



## APPLICATION OF MULTIDIMENSIONAL STATISTICAL MODEL FOR EVALUATION OF MEASURED DATA OBTAINED FROM TESTING OF THE HCCI ENGINE PROTOTYPE

Michal PUŠKÁR, Michal FABIAN, Tomáš TOMKO

Faculty of Mechanical Engineering, Technical University of Košice, Letná 9, 042 00 Košice, Slovakia

e-mail address: [michal.puskar@tuke.sk](mailto:michal.puskar@tuke.sk), [michal.fabian@tuke.sk](mailto:michal.fabian@tuke.sk), [tomas.tomko@tuke.sk](mailto:tomas.tomko@tuke.sk)

### Abstract

Evaluation of the measured data, which are obtained from testing of the HCCI engine prototype, is a complex process taking into consideration questions of the vibration diagnostics. The complexity of this procedure is caused due to a necessity to evaluate a large amount of the various data. One of the solution possibilities is application of a multidimensional statistical model within the evaluation process in order to obtain more accurate information about the actual operational condition concerning the analysed engine.

Keywords: diagnostics; engine prototype; statistical model

## 1. INTRODUCTION

From the perspective of understanding the context of the statistical evaluation of measured data itself within the amplitude and frequency of the engine Honda GX 25 (Fig.1) during the application of the statistical model may experience irregularities such as: is the liner regression model suitable, how much percent of the measured data of the statistical model can be explained by this model, how much percent of the model is strong, are the measured data affected by the multikolinearity, autokorelation or heteroscedasticity, is ensured by the normality of residues, what are the possible errors of the model and how to find out. These questions very resolutely answered by the methodology which the procedure is described in this work. To have these questions answered by the correct application of the linear regression model with two variables is effective using R-studio software [1, 2].



Fig. 1. Honda GX 25

## 2. REGRESSION MODEL AND METHOD

### 2.1. Calculate the regression coefficients $\beta$

The first step in the progressive realization of the linear regression model with two variables is to calculate the regression coefficients as:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (1)$$

In view of the equation (1) is clear that software R - studio have to calculate an inverse matrix multiplying the transposed matrix of matrix X and XT again multiplied by the transposed matrix XT and matrix y. The problem of equation (1) is easy to define in the program environment of R - studio using the following command:

```
>B1<-solve(t(X)÷÷_(X))÷÷÷t(X)÷÷÷y (2)
```

This operation assures linear model parameter estimation which will enter into vector shape. Output from R - Studio is a vector of estimation parameters in the following form [3, 4]:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.602e+01  7.019e+00  3.708 0.000221 ***
speed_data  4.216e-04  1.259e-03  0.335 0.737869
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.2. Random disturbances and their dispersion in the form of a variation – co variation matrix

The next step is to define the problems of random disturbances in the environment program of R - Studio. The difficulty of this step is the fact that the variance must be expressed in the variation - covariance matrix. From this point of view it is necessary to define the following problem expressed by the equations (3) and (4):

$$\text{var}(\hat{\beta}) = S^2 (X^T X)^{-1} \quad (3)$$

$$S^2 = \frac{e^T \cdot e}{n - (k + 1)} \quad (4)$$

In order to equation (3) to express through software R - studio is necessary to calculate the variance of random values according to equation (4). This equation refers to the proportion in which the numerator is the product of the transposed matrix of the vector of residual components of  $e^T$  ordinary matrix vector of residuals  $e$ . In the denominator there is the number of explanatory variables to increase the number by 1. Then the coefficient is subtracted from the number of observations  $n$  [5]. This problem can be defined in the program R - studio in following form:

$$S2 \langle -t(\text{resid}(B2)) \div \_ \div \text{resid}(B2) / \\ /(\text{dim}(\text{udaje}))[1] - \text{dim}(\text{udaje})[2] \rangle \quad (5)$$

