Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 45 (2020) 692–699

Transportation Research Procedia

www.elsevier.com/locate/procedia

AIIT 2nd International Congress on Transport Infrastructure and Systems in a changing world (TIS ROMA 2019), 23rd-24th September 2019, Rome, Italy

Investigating bus travel time and predictive models: a time series-based approach

Antonio Comi[a], Mykola Zhuk[b], Volodymyr Kovalyshyn[b], Volodymyr Hilevych[b]*

[a]*Department of Enterprise Engineering, University of Rome Tor Vergata, via del Politecnico 1, 00133 Rome, Italy*
[b]*Department of Transport Technologies, Lviv Polytechnic National University, St. Bandery 12, 79013 Lviv, Ukraine*

Abstract

Public transport agencies observe the travel time as one of the main parameters of urban transport performance. In particular, travel time forecasting is an important planning tool for public transport companies given that it can improve the quality of the planned services by reducing the gap between the actual and the planned travel times. In this paper, this relevance is discussed and based on the experimental evidences the goodness to use time series based approach is pointed out. In fact, among the large number of factors affecting the operation of public transport, most of them are shown to follow a given temporal pattern. The analysis is performed using data from automated vehicle monitoring of buses lines sharing the way with other traffic in Lviv (Ukraine). The results prove the goodness of such an approaches and the opportunity offered to operators to improve their services.

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

1. Introduction

In large historical cities, such as Lviv (Ukraine), overload on the road by traffic is a major problem facing widespread life. These problems are due to the growing number of private and commercial vehicles. In addition to other public services, buses play a key role in the city transport system. Therefore, to provide operators with performing tools for designing or revising service timetables could an effective tool for improving the attractiveness

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

10.1016/j.trpro.2020.02.109

of public services. In fact, the availability of timely and accurate information about bus travel time is significant, because it attracts more passengers and increases their satisfaction (Jeong and Rilett, 2004). Therefore, in order to provide the passenger with this type of information, there is an urgent need to develop models to forecast travel time with sufficient accuracy.

To accomplish this aim, a study was addressed to investigate bus travel times through time series methods, analyzing automated vehicle monitoring data of some buses operating in the city of Lviv, which share the lanes with other traffic users. Under such conditions, bus travel times are subject to a high degree of volatility, because their timing structures are similar to general traffic, which show seasonality and trend / cycle (Fusco et al., 2016; Comi et al., 2017a). Besides, the results of this study can help bus transport operators in improving their systems in terms of scheduling optimization and vehicle planning, as well as providing more accurate real-time information at stops.

The article is organized as follows. Section 2 briefly reviews long-term forecasting methods of bus travel time. Section 3 discusses the data available for this study. Section 4 presents analyses and results obtained. Section 5 draw conclusions and future research development.

2. Long-term bus travel time forecasting methods

Various authors have developed a variety of forecasting models to predict bus travel time (Pili et al., 2018; He et al., 2019). Widely used models of travel forecasting are based on regression model and time series.

Regression methods evaluate the value of dependent variables from the values of independent variables. The regression model can work in unstable driving conditions. Complex models such as regression of vector support, regression of k-nearest neighbor, regression of the project and artificial neural network are the most popular approaches to this problem. Since they are able to find complex nonlinear relationships between the target variable and the independent ones (Moreira et al., 2015; Shalaby and Farhan, 2003), they can work even traffic conditions are not stable. Such methods have been used by many authors (for example, Chen et al., 2004; Mendes-Moreira et al., 2012; Moreira-Matias et al., 2016), because they have a relative advantage in detecting which independent variables are smaller or more important for travel / forecasting travel time.

Methods based on time series, clarify the dependence on the travel time, which is observed in historical data (William and Hoel, 2003; Jeong, 2004; Billings and Jiann-Shiou, 2006; Suwardo et al., 2010; Comi et al., 2017a). The strength of methods based on time series is the high speed of computations due to the simple formulation of the algorithm and does not require a large number of variable bus operations: only the time which associated with the bus travel time (Moreira et al., 2015). They allow to specify the structure of the variability of travel time and to identify the impact on time (for example, daytime hours, days of the week, and periods of the year) that are related to the bus route. These models are certainly useful for revising the existing lines. If the line route and the structure remain the same and other variables (for example, the characteristics of the road network) do not change, then such analysis results may be useful in redefining more reliable charts and further planning of vehicles. Otherwise, regression models are generally better.

