on [129] and [130], and then, research on this topic is presented.

Setting an optimal frequency is a compromise between passenger demand and resources available. Let \( L_0 \) be the desired occupancy of the vehicles operating on a given high-frequency route of interest with \( n \) bus stops, i.e., \( \{b_1, \ldots, b_n\} \) during a

<table>
<thead>
<tr>
<th>Method Type</th>
<th>Methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Reference Works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning and Regression Methods</td>
<td>ANN, SVR, PPR</td>
<td>Good prediction results; Ability to discover non-linear relationships.</td>
<td>Low interpretability; High volumes of quality data are required.</td>
<td>[66], [88]</td>
</tr>
<tr>
<td></td>
<td>Trees-based Reg.</td>
<td>Non-parametric; High scalability and Interpretability.</td>
<td>Low Predictive Power.</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>kNN &amp; Kernel-based Reg.</td>
<td>Non-parametric; Handle missing data and Outliers.</td>
<td>Low interpretability; High volumes of data are required.</td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td>Trajectory-based KNN</td>
<td>Good approximation to the vehicle future’s trajectory.</td>
<td>Do not handle long-term TTP.</td>
<td>[104]</td>
</tr>
<tr>
<td></td>
<td>Other statistical/Reg. methods</td>
<td>Simplicity; High Interpretability.</td>
<td>Low Predictive Power; Non-linear relations are not dealt with.</td>
<td>[94], [103]</td>
</tr>
<tr>
<td>State-based and Time Series Models</td>
<td>Exponential Smoothing</td>
<td>Simplicity.</td>
<td>Long-term TTP and non-linear relations are not dealt with, Low Predictive Power.</td>
<td>[119]</td>
</tr>
<tr>
<td></td>
<td>ARIMA</td>
<td>High computational speed.</td>
<td>Long-term TTP is not dealt with.</td>
<td>[117]</td>
</tr>
<tr>
<td></td>
<td>Markovian Models</td>
<td>Ability to handle unknown system states.</td>
<td>Long-term TTP is not dealt with.</td>
<td>[109]</td>
</tr>
<tr>
<td></td>
<td>Kalman Filters</td>
<td>Ability to filter noisy data.</td>
<td>Long-term TTP is not dealt with.</td>
<td>[114]</td>
</tr>
<tr>
<td>Flow Conservation Equations and Traffic Dynamic Models</td>
<td>Dwell Time &amp; Flow Equations</td>
<td>Complex and Realistic Equations are applied.</td>
<td>Independent of the input data distributions; Expertise on traffic theory is required.</td>
<td>[86]</td>
</tr>
</tbody>
</table>

**Table VI**

High-Level Comparison of Short-Term TTP Models. Key: Reg.—Regression
time interval $T$ between two time instants $(t_i, t_j)$. Then, let $d_{b_k}$ be the average demand on a bus stop $b_k$ during such period $T$ and $N$ be the number of departures available in the same period. The optimal headway in this interval, i.e., $H(t_i, t_j)$, can be obtained as

$$H(t_i, t_j) = \min \left\{ \max \left( \frac{L_0 \times T}{\max (d_{b_k})}, \frac{T}{N} \right), H_0 \right\} : k \in \{1, \ldots, n\} \quad (12)$$

where $H_0$ represents the minimum service level on such period. Obviously, research can be employed to determine the demand levels $d_{b_k}$ based on AVL data.

The work in [73] presents a twofold methodology to set optimal headways. First, the APC data are clustered using hierarchical clustering. Each cluster corresponds to an optimal headway plan. Then, a classification tree is employed to discover rules to classify new instances (i.e., trips) into one of the available headway plans. A promising approach is introduced by Hadas and Shnaiderman [130]: the optimal frequency setting model presented is based on the theory of supply chain models. The AVL data are used to model the statistical distributions of both demand and TT. Although these works present useful insights on headway tuning, the authors believe that there is room to explore AVL data in this specific context.

It is well known that even an optimal schedule planning cannot handle all the problems that arise while the network is operating, particularly in high-frequency routes. The next section presents a summarized review of AVL-based methods to improve operational control in mass transit companies.

VI. AUTOMATIC STRATEGIES ON OPERATIONAL CONTROL

The large-scale introduction of AVL systems in the bus fleets around the globe opened new horizons to operational controllers. This technology made it possible to create highly sophisticated control centers to monitor all the vehicles in real time. However, this type of control often requires a large number of human resources, who make decisions on the best strategies for each case/trip. In the last decade, researchers have started to explore the historical AVL data to build automatic control strategies, which can maintain the buses on schedule while reducing the human participation on the decisions.