In this case I used the generic function *resid* (), which is in a standard offer program R-studio environment. This feature was chosen because of the purpose of the calculation of the residual component of the regression model itself. In addition to the generic function *resid* () in this step I used another generic function *dim* () [1]. This function determine the number of rows of a matrix with dimension  $n$  provided and the number of columns by using the dimension by  $k+1$ . The output of the R-studio is a program based on the defined command (5) variance of random failures. In this case, it's the value of 7.019. After a successful estimation of variance of random failures I have to the define a problem of S2 (3) to the environment program R-studio. Since this software understands the calculated variance matrix and a scalar variable is not as necessary in this step to express this matrix in the form of scalar. This transformation is in the program-defined R-studio using the function *as.vector* () in this form:

$$\rangle \text{sigma2} \langle -\text{as.vector}(S2) \rangle \quad (6)$$

After the expression of the scalar site of dispersion is necessary to estimate the variation - co variation matrix using the another generic function. This time I used the function *vcov* which is in an environment of R-studio software in this form:

$$\rangle \text{vcov}(B2) \quad (7)$$

The output from the R-studio is a variation - co variation matrix of dispersion folder of the measured values of time series.

### 2.3. Quantification of the coefficient of determination

The third and very important step in the application of a linear model with two variables is the quantification of the coefficient of determination which set out to answer a question about how many percent of the measured data they took my algorithms into account. I have defined this problem into R-studio. In the first place it is necessary to recall the fact that the coefficient of determination is working with three types of squares. Takes into account the total sum of squares, the residual sum of squares and explained sum of squares. These squares takes into consideration program R-studio using the function *r.squared*. This function is the issue of the quantification of the coefficient of determination which function in R-studio is in this form:

$$\text{summary}(B2)\$r.squared \quad (8)$$

In this case after the process of defining into the R-studio I calculated value 98.9 percent. This value defines the following statistics: Approximately 98.9 percent of the total variability values of the dependent variable is explained by the equation (14)[6, 7].

### 2.4. Detection of the vector of parameters in the form of interval

After the evaluation of the coefficient of determination it is necessary to go to the next step in which it is necessary to detect the interval of estimation  $\beta$ . The statistics understand this interval as a confidence interval in which, with 95 % of probability, are parameters  $\beta$ . The progress of this detection is implemented in an environment R-studio program using another generic functions.. In particular, it is a function of *confint* (). The function code () is in the environment programme R-studio the following shape:

$$\rangle \text{confint}(B2, \text{level} = 0.95) \quad (9)$$

The output of this calculation is a 95% confidence interval for the estimation of the parameters  $\beta$  in two-sided vector form.

### 2.5. Verification of the normal hypothesis of the model with two variables

With the proper application of this model arises from the nature of his hypothesis that the model must meet. This is a hypothesis which says that the residual component of the time of the order must follow normal distribution probability. For this reason, in order to verify this assumption need to use very well known statistical test called Jarqueov-Beraov test of normality. Of course this issue is in the program-defined R-studio. Define functions using the *jbTest* (), which is in the standard package of the program R-studio. This function enters the

calculation process in R-studio in the following code:

$$\rangle \text{jbTest}(\text{resid}(B2)) \quad (10)$$

This test takes into consideration two possible hypotheses that might occur during the evaluation. Zero hypothesis  $H_0$  is talking about, that the residual component shall enter into a normal probability distribution. Against this hypothesis to build alternative hypothesis  $H_1$ , which says that this component shall be other than a normal probability distribution. An important feature of this test is also called pointed out. p-value. This value is an essential indicator of the acceptance or rejection of the null hypothesis  $H_0$ . The output of this test is to value 0, p 739. The p-value is greater than the level of significance of  $\alpha$ , the null hypothesis could not be rejected 0.05. It is therefore possible that the hypothesis of a normal distribution of residual components regarded as satisfied [8-10].

## 2.6. Detection of possible defects of the model (Ramseyho test)

The most important part of this work is to verify if a linear regression model is defined correctly. Verify the accuracy of model specifications for the purposes of this thesis I used a very famous statistical test called Ramsey test specification errors. On the chosen level of significance  $\alpha = 0,05$  defined issue of Ramsey test in the program R-studio. Using generic function *resettest* () manages this program very quickly to assess whether the proposed model is defined correctly or not. Generic *resettest* () function used on the basis of model specifications, errors in the environment programme looks like R-studio as follows:

$$\rangle \text{resettest}(\text{model} - i, \text{power} = 2:3) \quad (11)$$

Test specifications for the chosen level of significance testing errors defined zero hypothesis  $H_0$  which talks about how that model is defined in the correct shape against the alternative hypothesis  $H_1$  which is considered the model for the incorrectly defined and recommended to implement the square expansion of the independent variable into this form:

$$y = \beta_0 + \beta_1 X \quad (12)$$

After using this test I failed to reject the zero hypothesis  $H_0$  in this step, in the level of significance  $\alpha$ , then the result of this test is a clear statistical proof that the proposed model meets all the statistical assumptions and may be considered as a model in correct form.