The above literature review indicates that there is a large number of studies on long-term forecasting of travel time. But only few studies have revealed the relationship between the indicators of transit systems and the variable time congestion (which follows from the overall variability of traffic). Their effects are significant in the development of the model for both long-term and short-term forecasting of travel time. Very few studies consider the time on the way from the terminal to the terminal (Cats, 2019; Cristóbal et al., 2019). The analysis is very relevant for transit operators who have to plan system schedules in order to limit service gaps during operations, while ensuring compliance with service obligations limits in terms of reliability (Cats, 2014). The complexity associated with the coincidence of various factors in the bus trip affects the bus time. Therefore, the needs for further research is growing.

The following sections contain the results of some analyses carried out, which combined AVM data on terminal-to-terminal travel time in the city of Lviv. This corridor, where buses share traffic lanes, are characterized by high variability. Attention can be subsequently paid to the recovery time, which is one of the main variables that must be taken into account when planning transit, which requires long-term forecasting. The analyzes were based on time series STL (seasonal and trend decomposition using Loess) decomposition, in order to recognize temporal patterns through the identification of trend/cycle, seasonality.

3. Data collected

The analyses synthetized below were performed using AVM data of the bus route #3A, operating in Lviv. The route connects two terminals (Figure 1): Terminal A - shopping center “King Cross Leopolis” (southern city); Terminal B - square “Rizni” (city center).

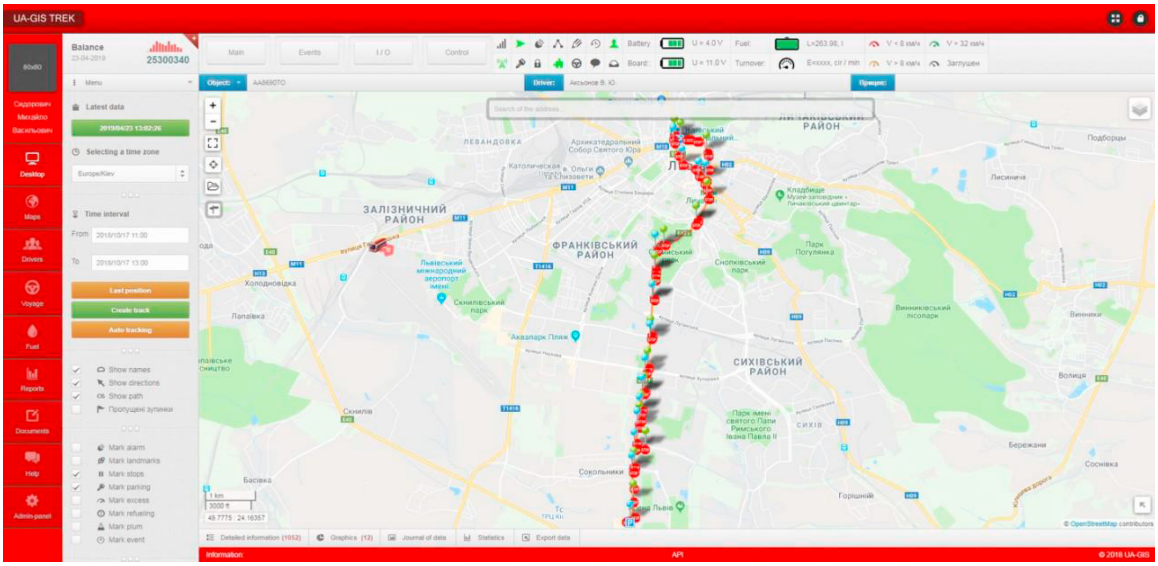


Figure 1. Stretch of the bus route #3A

Part of the route runs through the city center, where, in general, streets with high traffic and pedestrian flows. The other part of the route runs through the main street, where there are no obstacles to the movement of both private and public transport. Table 1 reports the main characteristics of bus line #3A. Data of bus travel time were obtained from the program “UA-GIS TREK” and are relative to the working days of each week in September and October 2018. This program is used by many Ukrainian (UA) companies to monitor bus services. The data of the bus route #3A was collected by monitoring of buses at the site http://track.ua-gis.com/ during for 10 weeks. The collected data include the time of departure from the initial stop and the time of arrival to the final stop for both directions. Then a first stage was performed to identify outliers in each time slices (external to 95-*th* percentile) and to solve the problem of missing data.

