Agronomy Research **17**(S1), 1080–1088, 2019
https://doi.org/10.15159/AR.19.043

Periodic polynomial regression analysis of urban driving characteristics

J. Kreicbergs and A. Grislis*

Riga Technical University, Faculty of Mechanical Engineering, Transport and Aeronautics, Department of Automotive Engineering, 1 Kalku Str., LV1658 Riga, Latvia
*Correspondence: juris.kreicbergs@rtu.lv

Abstract. Urban driving characteristics with a focus on energy consumption have been tested in Riga on three main city streets with inflexible coordinated traffic lights control. The aim of this article is to investigate periodic polynomial regression analysis method to analyse car urban driving parameters' change during weekday twenty-four hours to assess the influence of vehicle technologies on energy consumption in city driving, to map the energy demand on Riga city main street sections and to evaluate the traffic lights control on flow energetic characteristics. The tests have been done using GPS and OBD data loggers on a test car repetitively driven along a pre-planned route at around-the-clock hours. A regression analysis using periodic polynomials was developed and applied to evaluate the traffic flow characteristics with a given time shift. It was concluded that using polynomial regression function, the polynomial order has to be at least seven although a visual conformation of good regression line to the measured data has to be checked especially with lower orders. To evaluate the traffic conditions at a given 20 minutes to one hour shift the application of regression function is limited for the periods with fast changing traffic flow, especially after the end of rush hours when the usability of regression line for the given data has to be checked individually for tested street sections.

Key words: city driving, energy consumption, traffic flow.

INTRODUCTION

Energy consumption reduction has been an important issue for many industries. As shown by IEA (2018) passenger cars in IEA countries use above one fifth of the total energy. As classified by Fontaras (2017) the car energy consumption in urban traffic depends on vehicle technologies, environmental and traffic conditions and vehicle drivers actions. Modern vehicle technologies are focused first on safety, environment and performance, in many ways increasing energy requirements of the vehicle while certain technologies contribute to energy savings and it is important to understand the scope of usability of the technologies.

The research of vehicle energy consumption on three major Riga city streets was performed to assess the influence of vehicle characteristics and technologies on energy consumption in city driving, to map the energy demand on street sections and to evaluate the traffic lights control on flow energetic characteristics.

With the highly limited budget for the research, the traffic flow rate was measured by analysing the opposite traffic flow from the videos captured during the tests.

According to the test plan the cars return to the same street sections in the opposite direction after 10 to 11 km meaning the difference in time from 20 minutes to one hour. To evaluate the traffic flow at the time of test, it was decided to find regression lines that may allow to evaluate the traffic rate at least within the given time shift.

Traffic flow as a function of time of the day characteristics are documented by many researchers and are shown in Traffic Engineering textbooks like by Mannering et al. (2009), Garber & Hoel (2010), Roess et al. (2011) etc. Usually there are peaks for the morning and evening rush hours, calm traffic during the night and different transition periods between the peaks. The best regression line can be found if the functional relationship is exactly known. There is no one regression line that fits all the circumstances because traffic flow parameters depend on many factors, but it is clear that the function has to allow at least two peaks at it has its range of definition within 24 hours and it has to be periodic by this meaning that the end and the start of the day is a continuous process.

Periodic functions can be well analysed by Fourier series, but since most of the data processing was done in MS Excel, the closest built trendline fitting the data was polynomial. The function range does not have negative values and is limited by 24 hours, having value of 1 in MS Excel, therefore a truly periodic function is not needed, and the function has to give values in a single period only. Every new series in a Fourier transform allow the number of peaks doubled while every next order of polynomial adds just one extreme therefore it was decided to analyse polynomial regression lines where by setting certain limits the values and the first derivatives at the ends of the period may be equalized. As noted by Cobanovic et al. (2006) periodic regression is seldom included in syllabus of statistical course therefore there was made decision to make an exercise by finding the regression polynomial model for the performed urban driving tests. The found regression lines are not tested for other research data but well served for the analysis of the acquired data and may be extended to other driving characteristics data spread without essential time gaps along 24 hours of the day.

MATERIALS AND METHODS

Driving tests

The data have been taken from multiple driving tests performed in Riga city that have been described by Kreicbergs et al. (2018). The tests analysed in this paper were done with 2004 production year Opel Zafira with a 1.6i-16V petrol engine, driving more than 1000 km and 45 hours along a pre-planned route in the Riga city having extensive street sections with inflexibly coordinated traffic lights, by this enhancing repeatability of tests due to the time shifts between the traffic lights remaining constant for the whole duration of the tests. The route included three street sections that were driven along in both directions with a length between 2.3 and 2.8 km and having an inflexible traffic lights green wave (GW) in one direction and quite restricted traffic control against the green wave (AGW) in the other direction. The 25 km route was covered 41 times.

Fuel rate and car speed have been recorded by Auterra DashDyno SPD OBD logger. Car speed and GPS position were recorded by GPS logger RaceLogic DriftBox with a 10 Hz recording frequency. GPS speed data have been used for OBD speed calibration and for data synchronization, GPS position was used to split the driving route into sections allocated to the street grid.

Distance covered has been calculated from car speed data and logging timing data, then compared to car odometer recordings and distances calculated from GPS data. Fuel consumption FC in litres per 100 km calculated from the fuel rate and car speed. Traction energy consumption TW calculated from traction force and distance covered. The traction force is a sum of rolling resistance, air drag, grade force and inertia force, each calculated from car technical characteristics, car coast-down procedure and city streets vertical geometry data.

Periodical regression analysis

To perform the regression analysis of city driving parameter change during 24 hour period, up to eleventh degree of univariate polynomial equations n were tested:

$$P(x) = \sum_{i=0}^{n} (a_i \cdot x^i) \tag{1}$$

where x – time of the day, expressed as MS Excel value from 0 to 1 in a 24 hour period; a_i – variable at the i index of a data-set; N; n – degree of the polynomial.

To ensure the continuity of the regression polynomial, to get a polynomial with equalized endpoints, the values and the first derivative of the regression polynomial at midnight or any other time of the day have to be identical:

$$P(0) = P(1) \tag{2}$$

$$\dot{P}\,(0) = \dot{P}\,(1) \tag{3}$$

Solving Eq. 2:

$$a_1 = -\sum_{i=2}^{n} a_i, \tag{4}$$

and solving Eq. 3:

$$a_2 = -0.5 \cdot \sum_{i=3}^{n} (i \cdot a_i), \tag{5}$$

from Eqs 4 and 5 re-calculating a_1:

$$a_1 = \sum_{i=3}^{n} ((0.5i - 1) \cdot a_i), \tag{6}$$

transforming the Eq. (1) to:

$$P(x) = a_0 + \sum_{i=3}^{n} (a_i \cdot (x^i - 0.5 \cdot i \cdot x^2 + (0.5 \cdot i - 1) \cdot x)). \tag{7}$$

The least-squares method finds values of each a_i by minimizing the sum S of squared residuals:

$$S = \sum_{j=1}^{m} \left[y_j - a_0 - \sum_{i=3}^{n} \langle a_i \cdot \left(x_j^i - 0.5 \cdot i \cdot x_j^2 + \lfloor 0.5 \cdot i - 1 \rfloor \cdot x_j \right) \rangle \right]^2, \tag{8}$$

where m – number of measurements; y_j – value of the j_{th} measurement.

There are n-1 sought parameters a_i and n-1 partial differential equations:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = -2 \cdot \sum_{j=1}^{m} \left[y_j - a_0 - \sum_{i=3}^{n} \langle a_i \cdot \left(x_j^i - 0.5 \cdot i \cdot x_j^2 + \lfloor 0.5 \cdot i - 1 \rfloor \cdot x_j \right) \rangle \right] \\ \dfrac{\partial S}{\partial a_\mathrm{k}} = -2 \cdot \sum_{j=1}^{m} \left[\begin{matrix} \left\{ y_j - a_0 - \sum_{i=3}^{n} \langle a_i \cdot \left(x_j^i - 0.5 \cdot i \cdot x_j^2 + (0.5 \cdot i - 1) \cdot x_j \right) \rangle \right\} \cdot \\ \left(x_j^k - 0.5 \cdot \mathrm{k} \cdot x_j^2 + (0.5 \cdot \mathrm{k} - 1) \cdot x_j \right) \end{matrix} \right] \end{cases}, \quad (9)$$

where k – parameter a index; k values are from 3 to n.

Equating each differential equation to 0, solving the system of equations can be done by solving a matrix equation:

$$A = X^{-1} \cdot Y, \quad (10)$$

where $X =$

$$\begin{bmatrix} m & \sum_{j=1}^{m} P(3) & \sum_{j=1}^{m} P(4) & \cdots & \sum_{j=1}^{m} P(i) \\ \sum_{j=1}^{m} P(3) & \sum_{j=1}^{m} (P(3) \cdot P(3)) & \sum_{j=1}^{m} (P(4) \cdot P(3)) & \cdots & \sum_{j=1}^{m} (P(i) \cdot P(3)) \\ \sum_{j=1}^{m} P(4) & \sum_{j=1}^{m} (P(3) \cdot P(4)) & \sum_{j=1}^{m} (P(4) \cdot P(4)) & \cdots & \sum_{j=1}^{m} (P(i) \cdot P(4)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} P(k) & \sum_{j=1}^{m} (P(3) \cdot P(k)) & \sum_{j=1}^{m} (P(4) \cdot P(k)) & \cdots & \sum_{j=1}^{m} (P(i) \cdot P(k)) \end{bmatrix}; \quad (11)$$

$$A = \begin{bmatrix} a_0 \\ a_3 \\ a_4 \\ \vdots \\ a_k \end{bmatrix}; \; Y = \begin{bmatrix} \sum_{j=1}^{m} y_j \\ \sum_{j=1}^{m} (y_j \cdot P(3)) \\ \sum_{j=1}^{m} (y_j \cdot P(4)) \\ \vdots \\ \sum_{j=1}^{m} (y_j \cdot P(k)) \end{bmatrix} \quad (12)$$

where $P(r) = x_j^r - 0.5 \cdot r \cdot x_j^2 + (0.5 \cdot r - 1) \cdot x_j$.

Matrix calculations were done by means of MS Excel using VBA code to find regression polynomials with degrees from 3 to 10 for fuel consumption, average speed, traction energy consumption, braking energy, stoppage time and fuel consumption both for polynomial regression and polynomial regression with equalized endpoints. In total 75 sets of measurement data have been analysed.

To find an optimal polynomial degree for city driving parameters when analysed against the time of the day for each regression polynomial coefficients of determination R^2 and R^2 adjusted have been calculated and put in MS Excel charts for comparison. For R^2 adjusted the degree of freedom for polynomial is equal to the degree of the polynomial, for the polynomial with equalized endpoints the degree of freedom is n-2, since a_1 and a_2 are calculated from other polynomial constants (4) and (5). Besides the numerical evaluation of R^2 values, visual comparison of regression lines was done to see how well the regression lines fit the data.

RESULTS AND DISCUSSION

The driving tests were planned to be spread during all 24 hours of the day, but due to higher variations in traffic intensity in the evenings, slightly more tests were done in the evenings. Regression function may be influenced by gaps between the measurement points therefore the plan was to avoid essential time gaps between measurements even when no essential traffic condition changes were expected.

Comparison of PR and PRE regression lines

In the case with dense data during all hours there is no essential difference between the polynomial regression (PR) and polynomial regression with equalized endpoints (PRE). Traction energy consumption TW per 100 km on street sections where coordinated traffic lights ensure green wave is shown on Fig. 1. The difference in R^2 values between PR and PRE is close to 0.01.

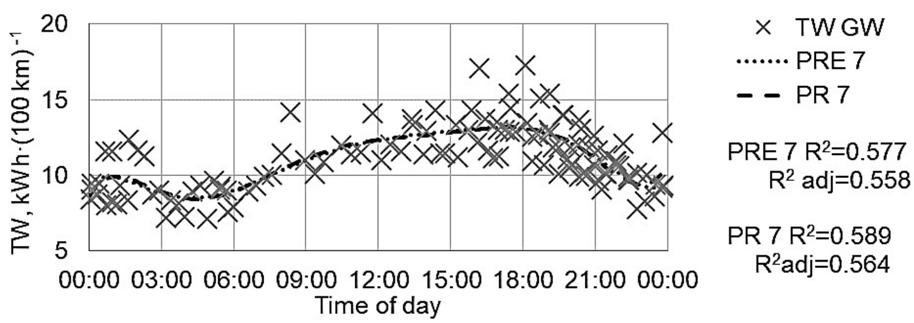

Figure 1. Car traction energy consumption on the test streets in the direction of the traffic lights green wave.

To visually see the difference in regression lines, the same graph is shown on Fig. 2 without data points and with smaller axis units. There is 1.5 kWh 100 km^{-1} or slightly above 10% difference in energy consumption regression line for PR just before and just after the midnight.

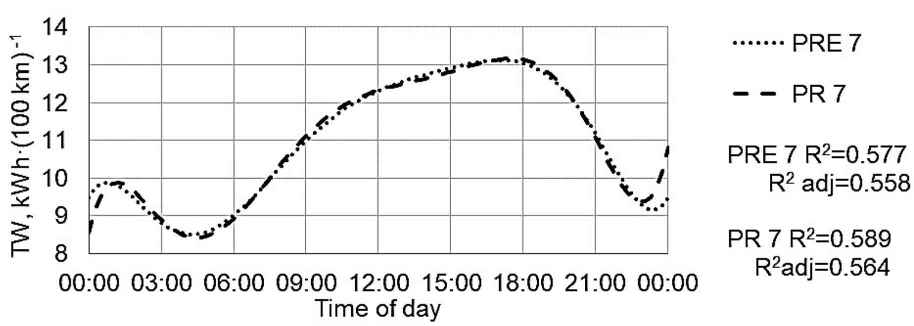

Figure 2. Car traction energy consumption regression lines on the test streets in the direction of the traffic lights green wave; PRE 7 – 7th order polynomial regression line with equalised endpoints, PR 7 – without equalised endpoints.

In certain cases, especially with less data around midnight like in Fig. 3, where fuel consumption FC is given for Čaka Street only in the direction of the green wave, the difference in R^2 values between PR and PRE is still quite small, but the fuel consumption for PR just before the midnight is more than double of the value just after the midnight. There is quite a small variance between regression lines where there is a small time gap in the dataset, but for the rest of the time PR and PRE lines are really close.

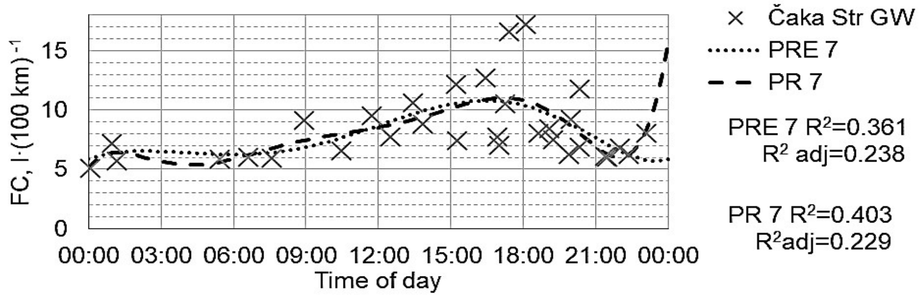

Figure 3. Car fuel consumption regression lines in the direction of the traffic lights green wave; PRE 7 – 7th order polynomial regression line with equalised endpoints; PR 7 – without equalised endpoints.

To find an optimal degree of polynomial for describing the change of various driving characteristics the PRE lines were calculated and plotted for polynomial orders from 3 to 10.

Optimal degree of the polynomial for driving characteristics

To evaluate the PRE regression lines the Excel VBA code also calculated R^2 and R^2 adjusted values for each line and for every characteristic a graph was plotted. The R^2 values for traction work on the test streets in both directions and for a total route are shown on Fig. 4. Different R^2 values for various streets do not describe how good the PRE line fits the data. The more the parameter changes during the day, the higher is R^2 value. It can be seen that for different street sections the increase of R^2 with the polynomial order is different, for some sections the values are close to the maximum already at 4 while for others quite noticeable increase is up to the polynomial order of 10, meaning that there is no equally good regression model for all datasets.

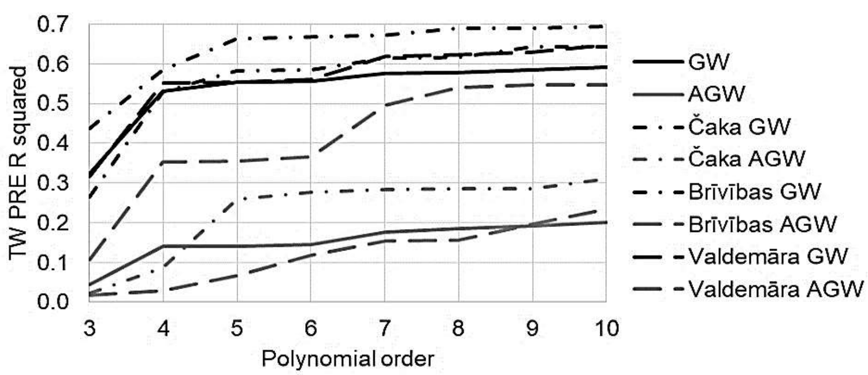

Figure 4. R squared values for traction work 3rd to 10th order polynomial regression lines on test streets; GW – in the direction of green wave; AGW – in the direction against the green wave.

The adjusted R^2 values shown along the R^2 value for fuel consumption data on Fig. 5 the growth up to 9th polynomial order for fuel consumption data on Valdemāra Street against the green wave direction while for some street sections the order of 6 seems using excessive polynomial degrees.

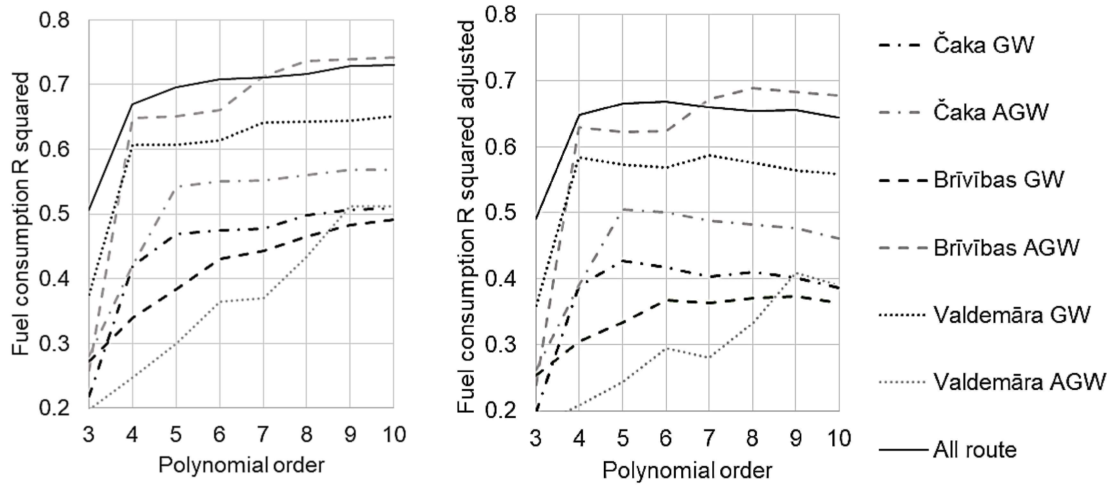

Figure 5. R squared and R squared adjusted values for fuel consumption 3rd to 10th order polynomial regression lines on test streets; GW – in the direction of green wave; AGW – in the direction against the green wave.

The visual inspection of regression lines were performed for all datasets. The PRE lines for traction work per 100 km for polynomial orders from 3 to 10 for GW direction leading out of the city for all three streets analysed are shown on Fig. 6. For the ease of analysis every next polynomial order line is shifted up by 1 kWh per 100 km. PRE line 3 depends on a_3 coefficient only, therefore the line is very simple and the corresponding R^2 value is essentially lower than for higher orders of polynomial.

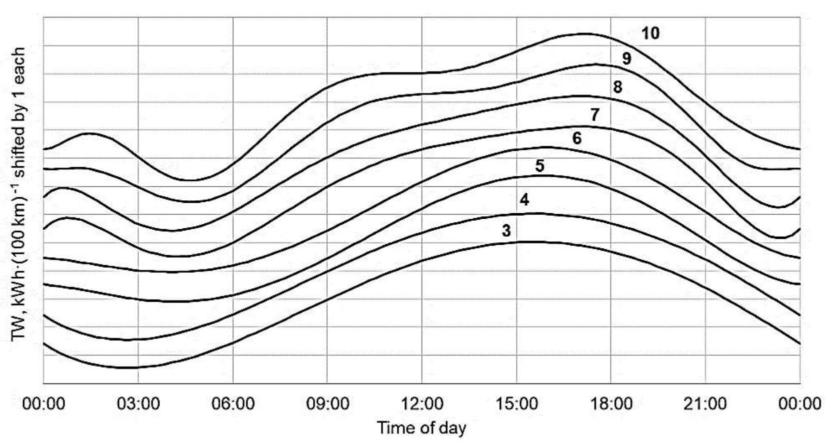

Figure 6. Polynomial regression lines with equalised endpoints for traction work per 100 km for polynomial orders from 3 to 10 in the direction of green wave.

From the shape of the regression line several periods for the traffic conditions can be found. All streets investigated have bus and trolleybus lines, therefore the lowest energy demand is while there is almost no public transport on streets between 2 am and 5 am. This period can be well seen from line 7. There can be seen only a minor peak for the morning traffic on lines 9 and 10 because the GW direction leads out of the city and the tests are done in the city centre while the traffic jams in the mornings are mostly on the approach to the centre. The highest energy demand is around 6 pm when people return from the work home and gradually diminishes up to 11 pm.

In the direction against the green wave towards the city centre shown on Fig. 7 there is higher morning peak detected from line 7 and due to multiple stops during the night time, the low peak is smaller and again can be observed from line 7.

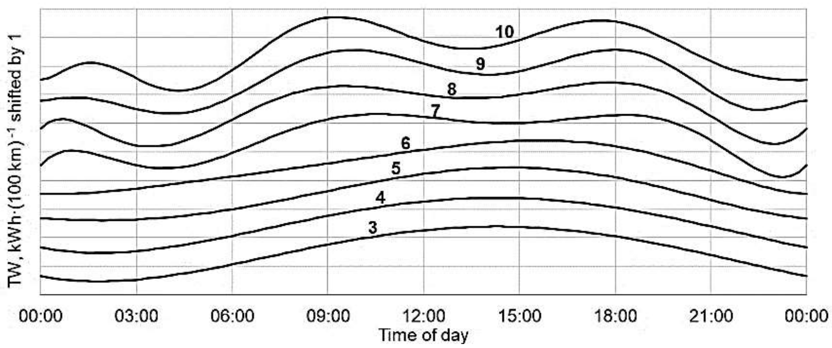

Figure 7. Polynomial regression lines with equalised endpoints for traction work per 100 km for polynomial orders from 3 to 10 in the direction against the green wave.

When comparing the regression lines against the data on certain occasions it can be seen that for certain times of day the line does not describe well the process. On Fig. 8 where the average speed on Brīvības street in the direction of green wave is calculated, it can be seen that just before 7 pm the speeds have reached the minimum and then in a short time jumped up for some 20 km per hour indicating that the traffic flow is instable and a special attention has to be devoted to evaluating the traffic flow at this time period. Since one of the goals for this particular regression analysis is to find traffic flow characteristics above 10 km from the actual measurement, between 6 pm and 8 pm the speeds cannot be evaluated from the regression line only.

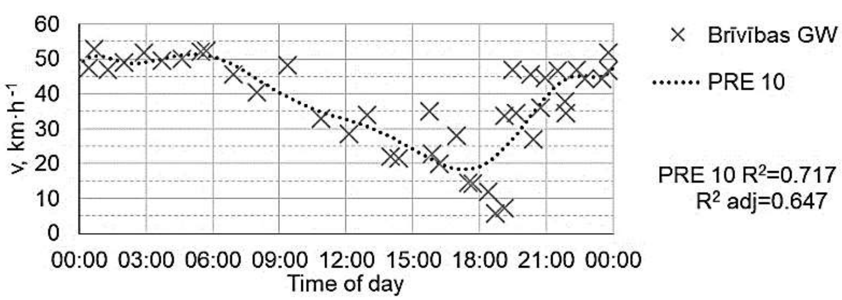

Figure 8. Average speed on Brīvības Street in the direction of green wave with 10th order polynomial regression line with equalised endpoints.

The same speed graph on the same street but against the direction of the green wave shown on Fig. 9 is much more stable without steep changes and the regression calculation would give a sensible result.

The regression analysis described cannot be extended to any traffic parameters analysis but is really helpful for the goals stated.

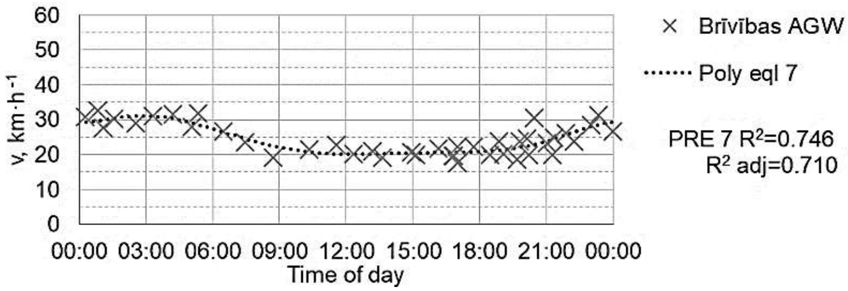

Figure 9. Average speed on Brīvības Street in the direction against the green wave with 7th order polynomial regression line with equalised endpoints.

CONCLUSIONS

The periodic polynomial regression analysis is not a regular textbook statistical analysis case while it is an enjoyable exercise to solve the equations with setting limits for function and the first derivative values to get a usable periodic function regression model.

The periodic polynomial regression function can be used for road traffic parameter change as a function of the time of the day analysis. The polynomial order has to be at least 7 although a visual conformation of good regression line to the measured data has to be checked especially for the lower polynomial orders.

The regression polynomials obtained can be used for the rate of flow estimates by measuring the oncoming traffic on most of the street sections analysed for most of the time, but due to high variance in traffic flow at certain conditions, any regression line would not be applicable for certain short time periods.

REFERENCES

International Energy Agency (IEA) 2018. Key world energy statistics. 40 pp.

Fontaras, G., Zacharof, N.G. & Ciuffo, B. 2017. Fuel consumption and CO2 emissions from passenger cars in Europe Laboratory versus real-world emissions. *Progress in Energy and Combustion Science* **60**, 97–131.

Mannering, F.L., Washburn, S.S. & Kilareski, W.P. 2009. *Principles of Highway Engineering and Traffic Analysis.* Wiley, USA, 398 pp.

Garber, N.J. & Hoel, L.A. 2010. *Traffic and Highway Engineering.* Cengage Learning, Stamford CT, 1229 pp.

Roess, R.P., Prassas, E.S. & McShane, W.R. 2011. *Traffic Engineering.* Pearson, Upper Saddle River NJ, 744 pp.

Cobanovic, K., Lozanov-Crvenkovic, Z. & Nikolic-Doric, E. 2006. Periodic regression. *International Association for Statistical Education Conference Proceedings* **7**, 1–4.

Kreicbergs, J., Gailis, M. & Grislis, A. 2018, Urban Real Driving Analysis with and without Coordinated Traffic Lights Control. In*: Transport Means 2018: Proceedings of the 22nd International Scientific Conference. Part 1, Lithuania, Trakai, 3-5 October, 2018.* Kaunas University of Technology, Kaunas, pp. 157–161.