## 2.7. Quantification of anticipated values of the model

One of the advantages of using a linear regression model in practice is the fact that this model offers the possibility of calculating the anticipated values. This issue is in the program-defined R-studio for the purpose of estimating the values of which can be used-the likelihood of the model to generate in the future. For this purpose it is necessary to calculate an estimate of the interval of values. It is used in this chapter, the command is called. *predict* (), which is in the argument folder of the confidence interval specified code, in which the value is calculated.

After the inclusion of this interval is the command to *predict* () in an R-studio (Fig.2) program used in the following form:

$$\rangle \text{exp}(\text{predict}(\text{model} - p - 2, \text{interval} = \text{"confidence"})) \quad (13)$$

The output of the R-studio, after the anticipated data (13) are defined, is the estimation interval of the variables X [11 -17].

## 3. RESULTS AND DISCUSSION

R-studio program application in the data measured in terms of the diagnosis of internal combustion engines is the inclusion of data from the measurement of the amplitude and speed of the engine Honda GX 25 (Fig.3), expressed in a linear regression model in the following form:

$$y = 2,602 + 4,216X \quad (14)$$

This equation has the following interpretation: the change of the parameter  $y$  on one drive will change the value of the  $X$  parameter to 4.216. The parameter  $y$  in this case is the magnitude and parameter  $X$  is speed of the engine Honda GX 25 used in the Shell eco marathon project. It is 4-stroke, single cylinder engine with displacement 25 cm<sup>3</sup>. Max. output is 0,81 KW/ 7000 rpm and max. torque is 1.25 Nm/ 5000 rpm.

This engine working in HCCI regime. Combustion of the homogenous charge is running all at once in the whole piston combustion area, i.e. the charge is combusted almost completely, without a rest. Exploitation of the gasoline using the HCCI technology is very efficient and it offers a significant reduction of the fuel consumption, together with the reduced exhaust gas emissions, whereas the NOx emission are almost neglecting.

