Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 27 (2017) 1179–1186

Transportation Research Procedia

www.elsevier.com/locate/procedia

20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary

Identifying Irregularity Sources by Automated Location Vehicle Data

Sara Mozzoni[a], Roberto Murru[a] and Benedetto Barabino[a]*

[a] *Department of Studies and Researches, CTM SpA, viale Trieste 159/3, Cagliari 09123, Italy*

Abstract

Irregularity is unavoidable in high frequency transit services due to the stochastic environment where bus services are operated. Therefore, identifying irregularity sources provides an opportunity to maintain planned headways. Previous research examined the irregularity sources by using scheduled and actual arrival (or departure) times at bus stops. However, as far as the authors' know, no studies analyzed the irregularity sources by comparing arrivals and departures headways between two consecutive bus stops. This analysis is relevant when buses run with short headways and it is difficult to maintain the planned timetable. This gap is addressed by an offline framework which characterizes the regularity over all bus stops and time periods and discloses systematic irregularity sources from collected Automated Vehicle Location (AVL) data by inferring information on headways only. Moreover, this framework selects preventive strategies, accordingly. This framework is tested on the real case study of a bus route, using about 15,000 AVL data records provided by the bus operator, CTM in Cagliari (Italy), whose vehicles are all equipped with AVL technologies. The experimentation shows that transit managers could adopt this framework for accurate regularity analysis and service revision.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: Automated Vehicle Location data, Automated Data Collection methods, Regularity measure, Irregularity sources.

Nomenclature

N	Set of all bus stops (including terminals)

---

*Corresponding author. Tel.: +39 070-209-1249; fax: +39-070-209-1222.
E-mail address: bbarabino@gmail.com

2352-1465
Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.
10.1016/j.trpro.2017.12.116

$A \subseteq N$	Set of terminals
J	Set of runs
Le	Set of legs
n	Bus stop index
j	Run index
$HDR_{n,(j,j-1)}$	Real departure headway at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
$HDS_{n,(j,j-1)}$	Scheduled departure headway at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
$HD_{n,(j,j-1)}$	Departure headway deviation at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
RAT_{ij}	Real arrival time at bus stop $n \in N$ for run $j \in J$
SDT_{ij}	Scheduled departure time at bus stop $n \in N$ for run $j \in J$
ART_{aj}	Available recovery time at terminal $a \in A$ for run $j \in J$
$HAR_{n,(j,j-1)}$	Real arrival headway at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
$HAS_{n,(j,j-1)}$	Scheduled arrival headway at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
$HA_{n,(j,j-1)}$	Arrival headway deviation at bus stop $n \in N$ between runs $j \in J$ and $j-1 \in J$
$HTSR_{n,(j,j-1)}$	Real time spent headway at bus stop $\mathrm{n} \in \mathrm{N \backslash A}$ between runs $j \in J$ and $\mathrm{j-1} \in \mathrm{J}$
$HTSS_{n,(j,j-1)}$	Scheduled time spent headway at bus stop $n \in N \backslash A$ between runs $j \in J$ and $j-1 \in J$
$HTS_{n,(j,j-1)}$	Headway time spent deviation at bus stop $n \in N \backslash A$ between runs $j \in J$ and $j-1 \in J$
$rrt_{(n-1,n)j}$	Real running time between stops $n-1 \in N$ and $n \in N$ for run $j \in J$
$srt_{(n-1,n)j}$	Scheduled running time between stops $n-1 \in N$ and $n \in N$ for run $j \in J$
$rs_{(n-1,n)j}$	Real speed between stops $n-1 \in N$ and $n \in N$ for run $j \in J$
$sms_{(n-1,n)j}$	Scheduled mean speed between stops $n-1 \in N$ and $n \in N$ for run $j \in J$
L^{HD}, L^{HA}, L^{HTS}	Minimum acceptable value of HD, HA, HTS, respectively
U^{HD}, U^{HA}, U^{HTS}	Maximum acceptable value of HD, HA, HTS, respectively
η	Congestion speed
ρ	Numerical threshold < 1
τ	Numerical threshold > 1
μ	Urban speed limit

1. Introduction

In high frequency transit systems, characterizing the regularity and understanding possible systematic irregularity sources is extremely desirable and provides an opportunity to keep buses running as planned. On the one hand, the analysis of irregularity has captured attention as theoretical models have investigated the impact of passengers on headway regularity along the route (*e.g.*, Bellei and Gkoumas, 2010; Islam *et al.,* 2014). However, transit operators do not eagerly adopt models since models cannot incorporate practical and operational considerations arising in a specific case study (*e.g.*, rules negotiated with trade unions). On the other hand, measuring regularity is technologically feasible by Automatic Vehicle Location (AVL) systems, which can collect abundance of disaggregated data on the delivered service (Moreira-Matias *et al.*, 2015). Many studies have so far used AVL data to analyze regularity by using inferential (*e.g.*, Diab *et al.*, 2015; Strathman *et al.*,2003) or descriptive models (*e.g.*, Barabino *et al.*, 2013a, 2013b; Feng and Figliozzi, 2011, 2015; Hammerle *et al.*, 2005; Kimpel *et al.*, 2008; Lin and Ruan, 2009; Lin *et al.,* 2008) to provide quantitative monitoring of service. However, a gap exists in the identification of systematic irregularity sources and their link to strategies. In fact, inferential models are suitable for long-term analysis but they do not provide details on where and when unreliability problems occur. Conversely, descriptive models provided a clear picture on the analysis of irregularity causes but they are not general, owing to the lack of an overall framework able to analyze the most systematic irregularity sources.

