V. Kovalyshyn
Lviv Polytechnic National University,
transport technologies department

FORECASTING OF THE TRAFFIC INTERVALS BETWEEN VEHICLES USING SOFTWARE R

The goal is to forecast time series of traffic intervals between vehicles on the main street under different conditions within one week and try to predict the intervals between vehicles for the future. To accomplish this goal it is decided to apply the modern computer Software R.
Key words: traffic intervals, vehicles, time series, drivers, traffic flow, software R.

В. В. Ковалишин
Національний університет "Львівська політехніка",
кафедра транспортних технологій

ПРОГНОЗУВАННЯ ІНТЕРВАЛІВ РУХУ МІЖ ТРАНСПОРТНИМИ ЗАСОБАМИ З ВИКОРИСТАННЯМ ПРОГРАМНОГО ПРОДУКТУ "R"

Спрогнозовано часові ряди інтервалів руху між транспортними засобами на головній вулиці за різних умов руху протягом одного тижня і визначено інтервали між транспортними засобами на майбутнє із застосуванням сучасного програмного продукту R.
Ключові слова: інтервали між транспортними засобами, автомобілі, часові ряди, водії, транспортний потік, програмний продукт R.

Problem formulation and analysis of well-known research and publications.

In large historical cities, such as Lviv, overload on the road by traffic is a major problem facing our widespread life. These problems are due to the growing number of private and commercial vehicles. Therefore, forecasting of the traffic intervals between vehicles on the main street will improve the quality of the transport system.

Drivers should have increased attention and good concentration during driving a vehicle in a dense traffic flow [1]. Driving in such flow a lot depends on the driver of the vehicle that is moving in front ("driver-leader"). Under such conditions, the visibility of the lane in front of the "car-leader" is limited. Therefore, the driver who moves behind is much more difficult to predict in advance the reasons for a possible decrease in speed or even an emergency stop of the car. The greatest danger in these conditions is the movement along the main streets, especially at the moment of acceleration at a distance of 50–100 meters behind the regulated intersection [1, 2].

Various authors have developed a variety of forecasting models to predict the traffic intervals between vehicles. Widely used models of travel forecasting based on time series and regression model.

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 [3]. They can work even traffic

conditions are not stable. Such methods have been used by many authors [4–5], because they have a relative advantage in detecting which independent variables are smaller or more important for forecasting traffic intervals between vehicles.

The analyzes were based on time series STL decomposition, in order to recognize time models, analyze the weekly data on the traffic flow and to identify the component of the trend, seasonality and cycle, which are very similar to different vehicles.

The main material. The study of time series of intervals between vehicles in the traffic flow was carried out on one of the streets of Lviv, which is regulated by traffic lights with a dividing line in a period from Monday to Friday. In that street, the intensity of the traffic flow is in the range from 900 to 1200 veh/h and the average share of passenger cars – 80 %.

Time series research was conducted using video, covering two lanes in one direction. The values of time series of traffic intervals between vehicles are received for each day of the week [6].

At the first stage a database of time series of intervals was downloaded into Software R [7]. Software R is a free software environment for statistical computing and graphics. It is recommend to use Software R with RStudio. Installing and run Software R:

ü download and install Software R [7];

ü run Software R and open Package "fpp2";

ü use command TV=scan("clipboard",sep=",") without push "Enter". Copy the data of distance time (TV). Return to R and click "Enter". The data should be loaded. To watch them: ts.plot(TV).

The time series of traffic intervals consists of 164 values for each day of the week (in the morning). To build a data time series in Software R use functions [1–2,6]:

$$ts.TV=ts(TV,frequency=164, start=c(0,1)); \qquad (1)$$

$$plot(ts.TV, ylab="Distance Time [Miliseconds]"); \qquad (2)$$

where

ü frequency – the number of observations per unit of time;

ü start – the time of the first observation;

ü ylab =Y axis label.

Based on these functions in Software R, the dependence "traffic intervals between vehicles – number of days" displayed graphically (Fig. 1).
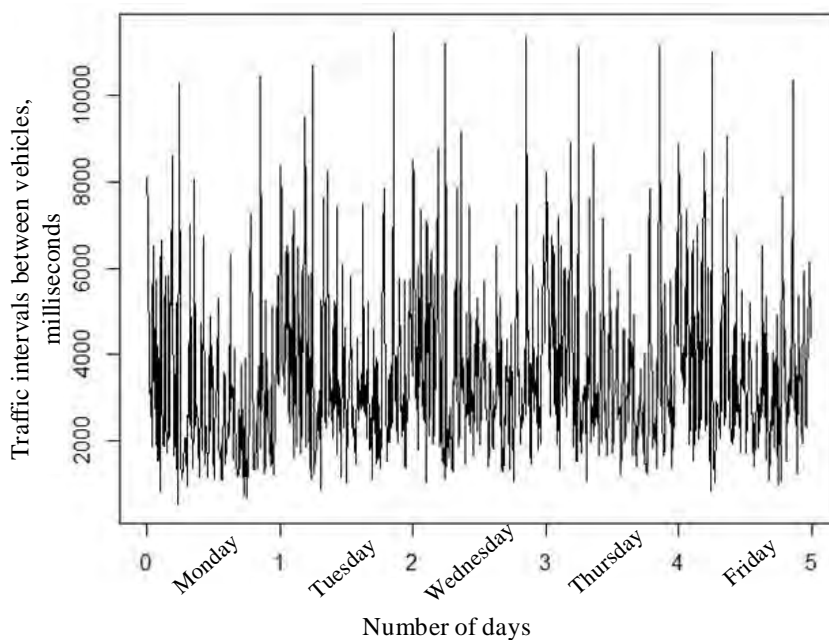

Fig. 1. Traffic intervals between vehicles – number of days

85

At the next stage it is carried out STL decomposition – Decompose a time series into seasonal, trend and irregular components using loess, acronym STL [8]:

$$\text{stl.TV=stl(ts.TV, t.window=164, s.window="periodic", robust=TRUE).} \tag{2}$$

where
- ù t.window – the span (in lags) of the loess window for trend extraction, which should be odd;
- ù s.window – either the character string "periodic" or the span (in lags) of the loess window for seasonal extraction. This has no default; robust – logical indicating if robust fitting be used in the loess procedure.

STL decomposition is used for forecasting of time series of traffic intervals between vehicles for the future period. The results of determining the components of the STL decomposition are shown in Fig. 2.
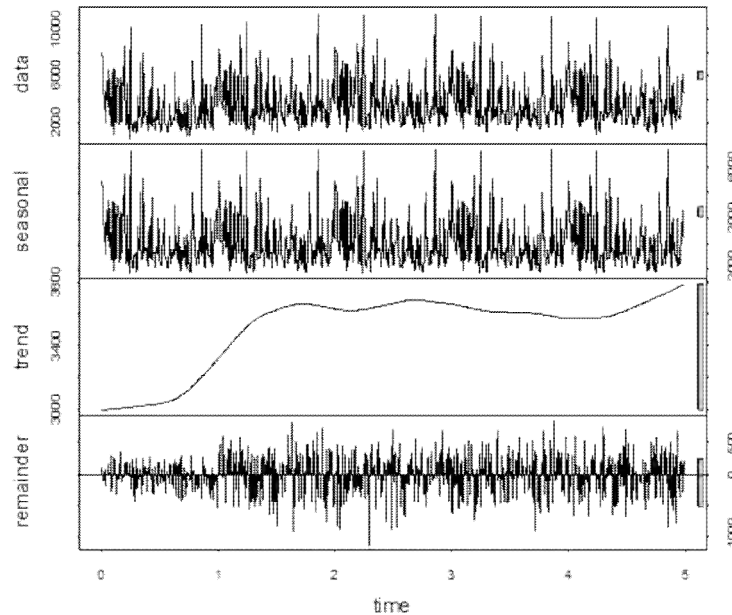

Fig. 2.STL decomposition of time series in Software R for all periods

Figures 2 shows that the trends/cycles are quite flattened with a small difference between maximum and minimum values – 0.8 second.
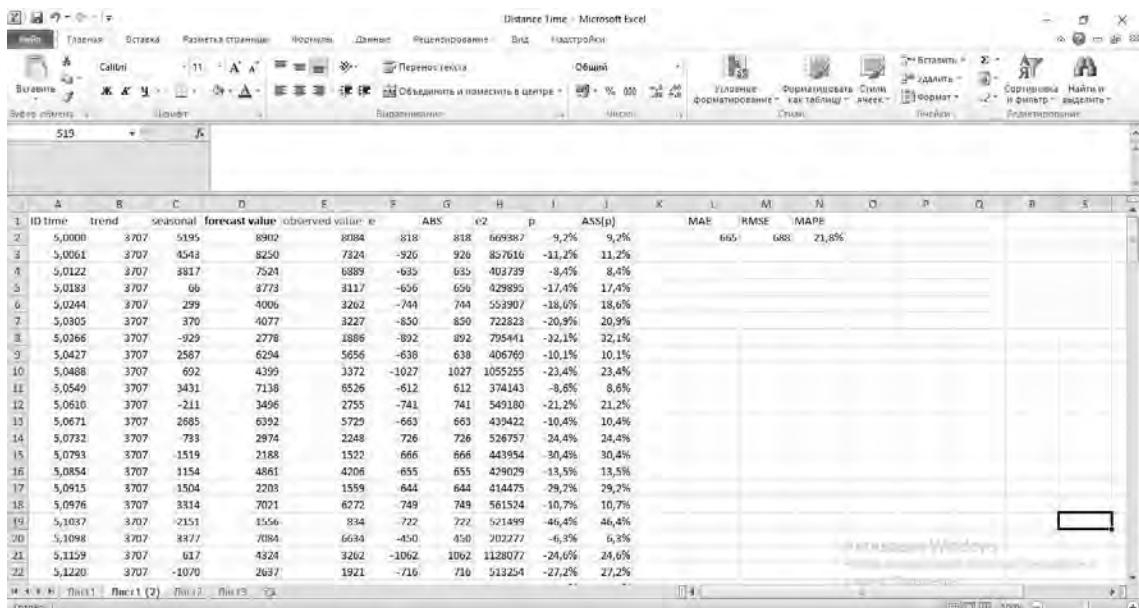

Fig. 3. Transferring the results of the forecasting from Software R to Microsoft Excel

86

To check the results of forecasting, the values (obtained during the STL decomposition of time series) from Software R is transferred to Microsoft Excel (Fig. 3).

Using the trend and the seasonal variable, we can get the predicted result of the time series of vehicle intervals in a certain period. For an example in Microsoft Excel, the forecast is scheduled for the next business day, Monday (Fig. 4).
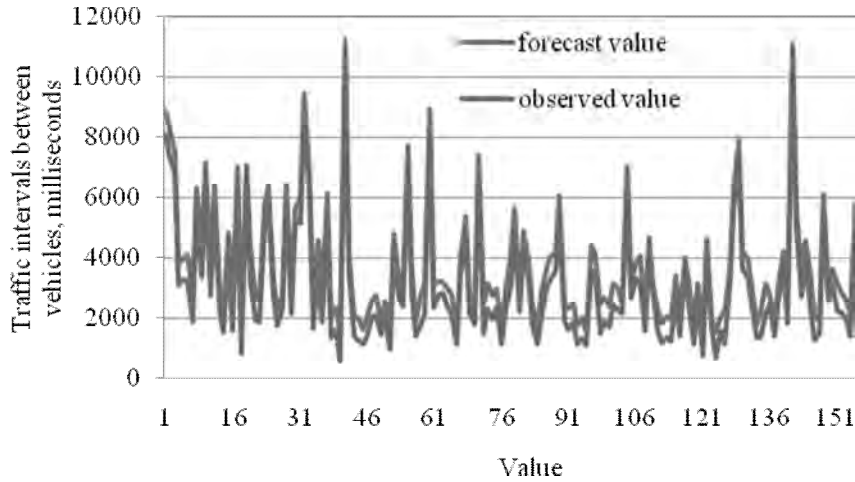

Fig. 4. Comparison of predicted and existing values
of time series of traffic intervals between vehicles

The results germinating from time series STL decomposition can be used for forecasting. It is thus important to evaluate forecast accuracy using genuine forecasts. 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 [9]. 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. The forecast error (unpredictable part of an observation) (e_{T+h}) can be written as:

$$e_{T+t} = Y_{T+h} - \hat{Y}_{T+h} \tag{3}$$

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

When comparing the obtained values with the existing, the average error is defined in milliseconds and percentages. As a result it is defined indicators such as: MAE (Mean Absolute Error), RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) [9]:

$$\text{Mean absolute error: MAE} = \text{mean}\left(\left|e_{T+h}\right|\right)$$
$$\text{Root mean squared error: RMSE} = \sqrt{\text{mean}\left(e_{T+h}^2\right)} \tag{4}$$
$$\text{Mean absolute percentage error: MAPE} = \text{mean}\left(\left|p_{T+h}\right|\right)$$

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

It is determined that MAE is 665 milliseconds, RMSE – 21.8 %, MAPE – 688 milliseconds. The results of forecasting can also be estimated using the graph.

Conclusions. On the basis of the obtained data it is possible to make such conclusions:

Ù in the data there is a pronounced dependence on the days of the week;

Ù the chosen model for predicting time series is fairly accurate, since the average error is 21 %.

The proposed model is universal and self-regulated. To improve the forecast, it is necessary to continue collecting data and refining the model. As more data the better is forecast.

Knowing the predicted intervals between vehicles, we can determine how much traffic, density, dynamic size and speed. Since these indicators are dependent.

1. Bolshinsky, E. and Freidman, R., 2012. Traffic flow forecast survey. Technion–Israel Institute of Technology–Technical Report. 2. Jeon, S. and Hong, B., 2016. Monte Carlo simulation-based traffic speed forecasting using historical big data. In: Future Generation Computer Systems, 65, 182–195.3. 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. 4. 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. 5. 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. 6. Hyndman, R. J. and Athanasopoulos, G., 2016. Forecasting: principles and practice.www.otexts.org. 7. https://cran.r-project.org. 8. 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). 9. Hyndman, R. J. and Athanasopoulos, G., 2018. Forecasting: principles and practice.Second edition. www.otexts.org.