Combining ITS and optimization in public transportation planning: state of the art and future research paths

Christina Iliopoulou* and Konstantinos Kepaptsoglou

Abstract

Intelligent Transportation Systems (ITS) applications in public transportation have allowed for automated data collection, which is particularly useful for planning and operations. While technological advancement of ITS has so far been extensive, their usage for developing relevant planning and operational tools is rather limited. Research on planning and operations of public transportation systems has not widely investigated the potential of combining optimization models with data originating from ITS. Such applications, which could benefit from such an approach include route planning, scheduling and resource allocation in real time. In this context, this paper investigates and critically discusses potential models and methodologies in public transport planning and operations, which can benefit from ITS data, highlights their potential and identifies possible research paths on that area. The overview of literature collectively points to a series of common challenges faced by transportation professionals and underlines the need for better decision support tools for ITS data.

Keywords: Optimization, Public transportation, ITS, Automatic vehicle location, Automatic passenger counting, Automatic fare collection

1 Introduction

The application of advanced communication, electronics and information technologies for improving the efficiency, safety, and reliability of transportation systems is commonly referred to as Intelligent Transportation Systems (ITS) [1]. ITS have enabled the automated collection of transportation data and their efficient transmission, allowing for better, more informed decisions, primarily in "real-time" operations. ITS data exhibit qualities of high volume and continuity in time, which introduce new opportunities in transportation research and practice.

Up to now, developments in ITS have mostly focused on the hardware side, with sophisticated data collection systems applied in daily operations. Nonetheless, software components exploiting ITS for planning and decision- making, have been developed to a lesser extent. For instance, ITS data exploitation for public transport strategic planning has only recently attracted attention, as sustainability has become a pressing issue of modern

times. Various ITS applications have enabled information collection on several fronts, such as performance of public transport, ridership and demand patterns [2, 3]. Examples include Automated Vehicle Location (AVL) systems, which aid monitoring of schedule adherence and permit more accurate development of schedules, electronic fare payment systems and automatic passenger counters, which allow for the collection of detailed ridership data, and computer-aided dispatch systems that help travel patterns to be tracked. As Chapleau et al. [2] note, "*smart card transactions data combined with AVL and GIS (Geographic Information Systems) constitute the ultimate survey for transit planning*". However, the exploitation of such data in order to improve strategic and operational planning of transportation systems has so far been overlooked by the research community [4].

Currently data availability offers numerous opportunities for analysis and extraction of information, yet a small fragment of that information is exploited [5]. In most cases, data transmitted by Geographical Position Systems (GPS) and other equipment are processed by

* Correspondence: christiliop2@gmail.com
School of Rural and Surveying Engineering, National Technical University of Athens, 5, Iroon Polytechniou str, 15770 Zografou, Greece

operators, while no commonly accepted decision support system exists for analyzing them [6]. Nevertheless, operations research methods have the potential to assist decision makers, by transforming huge data streams into meaningful information; these methods can be used not only to evaluate the performance of public transport, but also predict future conditions and generate solutions to planning problems.

The contribution of optimization models in ITS- supported decision making is three-fold. First, the advent of ITS data undoubtedly opens new research paths for optimization models in public transportation, allowing for the investigation of topics, which require the fine spatio-temporal granularity provided by such data (such as the identification of supply and individual mobility patterns) [3, 4]. Second, the availability of real-time information necessitates suitable modifications to assignment and trip planning algorithms, to account for passenger route choice behavior and handle the dynamic nature of data [4, 6]. Third, the lack of matching socio-economic and trip purpose attributes for trips captured through ITS records calls for the development of appropriate methods, which infer required information for model estimation [2, 3]. Evidently, optimization algorithms have a vital role in advancing the state-of-practice towards data-driven public transport planning.

The scope of this paper is thus to systematically and critically review the literature on optimizing public transport systems and services, using AVL and Automatic Passenger Count (APC)/Automatic Fare Collection (AFC) data. The literature on optimization models supported by ITS data has not been systematically reviewed so far, while relevant implementation and methodological issues have not received much attention. Furthermore, a comprehensive theoretical framework organizing such efforts is missing. As such, this study aims to fill research gaps, by systematically organizing existing work and identifying future research paths.

2 Literature review
The problem of planning efficient public transport systems subject to operational and resource constraints is not tractable and thus usually treated as a sequence of sub-problems solved at different stages [7]. There are four distinct stages: strategic, tactical, operational and real-time. At the strategic level, the design of the network and passenger assignment are typically examined as part of a long-term planning process. The tactical planning stage refers to determining operational characteristics of services, namely frequencies and timetables, while operational planning pertains to scheduling and dispatching problems. Finally, real-time applications deal with daily operations and refer to control strategies.

Herein, reviewed studies are classified as strategic, tactical, operational and real-time, based on the decomposition of public transport planning into stages, as proposed in [7]. Furthermore, as certain studies may fit under more than one category, these are classified depending upon their prevailing research focus.

2.1 Strategic level
The value and potential of ITS data for strategic planning has long been recognized [3, 4, 8]. Long-term transport planning usually exploits data derived from surveys; these lead to a static and confined picture of travel patterns, attributed to the long intervals between survey updates and limited samples [3–5]. In contrast, AFC data allow for monitoring individual travelers over long periods of time, thus contributing to an improved understanding of travel behavior mechanisms [3]. The exploitation of AFC and AVL data in this line of research allows for incorporating temporal and spatial demand variations and dynamic patterns into existing models [3, 5]. In this context, relevant studies have mostly dealt with calibration of transit assignment models. The main driver of this research direction has so far been the enhancement of accuracy in transit demand modeling.

2.1.1 Transit assignment
In traditional transit assignment models, passengers are assumed to have no information on actual vehicle arrival times and therefore, attractive path sets for passengers are derived based on the approximation of average traveler behavior [9, 10]. AVL systems however can provide passengers with actual information on vehicle arrivals, significantly affecting boarding decisions [9]. Furthermore, AFC data can reveal actual route choices and allow for constructing more accurate and diversified sets of potential traveler paths [10]. In this context, ITS data have been mostly used to calibrate transit assignment models and improve accuracy in route choice estimation.

AVL/AFC data can aid in realistically modeling headways and travel patterns, and therefore improve route choice models [9–11]. Often, headways are assumed to follow an exponential distribution, a hypothesis which simplifies transit assignment models, as it does not require a complete enumeration of all possible transit paths [9–11]. In this context, relevant research work focuses on using either AVL or AFC data for determining improved travel paths in transit networks [9–11] and for calibrating and/or validating transit assignment models [12–16].

In the same context, the use of AFC data as input in agent-based, microsimulation models has been investigated in the literature, as AFC transactions have the advantage of capturing the behavior of individual passengers at an improved spatio-temporal resolution [3]. Indeed, the disaggregate nature of AFC data permits the development

of direct demand models, which emulate travel demand dynamics, based on observed patterns and reduce modeling effort for agent-based simulation [17]. Relevant studies used AVL and AFC data as inputs for agent-based microsimulation models developed in the open-source platform MATSim [18]; these models were used for realistically capturing route choices [17], inferring daily activity patterns, in conjunction with socio-demographic and land-use data [19, 20], and assessing the impact of pricing policies on travel patters [21, 22].

Nevertheless, the inability to directly derive trip purpose and capture trips made on other transport modes confines the usefulness of AFC data [17, 19–22].

2.1.2 Network design

Traditionally, strategic network planning has been based on fixed demand and travel times representing average conditions, while the design process has relied on expected passenger flows derived from travel surveys, socio-demographic data and the application of transit assignment models [4, 23, 24]. The availability of observed demand and supply patterns from ITS streams presents a unique opportunity for transitioning to data-driven design in public transport. Depending on the nature of information available, revealed performance issues or mobility patterns can be exploited and appropriate design objectives can be defined for planning public transport networks. So far, few studies focus on adjusting bus route networks based on AVL data to improve performance [4]; these include bus route generation and schedules [23], optimal stop spacing [25] and inferring trip patterns along with bus network design [26].

2.2 Tactical level planning

Tactical level planning may largely benefit from longitudinal ITS data; APC/AFC and AVL data available over time, can capture frequent mobility patterns [2, 3, 8] and reliability issues [4], respectively. Indeed, AFC data aid in incorporating temporal and spatial demand variability in tactical planning, as well as assessing traveler response to service adjustments [2, 3]. Typically, in tactical-level decisions, demand is assumed to be a-priori known [27], yet in the presence of AFC data, several studies attempt to estimate origin – destination (OD) matrices in the context of timetable/frequency/level of service adjustments. Such studies are characterized as tactical, as the main driver is the improvement of the service offered to passengers [7].

2.2.1 Optimal timetabling

Outcomes of studies focusing on optimal timetabling studies are highly related to data availability and detail level. For instance, APC data have been used to distinguish homogeneous bus ridership patterns and determine distinct bus headways [28] and loop detector data

have been exploited for generating optimal bus schedules assuming constant headways [29]. On the other hand, the existence of AVL data contributes into explicitly considering travel time and headway variability throughout the day in timetable design [30, 31].

Temporal demand patterns have also been extracted using APC/AFC data and incorporated in multi-period timetabling optimization models, to account for demand variation over time [32–35]. Such patterns were inferred from AFC data and incorporated in timetabling optimization models [32, 34], while historical AVL and APC data were exploited to obtain reliable bus dispatching headways [33] or generate optimally coordinated timetables [35].

Overall, the lack of passenger arrival information has been a limiting factor for timetabling studies; researchers have so far resorted to the use of widely accepted assumptions on passenger arrivals at bus stops; alternatively, waiting times can be accurately estimated by using video footage, crowdsourced mobile application data [32] or by subtracting vehicle arrival and AFC timestamps [35].

2.2.2 Origin-destination and transfer inference

Service improvement decisions are contingent upon the availability of route load profiles and preferred route choices by transit users [3]. Regular travel surveys, albeit of limited temporal and spatial coverage, provide full trip details, including actual trip origins and destinations [2, 3]. On the contrary, extended AFC datasets can reveal ridership patterns over a long timeframe for the entire service network, yet a series of enrichment and inference methods are required in this case to deduce linked trips and journey edges [2, 21]. A popular field for these applications is that of bus systems without exit control; in such cases, the alighting stop must be inferred in order to generate trip sequences [36]. These studies may be characterized as tactical, as they can be used for service adjustments and better management of passenger flows [3]. The contribution of optimization methods is rather significant in this research area, as ridership estimation relies on the enumeration of feasible paths, which obviously leads to computationally intractable problems. Thus, the development of suitable and computationally tractable optimization models has allowed for inferring trip patterns from AFC data, while also exploiting large amounts of temporal information [37].

Several studies have developed algorithms to estimate origin-destination (OD) related data and structures using AFC data: Trépanier et al. [38] exploited AFC data to account for similarities between trips over successive days and identify transit alighting points, while Munizaga and Palma [39] combined AFC and AVL data to describe travel patterns for metro and bus trips. Other efforts

focused on using AFC/APC data for modifying the iterative proportional fitting (IPF) method [40], which has been widely applied for estimating route-level OD matrices from boarding and alighting counts [41, 42]. In detail, as a seed OD matrix is required for IPF implementation, Ji et al. [42, 43] derived such a matrix within hybrid IPF-based methodologies, using APC data. The simultaneous presence of AFC and AVL/farebox data has also been exploited within rigorous estimation algorithms to overcome the difficulty of distinguishing short activities from transfers when trying to identify linked trips [43–47].

In the same research direction, researchers attempted to model route choice under known trip origins and destinations for estimating passenger flows. A main contribution of AFC data in this case is the imputation of passenger behavioral choices. This allows for readjusting optimization objectives and quantifying the disutility of factors such as transfers and waiting times. Related studies have so far referred to urban railway networks, due to the availability of both entry and exit point AFC transactions in them and involved route choice modeling [48–50] and the identification of flows in network transfer points [51–54].

2.2.3 Activity modeling

The high spatio-temporal resolution of AFC data gathered over long time periods creates an advantageous setting for exploring the underlying mechanisms of travel behavior compared to traditional survey collection methods [3]. Nonetheless, AFC data do not capture socio-economic and trip purpose attributes, contrary to household and on-board travel surveys [19–21]. To overcome this limitation and improve the understanding of passenger behavior, some tactical-level studies have focused on devising appropriate methodologies for the identification of activity patterns [44]. In contrast to rule-based approaches, rigorous methodologies can yield more robust estimates for home locations and trip purposes [37, 55].

Most studies on activity and pattern detection have adopted segmentation approaches for the identification of homogeneous groups of transit users and frequent travel patterns using AFC data. Indeed, the presence of longitudinal geospatial data has directed research attention into clustering algorithms, the application of which is also congruent with market segmentation research and can serve a variety of policy-oriented questions [3]. A variety of clustering methods have been explored so far. Agglomerative hierarchical clustering has been employed to determine periods of homogeneous flow [56] and distinguish users with similar temporal behavior [57–59]. Similarly, a large body of literature has applied k-means clustering to identify regular spatial and temporal patterns [60–62] and understand social interactions between transit users [63].

The suitability of the Density-Based Scanning Algorithm with Noise (DBSCAN) for mining temporal and spatial travel patterns has also been recognized in the respective literature [64, 65], while modified versions of the algorithm have been devised to improve performance [66] and estimate residence and workplace locations of users [67]. As a general note, bi-level clustering procedures have been employed to treat the spatial and temporal nature of ITS data [68].

The aforementioned approaches utilize classic clustering methods which largely depend upon the specification of parameters, the specification of which warrants an extensive analysis on its own. Aiming to overcome these challenges, El Mahrsi et al. [69] used generative model-based clustering to investigate passengers' temporal patterns and station usage patterns. Furthermore, in most studies, clustering methods are mostly applied to isolate spatial and temporal clusters and in some cases, statistics are utilized to estimate spatio-temporal relationships. Qi et al. [68] pointed out that spatial or temporal travel patterns are incomplete, as the dimensions of time and space cannot be treated separately and proposed a suitable, three-step methodology to discern regional mobility patterns using ITS data. Finally, the increased computational complexity of clustering methods renders them inapplicable for large-scale real-world transit networks. In this context, Kieu et al. [70] devised a spatial clustering algorithm to generate user clusters with similar spatial and behavioral features and highlighted its superior performance over existing methods. As a final remark, the growing research attention towards the application of unsupervised methods [68, 69] and spatial analytics [67, 70] highlights the potential contribution of these methods in the field of activity detection.

2.3 Operational level

Operational-level planning refers to vehicle scheduling, driver rostering, maintenance planning, as well as parking and dispatching [7]. Associated planning decisions benefit from the AVL data availability, as incorporation of service reliability and trip time variation into typical approaches can yield improved optimization models for these planning tasks [4]. Still, few studies on operational decisions have exploited AVL data, while so far, the only problem addressed has been the generation of optimal vehicle schedules. The associated Vehicle Scheduling Problem (VSP) is that of the optimal allocation of vehicles to trips, based on precompiled timetables, yet in the presence of AVL data, operators can devise more robust vehicle schedules based on observed trip times [71–74]. Indeed, AVL data have allowed for extracting periods of homogeneous running time [72] and trip time probability distributions [73, 74] to determine reliable vehicle schedules that enhance service reliability;

computationally efficient heuristic solution approaches have been proposed to handle the increased problem complexity. Evidently, there are still ample grounds for research on the different sub-problems faced by operators in the operational planning stage. The availability of APC/AFC data can additionally allow for addressing associated problems through the perspective of both passengers and operators in multi-objective solution frameworks.

2.4 Real-time operations

AVL have been widely applied for real-time control of public transportation systems and particularly for alleviating bus bunching, large waiting times at stops and so on [4]. Real-time bus location data permit the provision of dynamic route guidance and traveler information, contributing to reduced waiting times and an overall enhanced user experience [5].

2.4.1 Real-time trip planning

The advent of AVL data has enabled the incorporation of real-time information in trip planning models. In the presence of real-time information, computationally intensive transit planning models may be unsuitable to quickly generate optimal paths, while inherent assumptions on fixed travel times and transit on-time performance should be modified as well [75, 76]. Indeed, itinerary planning applications based on published transit schedules are subject to inaccurate predictions since waiting and transfer times are naturally time-dependent, thus require appropriate modifications to be used in the real-time planning horizon [76]. Under this context, research efforts have been directed towards efficient trip planning models, which explicitly incorporate real-time AVL data in order to accurately represent bus arrival times.

A few studies have focused on the development of modified shortest path algorithms in order to take into account bus arrival information. Hickman [76] exploited historical AVL records to derive on-time arrival probabilities and determine possible passenger itineraries. Using real-time GTFS data, Chen et al. [75] proposed a reliability-based online trip planning model which explicitly considered schedule adherence and travel time uncertainty. Capitalizing on the availability of different data sources, Tien et al. [77] harnessed real-time AVL data and real-time user location traces provided by mobile devices to generate tailored trip plans.