Table 1. Main characteristics of bus line #3A

operating time	Working hours: from 6:15 (terminal A)/ 7:00 (terminal B) to 23:00 headway: 7-15 min
route length of the route (for direction)	18.54 km (9.26 km / 9.28 km)
route time duration	30-60 min
number of stops	17/17 stops (to the city centre / from the city centre)
distance between stops	from 350 m to 1000 m
number of buses on the route	12-14
type of buses on the route	12-meter low-floor

Figure 2 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 that to city center two peak hours are present (i.e. one in the morning between 8:00 and 9:00 and one in the evening about at 6:00 pm), while from the city center the peak hour is in the evening about at 6:00 pm. In particular, for both directions, the variance is higher in the morning and evening peak hours than in the midday.

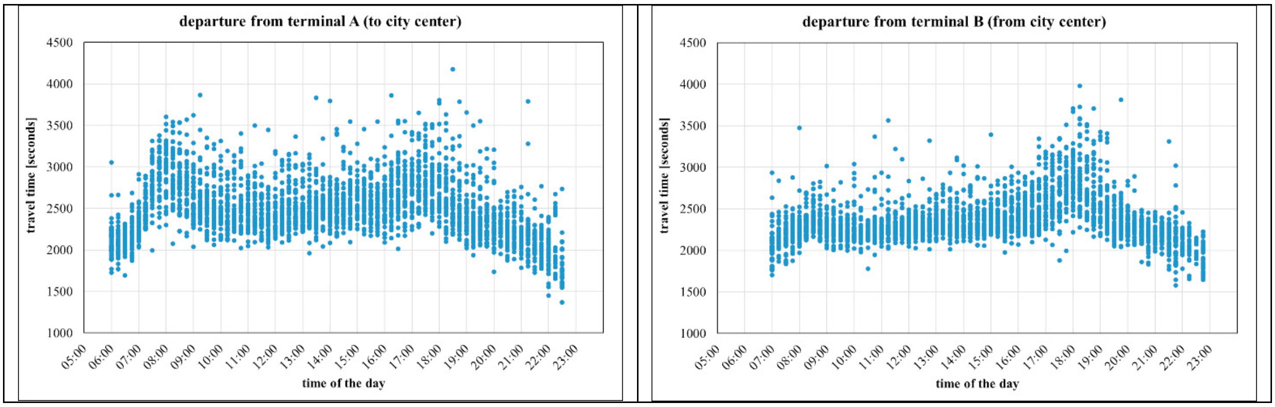


Figure 2. Hourly fluctuation of travel time of analyzed bus line #3A (working days).

4. Bus travel time analysis

Bus travel time (*TV*) from a terminal to a terminal can be represented as the sum of the running time (*RT*) and the dwelling time (*DT*). Running time depends on: flow speeds, link flow and flow composition; infrastructure link characteristics; functional link characteristics; context conditions, e.g. weather conditions, while dwelling time depends on alighting, boarding and on-board users, bus features.

Therefore, the variability of bus travel times can be analyzed 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 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). If an additive relationship is assumed follows:

$$Y_t = f\left(T_t, S_t, E_t\right) = T_t + S_t + E_t \tag{1}$$

where Y_t is the observed value at time t.

Then, the results germinating from time series decomposition can be used for forecasting. It is thus important to evaluate forecast accuracy. Consequently, the size of the residuals is not a reliable indication of how large true forecast errors are likely to be. The accuracy of forecasts can only be determined by considering how well a model performs on new data that were not used when fitting the model (Hyndman and Athanasopoulos, 2018). Therefore, the available data need to be divided into two portions, training and test data, where the training data is used to estimate any parameters of a forecasting method and the test data is used to evaluate its accuracy. Because the test data is not used in determining the forecasts, it should provide a reliable indication of how well the model is likely to forecast on new data. Subsequently, a forecast error (unpredictable part of an observation) is the difference between an observed value and its forecast. The “error” (e_{T+h}) can be written as

$$e_{T+t} = Y_{T+h} - \hat{Y}_{T+h} \tag{2}$$

where the training data is given by $\{Y_1, \ldots, Y_T\}$ and the test data is given by $\{Y_{T+1}, Y_{T+2}, \ldots Y_{T+h}, \ldots\}$, while the forecast is given by $\{\hat{Y}_{T+1}, \hat{Y}_{T+2}, \ldots, \hat{Y}_{T+h}, \ldots\}$.

