Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 27 (2017) 101–108

Transportation Research Procedia

www.elsevier.com/locate/procedia

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

Bus travel time variability: some experimental evidences

Antonio Comi[a]*, Agostino Nuzzolo[a], Stefano Brinchi[b], Renata Verghini[b]

[a] *Department of Enterprise Engineering, University of Rome Tor Vergata, via del Politecnico 1, 00133 Rome, Italy*
[b] *Rome Mobility Agency (Roma Servizi per la Mobilità), Via L. Gaurico 9, 00143 Rome, Italy*

Abstract

Bus travel time analysis is essential for transit operation planning. Then, this topic obtained large attention in transport engineering literature and several methods have been proposed for investigating its variability. Nowadays, the availability of large data quantities through automated monitoring allows more in-depth this phenomenon to be pointed out with new experimental evidence. The paper presents the results of some analyses carried out using automatic vehicle location (AVL) data of bus lines and automated vehicle counter (AVC) data on some corridors in the urban area of Rome where the bus services are mixed with other traffic and travel times are subject to high degrees of variability. The results show the effect of temporal dimension and similarity between travel time and traffic temporal patterns, and could open the road for the improvement of the short-term forecasting methods, too.

Keywords: travel time analyses; bus travel time; automated traffic data; traffic data processing; travel speed analysis; time series

1. Introduction

Transit operators face with relevant travel time variability issues related to the design of new transit lines or to the improvement of the performances of existing lines, especially when transit vehicles share the lanes with other traffic components. In this case, bus travel times are subject to high degrees of variability, as their temporal patterns are similar to general traffic ones, whose investigations show seasonality and trend/cycle (Fusco *et al.*, 2016). Therefore, reliable long-term travel time forecasts availability is one of the most relevant attributes for transit operators, to be used for improving transit service planning in terms of optimization of timetable and vehicle scheduling.

To accomplish this, a study was addressed to investigate bus travel times and traffic patterns through time series methods, analyzing automatic vehicle location (AVL) and automated vehicle counter (AVC) data of some bus lines,

* Corresponding author. Tel.: +39 06 7259 7061; fax: +39 06 7259 7061.
E-mail address: comi@ing.uniroma2.it

2352-1465 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.
10.1016/j.trpro.2017.12.072

which share the lanes with other traffic components, in the city of Rome. The main objectives of the study were to analyze to what extent bus travel time patterns and general traffic patterns are similar and to what extent bus travel time variability is explained by temporal congestion variability and by all other variables.

The results are reported in this paper, which is organized as follows. A brief state of art of *long term* bus travel time forecasting methods is reported in Section 2. Section 3 synthetizes the data available, while Section 4 reports the analyses performed and the results obtained. Finally, Section 5 draws conclusions and further research development.

2. Long-term bus travel time forecasting methods

A variety of methods to obtain *long-term* bus travel time forecasts has been developed over the years. The most widely used (Bolshinsky and Freidman, 2012; Kieu *et al*., 2012; Fan *et al.*, 2015; Moreira *et al.*, 2015; Nuzzolo and Comi, 2016; Yu *et al.*, 2017) refer to: regression methods and time series methods.

Regression methods require a mathematical function to explain a dependent variable (i.e. travel time) with a set of independent variables. Complex models such as support vector regression, k-nearest neighbor regression, project pursuit regression and artificial neural network are the most popular approaches to this problem due to their ability to find complex non-linear relationships between the target variable and the independent ones (Moreira *et al.*, 2015). They are able to work satisfactorily even if traffic conditions are not stable. Such methods have been used by many authors (e.g. Chen *et al.*, 2004; Mendes-Moreira *et al.*, 2012; Moreira-Matias *et al.*, 2016), because they have a relative advantage in revealing which independent variables are less or more important for reproducing/predicting travel times.

Time series-based methods enlighten travel time pattern dependency on time processing observed historical data (William and Hoel, 2003; Jeong, 2004; Billings and Jiann-Shiou, 2006; Suwardo *et al.*, 2010). The strength of time series based methods are high computation speed due to simple formulation of the algorithm and do not need large number of bus operation variables: only time related bus travel times. They allow the variability structure of travel time to be pointed out and to reveal effects along time (e.g. day hours, weekday, year period), which are relevant in bus route sharing the lanes with other traffic components. These models, of course, are useful for re-scheduling of existing lines. If the route of the line and the structure of time dependency remain the same and the other variables, for example the road network characteristics, do not change, such an analysis results can be usefully applied for the re-definition of more reliable timetables and subsequent vehicle scheduling. Otherwise, regression models are in general to be preferred.