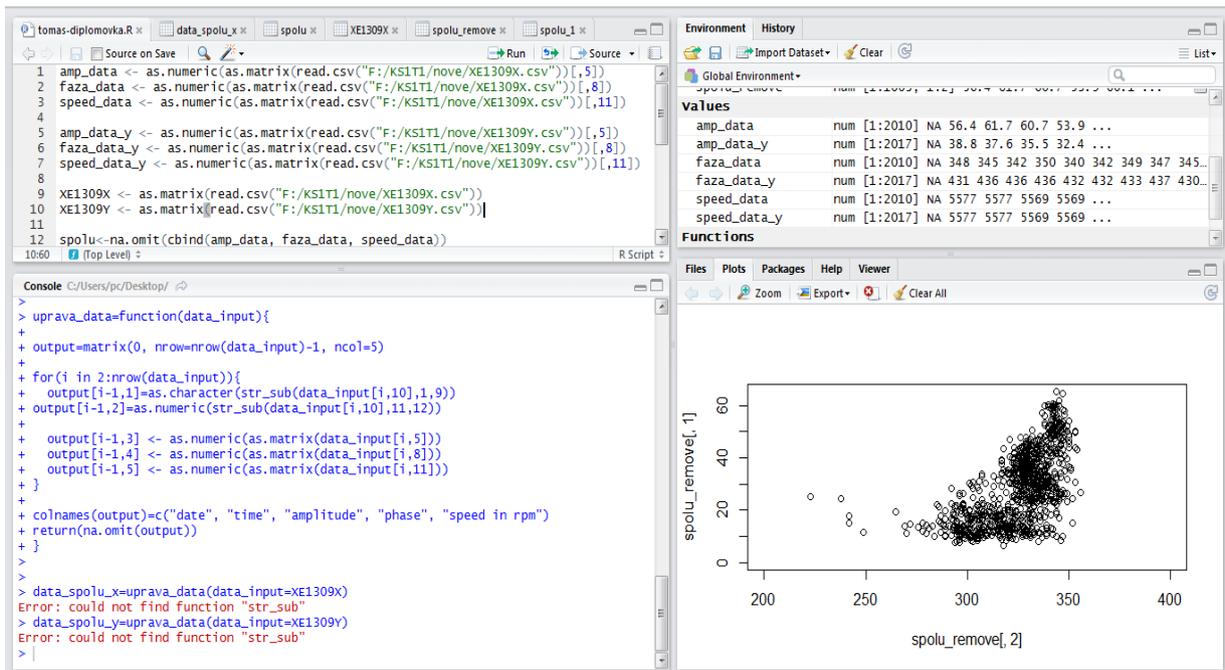


Fig. 2. R-studio software

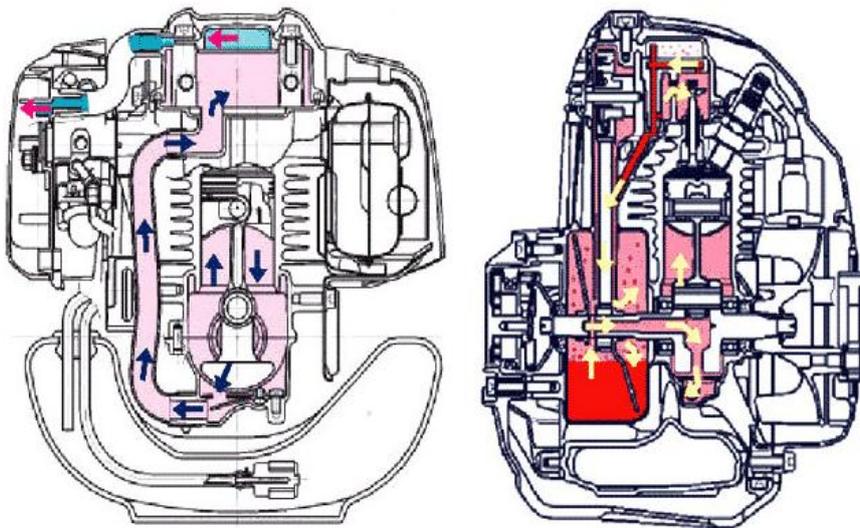


Fig. 3. Visualization of Honda GX 25 motor

Equation (14) is not the only output after application of a linear model with two variables. This methodology is also statistically confirmed the fact that the model (14) is based on the Ramseyho test specifications defined in the right form errors. While this equation is able to include 98.9 percent of all measured values, which clearly declares the value of the coefficient of determination [18].

#### 4. CONCLUSIONS

This research and development resulted in a significant reduction of the fuel consumption

compared to the standard motor system. The engine speed level was increased there was also heightened the maximal value of the experimental vehicle speed.

The results of our research work were introduced also internationally, namely at the international competition Shell Eco-marathon 2016 in London. Our experimental vehicle, which was equipped with the innovated driving unit, presented itself by a reliable operation and very favourable results (Fig. 4).



Fig. 4. Experimental vehicle during the international competition Shell Eco-marathon in London

## SOURCE OF FUNDING

This paper was written in the framework of Grant Projects: APVV-16-0259 “Research and development of combustion technology based on controlled homogenous charge compression ignition in order to reduce nitrogen oxide emissions of motor vehicles”, VEGA 1/0473/17 “Research and development of technology for homogeneous charge self-ignition using compression in order to increase engine efficiency and to reduce vehicle emissions”, VEGA 1/0198/15 “Research of innovative methods for emission reduction of driving units used in transport vehicles and optimisation of active logistic elements in material flows in order to increase their technical level and reliability” and KEGA 041TUKE-4/2017 “Implementation of new technologies specified for solution of questions concerning emissions of vehicles and transformation of them into the educational process in order to improve quality of education.”