Building on these studies, this paper investigates descriptive models and further develops methods to identify and analyze the spatial-temporal sources of low regularity in order to help bus operators select possible preventive strategies, since AVL archived data are used. More in detail, this paper proposes a framework which outperforms an existing structure for the analysis of time reliability (Barabino *et al.*, 2017a), because it proposes a diagnosis of regularity for any high frequency route from archived AVL data. This framework: 1) characterizes bus stops and time periods where regularity is not met; 2) discovers the irregularity sources starting from irregularity effects and originating causes, and selects quantification-based strategies which may be applied successfully to the case at hand. Although part 1) uses algorithms successfully applied in Barabino *et al.* (2013a), (2013b), (2017a) and (2017b) for the reliability characterization, Part 2) advances data analysis procedures over all bus stops and time periods, whereas several studies focused on a limited subset of bus stops and time periods (*e.g.*, Lin *et al.*, 2008; Lin and Ruan, 2009, and Horbury, 1999). It differs from Feng and Figliozzi (2015) and Hammerle *et al.* (2005) as this framework provides a quantitative analysis of data instead of time-space trajectory graphs, and from Feng and Figliozzi (2011) as the proposed framework provides analysis of bunching and large gaps together. Moreover, Part 2) removes the choice of finding irregularity sources by quantifying scheduled and actual times as if the actual service adheres to the scheduled time, it also adheres to the scheduled headway. This *modus operandi* may be tricky in high-frequency routes when buses run with short headways and it is problematic to maintain the planned timetable, which is uninteresting for passengers. In fact, in high frequency routes, passengers are supposed to arrive at bus stops independently from published timetables and actual vehicle arrivals, and are usually supposed to board on the first arriving vehicle. As passengers aim to minimize their waiting time and the waiting time depends on headways, it may be more coherent to analyze the irregularity causes and sources by comparing arrivals and departures headways between two consecutive bus stops, unlike in Barabino *et al.* (2017a). To the best of our knowledge, such a detailed analysis of comparison of headways at two consecutive stops has never been done to detect irregularity sources. Moreover, control dashboards present the outcomes by tables organized in time and space attributes. This paper is organized as follows. Section 2 proposes a framework to analyze regularity and detect irregularity sources. Section 3 illustrates its experimentation on a real case study. Section 4 draws conclusions and future research.

2. Methodological Approach

The proposed framework is organized in two parts. Part 1) characterizes the regularity and Part 2) identifies the magnitude of irregularity problem sources (which can be clustered in Improper Service Design (ISD), Drivers and/or Supervisors Failures (D&SF), Uncertainties in Passengers Volumes (UPV) and Uncontrollable External Factors (UEF) according to Ceder (2007), and selects suitable strategies for their mitigation.

2.1. The characterization of irregularity

Consider a high frequency route. First, its AVL data are collected from a database on the provided service. The main data provided by AVL system include date, route, trip number, bus stop code and order, actual and scheduled transit times and finally, the time spent in a pre-defined area around each bus stop, or the dwell time, depending on the specific AVL architecture. Second, since AVL data are not ready-for-use as are, the framework performs proper handling on data anomalies to account for Bus Overtakings (BO) and distinguishes between missing data points - *i.e.*, Technical Failures (TF) or Incorrect Operation in the Service (IOS) - according to Barabino *et al.* (2013a), (2013b), (2017a), (2017b) and (in press). Third, the actual and the scheduled headways are computed as the difference between two consecutive bus arrival (or departure) times. The actual headways need different analysis, depending on the occurrence of TF and IOS. TF cannot be used to compute real headways owing to the missing information on transit occurring for real. Conversely, actual headways are computed in the case of IOS because passengers suffer these headways for real. Four, according to Kittelson & Associates *et al.* (2013), the regularity measure is obtained by the coefficient of variation of headway (C_{vh}), which is computed as the ratio between the standard deviation of the differences between actual and scheduled headway and the average of scheduled headway. The C_{vh} is also calculated for all bus stops and time periods and is matched with a LoS in order to show which segments of the route do not attain a sufficient regularity level. Fifth, if LoSs report a sufficient mark denoted by A, B or C, the service is considered as acceptable and no further analysis is required. Conversely, if LoSs report an insufficient mark denoted by D, E or F, the service needs further investigation to understand the possible irregularity source. Sixth, AVL processed data are represented effectively by control dashboards organized in space and time attributes to show which time periods and bus stops of the route contain most of the problems and deserve further analyses.