The literature review points out that there is a quite extensive literature on long term travel time prediction, but few studies have enlightened the relationships between performances of transit systems and temporal congestion variability (which derives from total traffic variability), whose effects are significant in developing models for both long and short-term travel time prediction. Further, very few studies consider terminal-to-terminal travel time, whose analyses is quite relevant for transit operators that have to plan system in order to define timetable and bus scheduling able to limit, during operations, deficiencies in service providing respecting constraints of service commitment in terms of reliability (Cats, 2014). Therefore, due to the complexity affecting bus travel time related to the randomness of various factors of the bus journey, the need of further investigations germinates.

The two subsequent sections present the results of some analyses carried out combining automatic vehicle location (AVL) data of bus lines and automated vehicle counter (AVC) data in the urban area of Rome, on some corridors where buses share the lanes and the bus travel times are characterized by high variances. The attention was paid to terminal-terminal travel time, which is one of the main variables to be considered in transit planning, requiring long-term forecasting. The analyses were based on *time series* decomposition methods, aiming to recognize time patterns, analyzing three-month data relative to some bus lines and to automated vehicle counts of the same corridors, and discovering component of trend, seasonality and cycle very similar for bus travel time and general traffic flow.

3. Data collected

The analysis described below were performed combining automatic vehicle location (AVL) data of two bus lines and automated vehicle counter (AVC) data in the urban area of Rome. Table 1 summarizes the characteristics of bus service lines (A and B), while Figure 1 plots the observed travel times with respect to time of the day during working days (from Monday to Friday). The high variance according to the time of the day can be pointed out showing similar shape if travel is to or from the city center. In particular, for both the investigated bus lines, the variance is higher in the morning peak hour than in the afternoon.

In addition, data were also available on automatic counter sections. They allowed the average speed and the number of vehicles travelling through these sections to be obtained and hence the traffic patterns to be investigated.

According to these data, in the following, aiming to verify receptiveness in travel time determinants, before time series components of bus travel time data are studied, hence traffic count data are analyzed to reveal to what extent their patterns are similar.

Table 1. Main characteristics of bus analyzed services

Line	Daily average travel time (to the city center) [minutes]	Daily average travel time (from the city center) [minutes]	Travel length [km]	Peak-hour headway [minutes]	Off-peak hour headway [minutes]	One-way number of stops	Observation period (T) [weeks]
A	39.5	41.7	11	20	30	37	8
B	57.5	58.4	23	15	30	60	12

