A Web Portal for Reliability Diagnosis of Bus Regularity

Benedetto Barabino[1], Carlino Casari[2], Roberto Demontis[2], Cristian Lai[2], Sara Mozzoni[1], Antonio Pintus[2], and Proto Tilocca[3]

[1] Technomobility s.r.l. - Cagliari - Italy
bbarabino@gmail.com , sara.mozzoni@technomobility.it
[2] CRS4, Center for Advanced Studies, Research and Development in Sardinia - Pula (CA) - Italy
casari@crs4.it, demontis@crs4.it, cristian.lai@crs4.it, pintux@crs4.it
[3] CTM S.p.A. - Cagliari - Italy
proto.tilocca@ctmcagliari.it

Abstract. In high frequency transit services, bus regularity - i.e. the headway adherence between buses at bus stops - can be used as an indication of service quality, in terms of reliability, by both users and transit agencies. The Web portal is the entry point of a Decision Support System (DSS), contains an environment designed for experts in transport domain. The environment is composed of tools developed to automatically handle Automatic Vehicle Location (AVL) raw data for measuring the Level of Service (LoS) of bus regularity at each bus stop and time interval of a transit bus route. The results are represented within easy-to-read control dashboards consisting of tables, charts, and maps.

1 Introduction

Nowadays there is a growing interest in the measurement of public transport service quality, which is a key factor for both users and transit agencies [1]. A relevant element of quality of service is reliability, viewed as the capability of transit operators to meet the expectations raised by the service offer in terms of multidimensional aspects such as time, passenger loads, vehicle quality, and so on [2]. In high frequency bus services, where scheduled headways between buses are 10/12 minutes (e.g. [3], [4], [5], [6]), one of the main aspects of reliability is regularity, which is faced in this paper. High quality evaluation of regularity means working on huge amounts of data, which must be collected and normalized before processing to avoid misleading information. Moreover, for efficient monitoring, it is necessary to be able to process the huge amount of data, present the results in a user-friendly way and guarantee a fast and pervasive access to them.

In this paper we propose: i) the implementation of a methodology to evaluate regularity starting from data collected by Automatic Vehicle Location (AVL); ii) a Web based system specifically designed to support experts in transport engineering domain for evaluating regularity issues.

Currently, AVL technology can collect the raw data for detailed analysis, but its use requires addressing challenges such as missing data points and possible bus overtakings. Moreover, while regular methods of bus operators are thought to operate in data-poor environments, new methods must be developed to exploit the rich-data environments provided by AVL. It is therefore important to develop new methods suitable to handle these data.

Ability to process data and quickly present results is a crucial factor for the efficient management of the service at hand (i.e. decisions based on data). Usually, in transit agencies the executives use spreadsheets to face this challenge. Raw AVL data are first downloaded in a standard PC, separated according to routes. Next, the value of regularity, per route and time period is calculated using formulas. Finally, results of the processed data (route direction, time periods and the value of regularity)are presented in a table. However, these activities are very time- and energy-consuming, because usually performed manually. Besides, results are available only locally. Thus, the need for fast procedures to effectively process AVL raw data, quickly present results and guarantee their access from everywhere. Hence, in order to shed additional light into the diagnosis of service regularity, the authors, making reference to their previous works ([7], [8]), implement in a Web application a method to derive accurate measure of bus regularity. Such method is expected to improve the quality and regularity of transit operators measurements which are too often made at a limited number of check points, on selected routes, and at limited time intervals. One more point, based on these measures, managers will be able to prioritize actions and/or give recommendations to improve the service. Last, but not least, thanks to the possibility to perform fast AVL data processing and thanks to their easy accessibility (as long as a Web connection is available), the workload of transit agencies will be reduced.

This paper is organized as follows. In Section 2, we motivate the choice of the regularity indicator, describe the challenges derived from AVL technologies, and mention a number of Web existing tools. In Section 3, we propose a methodology to evaluate regularity. In Section 4 the Web application and its control dashboards are presented. In Section 5, we present conclusions and research perspectives.