The provision of information on alternative modes and possible connections is reasonably more attractive to passengers yet requires the integration of additional data sources. Under this scope, multi-modal trip planning systems using real-time GPS data from portable devices along with real-time traffic data [78] and data from passengers' mobile phones [79] have been presented in the literature. In general, although ITS data are indispensable for the development of accurate itinerary planners, without information on traffic conditions and alternative travel options, such applications remain mainly targeted towards regular public transport users. As such, these applications can greatly benefit from data integration and web crawling methods to merge different data streams.

2.4.2 Real-time control

Prior to the wide deployment of ITS, control strategies were implemented by personnel located at designated control points; consequently, earlier control models assumed no-real time information, rendering respective results inapplicable in current ITS-supported transit systems [80]. The emergence of AVL systems has directed a lot of research towards models for optimal real-time control, capitalizing on the availability of online information [4, 80]. Generally, three types of control strategies may be distinguished: station control (holding and station-skipping), inter-station control and other strategies [81]. So far, several models for optimal bus holding considering real-time information have been proposed in the literature; the models presented in [82, 83] considered real-time bus arrival information, while other studies considered both online AVL data and real-time passenger demand estimates [84–87]. The holding control problem has been formulated through analytical models under deterministic [82] or stochastic vehicle travel times and passenger loads [83, 84] and through dynamic programming [85]. Several studies focused on predictive control, using AVL data to forecast vehicle arrival/departures within the optimization framework [88]; GA-based predictive control models featuring both holding and stop-skipping strategies were formulated in [81, 86]. Exploiting real-time availability of bus location information, rolling horizon mathematical programming models were proposed for holding control and appropriate heuristic solution frameworks, to handle increased computational loads [87, 89, 90]. In a different approach, Yu and Yang [91] used support vector machine regression to more accurately predict vehicle departure times per stop and subsequently employed GA optimization to determine the optimal holding time. A few studies directly exploited real-time APC/AFC data to model passenger flows in holding control optimization attempting to minimize travel times and delays due to holding [80, 92–94].

However, aforementioned studies did not actually determine optimal control strategies in a data-driven manner, but relied on the estimation of arrival times through prediction methodologies and simulation analysis to evaluate proposed models [80, 88]. In this context, of specific interest is the work in [95, 96], who explored the practical applicability of optimal holding control models proposed in the literature and underlined arising issues on the topic.

Table 1 summarizes existing publications utilizing ITS generated data to optimize transit planning:

3 Main findings and research gaps
The emergence of ITS challenges conventional decision support methods, while at the same time creates new research opportunities. This section identifies data-related and methodological issues, gaps in existing literature and discusses how ITS are shaping new pathways for developing ITS data driven models in public transport planning.

3.1 Practical challenges arising in ITS data exploitation
The review of existing literature has shed light on certain practical issues, which have so far hindered the widespread adoption of ITS-based models for public transportation planning and design. These include, but are not limited to:

- Additional data processing required: Many AVL and AFC systems do not archive data in a readily utilized manner, as they are primarily designed for system monitoring [8]. This means that additional data processing and analysis are required in order to render this data useful to transit planners [4, 5, 96].
- Lack of integration among various data sources: Cumbersome procedures are required, so that the inputs required by a planning/design model, specific practitioners' knowledge and the outputs of monitoring systems may be consolidated in a common framework.
- Different degrees of fleet penetration: While AVL systems are typically installed on entire bus fleets, the same is not true for APCs which may be deployed on 10–15% of the fleet [8, 46]. The availability of passenger demand data or lack thereof dictates the analysis that can be undertaken, as without APC/AFC the latter is inevitably limited to operational characteristics such as speed, delay and reliability.
- Current state of practice: The role of optimization-based approaches has been somewhat limited to supporting decision-makers rather than actually deciding, while most studies address "stylized" problem settings, lacking the degree of realism required in practice [6].
- Increased computational requirements: Planning models require the execution of more computationally intense tasks, while traditionally used well-known algorithms must be modified in the case of real time information [9].
- Operators' data-sharing policies: Certain operators have adopted a data-sharing stance, spurring ITS related research. This, however, is not the typical case, as limited data sample availability is often reported because of privacy concerns and operators' restrictions.

3.2 Research opportunities
Combining optimization and ITS generated data for public transport planning problems is a field with increasing attractiveness by the research community. Published work mostly deals with tactical or real-time problems, while the lack of studies investigating design-related and operational-level problems is observed [4].

3.2.1 Strategic level planning
Harnessing ITS data for the purpose of strategic-level planning contributes to shifting towards data-driven and demand responsive public transport service design. Of specific interest is the concept of transit network redesign [97]. While public transport network design has been one of the most popular fields for optimization methods [24], reformulating the associated problem in a data-driven framework is not that straightforward. Similarly, AVL data can provide insights on the actual performance of public transport networks, permitting the computation of performance metrics, which may be used as design objectives. Furthermore, the analysis and utilization of both AVL and APC/AFC data enables the inclusion of social considerations, such as equity and accessibility in a realistic design process.

By integrating AFC and AVL data into Agent-based Microsimulation models, various issues related to passengers' response to different policies may be explored, allowing for a more realistic representation of problems investigated [98]. In this context, diverse passenger preferences can be reproduced based on AFC, including temporal flexibility and sensitivity to fare and service changes, thus a series of strategic decisions, including fiscal policy, can be evaluated [22]. Furthermore, the incorporation of AVL/GPS data into agent-based systems can improve route choice and passenger behavior modeling accuracy [19] and handle interactions with other modes [17]. Overall, the strength of strategic analysis using ITS data lies in the actual representation of supply and demand, rendering potential long-term decisions significantly more impactful. However, further research is needed to explore how to exploit ITS data to restructure public transport networks and define appropriate problem formulations.

3.2.2 Tactical level planning
Tactical planning decisions can benefit by analyzing patterns of ridership and vehicle trajectories. Relevant studies have embedded statistical and simulation techniques within optimization frameworks to account for the stochastic nature of vehicle travel times captured through

Table 1 Overview of studies using ITS data by research purpose and planning level

Author	Date	Data	Research Purpose	Level
Yan et al.	2006	Loop Detector	Optimal bus route design	Strategic
Liu et al.	2017	AFC	Optimal bus route design	Strategic
Li and Bertini	2008	AVL	Optimal bus stop spacing	Strategic
Poon et al.	2004	AFC	Transit assignment	Strategic
Fung et al.	2005	AFC	Transit assignment	Strategic
Vuk and Hansen	2006	APC	Transit assignment	Strategic
Bouman et al.	2012	AFC	Transit assignment	Strategic
Schmöcker et al.	2013	AFC	Transit assignment	Strategic
Ordóñez Medina and Erath	2013	AFC	Transit assignment	Strategic
Lovric et al	2013	AFC	Transit assignment	Strategic
Zhu, et al.	2014	AFC	Transit assignment	Strategic
Li et al.	2015	AVL	Transit assignment	Strategic
Chen and Nie	2015	AVL	Transit assignment	Strategic
Fourie et al	2016	AFC, AVL	Transit assignment	Strategic
Ali et al.	2016	AFC	Transit assignment	Strategic
Tavassoli et al.	2018	AFC, GTFS	Transit assignment	Strategic
Hadas and Shnaiderman	2012	AVL, APC	Optimal frequency setting	Tactical
Patnaik et al.	2006	APC	Optimal headway setting	Tactical
Gkiotsalitis and Cats	2018	AVL, APC, GTFS	Optimal headway setting	Tactical
Yan et al.	2006	Loop Detector	Timetabling	Tactical
Mazloumi et al.	2012	AVL	Timetabling	Tactical
Yan et al.	2012	AVL	Timetabling	Tactical
Sun et al.	2014	AFC	Timetabling	Tactical
Wang et al.	2017	AFC	Timetabling	Tactical
Guo et al.	2017	AFC	Timetabling	Tactical
Kusakabe et al.	2010	AFC	Route choice modeling	Tactical
Zhou and Xu	2012	AFC	Route choice modeling	Tactical
Van der Hurk et al.	2013	AFC	Route choice modeling	Tactical
Trepanier et al	2007	AFC	OD flow estimation	Tactical
Mc Cord et al.	2010	APC	OD flow estimation	Tactical
Munizaga and Palma	2012	AVL, AFC	OD flow estimation	Tactical
Gordon et al.	2013	AVL, AFC	OD flow estimation	Tactical
Ji et al.	2014	APC	OD flow estimation	Tactical
Ji et al.	2015	APC	OD flow estimation	Tactical
Xu et al.	2016	AFC	OD flow estimation	Tactical
Sanchez-Martinez	2017	AFC, AVL	OD flow estimation	Tactical
Gordon et al.	2018	Farebox, AFC	OD flow estimation	Tactical
Ji et al.	2011	AFC	OD flow modeling	Tactical
Ma et al.	2013	AFC	OD flow modeling	Tactical
Hofmann et al.	2009	AFC	Transfer identification	Tactical
Hong et al.	2016	AFC	Transfer identification	Tactical
Yap et al.	2017	AFC	Transfer identification	Tactical
Han and Sohn	2016	AFC	Activity detection	Tactical
Goulet-Langlois et al.	2016	AFC	Activity detection	Tactical