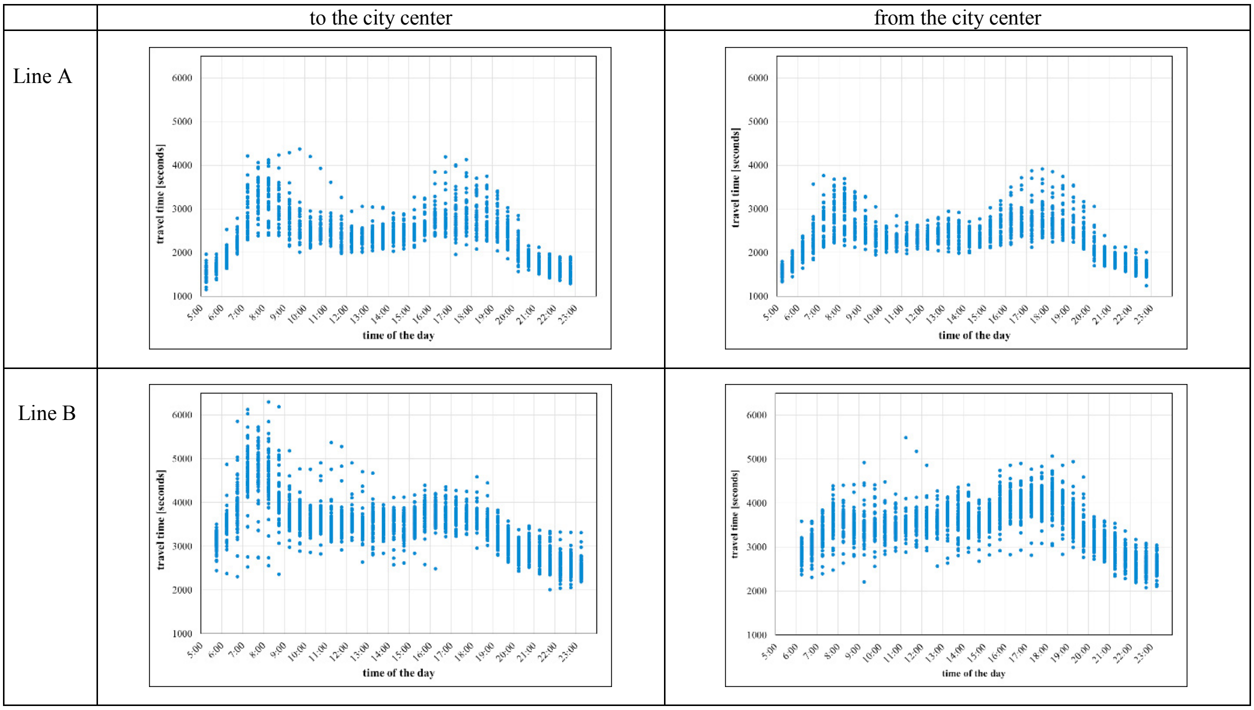


Figure 1. Hourly fluctuation of travel time of analyzed bus lines (working days).

4. Bus travel time analysis

Terminal-to-terminal bus travel time (*TV*) is formally expressed as:

$$TV = RT + DW \tag{1}$$

where

- *RT* is the running time, which depends on the below determinants:
  - *flow speeds*, which are related to link capacity, link flow and flow composition,

- *infrastructure link characteristics*, which are related to number of link lanes, lane length, signals, pedestrian crossings, type of parking, road work,
- *functional link characteristics*, which are related to interferences with other mobility components (e.g. shoppers, students),
- *context conditions*, e.g. weather conditions (raining, sunny, etc.);

• *DW* is the dwelling time, which depends on:
- *alighting*, *boarding* and *on-board* users,
- *bus features*, e.g. number of doors, lift operations.

Therefore, the variability of bus travel times can be analysed starting from the investigation of the fluctuations of the above identified determinants. Besides, such fluctuations can present a systematic nature (receptiveness) and time series based methods should be a powerful tool in such an analysis. Then, aiming to discover such a systematic nature in travel time fluctuations, the results obtained using time series methods are presented below.

A given time series Y_t can be considered as comprising three components: a seasonal component (S_t), a trend-cycle component (T_t, containing both trend and cycle), and a remainder component (E_t, containing anything else in the time series). Therefore, if an additive relation is assumed, follows:

$$Y_t = f\left(T_t, S_t, E_t\right) = T_t + S_t + E_t$$

Let Y_i denote the i-th observation and $\hat{Y}_i$ denote a modelled value of Y_i. The modelled error is simply

$$e_i = Y_i - \hat{Y}_i \tag{2}$$

which is on the same scale as the data.

To assess the modelled accuracy (Hyndman and Athanasopoulos, 2016), the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) are used.

The two-used scale-dependent measures are based on the *absolute errors* $\left|e_i\right|$ or *squared errors* $\left(e_i\right)^2$:

$$\begin{aligned} &\text{Mean absolute error: MAE} = \text{mean}\left(\left|e_i\right|\right) \\ &\text{Root mean squared error: RMSE}=\sqrt{\text{mean}\left(e_i^2\right)} \end{aligned} \tag{3}$$

The percentage error is given by $p_i = 100 \cdot e_i / Y_i$, which has the advantage of being *scale-independent*. The used measure is:

$$\text{Mean absolute percentage error: MAPE} = \text{mean}\left(\left|p_i\right|\right) \tag{4}$$

4.1. Bus travel time series decomposition

Figure 2 reports the decomposition of bus travel time series of line A and B in the months of February and March 2016 performed through the seasonal and trend decomposition using loess (STL; Cleveland *et al.*, 1990; Jeon and Hong, 2016) decomposition method implemented in R software. Trend/cycle and seasonality can be pointed out, in particular:

1. trends/cycles are quite flattened with a small difference between maximum and minimum values: less than 5% for line A and 2% per line B, at least in this test case, which considers three months;
2. the effects of daily seasonality emerge for all days with smaller values on Saturday and Sunday (Figure 3) deriving from lower congestion in the weekend than in the workdays;
3. seasonality is quite relevant for hours of the day because of variance of traffic flows and hence of the variance of road congestion. In fact, buses share the lanes and subsequently the traffic flow influences bus travel time;

4. seasonality is different for to or from the city center. Going to the city, high concentrations were revealed in the morning (e.g. due to concentration of arrival constraints at work or at school), while in the afternoon the effects along the hours are more distributed;
5. as reported in Table 2, the contribution of remainder (*E*) is low and in terms of variance is about 29% (line A) and 21% (line B), which are mainly due to the singular variability revealed in some days due to particular events concentrated in time and space than structural factors.

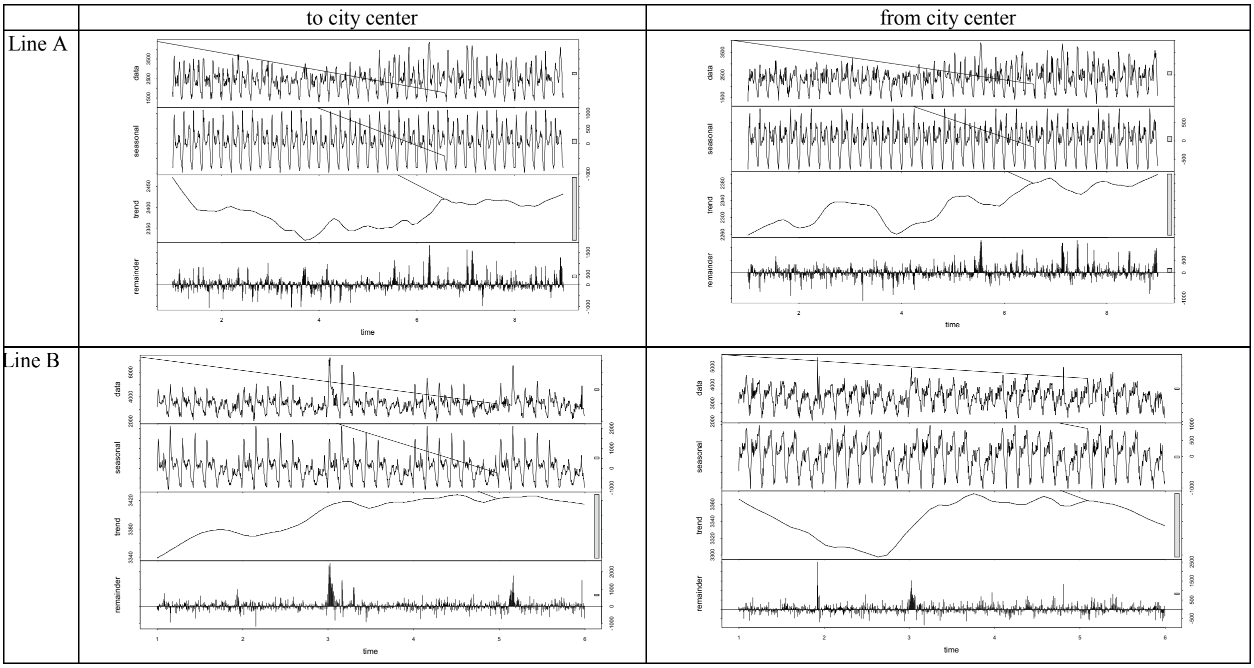


Figure 2. Travel time STL decomposition with weekly period (February – March 2016).