This section addresses the AVL-based automatic control strategies. First, the four corrective actions typically recommended by the controllers to the vehicle operators are defined. Then, a systematic review of these automatic control strategies is both presented and discussed.

A. Corrective Actions

There are four typical methods employed as real-time control strategies [18], [57].

1) **Bus holding:** It consists of forcing the driver to increase/reduce the dwell time on a given bus stop along the route.

2) **Speed modification:** This strategy forces the driver to set a maximum cruise speed on its course (lower than usual on that specific route).

3) **Stop-skipping:** Skip one or more route stops; also known as short-cutting when it requires a path change to reduce the original length of the route.

4) **Short-turning:** This complex strategy consists of causing a vehicle to skip the remaining route stops (usually at its terminus) to fill a large service gap in another route (usually, the same route but in the opposite direction). In a worst case scenario, the passengers may be subjected to a transfer.

Bus holding control strategies are the most classic way of maintaining the buses on time. However, this headway alignment is made by increasing the TT of the passengers running in the vehicle [18]. The same applies to speed-based techniques. Moreover, stop-skipping/short-turning techniques align the service headways at the cost of the passengers who have to wait at the stops that were skipped [131].

It is possible to divide the existing bus holding approaches into two main types [132]: 1) models that determine holding times on the basis of a mathematical control formulation with an explicit objective function, such as minimizing total passenger wait time, and 2) threshold-based control models, where buses are held at a control stop on the basis of the deviation of the current TT from the scheduled headway. These models may assume theoretical values of dwell time, TT, or passenger demand (i.e., deterministic models) or assume that such events occur randomly (i.e., stochastic models).

Short-turning actions can be also divided into two subgroups: 1) direct short-turning, where the passengers going in one direction may be transferred to other vehicles, while this one assumes the gap in the service without going to a terminal station, and 2) deadheading, where the vehicle runs directly to its terminus, skipping all stops along the way, regardless of the initial number of passengers running on the vehicle [131].

B. Review of Automatic Control Strategies

One of the first works to employ multiple control strategies was Eberlein et al. [133], which studies three types of control strategies (holding, short-turning, and stop-skipping). The authors considered a one-way loop light-rail transit network of two terminals and $n$ intermediate stations. In [134], the optimal holding time at each stop is formulated as a deterministic mathematical optimization problem. The continuous characteristics of this problem were approached by employing a sliding window. This methodology was easily adapted from the originally studied light train to a bus network by [135]. Hickman [136] presented an analytical model for optimizing the holding time at a given control point in the context of a stochastic vehicle operations model. The author formulated the problem as a convex quadratic program in a single variable, and it is solved using gradient techniques. Zhao et al. [137] presented a multiagent approach, which was based on a negotiation between the bus and the stop agent to address the optimal holding problem.

Fu and Yang [132] proposed a theoretical relationship to express the optimal holding time at a bus stop based on the current variation of the bus headways and the expected passenger waiting time. A similar approach was followed in [138], where the optimal holding finding problem is formulated according to deterministic variables such as the passenger arrival rate, the number of alighting passengers, or the regular dwell time at a bus stop. Then, heuristics were proposed to solve this optimization model. The authors concluded that multiple holding locations are beneficial for minimizing total passenger time, as opposed to the work in [134], which assumed that only
the original terminal should be used as control point. Again, it maintains the limitation of assuming deterministic variables.

The works in [139]–[141] also propose dynamic bus holding models to avoid headway irregularities in high-frequency routes. The methodology was tested using AVL-based simulations, which assumed stochastic distributions for the decision variables. Delgado et al. [142] also suggested preventing passengers from boarding by establishing maximum holding times to maintain the headway stable. A regression model is proposed in [143] to deal with the holding problem: an SVR-based method forecasts the early bus departure times from the next stop based on four input variables (time of day, segment, the latest speed on the next segment, and the bus speed on the current segment). Then, an optimization model is employed to determine the holding time in each (bus, station) pair.

A recent work [144] has proposed a system free to express the natural headway, abandoning the notions of schedule and predefined headway. The scheme presented tends to balance in the presence of disturbances in the dwell time. To test it, the authors started by employing an AVL-based simulation, and then, they reported the results of its deployment on a bus route in Atlanta, USA. Finally, Chen et al. [145] also presented an optimal bus holding time based on stochastic variables. Conversely to previous work on this topic, it explicitly considers passenger boarding activities during holding, reducing the waiting time for those who arrived during that period.