2 State of the Art

In high frequency services, regularity is a major aspect of service, and a classical topic for the transportation community. The major existing studies in the field, including details on the measure of regularity, AVL technology and existing Web tools, are presented in the following three subsections.

2.1 Measure of Regularity

Bus regularity can be measured by several indicators, which present pros and cons and denote the significant lack of a universal metric [5],[9]. The discussion about the several indicators used is not required in this paper, because already presented in [7] and [8]. However, to summarize, we look for an indicator which should satisfy these properties: ease communication (understandable and easy-to-read), objectivity (i.e. without subjective thresholds), customer oriented (penalizing longer waiting times), independence from data distributions and ranking well-established regularity levels. As discussed in [7] and [8], the Headway Adherence (HA) measured by the Coefficient of Variation of Headways (C_{vh}) proposed by [9] is a good indicator fitting the following requirements:

- although the C_{vh} is not of immediate understanding and communication, its values represent LoSs ranked in a well-established scale of regularity from A (the best) to F(the worst);
- it is objective; LoS thresholds are related to the probability that a given transit vehicle's actual headway is off headway by more than one-half of the scheduled headway;
- it is quite customer oriented; every trip is considered in the computation of the C_{vh}, to penalize long waiting times at bus stops. The output indicated the probability of encountering an irregular service, even if it is not a measure of severity of the irregularity;
- it does not require particular applicability conditions. Since bus operators sometimes schedule high frequency services irregularly, it is important to consider different headways in different time intervals;
- it can evaluate different regularity conditions and detect bunching phenomena.

2.2 AVL Technology and Regularity

Due to economic constraints and lack of technology, early experiences in the determination of regularity measures were performed at a few random or selected check points of route (e.g. [4],[6]). Typically, collected data were aggregated manually in time periods representing slack and peak hours in the morning and in the evening. This way of working generates restricted analysis and leads to limited conclusions. When data are aggregated from checkpoints to route level, one typically loses a considerable amount of information on the regularity between consecutive checkpoints. This procedure is rarely user-oriented, because passengers are mostly concerned with adherence to the headways at their particular bus stop (e.g. [10]). Hence, in order to provide the best possible service to passengers, measures should be performed at every stop of the bus route and for every investigated time period. In this way, performing regularity measures at all bus stops and time periods removes shortcomings deriving from choosing checkpoints and aggregating data in large time periods. Nowadays, relevant support is provided by AVL technology, because it can collect huge amounts of disaggregated data on different bus stops and time periods. Most important, if properly handled and processed, AVL data have the capability to show when and where the service was not provided as planned. However, there are two main criticalities which must be faced before being able to perform accurate regularity calculation. Indeed in case of not addressed criticalities, the calculation of regularity does not sufficiently reflect the service that customers experience and it can provide misleading information. These criticalities are:

1. Bus Overtaking (BO) which arises when the succeeding scheduled bus overtakes its predecessor in the route;
2. Missing data point, which consist of Technical Failures (TF) depending on AVL being temporarily out of work, and Incorrect Operations in the Service (IOS), such as missed trips and unexpected breakdowns.

Due to possible BO, buses might not arrive in the right order. For passengers whose aim is to board on the first useful arriving bus at bus stop, BO is irrelevant because the headway is the time elapsed between two consecutive buses, in which the last one may or may not be the scheduled bus. Hence, instead of tracking the bus (e.g. [11]), regularity measures should focus on transits (i.e. arrivals or departures) of the first bus arriving at the considered bus stop. TF and IOS result in missing data points, which are not recorded by AVL. Moreover, they result in temporal gaps. Hence, a crucial challenge is to recognize the type of missing data and handle the temporal gaps, because they have a different impact on users. The temporal gaps due to TF lead to an incorrect calculation of headways, because buses actually arrived at bus stop, but they were not recorded by AVL. Considering the temporal gaps due IOS is favorable because they are perceived by users as real. McLeod [12] provided insights, in order to determine temporal gaps due to TF and showed that less than 20% of missing data due to TF leads to good quality headway measures. In [7] and [8], in order to recognize and address BO, TF and IOS, a method has been proposed in the case of regularity analysis at the single route and at the whole bus transportation network, respectively. However, in this case two software applications are used to implement the method. Therefore, additional work must be done to implement the method by a single application in order to make AVL data a mainstream source of information when regularity calculation are performed.