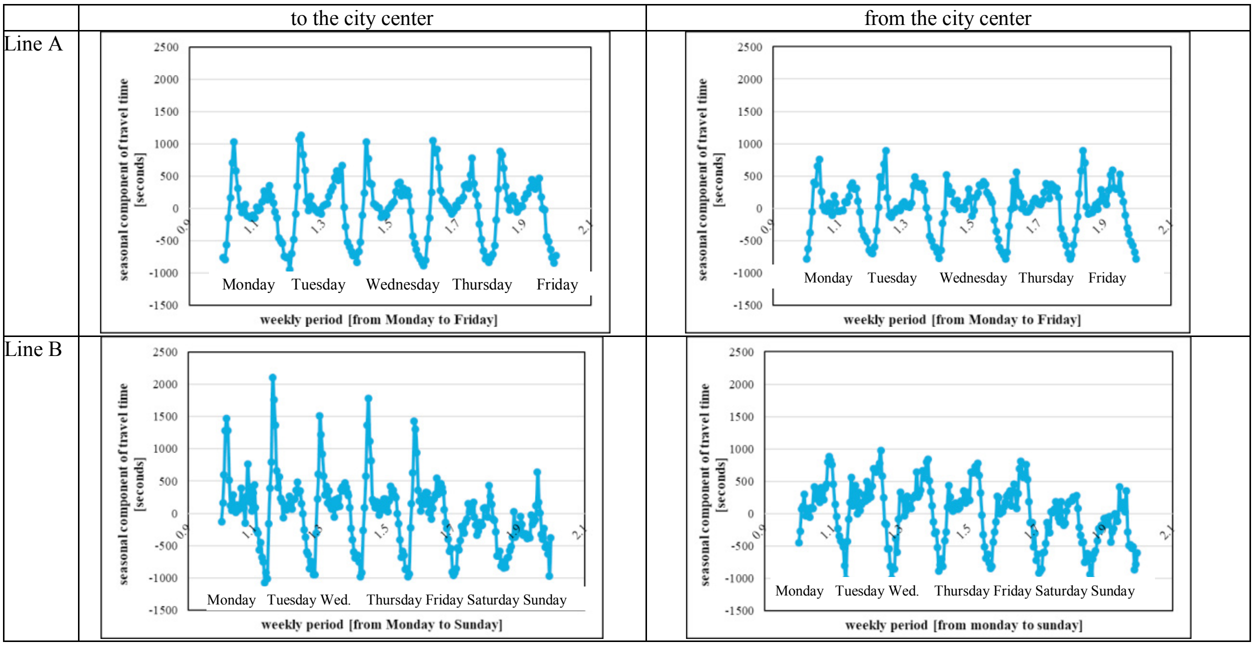


Figure 3. Hourly fluctuation of travel time in the time period for the analyzed bus lines.

Table 2. Variance (σ^2) and mean (μ) of the observed travel time (Y) and remainder (E).

	to the city center	from the city center
Line A	σ^2 [Y] = 297 754 μ [Y] = 2 412 seconds σ^2 [E] = 83 436 μ [E] = 30 seconds	σ^2 [Y]= 198 696 μ [Y] = 2 357 seconds σ^2 [E] = 58 329 μ [E] = 23 seconds
Line B	σ^2 [Y] = 405 878 μ [Y] = 3 435 seconds σ^2 [E] = 85 447 μ [E] = 32 seconds	σ^2 [Y] = 248 132 μ [Y] = 3 366 seconds σ^2 [E] = 52 089 μ [E] = 2 seconds

Finally, the accuracy in using only systematic components (i.e. trend/cycle - T_t - and daily/hourly seasonality - S_t) in reproducing travel time variability was evaluated. Therefore, the modelled (decomposition) error (e) can be assumed to be the remainder (E), i.e. $e \equiv E$. Figure 4 plots an extract of how reproduction works, while Table 3 synthetizes the above accuracy for whole investigated period of the two plotted routes (i.e. route of line A to the city center and line B from the city center). As synthetized by MAPE, which is smaller than 7.5%, trend/cycle and seasonality allow the main part of variance to be explained. It means an average error in reproducing travel time of about 3 minutes for line A to city center and 1.5 minutes for line B from city center. Besides, although the findings of these analyses refer mainly to the improvement of long-term travel time forecasting, they can open new research opportunities for short-term forecasting to develop methods and models that take the expected pattern into account.