Only a small portion of databased works have employed other preventive actions except changing the bus holding time. Daganzo and Pilachowski [146] proposed a multiagent system where each bus would cooperate with the following to negotiate a maximum cruise speed to maintain the headway reliable. Reference [147] proposed a dynamic discrete objective function that can detect disturbances in the headway regularity at each stop by employing a genetic algorithm. Stop-skipping and bus holding are suggested to the driver according to these events. The increase in the TT for the passengers on board caused by the holding strategy is taken into consideration in these last optimization models. Finally, Liu et al. [131] proposed to study the short-turning as a subproblem of stop-skipping. A mathematical model is proposed using cost-based variables, such as TT or passenger demand. Another issue that presents an overview of these automatic control frameworks.

By analyzing the existing literature on this topic, it is reasonable to conclude that this is still an open research field. Even if there are consistent studies on the holding problem, the four preventive actions were not regarded simultaneously in these works. Moreover, AVL data were used mainly as a proof of concept or to feed some statistical distributions on stochastic variables, such as TT or passenger demand. Another issue that is not as broadly discussed in the literature is the threshold definition [132] (for instance, the minimum level of headway accepted, i.e., \(H_0\) or the maximum service gap tolerated), to define when a control action should be adopted or not.

Despite the intrinsic chaotic characteristics of learning the headway instability problem, most techniques employed are mainly adapted batch learners or online models, which do not consider historical and real-time AVL data simultaneously. Even if such models are able to detect the concept drift often introduced by the unexpected events, which occur in the system, such as traffic jams or a massive demand, few works have reported their deployment in a real-world bus network.

VII. CHALLENGES AND OPPORTUNITIES

This survey presents the most significant contributions in AVL-based research to improve the service reliability on PT networks. This section discusses the challenges and opportunities to overcome and be explored in future research.

A. Issues and Challenges

There are five challenges (a–e) that remain for the research community with regard to both the evaluation and the improvement of the SP. These challenges are presented below.

1) Evaluating SP Reliability: It is possible to identify two main issues where further AVL-based research should be employed to improve the evaluation of SP reliability: (a) creating a unique evaluation indicator, considering the company’s perspective on the evaluation by including external factors in the evaluations or by developing cost-related evaluations and to (b) evaluate the reliability of the current schedule’s number and coverage. These subjects are described in the following.

The aforementioned four evaluation metrics are classical but widely used in evaluation studies. However, distinct metrics (which are highly correlated to the main ones) are continuously emerging. It is known that the importance of each one of these indicators depends on the frequency established in the route. However, to the author’s best knowledge, (a) there is no consensual, individual, and integrated reliability ratio. This gap in the literature leads to an important research question: Is it possible to build a consensual frequency-dependent reliability ratio based on these four main indicators?

The first step in building an SP is defining both the schedule’s number and day coverage. Then, a timetable is assigned to each schedule in a stepwise process already discussed in Section IV. This definition has an explicit impact on the definition of timetables. However, to the authors’ best knowledge, no research addresses the evaluation of whether the schedule’s number and coverage still suit the current demand patterns and network behavior. Consequently, a question arises: (b) Is it possible to assess whether the schedule’s number and coverage...
are suitable for the network needs based on historical AVL data?

2) Improving SP Timetabling: ANN is the state-of-the-art method on TTP due to its ability to solve complex nonlinear relationships. However, other methods seem to emerge to solve such regression problems, namely, the ensemble ones [88]. Two of the main ANN issues are the overfitting and the expertise required to estimate parameters; otherwise, a poor predictor is obtained. State-of-the-art ensemble methods such as bagging [148] and boosting [149] are capable of mitigating such limitations due to their well-known ability to reduce variance, avoiding overfitting, and mitigate bias. Therefore, a question arises: (c) Is it possible to consider ANN as the building block of an ensemble approach to the TTP-based regression problem? It is an open but promising machine learning question whose answers can be a relevant contribution to the TTP problem.

Recently, hybrid methodologies using both regression models and Kalman filters are being proposed for short-term TTP [71], [79], [113]. It is known that Kalman filters are inaccurate when handling distant forecast horizons [107]. However, improving SP timetables is a long-term TTP problem, where the regression methods are increasingly established as the most reliable approach [81], [83], [88]. A hypothesis arises for exploring the two types of models to improve both timetabling and the passengers’ perception on service quality: (d) Is it possible to provide better timetables and real-time adjustments to the arrival time provided by the ATIS by employing simultaneously regression models and speed-based Kalman filters? Further research can be conducted to provide an answer to these questions.