2.3 Web Regularity Tools

A key factor for the effective analysis of data is building intelligible performance reports. To date, there are few state of the art of modern Web platforms, specifically designed to providing a Decision Support System (DSS) focused on reliability diagnosis of bus regularity. There are a few research works focusing on Web-based, AVL data visualization, as in [13] or data analysis algorithms and techniques, including a very basic visualization of route paths and speeds using Google maps as in [14]. On the other hand, the current state of the art of information systems technologies includes mature and reliable tools. Mainstream commercial products, or Web frameworks released under Open Source licenses, are designed and documented to integrate with other systems in order to build complex and large-scale Web applications, usually thanks to the use of Application Programming Interfaces (APIs).

Some noteworthy product and framework categories are: business intelligence (BI) tools, reporting and OLAP systems, as Jaspersoft [4] or Pentaho [5]; Web portals development platforms and Content Management Systems (CMS), as the open-source ones like Entando [6] or Joomla [7]; database management systems (DBMS), like the well-known and broadly adopted MySQL [8] or PostgresSQL [9].

3 Methodology

In this section we summarize the method implemented in the Web Portal described in section 4. The method is taken from previous author's works ([7] and [8]) where further details can be found. The method addresses three main phases such as: to validate AVL data, to address criticalities in AVL raw data and to determine the value of C_{vh} in order to illustrate the LoS of regularity over space at every bus stops and route direction - and time - at every time period - in a bus transit network.

3.1 The Validation of AVL Raw Data

Specific attention must be paid to bus stops, because bus operators measure regularities at these points, where passengers board and get off. In this methodology, the relevant elements recorded by AVL at each bus stop for each high frequency route are: day, route, direction, actual and scheduled transit times.

When comparing the numbers of actual and scheduled transits, the lack of data might be observed due to IOS and TF. In this paper, we contemplate the situation where the transport service is good, according to historical data. As a result, few IOS are expected to occur. Therefore, missing data point are fundamentally TF which must be detected and processed in order to determine correct measures of headways. For this reason, we consider the following three main steps to accept or reject data related to days and months and validate a counting section.

STEP 1. Read daily AVL data at a bus stop of a specific route and check whether the number of recorded transits is larger than or equal to a certain percentage of scheduled transits. This percentage can be set equal to 80% of scheduled transits, because McLeod [12] showed that the estimation of headway variance is still good when 20% of data are missing. If a bus stop meets this criterion in that day, it is used for the next step.

STEP 2. Perform a chi-square test on the set of bus stops selected by STEP 1 to evaluate the approximation of the actual number of transits to scheduled

[4] http://www.jaspersoft.com
[5] http://www.pentaho.com
[6] http://www.entando.com
[7] http://www.joomla.org
[8] http://www.mysql.com
[9] http://www.postgresql.org

transits. A suitable significance value for this test is $\alpha = 0.05$. If a day satisfies this criterion, it is used for the next step as well as the bus stops of that day.

STEP 3. Collect all bus stops satisfying STEP 2 in a monthly list and compute the ratio between the number of bus stops in the monthly list and the total number of bus stops. If this ratio is larger than a threshold, all monthly data are supposed to be valid. Based on our experience in preliminary tests, we recommend the use of percentages larger than 60%, which is a good threshold value, in order to cover a significant number of bus stops in a route.

A detailed example of steps 1, 2 and 3 is reported in [7].

3.2 The Handling of Criticalities

Data validation is followed by the detection of criticalities in order to correctly calculate headways between buses. This phase is applicable both in case of few and of many unexpected IOS. As illustrated in section 2.2, three types of criticalities might occur: BO, TF and IOS. Since TF and IOS lead to missing data and temporal gaps, they can be addressed almost together. Gaps must be found and processed by comparing scheduled and actual transit times. Sophisticated AVL databases can be used to match scheduled transit time data (with no gaps) with actual transit time data (with possible gaps). As a result, to address all criticalities, the following steps are carried out:

STEP 4. Address BO by ordering the sequence of actual transit times at bus stops, because BO is irrelevant for the regularity perceived by users, who are not interested in the right schedule of buses.