## REFERENCES

1. Wackerly D, Mendenhall W, Scheaffer R. *Mathematical Statistics with Applications*. Seventh Edition, Belmont, USA. 2008.
2. Puškár M. Measuring method for feedback provision during development of fuel map in hexadecimal format for high-speed racing engines. *Measurement* 2014;50(1): 203-213.
3. Puškár M, Bigoš P. Method for accurate measurements of detonations in motorbike high speed racing engine. *Measurement*, 2012; 45(3): 529-534. <http://dx.doi.org/10.1016/j.measurement.2011.10.014>
4. Puškár M, Bigoš P, Puškárová P. Accurate measurements of output characteristics and detonations of motorbike high-speed racing engine and their optimization at actual atmospheric conditions and combusted mixture composition. *Measurement*, 2012; 45(5):1067-1076.
5. Puškár M, Bigoš P. Measuring of acoustic wave influences generated at various configurations of racing engine inlet and exhaust system on brake mean effective pressure. *Measurement*, 2013; 46(9): 3389-3400.
6. Czech P. Application of probabilistic neural network and vibration signals for gasket under diesel engine head damage. *Scientific Journal of Silesian University of Technology. Series Transport*, 2013; 78: 39-45.
7. Toman R, Polóni M, Chribik A. Preliminary study on combustion and overall parameters of syngas fuel mixtures for spark ignition combustion engine. *Acta Polytechnica*, 2017;57(1):38-48. <http://dx.doi.org/10.14311/AP.2017.57.0038>
8. Chribik A, Polóni M, Lach J, Ragan B. The effect of adding hydrogen on the performance and the cyclic variability of a spark ignition engine powered by natural gas. In *Acta Polytechnica*, 2014; 54(1): 10-14.
9. Wierzbicki S. Diagnosing microprocessor controlled systems. *Teka Komisji Motoryzacji i Energetyki Rolnictwa*, 2006; VI:183-188.
10. Czech P. Diagnosis of Industrial Gearboxes Condition By Vibration and Time-Frequency, Scale-Frequency, Frequency-Frequency Analysis. *Metalurgija*, 2012; 51(4): 521-524.
11. Magdolen L, Masaryk M. Flywheel storage energy, Conference Gepeszet 2012, Budapest, Hungary, Conference proceedings, Budapest University of Technology and Economy BME Budapest, 2012.
12. Komorska I. Modeling of vibration signal for reciprocating engine diagnostics. *Diagnostyka*, 2009; 2(50):23-26.
13. Niziński S, Wierzbicki S. Zintegrowany system informatyczny sterowania pojazdów. *Diagnostyka*, 2004; 30:47-52.
14. Czech P, et al. Application of the discrete wavelet transform and probabilistic neural networks in IC engine fault diagnostics. *JVE, Journal of Vibroengineering*, 2014; 16(4):1619-1639.

15. Blištan P, et al. Renewable energy sources and risk management, in: SGEM 2012: 12th International Multidisciplinary Scientific GeoConference: conference proceedings: Volume 4. Albena, Bulgaria - Sofia: STEF92 Technology Ltd., 2012: 587–594.
16. Czech P. Identification of leakages in the inlet system of an internal combustion engine with the use of Wigner-Ville transform and rbf neural networks. 12th International Conference on Transport Systems Telematics Location: Katowice Ustron, Poland. 2012. Edited by J. Mikulski. Telematics in the Transport Environment. Book Series: Communications in Computer and Information Science, 2012; 329: 414-422.
17. Kopilčáková L, Pauliková A: Technický metabolizmus v rámci orientovaného konštruovania. In: Manažérstvo životného prostredia. 8. konferencia so zahraničnou účasťou (recenzovaný zborník referátov). Bojnice, 2008:55-57.
18. Puškár M, Brestovič T, Jasminská N. Numerical simulation and experimental analysis of acoustic wave influences on brake mean effective pressure in thrust-ejector inlet pipe of combustion engine. In: International Journal of Vehicle Design, 2015; 67(1): 63-76.  
<http://dx.doi.org/10.1504/IJVD.2015.066479>.



**Ing. Tomáš TOMKO** is graduate at the Faculty of Mechanical Engineering of the Technical University in Košice. He was a PhD. student and member of shell eco prototype team on the Department of construction, Automotive and Transport Engineering.

Received 2017-08-03  
Accepted 2017-10-08  
Available online 2017-12-18



combustion engines and increasing their power output and reliability.

doc. ing. **Michal PUŠKÁR**, PhD, is an Associate Professor in the Department of Engineering for Machine Design, Automotive and Transport at the Faculty of Mechanical Engineering, Technical University of Košice. His current fields of research interest are single-track transport means, piston



systems in the processes of motorcar development and production, namely with the design of motorcar exteriors and interiors. His several graduates are already employed in the given automotive area.

doc. Ing. **Michal FABIAN**, PhD, is an Associate Professor in the Department of Engineering for Machine Design, Automotive and Transport at the Faculty of Mechanical Engineering, Technical University of Košice. He is dealing, from the year 1990, with the application of the CAX