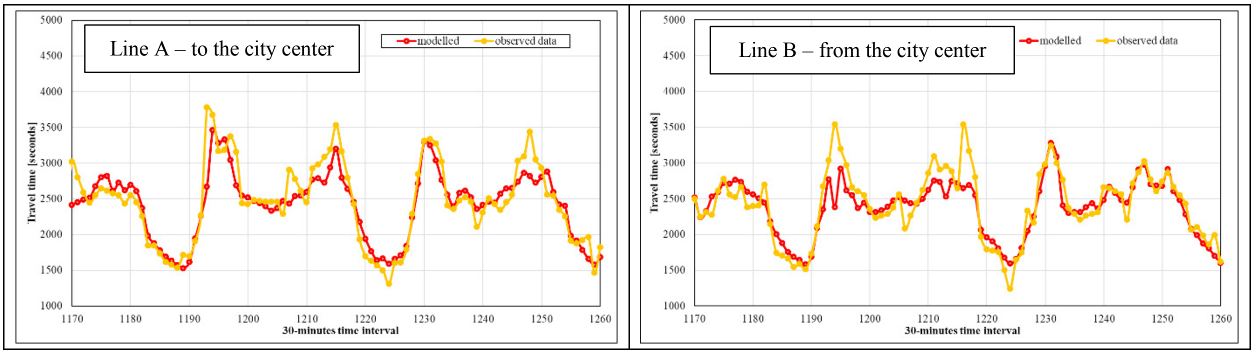


Figure 4. Example of travel time modelling for two routes of the investigated bus lines.

Table 3. Accuracy indicators relative to the whole period of analysis for a route of lines A and B

Line A – to city center	Line B – from city center
average e_i = -30.2 standard deviation of e_i = 288.7 MAE = 187.4 - RMSE = 290.3 - MAPE = 7.4%	average e_i = 15.3 standard deviation of e_i = 228.2 MAE = 139.8 - RMSE = 254.0 - MAPE = 4.2%

4.2. Bus travel time and congestion

The above identified characteristics of bus travel time and their dependences on the congestion were confirmed by the analysis of the time series of traffic counts even if in a road sections not along the bus routes. As shown in Figure 5, which reports the STL decomposition of flow and vehicle speed time series, the contribution of trend/cycle to variability is low, given that its profile is quite flattened along the day from the week after the beginning of the school time (9th of September), while a daily seasonality effect emerges in both direction (i.e. from and to the city center) showing similar patterns for working and weekend days. Comparing Figure 3 and Figure 6, very similar patterns of seasonality in both directions (i.e. to and from the city center) emerge.

These results confirm the findings obtained in other urban contexts. For example, Rajbhandari (2006), investigating similar types of data from some cities in New Jersey, obtained that on each day of the week, morning and evening peak hours showed higher travel times than other times of day. However, the peaks varied on different days, i.e. the mean of travel time increased from early morning towards the morning rush and in the afternoon peak, while decreased thereafter, which, as emerged from Rome data, is a very typical observation of traffic condition. Yetiskul *et al.* (2012) revealed the effects of temporal dimension through data of AVL of the city of Ankara. Among the days of the week, Wednesday exhibited the highest source of variation. Recurrent traffic congestion was consistently observed during peak hours in the morning with subsequent bus travel-time variation increase. Similarly, Hassan *et al.* (2016) analysed through regression methods AVL data of four routes of two bus lines (with maximum planned headway of 11 min). They also found significant effects due to departure time (i.e. time of day) and day, and driver's style in driving, which Rome data do not allow this last factor to be pointed out.