STEP 5. Fill up tables reporting unpredicted missed trips and unexpected breakdowns. Columns include the day, bus stops, route, direction, scheduled transit times and incorrect operation code, indicating whether it is neither a missed trip or a breakdown.

STEP 6. Consider the table of scheduled service on that day with the attributes of the table in STEP 5, whereas the incorrect operation code wil be neglected, because it is an unexpected event.

STEP 7. Match these tables, in order to generate a new table of scheduled transits with a new attribute indicating the incorrect operation code.

STEP 8. Detect TF and IOS. Match the table built in STEP 7 with data at the end of STEP 4 and detect possible gaps, when a transit between two recorded transits is missing.

STEP 9. Correct TF and IOS. Disregard gaps generated by TF, because no real headways can be derived as the difference between two consecutive transits. Keep the gaps generated by IOS, because these gaps are really perceived by users.

A detailed example of steps 4,8 and 9 is reported in [7] and [8].

3.3 The Calculation of Regularity LoS

Given a generic bus stop j at time period t along the direction d of a route r, once data have been validated and criticality have been addressed, we calculate the C_{vh} as follows:

$$C_{vh}^{j,t,d,r} = \frac{\sigma^{j,t,d,r}}{h^{j,t,d,r}} \tag{1}$$

where:

- $\sigma^{j,t,d,r}$ is the standard deviation of the differences between actual and scheduled headway at bus stop j, time interval t, direction d and route r; the values used for the evaluation span over a monthly planning horizon.
- $h^{j,t,d,r}$ is the average scheduled headway at bus stop j, time interval t, direction d and route r.

In many transit agencies, the standard time interval is one hour. Since transit agencies may add or remove some bus trips to better serve the changing demand ([14]), in this paper $h^{j,t,d,r}$ is computed as the average of headways of scheduled transits times, to account for these additional trips and possible gaps. As illustrated previously, Eqn. (1) provides results in a monthly planning horizon whose set is denoted by S. The elementary observation is denoted by $x_i (i = 1, n)$ and represents the precise headway deviation at the end of STEP 9. However, in order to provide monthly aggregated statistics for week and type of day, Eqn. (1) is also calculated for them, considering the sub-sets S_1 and S_2. The related elementary observations are denoted by $x_{1j} (j = 1, , m)$ and $x_{2k} (k = 1, , p)$ and represents the precise headway deviation at the end of STEP 9, when they are related to the week and type of day, respectively. Therefore, to summarize, the results provided by Eqn. (1) refer to the considered sets defined as:

S - the set representing the headway deviations (x_i) within the month.
S_1 - the set representing the headway deviations (x_{1i}) within the considered week in the month.
S_2 - the set representing the headway deviations (x_{2i}) within the considered type of day in the month.

The calculated values of C_{vh} can be converted into the LoS according to [9]. LoSs can be represented by dashboards as illustrated in the next section.

4 Web Portal

The web portal is the entry point of the DSS (hereafter the "system") composed of an environment designed for transit industry experts. The environment is designed to primarily handle AVL raw data for measuring the LoS of bus regularity at each bus stop and time period of a transit bus route. The web portal is powered by Entando, a Java Open Source portal-like platform for building information, cloud, mobile and social enterprise-class solutions. It natively combines portal, web CMS and framework capabilities. The portal provides dashboards that support experts with significant and useful data. A main feature of the dashboards is to identify where and when regularity problems occur. Dashboards are intended to show summaries and consist of tables, charts and maps. The items located on the dashboards include:

- regularity table - a table showing the LoS; the executive may select a route direction and a time period; the table shows for each bus stop of the route the LoS of regularity in different colors;
- regularity line chart - a multi-line chart showing one chart for each time period; the charts are distinguished by color and may be immediately compared;
- mapper - a Google Maps technology based interface; the map shows the path of a selected route direction. The bus stops associated with the route are represented with different colors depending on the LoS of regularity; the executive may interact with the map changing the time slot within the time period.