2.2. The identification of irregularity sources and the selection of possible strategies

2.2.1. Irregularities sources

Let A be the set of terminals and J the set of runs. For each terminal $a \in A$, the departures of two consecutive runs $j \in J$ and $j\text{-}1 \in J$ are considered in case of run beginning. Once the real and the scheduled departure headways are collected, since they may be different, one may compute their deviation as follows:

$$HD_{a(j,\,j\text{-}1)} = HDR_{\,a(j,\,j\text{-}1)} - HDS_{\,a(j,\,j\text{-}1)} \qquad \forall a \in N \backslash A,\ \forall j \in J \qquad (1)$$

If $HD_{\,a(j,\,j\text{-}1)} \approx 0$, the real headway complies with the scheduled headway. If $HD_{\,a(j,j\text{-}1)} \neq 0$, the analysis of the recovery time at the terminal is performed. Particularly, if $HD_{a(j,j\text{-}1)} > 0$, buses run with a long gap in the service, whereas if $HD_{\,a(j,j\text{-}1)} < 0$, buses run with a short gap (bunching) in the service. Since long and short gaps represent irregularity problems, in what follows, we focus on their analysis. In order to begin the next trip as scheduled and give operators a short break, drivers are provided with recovery times, but actual recovery times may be different from scheduled ones. Therefore, it is important to derive the available recovery time which is computed as:

$$ART_{aj} = SDT_{aj} - RAT_{a(j\text{-}1)} \qquad \forall a \in N \backslash A,\ \forall j \in J \qquad (2)$$

The analysis of the headway deviation at the terminal and that of the available recovery time are crucial for detecting irregularity sources, because they may result in additional irregularity in the next run. Their combined analysis helps understand if irregularity mainly depends on ISD or D&SF.

Nine different cases can be obtained, as shown in Table 1, according to the sign of $HD_{\,a(j,\,j\text{-}1)}$ and ART_{aj}.. For each entry in Table 1, one can compute the related magnitude in terms of percentage values. The notation ≈ 0 must be read as:

$$L^{HD} \leq HD_{a(j,\,j\text{-}1)} \leq U^{HD},\ L^{ART} \leq ART_{aj} \leq U^{ART} \qquad \forall a \in N \backslash A,\ \forall j \in J \qquad (3)$$

Table 1 shows that some critical combinations may occur if:

- $HD_{a(j,\,j\text{-}1)} < 0$ or $HD_{a(j,\,j\text{-}1)} > 0$ and $ART_{aj} < 0$, drivers do not have the available recovery time for the new run, then they start runs which fail to comply with scheduled headway. The problem may be ISD. At this stage, the traffic is not supposed to be an irregularity source for the new run since it already occurred in the previous one.
- $HD_{a(j,\,j\text{-}1)} < 0$ or $HD_{a(j,\,j\text{-}1)} > 0$ and $ART_{aj} > 0$, drivers have sufficient recovery time to start the new run which complies with scheduled headway, but they start bunched or overly spaced out. Therefore, in this case, the problem may depend on driver behaviour or the lack of public transport company supervision (D&SF).
- $HD_{a(j,\,j\text{-}1)} < 0$ or $HD_{a(j,\,j\text{-}1)} > 0$ and $ART_{aj} \approx 0$, drivers have little recovery time, so they may start their trips bunched or with a large gap if they want to have a break. However, in this case, there may also be a problem in ISD and/or D&SF. Besides, if it distinguishes better between ISD and D&SF, the following rules can be adopted. If $L^{ART} < ART_{aj} < 0$, the problem source is ISD, whereas if $0 < ART_{aj} < U^{ART}$, D&SF is detected.

Table 1 – Possible irregularity sources at the terminal

ART_{aj}	$HD_{a\,(j,j\text{-}1)}$ < 0	≈ 0	> 0
< 0	ISD	Ok	ISD
≈ 0	ISD or D&SF	Ok	ISD or D&SF
> 0	D&SF	Ok	D&SF

2.2.2. Irregularity sources en-route

Let N be the set of all bus stops (including terminals) and J the set of runs. In the case of bus stop $n \in N \backslash A$, a run $j \in J$ can arrive (depart) bunched, regular or with a long gap with the previous one. In order to understand the causes of irregularity, for each pair of consecutive bus stops n-1 $\in N \backslash A$ and $n \in N \backslash A$, and each pair of consecutive runs $j \in J$ and $j-1 \in J$, one derives the scheduled and the real headway, both in departure and arrival. Then, the headway deviations are calculated as follows:

$$HA_{n\,(j,\,j\text{-}1)} = HAR_{\,n\,(j,\,j\text{-}1)} - HAS_{\,n\,(j,\,j\text{-}1)} \qquad \forall n \in N \backslash A,\ \forall j \in J \qquad (4)$$

$$HD_{(n\text{-}1),(j,\,j\text{-}1)} = HDR_{(n\text{-}1),(j,\,j\text{-}1)} - HDS_{(n\text{-}1),(j,\,j\text{-}1)} \qquad \forall n \in N \backslash A,\ \forall j \in J \qquad (5)$$