3) Automatic Strategies on Operational Control: Recently, researchers have focused on building automatic and yet efficient control systems capable of monitoring headway regularity and of avoiding inconvenient events, such as BB. Surprisingly, there is not much research on this specific topic. Many of the existing works primarily focus on the optimal bus holding problem, thus disregarding the remaining corrective actions. Consequently, one challenge arises: (e) Is it possible to build a methodology that considers and selects one of the four known corrective control actions based on AVL/APC data? In fact, such methodology addressed two distinct problems that are not conveniently covered in the literature: 1) Is it possible to define an optimal control threshold? 2) How is it possible to choose the best corrective control action after the optimal control threshold is reached? These topics will certainly be addressed in the medium-term future research.

B. Research Opportunities

Three research fields that comprise opportunities for the short term with regard to both the evaluation and the improvement of the SP are described in the following.

1) Improving OP: Section I-A briefly revises the steps of the traditional OP. Although AVL-based research has recently emerged on improving route definition, most AVL-based works on OP focus on the SP. The state of the art relies on deterministic and cost-based models. The AVL data make it possible to perform a bottom-up OP evaluation, namely, correctly exploring the available resources or even reducing them if possible to meet the current demand. A complete AVL-based framework to redesign all the steps of the OP is an intelligent transportation system (ITS) that could be a research goal on this topic for the medium-term future.

2) Improving SP Timetabling: In terms of improving SP timetabling based on AVL data, four subjects emerged as research opportunities as a result of the extensive review of the existing literature: fine-tuning the schedule using the following: 1) long-term TTP and 2) optimal slack times; 3) building automatic methods to perform feature selection for TTP problems; and 4) performing before-and-after SP evaluations. These opportunities are now described in detail.

A large gap identified in the literature has to do with the AVL-based long-term TTP. As described in Section V-B, this problem addresses TTP prediction with a large time horizon and/or lead time. Long-term TTP may refer to a point forecast problem where the target variable is an arrival time for a given trip, or it may refer to predicting the function that describes the expected TT of a given route throughout a day, depending on the departure time, or even for a larger period. The regression models represent the most relevant slice of the state of the art on AVL-based short-term TTP. However, some works have also demonstrated their usefulness in long-term problems [88]. The AVL data make it possible to explore these models to improve the SP. Such approaches can present a breakthrough for this research area over the next decade.

As discussed in Section V-A, the slack time is introduced in the SP to handle variability in TT. Prior to deploying ADC systems in mass transit companies, computing that variability was a difficult task. However, the availability and the reliability of the historical AVL data used today represent a clear opportunity to improve the schedules using this well-known strategy. Moreover, only a few recent works have addressed this topic conveniently [55], [61], [63].

Although regression models are simple to apply, they suffer from several limitations in the context of TTP. The greatest limitation is that many variables in transportation are highly correlated [72], [75]. However, there is not much research on the automatic feature selection for TTP regression problems [16], [78], [150]. This step can be particularly important in large forecasting horizons (i.e., long-term TTP) to facilitate the training stage of complex regression models such as ANN or SVR. As starting point for future research on this topic, the reader can consult an introductory survey on automatic feature selection provided in [151].

Evaluating the changes performed on the SP is difficult prior to deployment. Although this survey discusses various works focused on improving the SP, not many of them evaluate the impact of the suggested changes. The before-and-after evaluation studies are crucial to quantify the relevance of these adjustments. To the authors’ best knowledge, there is only one AVL-based study of this type [49], [96]: [96] selected bus stops and estimated run times for new express services, whereas [49] evaluated the reliability of the route SP after deployment.

This kind of study cannot only encourage other companies to adopt AVL-based methodologies in their OP but also to attract researchers from other areas to explore data in this context. Consequently, this is a research topic that can be explored in the short term.
3) Automatic Strategies on Operational Control: To predict instability and unreliability in the network while the buses are operating is a difficult challenge. Not much research focuses on more than one preventive action. Moreover, the test beds employed are mainly proof of concepts because they use limited data collections both in space (i.e., number of routes) and time. However, many mass transit companies have large collections of data in their databases whose potential is far for being fulfilled. This family of problems is closely related to the online learning problems. Although they are common in TTP problems, machine learning techniques have rarely been applied to build automatic control strategies. Today, one of the most promising research areas is learning from data streams. Problem formulations, such as stream event detection or mining frequent item sets and association rules, can represent a breakthrough in the control area, and even more so if it is possible to combine online learning with patterns mined from historical data. Combining these two areas could be an interesting topic to explore in the near future. To learn more about these types of models, the reader should refer to the detailed survey provided in [106].