Although forecast accuracy can be measured by different metrics, the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) are used (Hyndman and Athanasopoulos, 2018):

$$\text{Mean absolute error: MAE} = \text{mean}\left(\left|e_{T+h}\right|\right) \tag{3}$$

Root mean squared error: $\text{RMSE}=\sqrt{\text{mean}\left(e_{T+h}^{2}\right)}$ (4)

Mean absolute percentage error: $\text{MAPE} = \text{mean}\left(\left|p_{T+h}\right|\right)$ (5)

where $p_{T+h} = 100 \cdot e_{T+h} / Y_{T+h}$ is the percentage error, which has the advantage of being *scale-independent*.

4.1. Bus travel time forecasting

As said, the data of TV time series refer to observed values during working days of September and October 2018. The time-series period was set in a week (five working days), then the time series consist of 335 values (from terminal A to terminal B) and 320 values (from terminal B to terminal A) for each week (the average values are determined every 15 minutes). The analyses below were performed using software R and the package fpp2 (Hyndman and Athanasopoulos, 2018). Both time series were decomposed using STL decomposition method (Cleveland et al., 1990; Jeon and Hong, 2016), where the training set consisted of 8 or 9 weeks, while the test data are relative to the week ahead, i.e. 9th and 10th week. The two cases were used in order to test if a stability of the method used along the different number of data available exists. The results of determining the components of the STL decomposition are shown in Figures 3-4.

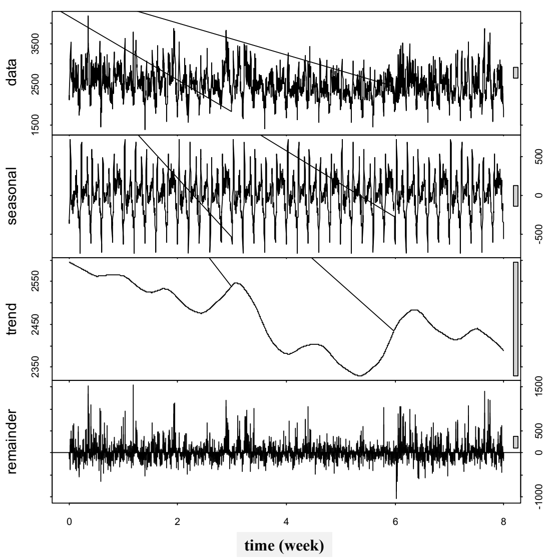
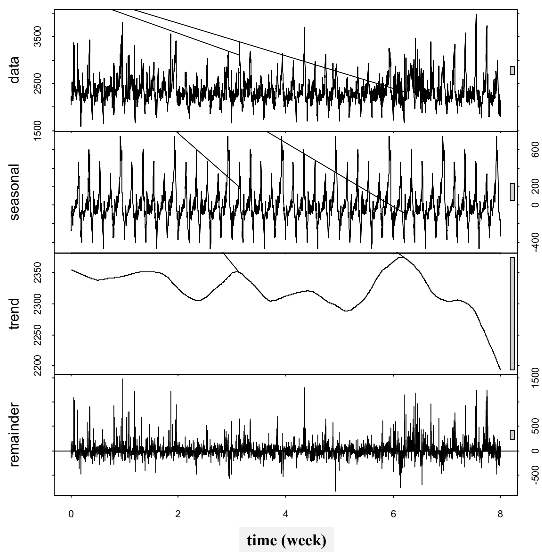
	From terminal A (to the city center)	From terminal B (from the city center)
Travel time STL decomposition of 8 weeks		

Figure 3. Bus travel time STL decomposition with weekly period of 8 weeks (September and October 2018).

Figures 3-4 show that the *trends/cycles* are quite flattened with a small difference between maximum and minimum values: 2.1% for bus travel time of 8 weeks and 2.4% for bus travel time of 9 weeks (departure from terminal A); 4.9% for bus travel time of 8 weeks and 2.9% for bus travel time of 9 weeks (departure from terminal B).

The effects of daily *seasonality* emerge for all days with differences according if to or from the city center, e.g. with larger values on Monday and Tuesday (departure from terminal A) and on Tuesday and Friday (departure from terminal B) (Figure 5). These results derive from different level of congestion occurring during the week in relation to the direction. Besides, 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. In addition, going to the city center, high concentrations were revealed in the morning (e.g. due to concentration of arrival constraints at work or at school), while moving from the city center such high concentrations were revealed in the evening (e.g. due to concentration of departures from work or shopping).

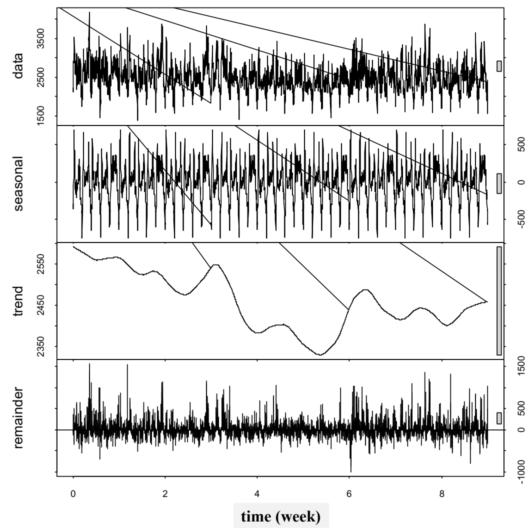
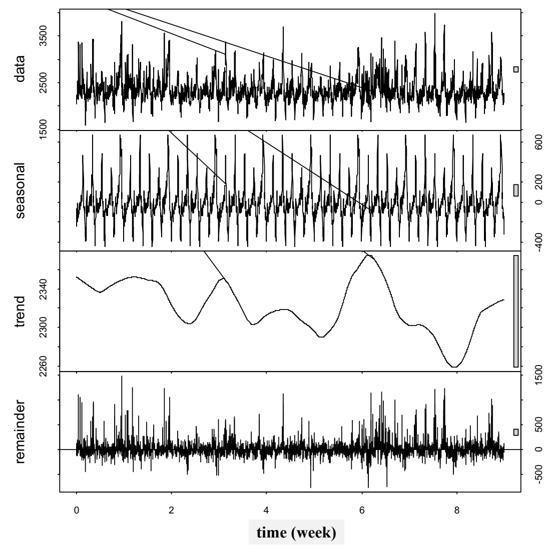
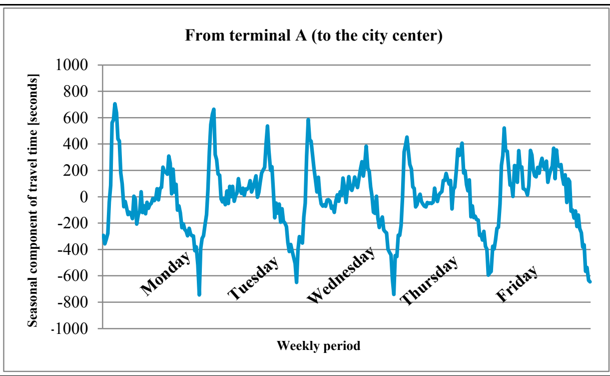
	From terminal A (to the city center)	From terminal B (from the city center)
Travel time STL decomposition of 9 weeks		

Figure 4. Bus travel time STL decomposition with weekly period of 9 weeks (September and October 2018).



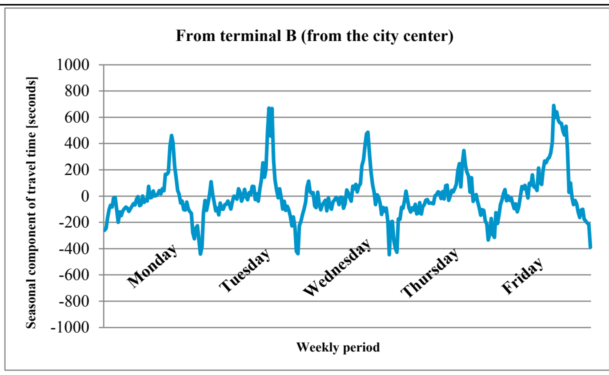


Figure 5. Hourly fluctuation of travel time in the time period for the analyzed bus line.

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 6 plots an extract of how reproduction works, while Table 2 synthetizes the above accuracy for whole investigated period of the two plotted routes (i.e. route to and from the city center) and the two datasets. As synthetized by MAPE, which is smaller than 7.9%, trend/cycle and seasonality allow the main part of variance to be explained. It means an average error in reproducing travel time of less than 2 minutes for route from terminal B (from city center) and less than 1 minute for route to 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 7 reports the comparison of forecasts for Monday from which it can be seen as the forecasted patterns fit quite well the observed values.

On the basis of the obtained data some findings can be pointed out:

- in bus services that share the way there is a pronounced dependence on the days of the week and of the time of the day;
- the chosen model for predicting time series is fairly accurate, since MAPE is smaller than 8% taking into account all periods, trend/cycle and seasonality allow the main part of variance to be explained;
- an average error in reproducing travel time is about 3 minutes for departure to the city center and about 2 minutes for departure from the 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.

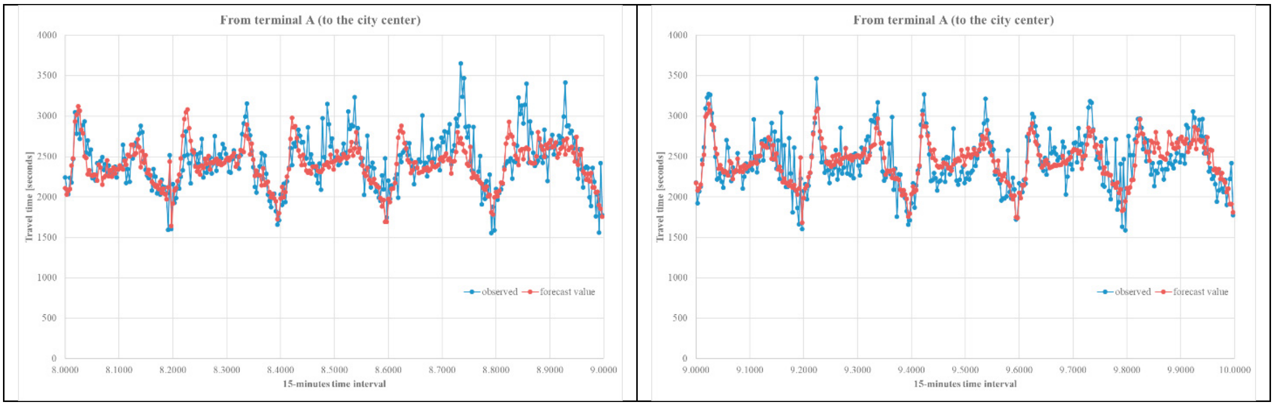


Figure 6. Example of travel time forecasting results for two datasets.

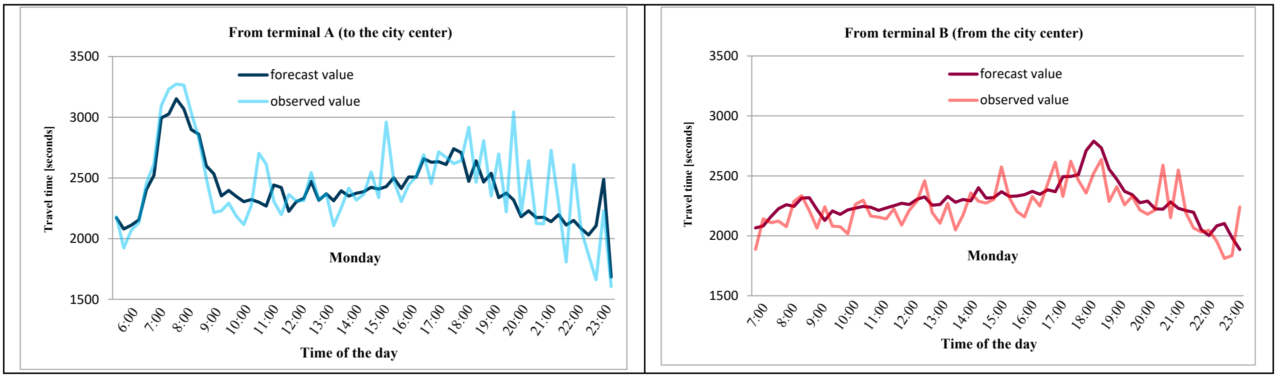


Figure 7. Example of travel time modelling for two directions.

Table 2. Accuracy indicators relative to the whole periods of analysis for the bus line #3A in Lviv

Number of forecasting weeks	From terminal A (to the city center)	From terminal B (from the city center)
9 weeks	average e_i = 53.7 standard deviation of e_i = 254.9 MAE = 191.7; RMSE = 260.2; MAPE = 7.9%	average e_i = 109.5 standard deviation of e_i = 192.6 MAE = 154.7; RMSE = 221.3; MAPE = 7.0%
10 weeks	average e_i = -2.8 standard deviation of e_i = 221.9 MAE = 171.6; RMSE = 221.6; MAPE = 7.1%	average e_i = -33.7 standard deviation of e_i = 173.6 MAE = 123.2; RMSE = 176.6; MAPE = 5.3%

Finally, 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 (Comi et al., 2017a and b), is a very typical observation of traffic condition. Yetiskul *et al.* (2012) revealed the effects of temporal dimension through automated vehicle location (AVL) data 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 Lviv data do not allow this last factor to be pointed out.

The proposed model is universal and self-regulated. This is evidenced by the fact that with an increase in the number of weeks studied, the prediction error remains quite constant (or decreases a little bit as expected). Therefore, the results obtained from this study can be used for planning new services on the same routes or to revise the existing timetables accordingly.

5. Conclusions

In this paper, the research of time series of bus travel times based on forecasting using STL decomposition was presented. The obtained results for Lviv can be used for designing transit service timetable and vehicle scheduling in re-planning of existing lines. 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. The results confirm the findings carried out in other world cities and the importance to design timetable according to time of the day and day of the week. Further developments in this study are in progress. They mainly concern additional investigations on new bus lines and determinants, and further types of automated data such as floating car data (e.g. data from private vehicles or taxi) for revealing (real-time) local traffic patterns, analysis of residuals deriving from decomposition method applying, in particular test of data autocorrelation and autoregressive integrated moving average (ARIMA) models, including the seasonal ARIMA (SARIMA) based regressions models and Neural Networks. Then, the profs and cons of each forecasting methods will be pointed out.

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*.

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

Cats., O., 2019. Determinants of bus riding time deviations: Relationship between driving patterns and transit performance. *Journal of Transportation Engineering Part A: Systems 145(1)*,04018078.

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.

Comi, A., Nuzzolo, A., Brinchi, S., Verghini, R., 2017a. Bus travel time variability: some experimental evidences. *Transportation Research Procedia 27*, pp. 101–108.

Comi, A., Nuzzolo, A., Brinchi, S. and Verghini, R. 2017b. Bus dispatching irregularity and travel time dispersion. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, DOI: 10.1109/MTITS.2017.8005632, pp. 856-860.

Cristóbal, T., Padrón, g., Quesada-Arencibia, A., Alayón, F., De Blasio G. et al., 2019. Bus Travel Time Prediction Model Based on Profile Similarity. *Sensors* 2019, *19*, DOI: 10.3390/s19132869.

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

He, P., Jiang, G., Lam, S. K. and Tang, D., 2019. Travel-Time Prediction of Bus Journey With Multiple Bus Trips. *IEEE Trans. on Intelligent Transportation Systems*.

Hyndman, R. J. and Athanasopoulos, G., 2018. *Forecasting: principles and practice*. Second edition. 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*. PhD thesis at Texas A&M University.

Jeong, SR. and Rillet, L.R.Hong, B.R., 201046. The Prediction of Bus Arrival Time using AVL dataMonte Carlo simulation-based traffic speed forecasting using historical big data. In: *Transportation Research Board 83rd Annual Meeting*, Washington, D.C.*F uture Generation Computer Systems, 65*, 182-195.

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.

Pili, F., Olivo, A. and Barabino, B., 2018. Evaluating alternative methods to estimate bus running times by archived automatic vehicle location data. *IET Intelligent Transport Systems 13(3),* pp. 523-530.

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

Shalaby, A. and Farhan, A., 2003. Bus travel time prediction model for dynamic operations control and passenger information systems. *Transp. Research Board 2*.

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)*, pp. 664-672.

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