Table 1 Overview of studies using ITS data by research purpose and planning level *(Continued)*

Author	Date	Data	Research Purpose	Level
Ma et al.	2017	AFC	Activity detection	Tactical
Zou et al.	2018	AFC	Activity detection	Tactical
Qi et al.	2018	AFC	Activity detection	Tactical
Agard et al.	2006	AFC	Pattern detection	Tactical
Morency et al.	2007	AFC	Pattern detection	Tactical
Agard	2009	AFC	Pattern detection	Tactical
Sun et al.	2013	AFC	Pattern detection	Tactical
El Mahrsi et al.	2015	AFC	Pattern detection	Tactical
Kieu et al.	2015	AFC	Pattern detection	Tactical
Kieu et al.	2015	AFC	Pattern detection	Tactical
Ghaemi et al.	2017	AFC	Pattern detection	Tactical
Zhao et al.	2017	AFC, AVL	Pattern detection	Tactical
Kieu et al.	2018	AFC	Pattern detection	Tactical
Shen et al.	2016	AVL	Vehicle Scheduling	Operational
Shen et al.	2016	AVL	Vehicle Scheduling	Operational
Shen et al.	2017	AVL	Vehicle Scheduling	Operational
Eberlein et al.	2001	AVI	Optimal Control	Real-Time
Zolfaghari et al.	2004	AVL	Optimal Control	Real-Time
Yu and Yang	2009	AVL	Optimal Control	Real-Time
Chen et al.	2013	AVL,APC	Optimal Control	Real-Time
Asgharzadeh and Shafahi	2017	AVL,APC	Optimal Control	Real-Time
Luo et al.	2017	AVL, AFC	Optimal Control	Real-Time
Berrebi et al.	2018	AVL,APC	Optimal Control	Real-Time
Berrebi et al.	2018	AVL	Optimal Control	Real-Time
Hickman	2003	AVL	Trip planning	Real-Time
Tien et al.	2011	AVL	Trip planning	Real-Time
Zhang et al.	2011	AVL	Trip planning	Real-Time
Li et al.	2012	AVL	Trip planning	Real-Time
Chen et al.	2016	GTFS	Trip planning	Real-Time

AVL records, and incorporated trip patterns exploiting APC/AFC data.

Besides timetabling, the extraction of traveler flows from AFC data allows for further tactical-level analyses, rendering origin-destination inference a prominent research path. The majority of earlier studies in this field utilized fixed sets of assumptions and rules, sequentially applied to select the most probable origins/destinations [36, 99–102]. In contrast, optimization-based methodologies using AFC and AVL data can capture the effect of service-related parameters on route-choice behavior, improve the understanding of passenger choices during service disruptions [49, 53], deduce missing information [37] and estimate the percentage of transit users not captured by AFC data [103]. Such studies have reported improved estimation accuracy, underlining the potential of

devoting more research effort towards optimization-based enrichment and validation processes [43, 47].

Spurred by the presence of geospatial data, as well as the need to circumvent the lack of socio-demographic and trip purpose information in ITS data, activity and pattern detection has been a topic investigated in the literature [104]. Well-known clustering methodologies have been employed to extract spatial and temporal patterns from AFC data. These rely on arbitrary thresholds and parameter values under some type of contextual information or user preferences. On the contrary, although harder to design and implement, model-based clustering methods can adapt to more complex data patterns and can be used in conjunction with travel demand simulation models [57, 69]. Along the same lines, machine learning algorithms can be applied prior to segmentation,

to transition from user-specified parameters to data-driven inference [104]. Spatial analysis can also be exploited to investigate the presence of spatial relationships between ridership patterns and service characteristics. The identification of these features can help correct potential biases and derive underlying mobility principles at different levels of aggregation. Such information may in turn be used within optimization models to define more appropriate design objectives for passenger-oriented service adjustments or simply to ensure computational feasibility in cluster-first/schedule-second schemes [20, 105].

3.2.3 Operational level planning
Overall, there is a lack of studies on operational planning decisions using AVL/AFC data. In general, if suitably processed, ITS data can be used to reduce costs and improve service level [106]. Specifically, because of AVL technology, flexible routing and paratransit can be incorporated into regular transit services, particularly for agencies operating in low density areas. Although a few studies use AVL data for vehicle scheduling, subsequent operational planning steps have so far been neglected. Like timetabling and scheduling, new problem formulations for dispatching and parking allocation are required to deal with travel time variability. This is very important, since several operational planning problems, such as vehicle parking and dispatching need to be addressed daily [7]. What is more, the discrete problems included in operational planning are computationally expensive [71]. In the case of the VSP for instance, which is a NP-hard problem, devising computationally efficient methods is a promising research area. Overall, given the complexity of multi-period scheduling and dispatching problems, the contribution of ITS supported optimization methods in the operational planning stage is expected to be significant [7].

3.2.4 Real-time operations
Real-time control strategies have significantly benefited from the existence of AVL and APC/AFC data [4, 27]. Several directions for improving real-time control algorithms may be identified in this case. Travel time prediction algorithms could aid flexible routing solutions to estimate how schedule deviations may alter running times [107]. Few such studies were identified [76, 77], indicating that this appears to be a promising research path, which could also include performance comparisons among different algorithms [96]. Furthermore, combining optimal control models and prediction methodologies is deemed a promising path [91], as existing studies typically use model-based predictions for arrival times, thus not performing a purely data-driven analysis [88]. As such,. Further, the availability of real-time passenger demand data can significantly improve the performance of control models in cases of overcrowding [87] and in

the context of transfer synchronization [80]. Finally, control strategies are almost exclusively verified using simulation, yet the implementation of a real-time holding method involves technical challenges that can be overlooked in a simulation environment [96]. Although it can be hard to convince agencies to allow experimentation [94], such experiments lead to valuable conclusions and advance both research and practice.

3.3 Research limitations
The advent of ITS data has undeniably enhanced modeling accuracy with respect to spatial and temporal characteristics of mobility and highlighted new research avenues along the way. Yet, the application of optimization techniques has been relatively slower as apart from technical challenges, a series of limiting factors are identified in the process of devising ITS-supported models. Prominent issues include the underlying data quality, the need for supplementary data sources and the increased computational burden faced by researchers.

3.3.1 Data quality considerations
Inevitably, benefits in modeling accuracy obtained by exploiting ITS data naturally depend on the quality of the data utilized [46, 105]. The latter is dictated by the technical specifications of the ITS system deployed [72] as well as the archiving process [8]. Indeed, benefits stemming from ITS-supported decision making are intertwined with the data reporting standards adopted by operators. For instance, older/less advanced AVL systems produce reports which contain vehicle trajectory data, lacking stop-level information [72], calling for matching algorithms to couple raw location data to route maps and schedules [8, 44]. Regardless of the type of ITS, a series of similar data manipulation procedures have been proposed to remove problematic entries and impute missing values [2, 4, 44, 96]. Still, the success of these methods is contingent on the underlying datasets, while operator-specific data archival practice results in peculiarities in captured data [8]. Mitigation of these concerns is mostly dependent on public transport operator policies, through maintaining quality control and post-processing procedures [8]. Interoperability is another key issue, as the adoption of common standards and input file specifications among agencies can advance both research and practice [5, 8]. Interestingly, in the realm of ITS-assisted operations, research progress largely depends on applied practice, thus the creation of various synergies between agencies, research institutions and software development are crucial.

3.3.2 Supplementary data requirements
The use of ITS data may undeniably provide answers on a broad spectrum of transportation research questions,

from long-term planning to real-time control strategies. Nevertheless, AFC and APC data have limitations for some analyses, as critical elements required for decoding traveler choices are lacking [19, 46, 105]. In this context, demand-related issues such as mode shift behavior and induced demand cannot be exclusively accounted for, by solely using ITS data [17]. Furthermore, if AFC data are available, passenger flows and activities may be inferred to some extent, yet through a series of associated processes. These include estimating alighting points through transaction sequences, linking trips based on spatio-temporal coincidence and imputing trip purposes based on location are most commonly employed [2].