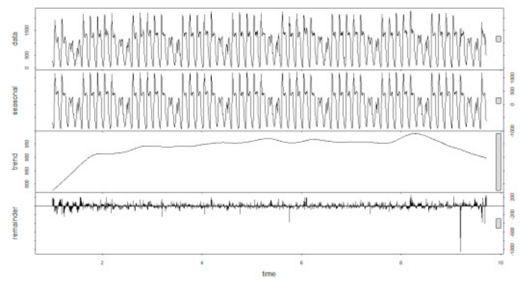
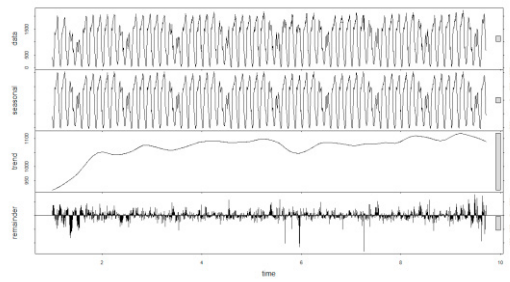
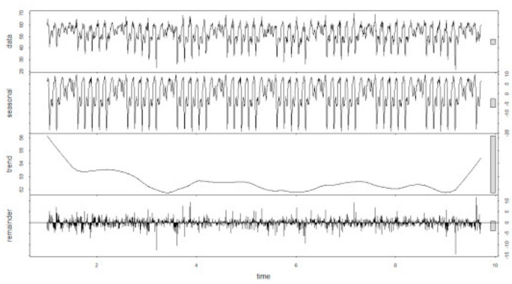
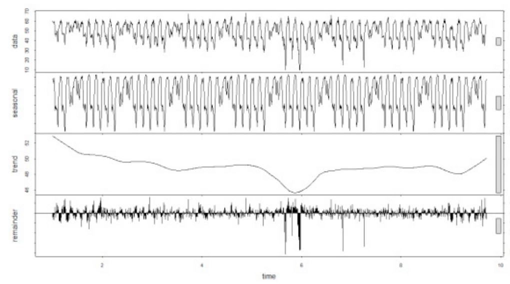
	to the city center	from the city center
Flows [veh/h]		
Average speed [km/h]		

Figure 5. Flow and speed STL decomposition of traffic counts with weekly period (September – October 2016).

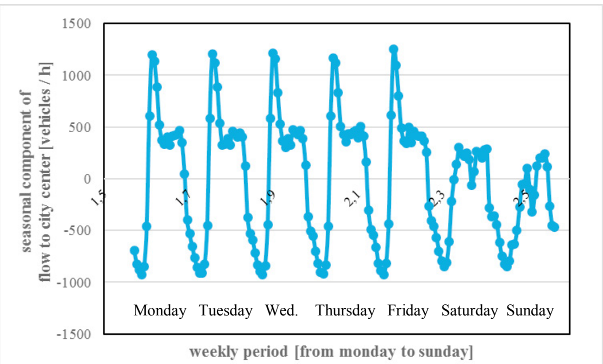
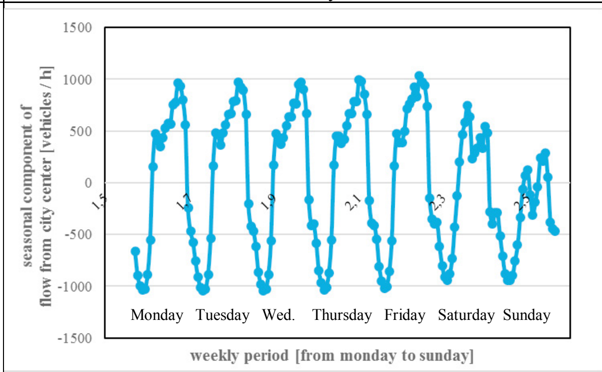
to the city center	from the city center
	

Figure 6. Seasonal component of flow time series at counter section (direction to and from the city center).

5. Conclusions

An investigation of the variability of bus travel times was carried out and the first results for the urban area of Rome, whose findings can be used for designing transit service timetable and vehicle scheduling in re-planning of existing lines, are presented. The focus was on systematic components and the results are mainly devoted to bus lines that share the lanes. The analyses were performed through time series methods, which allowed to recognize temporal patterns and to point out that also day of the week (i.e. Wednesday *vs* Thursday) has significant effects on travel time variability and have to be taken into account. Traditionally, such characteristics are neglected in short-term forecasting methods that, in general, do not take into accounts the expected temporal patterns. Besides, given that the investigated bus services share the lanes with other traffic, a comparison between bus travel time patterns and traffic ones, using also data from AVC system, was performed, and the travel time patterns were confirmed as very similar to the traffic count ones. Further developments in this study are in progress. They mainly concern additional investigations on new bus lines and traffic counters, and further types of automated data such as floating car data for revealing local traffic patterns, analysis of residuals deriving from decomposition method applying, in particular test of data autocorrelation and ARIMA models, including the seasonal ARIMA (SARIMA) based regressions models. Finally, the development of more performing models for long and short-term predictions that include the findings of these analyses (e.g. the expected temporal patterns), through the conjoint use of time series and regression methods, will be embarked.