4.1 Components of the System

The web portal is part of the whole system. The Figure 1 shows the components of the system.

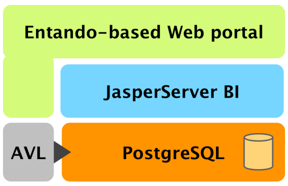

Fig. 1. Components of the system.

At the bottom, the AVL is the technology which collects raw data during the transport service. As illustrated in section 3, collected data deals with real measures including actual and scheduled bus transit time, bus information, route information, bus stops information. A pre-filtering process is necessary before storing in the database in order to harmonize data in case of non-homogeneous values. The database is a PostgreSQL instance extended with spatial and geographic features, PostGIS[10]. 451Research[11] estimates that around 30% of tech companies use PostgreSQL for core applications as of 2012. PostgreSQL is an Open Source solution that strongly competes with proprietary database engines and is supported by a consistent community of users. The use of the database in the system is twofold. It is first necessary to store the data collected by the

[10] http://www.postgis.net
[11] https://451research.com/

AVL. An entity-relationship diagram defines how data is scattered among the tables. The database is also necessary for the administrative activities required to manage the portal. An instance of JasperServer is responsible for dashboards creation. JasperServer is a Business Intelligence (BI) tool and a reporting and analytics platform. With JasperServer it is possible to create single reports or dashboards faster. The dashboards are pre-defined through the JRXML markup language and are available for integration in the system. Entando Web Portal provides the user interface.

4.2 Implementation of the method

The implementation of the method consists in four modules, that manage the four stages of the data flow process: data import; data processing; data pre-aggregation and data management.

All the modules use PostgreSQL functions in PL/pgSQL language to perform database's tasks and Entando modules in Java language to start and supervise the execution of each task as the Web application. The data are collected in a single database. A database schema is created for each month in which data is elaborated.

The main entity is represented as a table named "Bus_Stop". A Bus Stop entity contains date attributes, which are used to generate dynamically the path in use in a particular year and month. Moreover, a geometric attribute contains the polyline, which starts at the previous bus stop and ends at the considered bus stop. In this way, using spatial aggregate functions, one is able to merge polyline of consecutive bus stops per path code, in order to derive spatial characteristics (geometry) of the Path entity.

The first module contains the functions that implement the first phase of the methodology illustrated in Section 3: the raw data of a month are imported and validated. The system loads these primary data in two tables: the 'AVL' table where each row contains real transit at bus stops and the 'Scheduler' table which contains the scheduled transit. Then the system validates the 'AVL' by applying the three steps described in Section 3.1. The parameters of transits percentage (80%), the chi-square test value ($\alpha = 0,05$) and the threshold of the percentage ratio between the number of bus stops that pass the chi-square test and the total number of bus stops (60%) can be changed by the analyst.

The second module contains the tools that implement the second phase of the methodology. The module generates the "Differences" facts table. This table contains the difference between actual and scheduled headways between two consecutive buses as measure and two multi level dimensions: year, month, day and time slot are the temporal dimension; route, path and bus stop are the logical dimension.

The third module implements the third phase of the methodology. It generates the "Regularity" facts table that contains the regularity measures evaluated over three distinct type of day aggregations: by week (4 measures), by day of week (7 measures) and by the entire month. The pre-aggregation uses the eqn.(1) to calculate the C_{vh} measure over each set of samples defined in Section 3.3. The fourth module contains a set of functions to manage data and reports. The System can also manage the JasperServer configuration for its connection to the database and regularity reports definition via the Entando/JasperServer Connector. Currently, a Mondrian Olap Cube for "regularity" or "differences" facts tables is not yet configured.