Among required procedures, alighting point estimation is the first and most important step for OD inference. This process requires the definition of arbitrary thresholds for spatial and temporal proximity, reasonably resulting to the inability of linking a significant portion of individual trips [39]. In this case, the validation of inference methodologies is contingent on the availability of actual passenger counts [46, 53]. When survey data is lacking, the inclusion of historical OD flow or onboard survey data [42] and the comparison of different estimation methodologies are alternative options to assess consistency of results [46]. Similarly, transfer identification is dominated by rules on maximum journey duration and elapsed time thresholds [102, 105]. Cross-referencing AVL and AFC records can generally allow for higher precision in the estimation of bus to bus or bus to metro transfers [102]. As a step towards decreasing reliance on external data sources, the possibility of endogenous validation has been proposed for checking the validity of estimation of users' home location and trip distances [105, 108]. Nonetheless, exogenous validation is still required for behavior-related parameters such as willingness to walk [108].

Along the same lines, activity identification is often conducted based on temporal windows linked to anticipated work/study schedules and/or spatial proximity to points of interest. This approach obviously renders generated results largely dependent on subjective assumptions about typical passenger behavior [69, 104]. Point of interest and land use data are generally easy to obtain and perhaps the most widely used data source for characterizing trip purposes [37]. If AFC records are linked to fare types, a crude segmentation of users based on age and occupation may allow for more insightful conclusions [2]. Alternatively, activities can be assigned based on archived socio-demographic and census data [19], while onboard complementary surveys are naturally the most informative data source, yet sample rates are typically low [100, 105].

Last, the availability of AFC data does not itself guarantee an accurate depiction of ridership patterns; apart from data quality and completeness, market penetration for the operators is critical for the modeling accuracy achieved [17]. A notable consideration refers to the issues of user noninteraction and fare evasion, which can lead to underestimating transit flows and may only be captured by questionnaires and manual surveys [39, 103].

3.3.3 Computational effort

Collectively, researchers have agreed upon increased computational costs associated with (a) processing ITS data and (b) specifying optimization models across all planning stages [29, 42, 43, 78, 94]. Optimization formulations accounting for variability in input data, either through statistics or simulation-based evaluation of objectives reasonably entail the execution of additional processes [31, 51, 54, 55, 85, 91]. Especially agent-based simulation models require significant efforts for calibration and validation [18, 20]. Clustering approaches are also subject to the large computational cost of processing vast amounts of transactions [59, 65, 70]. Route choice modeling faces similar challenges, as the incorporation of information provision to passengers via ATIS is captured through time-expanded transportation networks, increasing the dimensionality of the underlying path selection problem [9, 10, 39, 77, 78]. Reasonably, these issues are exacerbated in the real-time planning horizon, as results must be generated in a timely manner [94].

A direct approach to computational effort considerations is obviously the use of high computing power, yet access to equipment of such specifications is among all subject to budget availability. Distributed and cloud computing is an efficient and cost-effective alternative, as it allows for performing different procedures simultaneously, thus greatly reducing processing times [12]. However, identifying the tasks to be parallelized is not straightforward, while computer science skills are required to a certain extent [6].

Recognizing the contribution of optimization models in solving transportation problems, the shift from mathematical programming formulations towards powerful heuristic/metaheuristic algorithms is a promising strategy. So far, efforts have employed mathematical programming and heuristic approaches, despite the abundance of metaheuristics for transit planning [24, 29, 73]. There are various opportunities for such applications. Adaptive metaheuristic and dynamic programming algorithms can be applied to efficiently handle dynamic real-time problems [6], while population-based methodologies can produce optimal solutions in a fraction of the time required by integer solvers for multi-period problems [29, 35]. Still, transforming time-varying data into appropriate encoding schemes for metaheuristics is not straightforward, while it is computationally infeasible to manipulate solutions which occupy

too much computer memory [91]. Research is thus needed towards translating ITS data into suitable forms which can be inputs to metaheuristic frameworks, as well as devising hybrid algorithmic frameworks.

4 Emerging trends

The era of big data has cultivated a new reality in transportation planning. Besides ITS, new data sources have become available for capturing travel behavior mechanisms and estimating relevant transportation models. The overall explosion of data has in turn led to the exploration of automated planning frameworks, offering a streamlined process for data manipulation.

4.1 Emerging data sources

Emerging data sources stemming from the ubiquitous penetration of internet-based devices may be exploited on their own or in conjunction to ITS data, to facilitate transportation planning. Most notably, mobile phone data have been at the core of relevant efforts due to their broad spatial and temporal coverage and the possibility of real-time updating, which can lead to more robust and responsive transport models [109]. These data refer to Call Demand Records (CDR) or sightings records, depending on whether a trace is generated when a person uses their phone to text/call or simply when the phone connects to the network [110]. Along with their undeniable advantages, mobile data come with a unique set of challenges. Researchers have collectively distinguished the most prominent issues faced when dealing with mobile phone data, namely oscillation/false displacement and location uncertainty [110–112]. Despite these issues, the immense research opportunities arising from mobile phone data have spurred efforts, mainly in the computer science field, towards methods and algorithms for overcoming the difficulty of accurately estimating user locations and consequently, travel behavior models.

In contrast to ITS data, mobile phone data present the major advantage of tracking users across all transport modes and capturing a larger spectrum of activities. While the event-driven nature of mobile phone data might not allow for link travel time estimation, the high penetration rates and long recording periods hold potential for estimating passenger flows [113]. Capitalizing on the latter, a series of research efforts in passenger flow estimation from mobile phone traces have been published recently [109, 111, 112, 114–116]. Nonetheless, these studies have either entirely neglected mode choice [109, 111, 112, 115] or solely focused on vehicle trips [114, 116], due to some limiting factors. Indeed, vehicle trips may be validated through odometer readings [109], known speed-space profiles [117], usage rates in geographical units corresponding to home locations of users [116] or observed traffic counts [114, 116].

For public transportation planning, OD matrices generated from mobile data must be post-processed to obtain mode-specific trip tables [111]. So far, studies on mode-choice inference from mobile data are scarce [118], as researchers have underlined the complexity of such a task [116]. Indeed, travel mode identification from mobile phone records requires the use of multiple data sources in conjunction to speed estimation and trip matching algorithms [117, 119]. In terms of strategic planning, data-driven transit network design has been examined in [120, 121] using large-sample trajectory data to (a) identify frequent mobility patterns (ignoring mode choice) from mobile phone data and (b) generate public transport routes. Mobile phone data has the potential to facilitate microsimulation modeling, including activity-based and agent-based modeling based on complex network theory [122]. They can also serve as supplementary data sources for AFC data to determine the locations visited by an individual between successive transaction records [21, 55]. However, like ITS data, mobile phone data lack semantic information, such as socio-economic attributes and trip purpose [109, 111]. In this context, segmentation approaches have been used along with sets of rules and assumptions for activity inference and trip distribution [109, 111]. Since clustering approaches do not offer insights on the type of activity performed, the frequencies of visits, land use patterns and empirical rules can be exploited to impute the most probable work/home locations activity types [118].

Still, the former approaches refer to OD matrices which are not mode-specific, thus an additional step would still be required for discerning public transport trips. As an alternative approach, web-based and social media information can be combined with AFC or other ITS data to infer trip purpose and mode information, particularly for special events [123]. Mode inference and trip chaining can be performed based on data from GPS-tracking devices, such as car navigation systems [55]. Crowdsourced data can be helpful in providing quality metrics for services offered or collecting information on facilities such as bike paths [113, 117, 123]. These data can provide insights on the factors driving passenger route choices [32] and enhance estimation accuracy [77]. Further, by exploiting crowdsourced data in conjunction with spatial data, the use of additional variables can be permitted to detect the type of activities performed [104]. Still, such data are drawn from very specific user groups, thus inherently suffer from sampling bias and should be carefully interpreted [122].

Overall, the complexity and extensive data requirements to infer public transport trips have reasonably hindered the application of mobile phone data for the purposes of public transportation planning. So far, their use for operational and real-time planning faces major

challenges and entirely relies on the progress made at the previous planning stages.

4.2 Integrated transit modeling

The disaggregate nature of AFC transactions and the presence of trajectory data calls for new data mining methods and algorithms, as well as advanced statistical inference techniques [19, 20, 38, 122]. Responding to the overarching need for better decision support tools for ITS data, there exists some work on the development of data-driven platforms for public transportation planning [5, 124]. The latter integrate data mining methods, regression models and visualization techniques to assist in performance monitoring, predict and evaluate potential impact of different transit strategies and provide a more comprehensive understanding of network dynamics overall. Unsupervised machine learning tools can be employed to classify activities based on AFC data without any preconceptions on activity types [55], identify mobility patterns [68, 69] or detect performance issues for which no prior knowledge exists [88]. Data-driven optimization models may be employed following automated data cleaning and processing and optimized design parameters can be readily available to planners, operators and administrative staff. Regardless of the level of sophistication in associated models, the commercialization of such tools can directly contribute towards the wider adoption of ITS-enabled analyses.

5 Conclusions

Optimization models have been useful planning tools for decades and are utilized to solve problems at every stage of the public transport planning process. The explosion of data stemming from ITS systems calls for a readjustment of such models to incorporate actual knowledge of passenger demand patterns and bus arrival times. The literature is slowly shifting towards the adoption of data-driven planning approaches, introducing a new era in transit planning.

Indeed, planners and engineers must extend the capabilities of current models to adapt to the challenges posed by the wealth of available data. Moving forward, the success of ITS-based public transport planning lies on the integration of traditional transport planning, advanced computer science algorithms and data mining techniques. Collectively, however, these issues put additional pressure on transportation research to understand and implement computer science algorithms and tools. It is relatively uncertain to expect that the transportation community can independently handle this challenge, but standardization of main data processing steps and commercialization of necessary tools may be an encouraging step in this

direction. Research progress may be achieved by open-sourcing relevant software and creating publicly available resources for dealing with big data manipulation. Above all, the cooperation between the fields of computer science, advanced statistics and transportation planning is considered indispensable in the face of the big data era.

In order to achieve the transition to data-driven planning, existing and well-known algorithms and models including transit assignment, route design and shortest path algorithms must be suitably modified. Particularly in the context of strategic planning and demand-oriented improvements, the determination of an appropriate data manipulation strategy to incorporate ITS data into optimization frameworks in a meaningful and computationally feasible manner is not trivial. Up until now, there is no clearly defined path for translating ITS data streams in meaningful inputs, thus comparative analyses between different approaches are needed to identify the most efficient strategies.

The overview of the literature underlines that no data source is independently adequate for efficiently applying transportation-related optimization models and algorithms. Research efforts should be devoted to automated validation procedures, through the application of advanced artificial intelligence techniques to discover and correct inconsistencies in the data sets. Such applications could be validated against survey estimates to derive the most efficient inference methodologies, giving rise to a new design paradigm.

To conclude, the relationship between optimization and public transport planning, although being constantly redefined, remains indispensable and will continue to evolve in parallel with the emerging significance of the role of transit systems [7]. With the advent of big data, the contribution of optimization models in public transport planning is multifaceted and is manifested in various problem-solving stages, from parameter calibration to results' validation. It is thus expected that data-driven public transport planning will be the mainstream approach in a few years, following the introduction of ITS systems on urban centers under the sustainable mobility paradigm.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Smith, B. L., & Venkatanarayana, R. (2005). Realizing the promise of intelligent transportation systems (ITS) data archives. *Journal of Intelligent Transportation Systems, 9*(4), 175–185.
2. Chapleau, R., Trépanier, M., & Chu, K. K. (2008). The ultimate survey for transit planning: Complete information with smart card data and GIS. In *Proceedings of the 8th international conference on survey methods in transport: Harmonisation and data comparability* (pp. 25–31).
3. Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies, 19*(4), 557–568.
4. Moreira-Matias, L., Mendes-Moreira, J., Sousa, J. F., & de Gama, J. (2015). Improving mass transit operations by using AVL-based systems: A survey. *IEEE Transactions on Intelligent Transportation Systems, 16*(4), 1636–1653. https://doi.org/10.1109/TITS.2014.2376772.
5. Ma, X., & Wang, Y. (2014). Development of a data-driven platform for transit performance measures using smart card and GPS data. *Journal of Transportation Engineering, 140*(12), 04014063.
6. Crainic, T. G., Gendreau, M., & Potvin, J. Y. (2009). Intelligent freight-transportation systems: Assessment and the contribution of operations research. *Transportation Research Part C: Emerging Technologies, 17*(6), 541–557.
7. Desaulniers, G., & Hickman, M. D. (2007). Public transit. *Handbooks in Operations Research and Management Science, 14*, 69–127.
8. Furth, P. G., Hemily, B., Muller, T. H., & Strathman, J. G. (2006). *Using archived AVL-APC data to improve transit performance and management*. Washington: Transit Cooperative Res program (TCRP) Report 113, Transp Res Board.
9. Chen, P. W., & Nie, Y. M. (2015). Optimal transit routing with partial online information. *Transportation Research Part B: Methodological, 72*, 40–58.
10. Schmöcker, J.-D., Shimamoto, H., & Kurauchi, F. (2013). Generation and calibration of transit hyperpaths. *Procedia - Social and Behavioral Sciences, 80*, 211–230.
11. Li, Q., Chen, P. W., & Nie, Y. M. (2015). Finding optimal hyperpaths in large transit networks with realistic headway distributions. *European Journal of Operational Research, 240*(1), 98–108.
12. Zhu, W., Hu, H., & Huang, Z. (2014). Calibrating rail transit assignment models with genetic algorithm and automated fare collection data. *Computer-Aided Civil and Infrastructure Engineering, 29*(7), 518–530.
13. Poon, M. H., Tong, C. O., & Wong, S. C. (2004). Validation of a schedule-based capacity restraint transit assignment model for a large-scale network. *Journal of Advanced Transportation, 38*(1), 5–26.
14. Fung, S., Tong, C., & Wong, S. (2005). Validation of a conventional metro network model using real data. *Journal of Intelligent Transportation Systems, 9*(2), 69–79.
15. Vuk, G., & Hansen, C. O. (2006). Validating the passenger traffic model for Copenhagen. *Transportation, 33*(4), 371–392.
16. Tavassoli, A., Mesbah, M., & Hickman, M. (2018). Application of smart card data in validating a large-scale multi-modal transit assignment model. *Public Transport, 10*(1), 1–21.
17. Fourie, P. J., Erath, A., Ordonez, S., Chakirov, A., & Axhausen, K. W. (2016). Using Smartcard Data for Agent-Based Transport Simulation. In *J.-D. S. a. F. Kurauchi (Ed.), Pub Transp Planning with Smart Card Data*. London: Taylor & FranciS.
18. Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N., & Nagel, K. (2009). MATSim-T: Architecture and simulation times. In *Multi-agent systems for traffic and transportation engineering* (pp. 57–78). Hershey, PA: IGI Global.
19. Ali, A., Kim, J., & Lee, S. (2016). Travel behavior analysis using smart card data. *KSCE Journal of Civil Engineering, 20*(4), 1532–1539.
20. Ordóñez Medina, S. A., & Erath, A. (2013). Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore. *Transportation Research Record, 2344*(1), 20–30.
21. Bouman P, Van der Hurk E, Kroon L, Li T, Vervest P (2013) Detecting activity patterns from smart card data, paper presented at the BNAIC 2013: Proceedings of the 25th Benelux conference on artificial intelligence, Delft, The Netherlands, 7-8, 2013.
22. Lovrić, M., Li, T., & Verves, P. (2013). Sustainable revenue management: A smart card enabled agent-based modeling approach. *Decision Support Systems, 54*(4), 1587–1601.
23. Yan, S., Chi, C.-J., & Tang, C.-H. (2006). Inter-city bus routing and timetable setting under stochastic demands. *Transportation Research Part A: Policy and Practice, 40*(7), 572–586.
24. Kepaptsoglou, K., & Karlaftis, M. (2009). Transit route network design problem:Review. *Journal of Transportation Engineering, 135*(8), 491–505.
25. Li, H., & Bertini, R. L. (2008). Optimal bus stop spacing for minimizing transit operation cost. In *Traffic and transportation studies* (pp. 553–564). https://doi.org/10.1061/40995(322)51.
26. Liu, Y., Liu, C., Yuan, N. J., Duan, L., Fu, Y., Xiong, H., Xu, S., & Wu, J. (2017). Intelligent bus routing with heterogeneous human mobility patterns. *Knowledge and Information Systems, 50*(2), 383–415.
27. Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological, 77*, 38–75.
28. Patnaik, J., Chien, S., & Bladikas, A. (2006). Using data mining techniques on APC data to develop effective bus scheduling plans. *Journal of Systemics, Cybernetics and Informatics, 4*(1), 86–90.
29. Mazloumi, E., Mesbah, M., Ceder, A., Moridpour, S., & Currie, G. (2012). Efficient transit schedule design of timing points: A comparison of ant colony and genetic algorithms. *Transportation Research Part B: Methodological, 46*(1), 217–234.
30. Hadas, Y., & Shnaiderman, M. (2012). Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. *Transportation Research Part B: Methodological, 46*(8), 1068–1084.
31. Yan, Y., Meng, Q., Wang, S., & Guo, X. (2012). Robust optimization model of schedule design for a fixed bus route. *Transportation Research Part C: Emerging Technologies, 25*, 113–121.
32. Wang, Y., Zhang, D., Hu, L., Yang, Y., & Lee, L. H. (2017). A data-driven and optimal bus scheduling model with time-dependent traffic and demand. *IEEE Transactions on Intelligent Transportation Systems, 18*(9), 2443–2452.
33. Gkiotsalitis, K., & Cats, O. (2018). Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. *Transportation Research Part C: Emerging Technologies, 88*, 187–207.
34. Sun, L., Jin, J. G., Lee, D.-H., Axhausen, K. W., & Erath, A. (2014). Demand-driven timetable design for metro services. *Transportation Research Part C: Emerging Technologies, 46*, 284–299.
35. Guo, X., Sun, H., Wu, J., Jin, J., Zhou, J., & Gao, Z. (2017). Multiperiod-based timetable optimization for metro transit networks. *Transportation Research Part B: Methodological, 96*, 46–67.
36. Nassir, N., Khani, A., Lee, S., Noh, H., & Hickman, M. (2011). Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. *Transportation research record, 2263*, 140–150. https://doi.org/10.3141/2263-16.
37. Zou, Q., Yao, X., Zhao, P., Wei, H., & Ren, H. (2018). Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway. *Transportation, 45*(3), 919–944.

38. Trépanier, M., Tranchant, N., & Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems, 11*(1), 1–14.

39. Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies, 24*, 9–18.

40. Ben-Akiva, M. E., Macke, P. P., & Hsu, P. S. (1985). Alternative methods to estimate route-level trip tables and expand on-board surveys. *Transportation Research Record, 1037*, 1–11 http://onlinepubs.trb.org/Onlinepubs/trr/1985/1037/1037-001.pdf.

41. McCord, M., Mishalani, R., Goel, P., & Strohl, B. (2010). Iterative proportional fitting procedure to determine bus route passenger origin-destination flows. *Transportation Research Record, 2145*, 59–65. https://doi.org/10.3141/2145-07.

42. Ji, Y., Mishalani, R. G., & McCord, M. R. (2014). Estimating transit route OD flow matrices from APC data on multiple bus trips using the IPF method with an iteratively improved base: Method and empirical evaluation. *Journal of Transportation Engineering, 140*(5), 04014008.

43. Ji, Y., Mishalani, R. G., & McCord, M. R. (2015). Transit passenger origin–destination flow estimation: Efficiently combining onboard survey and large automatic passenger count datasets. *Transportation Research Part C: Emerging Technologies, 58*, 178–192. https://doi.org/10.1016/j.trc.2015.04.021.

44. Nassir, N., Hickman, M., & Ma, Z.-L. (2015). Activity detection and transfer identification for public transit fare card data. *Transportation, 42*(4), 683–705. https://doi.org/10.1007/s11116-015-9601-6.

45. Gordon, J., Koutsopoulos, H., Wilson, N., & Attanucci, J. (2013). Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transportation Research Record, 2343*, 17–24.

46. Gordon, J. B., Koutsopoulos, H. N., & Wilson, N. H. (2018). Estimation of population origin–interchange–destination flows on multimodal transit networks. *Transportation Research Part C: Emerging Technologies, 90*, 350–365.

47. Sánchez-Martínez, G. E. (2017). Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. *Transportation Research Record, 2652*, 1–7.

48. Kusakabe, T., Iryo, T., & Asakura, Y. (2010). Estimation method for railway passengers' train choice behavior with smart card transaction data. *Transportation, 37*(5), 731–749. https://doi.org/10.1007/s11116-010-9290-0.

49. Van der Hurk v E, Kroon L, Maroti G, Vervest P. (2014). Deduction of Passengers' route choices from smart card data. *IEEE Transactions on Intelligent Transportation Systems, 16*(1), 430–440.

50. Zhou, F., & Xu, R. H. (2012). Model of passenger flow assignment for urban rail transit based on entry and exit time constraints. *Transportation Research Record, 2284*(1), 57–61.

51. Hofmann, M., Wilson, S. P., & White, P. (2009). Automated identification of linked trips at trip level using electronic fare collection data. In *Presented at the transportation res board 88th annual meeting transportation res board* https://trid.trb.org/view.aspx?id=881728.

52. Hong, S. P., Min, Y. H., Park, M. J., Kim, K. M., & Oh, S. M. (2016). Precise estimation of connections of metro passengers from smart card data. *Transportation, 43*(5), 749–769.

53. Yap, M. D., Cats, O., van Oort, N., & Hoogendoorn, S. P. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. *Transportation Research Procedia, 27*, 1042–1049.

54. Xu, X., Liu, J., Li, H., & Jiang, M. (2016). Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. *Transportation Research Part E-Logistics & Transportation Review, 87*, 130–148.

55. Han, G., & Sohn, K. (2016). Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. *Transportation Research Part B: Methodological, 83*, 121–135. https://doi.org/10.1016/j.trb.2015.11.015.

56. Ji, Y., Mishalani, R., McCord, M., & Goel, P. (2011). Identifying homogeneous periods in bus route origin-destination passenger flow patterns from automatic passenger counter data. *Transportation Research Record, 2216*, 42–50. https://doi.org/10.3141/2216-05.

57. Ghaemi, M. S., Agard, B., Trépanier, M., & Partovi, N. V. (2017). A visual segmentation method for temporal smart card data. *Transportmetrica A: Transport Science, 13*(5), 381–404.

58. Goulet-Langlois G, Koutsopoulos HN, , Zhao J (2016) Inferring patterns in the multi-week activity sequences of public transport users. Transportation Research Part C: Emerging Technologies, 64:1–16.

59. Agard, B., Morency, C., & Trépanier, M. (2006). Mining public transport user behaviour from smart card data. *IFAC Proceedings Volumes, 39*(3), 399–404.

60. Morency, C., Trépanier, M., & Agard, B. (2007). Measuring transit use variability with smart-card data. *Transport Policy, 14*(3), 193–203.

61. Agard, B. (2009). Mining smart card data from an urban transit network. In *Encyclopedia of data warehousing and mining* (2nd ed., pp. 1292–1302). Hershey, PA: IGI Global.

62. Zhao, J., Qu, Q., Zhang, F., Xu, C., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems, 18*(11), 3135–3146.

63. Sun, L., Axhausen, K. W., Lee, D. H., & Huang. (2013). Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences, 110*(34), 13774–13779.

64. Ma, X., Wu, Y.-J., Wang, Y., Chen, F., & Liu, J. (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies, 36*, 1–12.

65. Kieu, L. M., Bhaskar, A., & Chung, E. (2015). Passenger segmentation using smart card data. *IEEE Transactions on Intelligent Transportation Systems, 16*(3), 1537–1548.

66. Kieu, L. M., Bhaskar, A., & Chung, E. (2015). A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies, 58*, 193–207.

67. Ma, X., Liu, C., Wen, H., Wang, Y., & Wu, Y. J. (2017). Understanding commuting patterns using transit smart card data. *Journal of Transport Geography, 58*, 135–145.

68. Qi, G., Huang, A., Guan, W., & Fan, L. (2018). Analysis and prediction of regional mobility patterns of bus Travellers using smart card data and points of interest data. *IEEE Transactions on Intelligent Transportation Systems, 20*(4), 1197–1214.

69. El Mahrsi, M. K., Côme, E., Oukhellou, L., & Verleysen, M. (2017). Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems, 18*(3), 712–728.

70. Kieu, L. M., Ou, Y., & Cai, C. (2018). Large-scale transit market segmentation with spatial-behavioural features. *Transportation Research Part C: Emerging Technologies, 90*, 97–113.

71. Ceder, A. (2007). *Public transit planning and operation: Modeling, practice and behavior.* Boca Raton: CRC press.

72. Shen, Y., Xu, J., & Zeng, Z. (2016). Public transit planning and scheduling based on AVL data in China. *International Transactions in Operational Research, 23*(6), 1089–1111.

73. Shen, Y., Xu, J., & Li, J. (2016). A probabilistic model for vehicle scheduling based on stochastic trip times. *Transportation Research Part B: Methodological, 85*, 19–31.

74. Shen, Y., Xu, J., & Wu, X. (2017). Vehicle scheduling based on variable trip times with expected on-time performance. *International Transactions in Operational Research, 24*(1–2), 99–113.

75. Chen, Y., Yang, S., Hu, M., & Wu, Y. J. (2016). A reliability-based transit trip planning model under transit network uncertainty. *Public Transport, 8*(3), 477–496.