As both bunching and large gaps in arrival (or departure) at each bus stop $n \in N \backslash A$ n represent irregularity problems, in what follows, we focus on their analysis which involves the selected bus stop, $n \in N \backslash A$ and the previous one, n-1 $\in N \backslash A$. More in detail, one computes $HA_{n\,(j,\,j\text{-}1)}$ and $HD_{(n\text{-}1),(j,\,j\text{-}1)}$ according to eqns. (4) and (5), respectively, and analyse their values. Besides, if $HA_{n\,(j,\,j\text{-}1)} \approx 0$, any further analysis is not required as the regularity is assured (*i.e.*, regular headways are kept at the bus stop $n \in N \backslash A$). Conversely, if $HA_{n\,(j,\,j\text{-}1)} \neq 0$, a combined analysis of $HD_{(n\text{-}1),(j,\,j\text{-}1)}$ at bus stop

n-1 $\in N\backslash A$, and the HA_n at bus stop n-$\in N\backslash A$ is performed, and six situations may occur according to their signs. In these situations, the ">"means that the real headway is amplified with respect to the scheduled one, that is, the arriving (or departing) buses are moving away from each other. The "<"means that the real headway is lesser than the scheduled one, that is, the arriving (or departing) buses are bunched. The notation ≈ 0 must be read as:

$$L^{HA} \le HA_{n(j,\,j\text{-}1)} \le U^{HA};\ L^{HD} \le HD_{(n\text{-}1),(j,\,j\text{-}1)} \le U^{HD} \quad \forall n \in N\backslash A,\ \forall j \in J \tag{6}$$

Before analyzing these possible combinations, the proposed method performs an analysis of real and scheduled headways time spent and its deviations at bus stop n-1 $\in N\backslash A$, to understand if the irregular departure may depend on passenger volumes. The dwell time can provide information on the volumes of boarding and alighting passengers, but the time spent may be related to passenger volumes in the case of less advanced transit operators that do not automatically count passengers or when AVL architectures are conceived to record the time spent in the proximity of bus stops instead of the dwell time. In detail, once the real and the scheduled time spent headways are collected, one may compute their difference as follows:

$$HTS_{(n\text{-}1),(j,\,j\text{-}1)} = HTSR_{(n\text{-}1),(j,\,j\text{-}1)} - HTSS_{(n\text{-}1),(j,\,j\text{-}1)} \qquad \forall n \in N\backslash A,\ \forall j \in J \tag{7}$$

Three cases can occur according to the sign of $HTS_{(n\text{-}1),(j,\,j\text{-}1)}$. If $HTS_{(n\text{-}1),(j,\,j\text{-}1)} \approx 0$, the time spent at bus stop $n \in N\backslash A$ does not affect the irregular departure. The notation ≈ 0 must be read as follows:

$$L^{HTS} \le HTS_{(n\text{-}1),(j,\,j\text{-}1)} \le U^{HTS} \qquad \forall n \in N\backslash A,\ \forall j \in J \tag{8}$$

Conversely, if $\text{HTS}_{(n\text{-}1),(j,\,j\text{-}1)} < 0$ or $\text{HTS}_{(n\text{-}1),(j,\,j\text{-}1)} > 0$, then the irregular departure could depend on passengers as the volume of boarding and alighting is probably lower or greater than the expected one and thus, UPV may occur. The impact of passengers could optionally be observed by a correlation analysis between the time spent at bus stop n-1 $\in N$ and the lateness at bus stop $n \in N$, as a large headway of a vehicle often results in a short headway in the following. In order to understand the possible sources that generate irregularities along the route, one can examine together, headway arrival deviations at bus stop $n \in N\backslash A$, headway departure deviations at bus stop n-1 $\in N\backslash A$ and headway time spent deviations at bus stop n-1 $\in N\backslash A$, and combine all these cases. Table 2 shows these 18 critical cases that may occur. The different cases can be clustered according to common characteristics related to the sign of all deviations and problem sources. The magnitude of clustered problems is expressed in terms of percentage values. In what follows, the discussion of these cases is provided.

Table 2: Possible irregularity sources along the route

	$HA_{n\,(j,\,j\text{-}1)} < 0$			$HA_{n\,(j,\,j\text{-}1)} > 0$		
$HTS_{(n\text{-}1)\,(j,\,j\text{-}1)}$ / $HD_{(n\text{-}1)\,(j,\,j\text{-}1)}$	< 0	≈ 0	> 0	< 0	≈ 0	> 0
< 0	UPV	D&SF	D&SF	D&SF or ISD or UEF	D&SF or ISD or UEF	D&SF or ISD or UEF
≈ 0	D&SF	D&SF	D&SF	D&SF or ISD or UEF	D&SF or ISD or UEF	D&SF or ISD or UEF
> 0	D&SF	D&SF	D&SF	D&SF or ISD or UEF	D&SF or ISD or UEF	UPV

More in detail:

- if $HA_{n(j,\,j\text{-}1)} < 0$ (*i.e.*, buses arrive bunched to bus stop $n \in N\backslash A$) and $HD_{(n\text{-}1),(j,\,j\text{-}1)} > 0$ (*i.e.*, buses depart with gap from bus stop n-1 $\in N\backslash A$) or if $HA_{n(j,\,j\text{-}1)} < 0$ and $HD_{(n\text{-}1),(j,\,j\text{-}1)} \approx 0$ (*i.e.*, buses depart regularly from bus stop n-1 $\in N\backslash A$), the problem sources may depend on D&SF. Indeed, it is supposed that $HTS_{(n\text{-}1),(j,\,j\text{-}1)}$ at bus stop n-1 $\in N\backslash A$ does not affect the regularity. Thus, the irregularity source can be most likely due to the operator's driving style that may be too "sporty" or lack of supervision from coordinators.
- if $HA_{n(j,\,j\text{-}1)} < 0$ and $HD_{(n\text{-}1),(j,\,j\text{-}1)} < 0$ (*i.e.*, buses depart bunched from bus stop n-1 $\in N\backslash A$), the problem sources may also depend on D&SF. Indeed, the irregularity is constant along the route between bus stops n-1 $\in N\backslash A$ and $n \in N\backslash A$. Thus, it is likely that the scheduled travel time is suitable. Besides, there are three possible cases; where $HTS_{(n\text{-}1),(j,\,j\text{-}1)}$ is larger, lesser or $\approx$ than the expected time, thus a further analysis follows. If $HTS_{(n\text{-}1),(j,\,j\text{-}1)} > 0$, the following bus stayed at bus stop longer than the previous, however buses have been bunched; therefore, the problem may be also related to D&SF. If $HTS_{(n\text{-}1),(j,\,j\text{-}1)} \approx 0$, the time spent is regular but the buses are bunched, and also in this case, the problem may be due to D&SF. Conversely, if $HTS_{(n\text{-}1),(j,\,j\text{-}1)} < 0$, means that the following bus stayed at bus stop lesser than the expected time, thus buses depart bunched. In this case, the irregularity source may also depend on UPV, as few passengers are supposed to be at bus stop n-1 $\in N\backslash A$.

- if ***a)*** $HA_{n(j,j-1)} > 0$ (*i.e.*, buses arrive with a gap to bus stop $n \in N \backslash A$), and $HD_{(n-1),(j,j-1)} < 0$ (*i.e.*, buses depart bunched from bus stop n-1 $\in N \backslash A$) or if ***b)*** $HA_{n(j,j-1)} > 0$ and $HD_{(n-1),(j,j-1)} \approx 0$ (*i.e.*, buses depart regularly from bus stop n-1 $\in N \backslash A$) or ***c)*** if $HA_{n(j,j-1)} > 0$ and $HD_{(n-1),(j,j-1)} > 0$ (*i.e.*, buses depart with gap from bus stop n-1 $\in N \backslash A$), the problem sources may depend on D&SF, ISD, UEF or UPV. More in detail, for cases ***a)*** and ***b)***, $HTS_{(n-1),(j,j-1)}$ at bus stop n-1 $\in N \backslash A$ does not affect the regularity. Indeed, despite departures being bunched or regular (*i.e.*, $HD_{(n-1),(j,j-1)} \approx 0$), buses arrive at bus stop $n \in N \backslash A$ with a large gap. In both these cases, the irregularity source may be due to D&SF, ISD or UEF. Therefore, the irregularity source cannot be identified by only this analysis, thus a further refinement is suggested. Differently, for case ***c)***, $HTS_{(n-1),(j,j-1)}$ at bus stop n-1 $\in N \backslash A$ can affect the regularity in one case. Indeed, if $HTS_{(n-1),(j,j-1)}>0$, means that the following bus stayed at bus stop greater than the expected time, thus buses may depart with a long gap. In this case, the irregularity source may depend on UPV, as many passengers are supposed to be at bus stop n-1 $\in N \backslash A$. In the remaining cases, if $HTS_{(n-1),(j,j-1)} <0$ (*i.e.*, the following bus stayed at bus stop lesser than the previous one) or if $HTS_{(n-1),(j,j-1)} \approx 0$, buses are not unable to fill the gap which is maintained along the route. Therefore, as in cases ***a)*** and ***b)***, the irregularity source may be due to driver's behavior or wrong scheduled travel time, and/or scheduled time spent or uncontrollable external factors (*i.e.*, D&SF, ISD or UEF). When the irregularity source is identified as ambiguous (*i.e.*, it presents a combination of possible sources), an analysis of the speed between consecutive bus stops is suggested, as the speed can provide information about the running time.

2.2.3. Refinement of irregularity sources en-route by speed analysis

In order to distinguish among ambiguous irregularity sources, the proposed method performs an analysis of the speed along the leg $le \in Le$ from n-1 $\in N$ to $n \in N$. Next, the method tries to understand if problems along the leg $le \in Le$ from n-1 $\in N$ to $n \in N$ depend on ISD, D&SF or UEF. The magnitude of these causes is evaluated by percentage values. Two different speeds are considered: the real speed between stops n-1 $\in N$ to $n \in N$ for each run $j \in J$; and the scheduled mean speed between bus stops n-1 $\in N$ to $n \in N$. For a fixed time period t, they are computed as follows:

$$rs_{(n-1,n)j} = \frac{le_{n-1,n}}{rrt_{(n-1,n)j}} \qquad \forall n \in N,\ \forall j \in J \tag{9}$$

$$sms_{n-1,n} = \frac{le_{n-1,n}}{\frac{\sum_{j=1}^{M} srt_{(n-1,n)j}}{M}} \qquad \forall n \in N,\ \forall j \in J \tag{10}$$

Some problems can be detected by the value of $rs_{(n-1,n)j}$ and appropriate threshold parameters $\rho < 1$ and $\tau > 1$. More precisely if: **a)** $rs_{(n-1,n)j} \leq \eta$ (*i.e.*, the minimum acceptable speed), the unreliability source is probably UEF; **b)** $\eta < rs_{(n-1,n)j} \leq \rho$, buses run beyond the minimum acceptable speed but they cannot reach the planned speed. The unreliability source is probably ISD; **c)** $\rho * sms_{n-1,n} < rs_{(n-1,n)j} \leq \tau * sms_{n-1,n}$, there is no problem disclosed by the speed analysis because the real speed is close to the planned one; **d)** $rs_{(n-1,n)j} > \tau * sms_{n-1,n}$ or larger than the urban speed limit μ, buses run beyond the acceptable speed. This slack may depend on the too "sporty" driving style, thus the unreliability source is probably D&SF.

2.2.4. Possible preventive strategies

Since the framework runs offline, we focus only on preventive strategies which can be divided into priority and operational ones, according to Barabino *et al.* (2017a). A possible link between irregularity sources and strategies shows that if the irregularity source is: 1) D&SF, possible strategies are operator training, incentives and penalties, as well as supervision; 2) ISD, possible strategies are schedule adjustments (*e.g.*, running time, recovery time); 3) UPV, possible strategies are schedule adjustments (*e.g.,* dwell time and/or time spent) and/or improving vehicle access; 4) UEF, possible strategies are exclusive lanes, route re-design and signal priority.

3. Application in a Real Case

The experiment was conducted on the major bus operator (*i.e.*, CTM) from Cagliari, a coastal Italian city with 0.4 M inhabitants. For the sake of synthesis, the proposed method is tested on the westbound direction of a route about 8 km long with 6 bus stops, which links the university and a hospital center with the city. The route has been chosen because of these characteristics which are supposed to point out different problem sources. Its headway is 10 minutes from 7.00 to 9.59 and from 12.00 to 14.59, thus the route is evaluated in terms of regularity only in these time periods. This route-direction can be divided into two parts. A fast flow exurban road characterizes the former (bus stops 1 and 2); the bus can travel at a good commercial speed without interferences with vehicle looking for parking or pedestrian flows. Two-way streets characterize the latter within the city in mixed-traffic conditions. The AVL data were collected during weekdays of March 2016. In this analysis, 15,034 AVL data were collected from 07.00 to 21.59, but we focus only from 7.00 to 9.59 and from 12.00 to 14.59, when the route operates with high frequency. At the end of Part 1), we processed 3,416 AVL data and the outcomes are represented in the control dashboard of Fig. 1, where rows represent bus stops, and columns represent time periods. Each entry represents the LoS at that bus stop and at that time period, according to the values obtained by the C_{vh} (Kittelson & Assoc, Inc *et al.*, 2013).

Westboud direction

Part	Bus stop	7.00 7.59	8.00 8.59	9.00 9.59	12.00 12.59	13.00 13.59	14.00 14.59
1	1	E	E	D	C	D	D
	2	F	E	D	C	C	C
2	3	F	E	E	F	E	F
	4	E	E	E	F	F	F
	5	E	E	E	F	F	F
	6	E	E	E	E	F	E

Fig. 1 - The characterization of regularity LoS – Part 1

For example, LoSs from A to C means that regularity can be considered satisfied ($C_{vh} \leq 0.40$). Fig. 1 shows that the regularity is critical in all time periods, particularly from 12.00 to 14.59 where LoS E and LoS F are frequently observed. Moreover, in this route direction, the analysis shows that vehicles run bunched in Part 2 (from bus stops 3 to 6), when they operate in the city. As LoSs E and F are consistently observed, along the route, the experimentation of Part 2) was carried out considering all time periods. The outcomes are shown in Fig. 2 where a dashboard includes both at the terminal and en-route analysis for the selected route-direction.

				TERMINAL AND EN- ROUTE ANALYSIS						REFINEMENT WITH SPEED ANALYSIS						
				Time period							Time period					
Part	Bus stop	Leg Code	Source	7.00 7.59	8.00 8.59	9.00 9.59	12.00 12.59	13.00 13.59	14.00 14.59	Source	7.00 7.59	8.00 8.59	9.00 9.59	12.00 12.59	13.00 13.59	14.00 14.59
1	1		OK	43.90%	56.38%	34.04%	78.41%	67.37%	17.95%							
			D&SF	14.64%	4.26%	41.49%	19.32%	23.16%	**80.77%**							
			ISD	41.46%	39.36%	24.47%	2.27%	9.47%	1.28%							
		T1R	OK	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	OK	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			D&SF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	D&SF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			ISD	-	-	-	-	-	-	ISD	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			UPV	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	UPV	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			ISD/D&SF/UEF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	UEF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
	2															
		T2R	OK	8.33%	12.50%	13.04%	20.41%	22.33%	13.33%	OK	8.33%	16.07%	13.04%	20.41%	25.24%	13.33%
			D&SF	**50.00%**	**57.29%**	**46.74%**	**46.94%**	**44.66%**	**53.33%**	D&SF	**70.83%**	**62.05%**	**63.22%**	**60.20%**	**52.43%**	**65.52%**
			ISD	-	-	-	-	-	-	ISD	4.17%	3.57%	2.55%	2.04%	4.85%	4.06%
			UPV	4.17%	5.21%	9.79%	2.04%	1.94%	2.86%	UPV	4.17%	5.21%	9.79%	2.04%	1.94%	2.86%
			ISD/D&SF/UEF	37.50%	25.00%	30.43%	30.61%	31.07%	30.48%	UEF	12.50%	13.10%	11.40%	15.31%	15.54%	14.23%
2	3															
		T3R	OK	10.81%	11.11%	12.35%	15.11%	14.77%	3.30%	OK	10.81%	11.11%	12.35%	15.11%	18.18%	6.87%
			D&SF	**54.05%**	**52.22%**	**49.38%**	**45.35%**	**52.27%**	**43.96%**	D&SF	**54.05%**	**52.22%**	**49.38%**	**45.35%**	**52.27%**	**43.95%**
			ISD	-	-	-	-	-	-	ISD	24.32%	27.78%	33.33%	31.40%	23.87%	39.29%
			UPV	10.82%	8.89%	4.94%	8.14%	5.68%	9.89%	UPV	10.82%	8.89%	4.94%	8.14%	5.68%	9.89%
			ISD/D&SF/UEF	24.32%	27.78%	33.33%	31.40%	27.27%	42.86%	UEF	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	4															
		T4R	OK	12.00%	25.00%	22.55%	31.37%	18.45%	16.30%	OK	33.33%	42.97%	27.90%	31.37%	23.74%	28.48%
			D&SF	**56.00%**	**40.63%**	**39.22%**	**37.25%**	**33.01%**	**43.48%**	D&SF	**66.67%**	**46.62%**	**57.93%**	**37.25%**	**38.31%**	**43.48%**
			ISD	-	-	-	-	-	-	ISD	0.00%	0.00%	5.35%	16.67%	18.53%	18.26%
			UPV	0.00%	10.41%	8.82%	14.71%	19.42%	9.78%	UPV	0.00%	10.41%	8.82%	14.71%	19.42%	9.78%
			ISD/D&SF/UEF	32.00%	23.96%	29.41%	16.67%	29.13%	30.43%	UEF	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	5															
		T5R	OK	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	OK	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			D&SF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	D&SF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			ISD	-	-	-	-	-	-	ISD	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			UPV	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	UPV	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
			ISD/D&SF/UEF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	UEF	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
	6															

Fig. 2. Analysis of irregularity sources at terminal, en-route and speed analysis. Westbound direction – Part 2

In addition, this dashboard shows the results of speed analysis which allow the dissolving of the ambiguities arising from nested cases as shown in the previous section. More in detail, each entry can indicate: a) the relative percentage of occurrences of each irregularities, which are shown with black background when the occurrences are larger than 50% (*i.e.*, they are supposed to be systematic criticalities) and with grey background when they result with the highest relative occurrence; b) the absence of a datum (the symbol *n/a* is used). The analysis at the terminal shows almost regular situations. Besides, the time period 7.00 - 8.59 has a high relative percentage of ISD, even if it is not the prevailing source. Conversely, from 14.00 to 14.59, unambiguous problem sources on D&SF arose. In this case, drivers have sufficient recovery time that would enable them to start the new run and comply with the scheduled headway, but they extend this recovery time (they start the run overly spaced out) or decrease it excessively (they start the run bunched). Therefore, the irregularity problem depends on driver behaviour or on the lack of supervision. The en-route analysis shows that the main problem sources are D&SF in all times and in all legs. The refinement through speed analysis would only be necessary in cases where the percentages of the prevailing sources are lesser than 50%, but it is shown as an example in all time periods and legs. For example, focusing on leg T4R, from 9.00 to 9.59, the first analysis detects a 29.41% of possible ambiguous sources (*i.e.*, ISD/D&SF/UEF). Through speed analysis, this percentage is distributed among D&SF, ISD and UEF and identifies operator's driving style as the main source. Thus, from bus stop 3 to 5, buses run beyond the planned speed. Although UPV is not the most frequent irregularity source in this route, we also carried out a correlation analysis between the time spent at the considered bus stop and the lateness at the following bus stop. As expected, a significant correlation was observed, even if the coefficient of correlation did not take high values. The recommended strategy could be to set up operator training and/or to check their behaviour better by AVL supervisors.