VIII. CONCLUSION AND FUTURE TRENDS

This paper has revised the location-based ITS applications for improving both the OP and the control of mass transit transportation networks. Over the last decade, various relevant contributions have emerged on this topic. The spatiotemporal features of this type of data provided novel opportunities to reveal underlying patterns on unexpected behaviors that are deteriorating the quality of the service. These data are now affordable and widely available as a standard in every medium/ large-sized mass transit company.

Such innovation revolutionized the way to improve both operational planning and control in these networks. The theoretical traffic models, which were the state of the art for improving OP during the 1980s/1990s are now being progressively replaced by complex yet efficient statistical and machine learning models, such as support vector machines or ANN, for instance.

It is even more important to provide real-time information to the passengers about what is happening in the network (i.e., on-the-spot information on arrival times). More than building an exact but time-consuming prediction on arrival time, the researchers have focused on building simple frameworks capable of learning from location-based data streams and of providing predictions with low uncertainty.

The AVL-based improvements to planning and control are becoming increasingly mature. The existing evaluation studies are still mainly proofs of concept focused on the passengers' perspective, whereas improvement techniques typically rely on short-term TTP models, ignoring relevant planning problems such as the long-term TTP or finding the optimal slack time to add to each schedule timepoint. On the other hand, the control area focuses on bus optimal holding in control stations, disregarding other control actions. Some challenges related to these gaps in the literature have been addressed by this survey, along with some research lines that can be explored in the current state of the art.

Fig. 5 illustrates a reference map with a timestamped network of regression-based works on TTP. The dashed lines represent the relationships between these papers. As it is easily observable, a subset of these works forms a tree-like directed graph, where the root is the landmark work of Chien et al. [66]. It was one of the first works to take full advantage of GPS data after its accuracy increase in 2000 [21]. It influenced many researchers to explore other types of AVL-based methods to improve the TTP. The authors expect an increase on these graph’s complexity and depth in the next few years. The high availability of reliable data reporting the vehicle operations in real time pushes up this trend. Moreover, research works on other topics (e.g., automatic control strategies) hereby approached can also form these types of graphs. Consequently, it is expectable that data-driven models will prove themselves as state-of-the-art methods for improving PT reliability. More than ever, the AVL data are a real-time stream. Such availability, along with the expansion of urban centers, can progressively change the traditional focus on planning to an autonomous data-driven real-time control, which may reduce the manpower required for those tasks.

REFERENCES

MOREIRA-MATIAS et al.: SURVEY: IMPROVING MASS TRANSIT OPERATIONS BY USING AVL-BASED SYSTEMS


Lúis Moreira-Matias received the M.Sc. degree in informatics engineering from University of Porto, Porto, Portugal, in 2009. He is currently working toward the Ph.D. degree in machine learning in the Faculty of Engineering, University of Porto.

His research interests include supervised learning, online learning, and public transportation planning and control problems.

Mr. Moreira-Matias won a European Data Mining competition at a Research Summer School, Technical University of Dortmund, in 2012.

João Mendes-Moreira received the Ph.D. degree in engineering sciences from University of Porto, Porto, Portugal.

He is an Assistant Professor with the Department of Informatics Engineering, Faculty of Engineering, University of Porto and a Researcher with the Laboratory for Artificial Intelligence and Decision Support (LIAAD), INESC TEC, Porto. His research focuses on the application of machine learning to transport planning problems.

Jorge Freire de Sousa received the Ph.D. degree in engineering sciences from University of Porto, Porto, Portugal.

He was a Member of the Board of the Public Transport Company of Porto (STCP) in 1998–2002 and 2006–2012. He currently teaches in the Department of Industrial Engineering and Management, Faculty of Engineering, University of Porto. He is also currently a Research Member of the Industrial Engineering and Service Management Unit (UGEI), a research unit of INESC TEC, Porto, where his main research interests include transportation, decision support systems, and applied operations research.

João Gama received the Ph.D. degree in computer science from University of Porto, Porto, Portugal, in 2000.

He is a Researcher with the Laboratory of Artificial Intelligence and Decision Support (LIAAD), INESC TEC, Porto and also with the Faculty of Economics, University of Porto. He has authored the book Knowledge Discovery from Data Streams (Chapman & Hall/CRC Press). His main research interest is learning from data streams.