Proceedings of the 10th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'10), 20–23 October 2010, Riga, Latvia, p. 413-419. ISBN 978-9984-818-34-4
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SELECTION OF THE DIAGNOSTIC PARAMETERS FOR CONTROLLING AND DIAGNOSING THE GAS TURBINE ENGINE USING REGRESSION ANALYSIS

Sergey Yunusov[1], Eugene Kopytov[2], Vladimir Labendik[3]

Transport and Telecommunication Institute
Lomonosova str. 1, Riga, LV-1019, Latvia
[1] Ph.: +371 26435655. E-mail: yunusov@inbox.lv
[2] Ph.: +371 29257215. E-mail: kopitov@tsi.lv
[3] Ph.: +371 29754804. E-mail: lvp@tsi.lv

The work suggests the application of the regression analysis for selecting the parameters in case of controlling and diagnosing the gas-turbine engines under the conditions of maintenance. The existing approaches to the determination of the diagnostic parameters choice are based on the analysis of the registered flight and ground information on the engine functioning and its damages. The analysis of the registered information and damages is carried on, as a rule, by the specialists-experts using the heuristic methods in contrast to the formal solution method being based on the exact mathematical models. The application of the regression analysis allows to reveal the connections between the engine parameters by means of the existing statistical data without complex calculations and to determine the choice of parameters possessing the most considerable diagnostic value.

***Keywords**: gas-turbine engine, diagnostic information, diagnostic parameters, regression analysis*

1. Introduction

Aircraft engine is the complex technical object, where the number of factors influencing one way or another gas turbine engine (GTE) functioning, is too great to be taken into account. The number of factors being measured in the modern aircraft engines can achieve the range from some tens to hundreds, not all these parameters including the necessary diagnostic information. Under the conditions of maintenance it is necessary to emphasize only the parameters with the most significant information on the engine state. The formation of the diagnostic parameters and properties choice is usually accomplished basing on the analysis of the registered information by means of special facilities of registering the flight and ground data, taking into consideration the peculiarities of the engine structure as the object of diagnostics, i.e. the analysis of the functional connections of the aggregates, systems, engine damages and the relationships between its performance and present state [1]. The fragment illustrating the cause-and-effect relations in case of occurred damages is showed in Table 1.

Table 1. