FORECASTING TRAFFIC FLOW AT THE INTERSECTION BASED ON CYCLICAL FLUCTUATIONS

GRZESICA Dariusz

Krakow University of Technology, Faculty of Civil Engineering, Cracow, Poland, EU

Abstract

Forecasting future volume of traffic at the intersection may lead to increase the drivers safety on city roads. The possibility of realization forecasting methods for the transportation system is an important element of traffic management in the city. In the article the forecast for future periods on the basis of cyclical fluctuations in the traffic flow using the method of harmonic analysis will be calculated. The forecast data obtained from the intersection of the Słowackiego and Łobzowska street were registered with an interval of time during one week. Cyclical fluctuations are typical for the period exceeding one year. The article aims to prove the application of harmonic analysis for hourly periods in forecasting traffic flow at the intersection. The forecast is confirmed using mean absolute percentage error for long-time periods.

Keywords: Forecasting traffic flow, cyclical fluctuations, spectral analysis

1. INTRODUCTION

When trying to explain changes in the time series, it is necessary to perform its decomposition, i.e. the extraction of the elements through which it is characterized. Splitting the time series into separate components it allows us to understand the fluctuations that occur in the studied phenomenon. The components are as follows:

- secular trend or long term variations,
- seasonal variations,
- cyclical variations,
- irregular variations [1].

To isolate the impact of cyclical fluctuations in the model-building process and its consideration in the forecasting process improves prediction accuracy.

In the analysis of time series, you notice it changes, cyclical, periodic. They occur at the same stage of change for equal intervals of time. The time period for the development of all phases, is called the oscillation period or cycle. Cyclical fluctuations are difficult to simulate because of the length of cycles and amplitude of oscillations which are less regular than seasonal fluctuations. Cyclicity may occur in cycles of different frequencies, for example, it is high when fluctuations occur frequently; low, if fluctuations occur rarely. Cyclicity may also occur in cycles of different length (short, long) and different nature (regular and irregular). Therefore, it is understood as a change that is repeated in a regular, rhythmical, recurring at intervals.

In the transport field methods of spectral analysis that identifies cyclical fluctuations are rarely used. Usually the yearly changes are studied under trend, therefore, the changes that occur for period more than one year come under the category of cyclic changes [2]. Increasingly, scientists use the harmonic analysis method in search of cyclical fluctuations in periods shorter than one year. In work [3], he used it for a prediction in the field of transportation as real-time traffic flow. In turn, Zhang [4] as an element of a hybrid method use spectral analysis in short-term traffic flow forecasting. The efficiency of harmonic analysis for a short period of time based on product demand was demonstrated in work [5].

This work is an attempt to create predictions based on spectral analysis, i.e. one which takes into account the cyclical nature of the time series, as one of the elements of its decomposition. This method does not fully identify all aspects of cognitive time series and draws attention to the cyclical changes, as an element playing important role in the process of forecasting future values.

2. BACKGROUND SPECTRAL ANALYSIS

Spectral analysis is one of analysis, which is transforming the time series into the frequency domain. This method appeared thanks to J. Fourier, in 1807, proved that the values of the time series exhibit a certain periodicity, can be aproximated with any accuracy through the ranks with a correspondingly large number of elements. These series are called harmonic Fourier series.

Spectral analysis is a kind of converting time series into the sum of the sine waves and cosine [6]. Fourier analysis is essentially concerned with approximating a function by a sum of sine and cosine terms, called the Fourier series representation [7]. The model is constructed as a sum so-called harmonics. The quantity of harmonics for *n* observations is *n* / 2. First harmonic has a period equal to *n* - number of all observations, the second *n* / 2, the third *n* / 3, etc. Cyclicity contained in the time series can only be define by harmonic analysis [8].

An important element why using spectral analysis to construct a predictive model is to test the time series from the point of view of stationarity, i.e., constant in time average, variance and autocorrelation of the studied series [9]. To examine the stationarity of the time series the advanced Dickeya and Fuller test (ADF) developed by Dickeya and Fuller in 1979 will be used [10]. It removes the influence of autocorrelation in the series. In the case of instability, it is necessary to subject the differencing procedure in accordance with the formula $\Delta y_t = y_t - y_{t-1}$.

This procedure is used to achieve time series stationarity, however, as a rule, usually no more than three times. The conclusion about the stationarity of time series, most often performed for a significance level equal to *p* = 0.05.

3. FORECASTING TRAFFIC FLOW AT THE INTERSECTION USING HARMONIC ANALYSIS

Empirical data, components of time series are the number of cars passing through the intersection of Słowackiego and Łobzowska streets in the period from December 1, 2014 at 1:00 p.m. to 7 December 2014 at 23:00 in hourly periods. Data from each lane at the inlet, they are summed and represent the output of one of the four inlets , which will be considered individually. For prediction at each inlet the data of six days (142 observations), whereas data from the last day to verify forecast (24 observations) will be used.

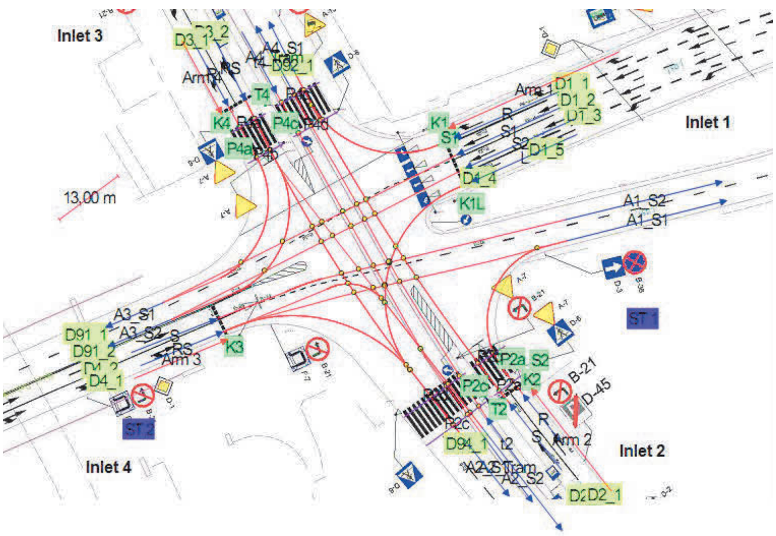

Figure 1 Four-way intersection for studied example (horizontal - Słowackiego street, vertically - Łobzowska street)

The time series graph on each of the four entrances to the intersection in the course of a week is shown in **Figure 2**.

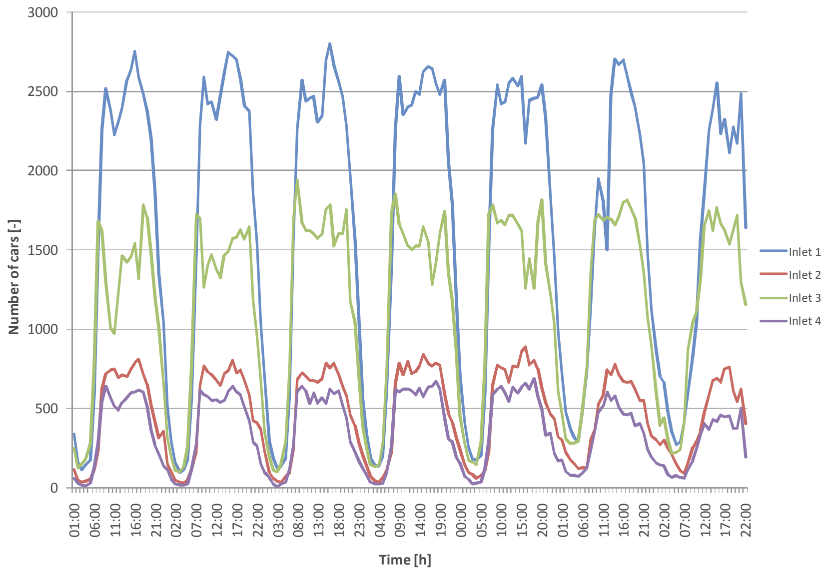

Figure 2 The time series of the number of vehicles at each of the four inlets to the intersection

First, the time series must be checked for the presence of stationarity. Calculated ADF test for all four inlets is sequentially: for inlet 1, p = 0.0000062, for inlet 2, p = 0.0000095, for inlet 3, p = 0.0000001, for inlet 4, p = 0.0000202.

The p coefficient of each inlets of intersection is below the level of significance p = 0.05, therefore, all time series are stationary and there is no need for their differentiation.

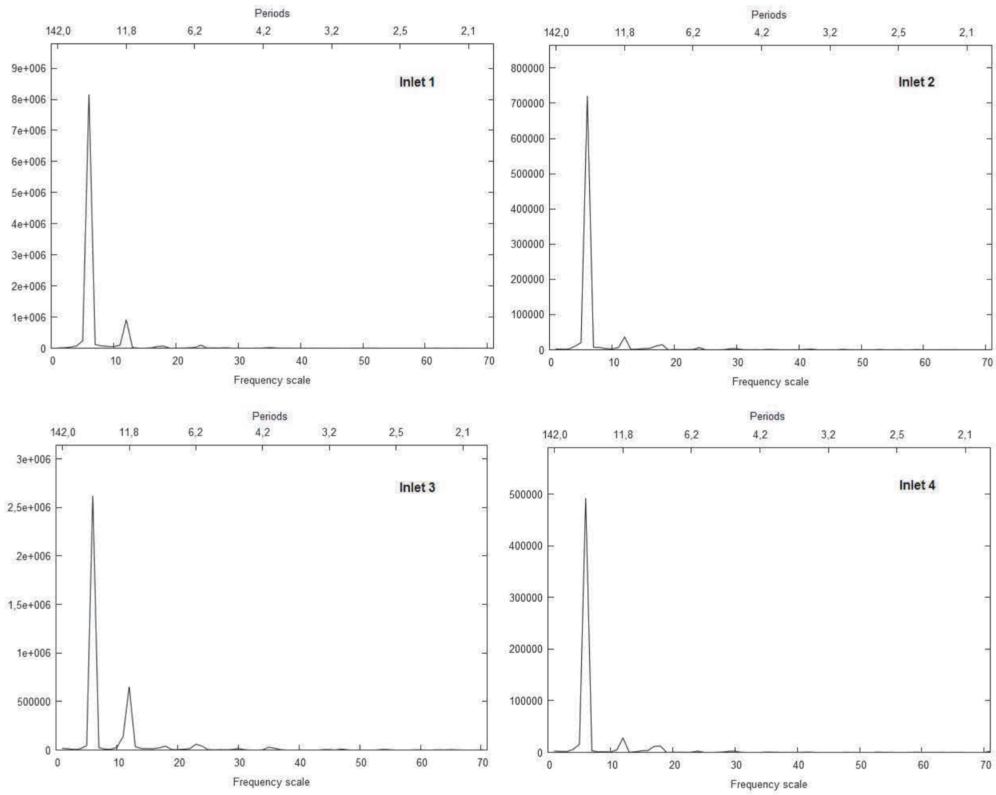

Figure 3 Periodograms of the time series for each of the inlets

The basis of spectral analysis is to examine if the time series contains cyclical changes. For this purpose, periodogram, which is designed for detection of periodic oscillations [11]. Periodogram of the number of cars passing through the intersection on any of the inlets is shown in **Figure 3**.

On the basis of the periodograms, we can conclude that the time series has cyclic fluctuations. In this regard, the harmonics in accordance with the formula (8) and (9) will be assigned. The values of some harmonics that have the largest share in explaining the time series are presented in **Table 1**. They will be required to build the model and then to determine forecast for future periods.

Table 1 Harmonics used to build the model

Number of harmonic	Designation	Inlet 1	Inlet 2	Inlet 3	Inlet 4
		Contribution [%]			
5	n/5	-	2.29	-	2.60
6	n/6	79.69	81.81	65.87	81.16
11	n/11	-	-	3.47	-
12	n/12	8.90	4.11	16.39	4.57

Source: own elaboration

The dominance of the harmonic No 6 on each of the inlets is a result of the occurrence of a traffic lights at the intersection. It affects the appearance of cycles in traffic.

Using harmonics from the **Table 1** and the model parameters in the process of converting a time series into a frequency domain the models were built. For each inlet models are as follows:

$$\boldsymbol{y_{R1}} = 1684{,}23 - 1023{,}68 \cdot sin\left(\frac{2\pi}{142}t\right) - 629{,}201 \cdot cos\left(\frac{2\pi}{142}t\right) - 380{,}486 \cdot sin\left(\frac{2\pi}{142}t\right) + 128{,}495 \cdot cos\left(\frac{2\pi}{142}t\right)$$

$$\boldsymbol{y_{R2}} = 476{,}73 + 45{,}09 \cdot sin\left(\frac{2\pi}{142}t\right) + 39{,}3 \cdot cos\left(\frac{2\pi}{142}t\right) - 299{,}32 \cdot sin\left(\frac{2\pi}{142}t\right) - 194{,}81 \cdot cos\left(\frac{2\pi}{142}t\right) - 67{,}47 \cdot sin\left(\frac{2\pi}{142}t\right) + 43{,}14 \cdot cos\left(\frac{2\pi}{142}t\right)$$

$$\boldsymbol{y_{R3}} = 1146{,}54 - 525{,}5 \cdot sin\left(\frac{2\pi}{142}t\right) - 432{,}68 \cdot cos\left(\frac{2\pi}{142}t\right) + 148{,}39 \cdot sin\left(\frac{2\pi}{142}t\right) - 49{,}22 \cdot cos\left(\frac{2\pi}{142}t\right) - 336{,}11 \cdot sin\left(\frac{2\pi}{142}t\right) + 48{,}\cdot cos\left(\frac{2\pi}{142}t\right)$$

$$\boldsymbol{y_{R4}} = 367{,}71 + 36{,}03 \cdot sin\left(\frac{2\pi}{142}t\right) + 38{,}6 \cdot cos\left(\frac{2\pi}{142}t\right) - 222{,}83 \cdot sin\left(\frac{2\pi}{142}t\right) - 193{,}53 \cdot cos\left(\frac{2\pi}{142}t\right) - 66{,}88 \cdot sin\left(\frac{2\pi}{142}t\right) + 20{,}89 \cdot cos\left(\frac{2\pi}{142}t\right)$$

Based on the constructed models the coefficient of determination R^2 was determined. It gives a proportion of the variance in the dependent variable that is predictable from the independent variable. The coefficient of determination is assigned by the formula:

$$R^2 = \frac{\sum_{i=1}^{n}(\check{y}_t - \bar{y})^2}{\sum_{i=1}^{n}(y_t - \bar{y})^2} \tag{1}$$

where:
y_t - actual value of the variable in period t,
$\check{y}_t$ - model value,
$\bar{y}$ - arithmetic mean of the dependent variable.

The determined coefficients of determination equal:

- for inlet 1, y_{R1} = 0.871,
- for inlet 2, y_{R2} = 0.947,
- for inlet 3, y_{R3} = 0.788,
- for inlet 4, y_{R4} = 0.912.

To create predictions for each of the inlets of the intersection constructed model can be used. Substituting the future variable of *t* in the formula, it is possible to forecast values for future periods. Of the 166 142 observations were used for making models, while the remaining observations were used to verify the forecast. Forecasts along with the actual data is shown in **Figure 4**.

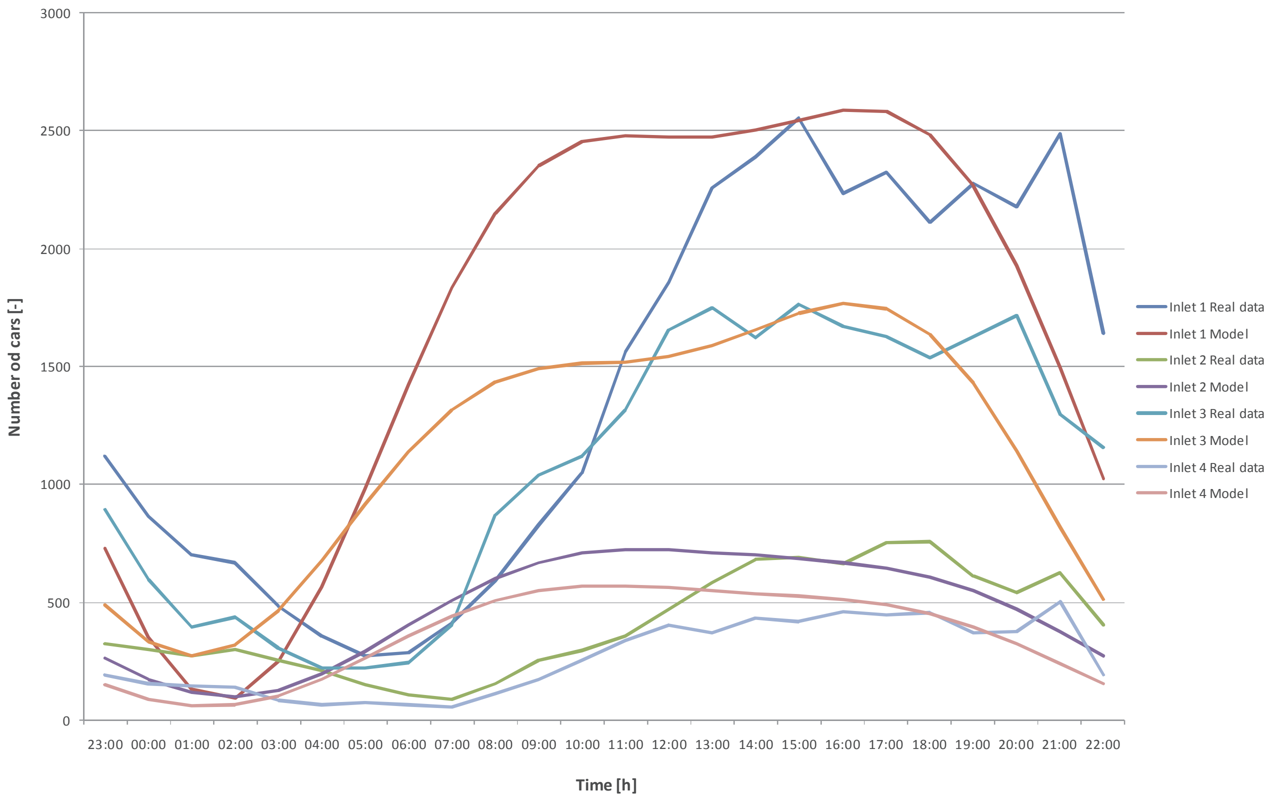

Figure 4 Actual data along with the forecast

Figure 4 shows, a relatively good coverage predictions with the actual data. It should be noted that the forecast was calculated for 24 periods ahead. In the case of long-horizon forecasting, we can conclude that the farther into the future, the forecast is fraught with even greater error. In the presented example, it is possible to observe that the cyclic factor, acting in the considered time series has consequences even in periods far from the beginning of the forecasting period. It is well illustrated by using mean absolute percentage error (MAPE) in **Figure 5.**

The largest forecast error for all inlets is from 6th to 14th period. The largest forecast error is 662% for the fourth inlet in the 9th period forecast, and the lowest of 0.31% for the first inlet in 17th of the forecast period. In **Figure 5** you can see that with the 14th period, the average forecast error for the subsequent periods is very small and ranges from 0.31 % to almost 20 % in long-term period of time.

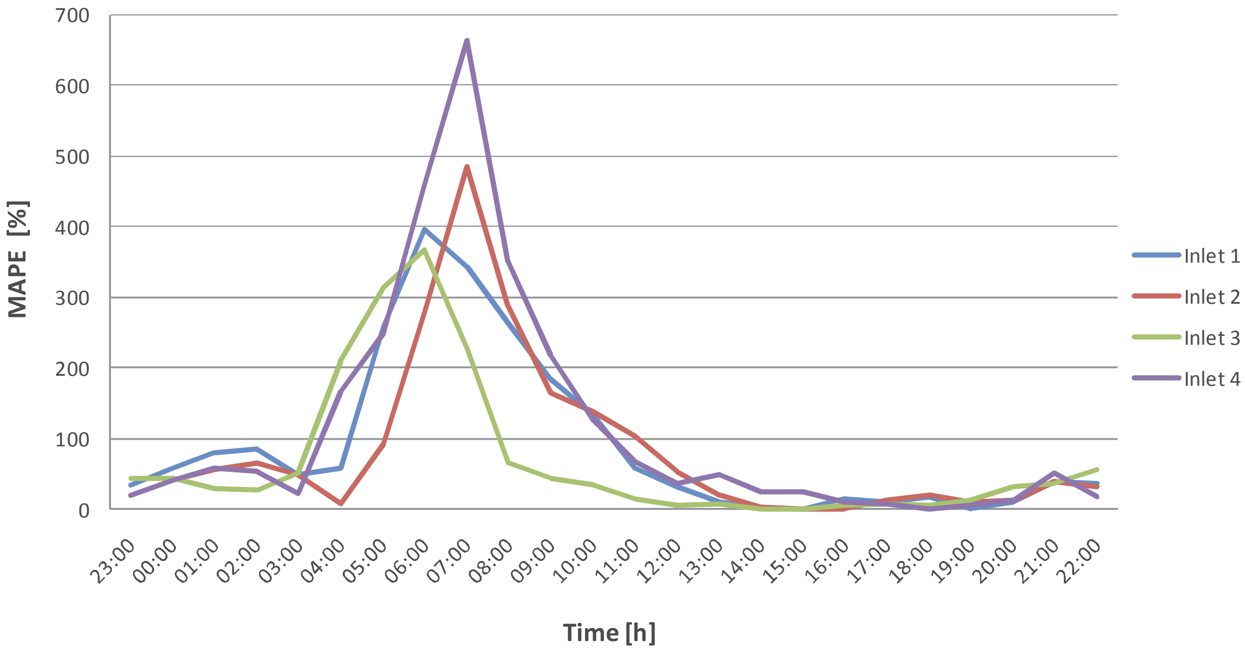

Figure 5 Mean absolute percentage error for 24 periods ahead

4. CONCLUSIONS

From the conducted research it follows that the number of vehicles at the traffic light intersection is characterized by cyclic changes. The largest share in explaining the time series have a harmonic no 12, that is, 142 / 12 (*n* / 12) means that the number of cars varies cyclically every 12 hours. This allowed to build a model for each of the inlets with a high adoption rate of the model on real data.

On the basis of the model the forecast for 24 periods into the future was created. The forecast has been tested on the basis of the mean absolute percentage error. The results are impressive, because the farther into the future, the forecast is more accurate. This is due to the work of the traffic lights at the intersection, which provides the cyclic changes of passing vehicles. Periods in which recorded the highest mean absolute percentage error are periods devoid of cyclical fluctuations and vice versa. The results show that on the basis of cyclic changes, you can achieve a relatively accurate long-term forecast.

Based on these results emphasize the usefulness of harmonic analysis in the long-term forecasting. Further studies to establish the length of time that you can predict the future based on the spectral analysis will be carried out.

REFERENCES

[1] GUPTA P.N. Comprehensive Business Statistics. Laxmi Publications Ltd, New Delhi 2005.

[2] PILLAI R. S. N. & BAGAVATHI Statistics (Theory & Practice). S. Chand Publishing, New Delhi 2008.

[3] TCHRAKIAN T. T., BASU B., O'MAHONY M. Real-Time Traffic Flow Forecasting Using Spectral Analysis. IEEE Transactions on Intelligent Transportation Systems, Vol. 13, Issue 2, 2012.

[4] ZHANG Y. ZHANG Y. HAGHANI A.. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. Transportation Research Part C-Emerging Technologies. Special Issue 43, 2014.

[5] GRZESICA D., WIĘCEK P. Advanced Forecasting Methods Based on Spectral Analysis. Procedia Engineering Volume 161. World Multidisciplinary Civil Engineering-Architecture-Urban Planning Symposium 2016. WMCAUS 2016.

[6] BLOOMFIELD P. Fourier analysis of time series: an introduction. John Wiley & Sons, 2004, pp. 3.

[7] CHATFIELD C. The Analysis of Time Series: An Introduction, Sixth Edition. CRC Press, 2016.

[8] KOOPMANS L. H. The Spectral Analysis of Time Series: Probability and Mathematical Statistics. Academic Press, New Mexico 1974, pp. 195.

[9] WIĘCEK, P., GRZESICA D., KUBEK D. The advanced prediction method of inventory costs in coal mine industry. In Proceedings of the Second International Conference on Traffic and Transport Engineering ICTTE, November 27-28, Belgrade 2014.

[10] DICKEY D.A., FULLER W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, No. 74., 1979.

[11] BOX G. E. P., JENKINS G. M., REINSEL G. C., LJUNG G. M. Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015, pp. 198.