4.3 Reliability Diagnostic Tools

AVL technology is installed in-busses. It records several data, such as actual arrival times at every bus stop in minutes and seconds. Such information belongs to the class of time-at-location data collected by vehicles and can be used for off-line analysis. As a vehicle finishes its service, it moves back to the depot, where data recorded during the daily shift are downloaded. Daily data are stored in a central database. The diagnostic tools are realized using some JasperServer functionalities. First, the database containing the facts tables is connected to the JasperServer business intelligence engine. Then, as shown in figure 2, three type of reports are created through the JasperServer tools: the regularity table, the multi-line charts and the mapper .

Each report shows the regularity measures of all bus stops of a particular route direction. For the sake of clarity in representation, the value of measures is represented by colours depending on LoS. The red colour represents LoS **F** ($C_{vh} > 0,75$, i.e. most vehicles bunched), the orange colour indicates LoS **E** ($0.53 < C_{vh} < 0.74$, i.e. frequent bunching), the yellow colour shows LoS **D** ($0.40 < C_{vh} < 0.52$, i.e. irregular headways, with some bunching). Other colour gradations mean LoS from **A** to **C** ($C_{vh} < 0.40$, i.e. satisfactory regularity). When LoS are not available, they are denoted by **null**.

The executive selects the route direction and the day aggregation type. Moreover, when a report is showed, the executive can select the time slot.

The regularity classes (see figure 2) of a bus stop in a selected time slot are represented as a table in the regularity table report, as a coloured poly-lines and icons in the mapper report or as a coloured lines in the multi line charts report. It is important to highlight that figure 2 is the result of different screens, than there is no exact correspondence among the colouring. In order to permit a map representation of bus stop and path, the geometries in the "Bus_Stops" table are transformed in the WGS84 projection and GeoJSON format using Postgis functionality and linked to the reports table.

5 Conclusion

In bus transit operators the measure of regularity is a major requirement for high frequency public transport services. Besides, it is necessary to properly account for the efficient monitoring of quality of service and for the perspectives of both bus operators and users. In this paper we have implemented a methodology to evaluate regularity starting from data collected by AVL, and proposed the integration of technologies in a web portal as an environment designed to