76. Hickman, M. (2003). Robust passenger itinerary planning using transit AVL data. In: The IEEE 5th International Conference on Intelligent Transportation Systems, Singapore (pp. 840–845). IEEE. https://doi.org/10.1109/ITSC.2002.1041329.

77. Tien DN, MacDonald T, Xu Z (2011) TDplanner: Public transport planning system with real-time route updates based on service delays and location tracking. In: Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd (pp. 1–5). IEEE. http://ieeexplore.ieee.org/abstract/document/5956479/

78. Li, J.-Q., Zhou, K., Zhang, L., & Zhang, W.-B. (2012). A multimodal trip planning system with real-time traffic and transit information. *Journal of Intelligent Transportation Systems, 16*(2), 60–69. https://doi.org/10.1080/15472450.2012.671708.

79. Zhang, L., Li, J. Q., Zhou, K., Gupta, S. D., Li, M., Zhang, W. B., Miller, M. A., & Misener, J. A. (2011). Traveler information tool with integrated real-time transit information and multimodal trip planning: Design and implementation. *Transportation Research Record, 2215*(1), 1–10.

80. Gavriilidou, A., & Cats, O. (2018). Reconciling transfer synchronization and service regularity: Real-time control strategies using passenger data. *Transportmetrica A: Transport Science, 15*(2), 215–243. https://doi.org/10.1080/23249935.2018.1458757.

81. Sáez, D., Cortés, C. E., Milla, F., Núñez, A., Tirachini, A., & Riquelme, M. (2012). Hybrid predictive control strategy for a public transport system with uncertain demand. *Transportmetrica, 8*(1), 61–86.

82. Sun, A., & Hickman, M. (2008). The holding problem at multiple holding stations. In *Computer-aided systems in public transport* (pp. 339–359). Berlin, Heidelberg: Springer. http://link.springer.com/chapter/10.1007/978-3-540-73312-6_17.

83. Berrebi, S. J., Watkins, K. E., & Laval, J. A. (2015). A real-time bus dispatching policy to minimize passenger wait on a high frequency route. *Transportation Research Part B: Methodological, 81*, 377–389.

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

85. Hadas Y, Ceder A (Avi) (2010) Optimal coordination of public-transit vehicles using operational tactics examined by simulation. Transportation Research Part C: Emerging Technologies, 18(6) : 879–895. https://doi.org/10.1016/j.trc.2010.04.002

86. Cortés, C. E., Sáez, D., Milla, F., Núñez, A., & Riquelme, M. (2010). Hybrid predictive control for real-time optimization of public transport systems' operations based on evolutionary multi-objective optimization. *Transportation Research Part C: Emerging Technologies, 18*(5), 757–769.

87. Sánchez-Martínez, G. E., Koutsopoulos, H. N., & Wilson, N. H. M. (2016). Real-time holding control for high-frequency transit with dynamics. *Transportation Research Part B: Methodological, 83*, 1–19.

88. Andres, M., & Nair, R. (2017). A predictive-control framework to address bus bunching. *Transportation Research Part B: Methodological, 104*, 123–148.

89. Zolfaghari, S., Azizi, N., & Jaber, M. Y. (2004). A model for holding strategy in public transit systems with real-time information. *International Journal of Transport Management, 2*(2), 99–110.

90. Eberlein, X. J., Wilson, N. H., & Bernstein, D. (2001). The holding problem with real–time information available. *Transportation Science, 35*(1), 1–18.

91. Yu, B., & Yang, Z. (2009). A dynamic holding strategy in public transit systems with real-time information. *Applied Intelligence, 31*(1), 69–80.

92. Chen, Q., Adida, E., & Lin, J. (2013). Implementation of an iterative headway-based bus holding strategy with real-time information. *Public Transport, 4*(3), 165–186. https://doi.org/10.1007/s12469-012-0057-1.

93. Luo, X., Liu, S., Jin, P. J., Jiang, X., & Ding, H. (2017). A connected-vehicle-based dynamic control model for managing the bus bunching problem with capacity constraints. *Transportation Planning and Technology, 40*(6), 722–740.

94. Asgharzadeh, M., & Shafahi, Y. (2017). Real-time bus-holding control strategy to reduce passenger waiting time. *Transportation Research Record, 2647*, 9–16.

95. Berrebi, S. J., Hans, E., Chiabaut, N., Laval, J. A., Leclercq, L., & Watkins, K. E. (2018). Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C: Emerging Technologies, 87*, 197–211.

96. Berrebi, S. J., Crudden, S. Ó., & Watkins, K. E. (2018). Translating research to practice: Implementing real-time control on high-frequency transit routes. *Transportation Research Part A: Policy and Practice, 111*, 213–226.

97. Fan, W., & Machemehl, R. (2011). Bi-level optimization model for public transportationnetwork redesign problem: Accounting for equity issues. *Transportation Research Record, 2263*, 151–162.

98. Chen, B., & Cheng, H. H. (2010). A review of the applications of agent technology in traffic and transportation systems. *IEEE Transactions on Intelligent Transportation Systems, 11*(2), 485–497.

99. Barry, J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in new York City with automated fare system data. *Transportation Research Record, 1817*, 183–187.

100. Wang, W., Attanucci, J., & Wilson, N. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation, 14*(4), 131–150. https://doi.org/10.5038/2375-0901.14.4.7.

101. Zhao, J., Rahbee, A., & Wilson, N. H. M. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering, 22*(5), 376–387. https://doi.org/10.1111/j.1467-8667.2007.00494.x.

102. Seaborn, C., Attanucci, J., & Wilson, N. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation research record, 2121*, 55–62.

103. Sánchez-Martínez, G. E. (2017). Estimating Fare Noninteraction and Evasion with Disaggregate Fare Transaction Data. *Transportation Research Record, 2652*, 98–105.

104. Ectors, W., Reumers, S., Lee, W. D., Choi, K., Kochan, B., Janssens, D., Bellemans, T., & Wets, G. (2017). Developing an optimised activity type annotation method based on classification accuracy and entropy indices. *Transportmetrica A: Transport Science, 13*(8), 742–766.

105. Devillaine, F., Munizaga, M., & Trépanier, M. (2012). Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record, 2276*, 48–55.

106. Ghiani, G., Guerriero, F., Laporte, G., & Musmanno, R. (2003). Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research, 151*(1), 1–11.

107. Okunieff, P. E. (1997). *AVL systems for bus transit: A synthesis of transit practice.*

108. Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies, 44*, 70–79.

109. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Jr., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies, 26*, 301–313.

110. Wang, F., & Chen, C. (2018). On data processing required to derive mobility patterns from passively-generated mobile phone data. *Transportation Research Part C: Emerging Technologies, 87*, 58–74.

111. Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies, 58*, 240–250.

112. Ma, J., Li, H., Yuan, F., & Bauer, T. (2013). Deriving operational origin-destination matrices from large scale mobile phone data. *International Journal of Transportation Science and Technology, 2*(3), 183–204.

113. Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., & Yang, L. (2016). Big data for social transportation. *IEEE Transactions on Intelligent Transportation Systems, 17*(3), 620–630.

114. Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies, 40*, 63–74.

115. Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies, 38*, 122–135.

116. Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies, 58*, 162–177.

117. Kelen, C., Vilarino, P., & Christou, G. (2017). Advanced demand data collection technologies for multi modal strategic modelling. *Transportation Research Procedia, 27*, 1058–1065.

118. Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies, 68*, 285–299.

119. Ge, Q., & Fukuda, D. (2016). Updating origin–destination matrices with aggregated data of GPS traces. *Transportation Research Part C: Emerging Technologies, 69*, 291–312.

120. Pinelli, F., Nair, R., Calabrese, F., Berlingerio, M., Di Lorenzo, G., & Sbodio, M. L. (2016). Data-driven transit network design from mobile phone trajectories. *IEEE Transactions on Intelligent Transportation Systems, 17*(6), 1724–1733.

121. Berlingerio, M., Calabrese, F., Di Lorenzo, G., Nair, R., Pinelli, F., & Sbodio, M. L. (2013). AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 663–666). Berlin Heidelberg: Springer.

122. Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society, 11*, 141–155.

123. Pereira, F. C., Rodrigues, F., & Ben-Akiva, M. (2015). Using data from the web to predict public transport arrivals under special events scenarios. *Journal of Intelligent Transportation Systems, 19*(3), 273–288.

124. Othman, N. B., Legara, E. F., Selvam, V., & Monterola, C. (2015). A data-driven agent-based model of congestion and scaling dynamics of rapid transit systems. *Journal of Computer Science, 10*, 338–350.