4. Conclusions and Future Work

This paper improved the state of the art for any high frequency route by a new offline framework. The main contributions of this paper are the categorization and analysis of headway irregularity at the terminal and at pairs of consecutive stops. The framework shows where and when irregularity occurs by accurately processing AVL data, discloses who is responsible for the irregularity starting from what irregularity effects are and why irregularity occurs. This framework is tested on a real route using about 15,000 AVL data records provided by the bus operator CTM in Cagliari (Italy). This new framework results in significant time and energy savings in the investigation of large datasets. Control dashboards show clear and synthetic outcomes from the analysis of irregularities. The tuning of some thresholds and the validation of the framework will enable the evaluation of specific strategies and testing of outcomes by the support of several bus operator departments.

References

Barabino B, Di Francesco M., Mozzoni S. (2017a). An Offline Framework for the Diagnosis of Time Reliability by Automatic Vehicle Location Data. IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 583-594.

Barabino B., Lai C., Casari C., Demontis R., Mozzoni S. (2017b). Rethinking Transit Time Reliability by Integrating Automated Vehicle Location Data, Passenger Patterns, and Web Tools. IEEE Transactions on Intelligent Transportation Systems, vol.18, no. 4, pp. 756-766.

Barabino B., Di Francesco M., Mozzoni S. (in press). Time reliability measures in bus transport services from the accurate use of Automatic Vehicle Location raw data. Quality and Reliability Engineering International, DOI: 10.1002/qre.2073.

Barabino B., M. Di Francesco, S. Mozzoni (2013a). Regularity analysis on bus networks and route directions by Automatic Vehicle Location raw data. IET Intelligent Transport Systems, vol. 7, no. 4, pp. 473-480.

Barabino B., M. Di Francesco, S. Mozzoni (2013b). Regularity diagnosis by Automatic Vehicle Location raw data. Public Transport, vol. 4, no. 3, pp. 187-208.

Bellei G., Gkoumas K. (2010). Transit vehicles' headway distribution and service irregularity. Public Transport vol. 2 n°2, pp. 269-289.

Ceder, A. (2007). Public Transit Planning and Operation. Theory, Modelling and Practice. Oxford, UK: Butterworth-Heinemann, Elsevier, 2007.

Diab E. I., Bertini R., El-Geneidy A. M. (2015). Bus transit service reliability: Understanding the impacts of service overlapping on service operations. Proceeding of Conference on Advanced Systems in Public Transport. Rotterdam 19-23, July, 2015.

Feng W., Figliozzi M. (2015). Empirical Analysis of Bus Bunching Characteristics Based on Bus AVL/APC Data. Transportation Research Board 94th Annual Meeting. No. 15-5138. 2015.

Feng W., Figliozzi M. (2011). Using Archived AVL/APC Bus Data to Identify Spatial-Temporal Causes of Bus Bunching. Compendium of papers of 90th Transportation Research Board, Annual Meeting, Washington, D.C., 2011.

Hammerle, M., M. Haynes, S. McNeil (2005). Use of Automatic Vehicle Location and Passenger Count Data to Evaluate Bus Operations. Transportation Research Record, pp. 27-34.

Horbury, A. X. (1999). Using non-real-time automatic vehicle location data to improve bus services. Transportation Research Part B, vol. 33, no. 8, pp. 559-579.

Kittelson & Assoc, Inc, Parsons Brinckerhoff, Inc, KFH Group, Inc, Texas A&M Transportation Institute, Arup (2013). Transit capacity and quality of service manual. Third Edition TRB, Washington, DC.

Kimpel, T. J., Strathman J. G., and Callas S. (2008). Improving Scheduling Through Performance Monitoring. Computer-Aided Systems in Public Transport. Vol. 600, pp. 253-280.

Islam M.K., Vandebona U., Dixit V.V.and Sharma A. (2014). A Bulk Queue Model for the Evaluation of Impact of Headway variation and Passenger Waiting Behaviour on Public Transit Performance. IEEE Transaction on Intelligent Transportation Systems vol. 15, n°6, pp. 2432-2442.

Moreira-Matias, L., Mendes-Moreira J., de Sousa J.F., Gama J. (2015). Improving Mass Transit Operations by Using AVL-Based Systems: A Survey. IEEE Transaction on Intelligent Transportation Systems vol.16, no.16, pp.1636 - 1653.

Lin J. and Ruan M. (2009). Probability-based bus headway regularity measure. IET Intelligent Transport Systems, vol. 3, no. 4, pp. 400-408.

Lin, J., M.L. Wang, P. D. Barnum (2008). A Quality Control Framework for Bus Schedule Reliability. Transportation Research Part E, vol.44, no 6, pp. 1086-1098.

Strathman J.G., Kimpel T.J., Callas S. (2003). Headway deviation effects on bus passenger loads: Analysis of Tri-Met's archived AVL-APC data. Technical Report PR126, Center for Urban Studies, Portland, OR.