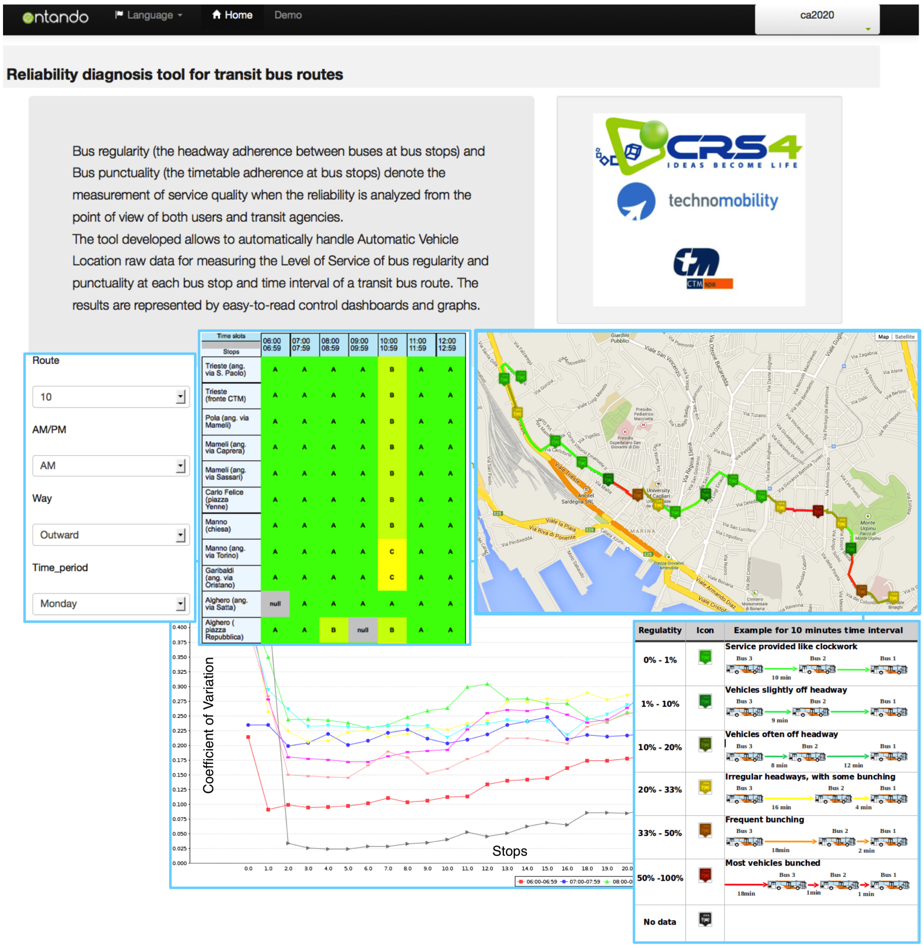

Fig. 2. Diagnostic tools.

support bus transit operators experts in evaluating regularity issues. This paper shows that it is possible to handle huge AVL data sets for measuring bus route regularity and understand whether a missing data point is a technical failure or an incorrect operation in the service, providing a detailed characterization of bus route regularity at all bus stops and time periods by AVL technologies. The web portal ensures tool access from everywhere and anywhere. This procedure results in significant time and energy savings in the investigation of large data sets.

The next step will be to extend the web portal for both operators and users, then at a later stage for transit agencies and passengers. Illustrating the practical effectiveness of this procedure will be important to implement a real case study. User-friendly control dashboards help to perform an empirical diagnosis of performance of bus route regularity. Transit managers can use easily-understood representation and control rooms operators following buses in real time, to focus on where and when low regularities are expected to occur. Moreover, possible cause of low level of service will be investigated, in order to put the bus operator in the position of selecting the most appropriate strategies to improve regularity. In addition, the method and the integration of technologies will be adapted for the measurement of punctuality in low-level frequency services.

References

1. European Committee for Standardization, B.: Transportation logistics and services. european standard en 13816: Public passenger transport service quality definition, targeting and measurement. (2002)
2. Ceder, A.: Public Transit Planning and Operation: Theory, Modelling, and Practice. Butterworth-Heinemann, Oxford, England (2007)
3. Marguier, P., Ceder, A.: Passenger waiting strategies for overlapping bus routes. Transportation Science **18(3)** (1984) 207–230
4. Nakanishi, Y.: Bus performance indicators. Transportation Research Record **1571** (1997) 3–13
5. Board, T.R.: A Guidebook for Developing a Transit Performance-measurement System. Report (Transit Cooperative Research Program). National Academy Press (2003)
6. Trompet, M., Liu, X., Graham, D.: Development of key performance indicator to compare regularity of service between urban bus operators. Transportation Research Record: Journal of the Transportation Research Board **2216**(1) (12 2011) 33–41
7. Barabino, B., Di Francesco, M., Mozzoni, S.: Regularity diagnosis by automatic vehicle location raw data. Public transport **4(3)** (2013) 187–208
8. Barabino, B., Di Francesco, M., Mozzoni, S.: Regularity analysis on bus networks and route directions by automatic vehicle location raw data. Forthcoming in IET Intelligent Transport Systems (2013)
9. Board, T.R.: Transit capacity and quality of service manual, 2nd ed. Transit Cooperative Research Program Report 100. Transportation Research Board (2003)
10. Kimpel, T.: Time Point-level Analysis of Transit Service Reliability and Passenger Demand. Portland State University (2001)
11. Mandelzys, M., Hellinga, B.: Identifying causes of performance issues in bus schedule adherence with automatic vehicle location and passenger count data. Transportation Research Record: Journal of the Transportation Research Board **2143**(1) (12 2010) 9–15
12. McLeod, F.: Estimating bus passenger waiting times from incomplete bus arrivals data. J. Oper. Res. Soc. **58(11)** (2007) 1518–1525
13. Oluwatobi, A.: A gps based automatic vehicle location system for bus transit . Academia.edu (2013)
14. Cortés, C.E., Gibson, J., Gschwender, A., Munizaga, M., Zuniga, M.: Commercial bus speed diagnosis based on gps-monitored data. Transportation Research Part C: Emerging Technologies **19**(4) (2011) 695 – 707