References

Billings, D. and Jiann-Shiou, Y., 2006. Application of the ARIMA Models to Urban Roadway Travel Time Prediction - A Case Study. Paper presented at the Systems. In: *IEEE International Conference on Man and Cybernetics*.

Bolshinsky, E. and Freidman, R., 2012. *Traffic flow forecast survey*. Technion–Israel Institute of Technology–Technical Report.

Cats, O., 2014. Regularity-driven bus operation: Principles, implementation and business models. In: Transport Policy 36, pp. 223-230.

Chen, M., Liu, X., Xia, J., & Chien, S., 2004. A Dynamic Bus-Arrival Time Prediction Model Based on APC Data. In: *Computer-Aided Civil and Infrastructure Engineering, 19(5)*, pp. 364-376.

Cleveland, R. B., Cleveland, W. S. and Terpenning, I., 1990. STL: A seasonal-trend decomposition procedure based on loess. In: *Journal of Official Statistics, 6(1)*, 3.

Fan, W., Gurmu, Z., 2015. Dynamic Travel Time Prediction Models for Buses Using Only GPS Data. In: *International Journal of Transportation Science and Technology 4 (4)*, pp. 353–366.

Fusco, G., Colombaroni, C. and Isaenko, N., 2016. Short-term speed predictions exploiting big data on large urban road networks. In: *Transportation Research Part C: Emerging Technologies 73*, pp.183–201.

Hassan, S. M., Moreira-Matias, L., Khiari, J. and Cats, O. Feature Selection Issues in Long-Term Travel Time Prediction. In: *Advances in Intelligent Data Analysis XV* - International Symposium on Intelligent Data Analysis, Springer, pp. 98-109

Hyndman, R. J. and Athanasopoulos, G., 2016. *Forecasting: principles and practice*. www.otexts.org.

Jeon, S. and Hong, B., 2016. Monte Carlo simulation-based traffic speed forecasting using historical big data. In: *Future Generation Computer Systems, 65*, 182-195.

Jeong, R. H., 2004. *The prediction of bus arrival time using automatic vehicle location systems data*. Ph. D. Thesis, Texas A&M University.

Kieu, Le Minh, Bhaskar, Ashish, and Chung, E., 2012. Benefits and issues for bus travel time estimation and prediction. In: *Australasian Transport Research Forum 2011 Proceedings*, Perth, WA.

Mendes-Moreira, J., Jorge, A. M., de Sousa, J. F. and Soares, J., 2012. Comparing state-of-the-art regression methods for long term travel time prediction. In: *Journal Intelligent Data Analysis archive 16 (3)*, pp. 427-449.

Moreira-Matias, L., Cats, O., Gam, J., Mendes-Moreira, J., and Freire de Sousa, J., 2016. An online learning approach to eliminate Bus Bunching in real-time. In: *Applied Soft Computing 47*, pp. 460–482.

Moreira-Matias, L., Mendes-Moreira, J., de Sousa, J. F., Gama, J., 2015. Improving Mass Transit Operations by Using AVL-Based Systems: A Survey. In: *IEEE Transactions on Intelligent Transportation System*, DOI 10.1109/TITS.2014.2376772.

Nuzzolo, A. and Comi, A., 2016. Advanced Public Transport and ITS: new modelling challenges. In: *Transportmetrica A: Transport Science 12 (8)*, DOI: 10.1080/23249935.2016.1166158, Taylor & Francis, pp. 674 – 699.

Rajbhandari, R., 2006. *Bus arrival time prediction using stochastic time series and Markov chains*. Ph. D. Thesis, Faculty of New Jersey Institute of Technology.

Suwardo, Madzlan, N. and Ibrahim, K., 2010. ARIMA models for bus travel time prediction. In: *The Journal of the Institution of Engineers, Malaysia, vol. 71(2)*, pp. 49-58.

Williams, B. M. anf Hoel, L. A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. In: *Journal of Transportation Engineering, 129(6)*, 664-672.

Yetiskul, E. and Senbil, M., 2012. Public bus transit travel-time variability in Ankara (Turkey). In: *Transport Policy 23*, pp. 50–59.

Yu, Z., Wood, J. S., Gayaha, V. V., 2017. Using survival models to estimate bus travel times and associated uncertainties. In: *Transportation Research Part C: Emerging Technologies 74*, pp. 366–382.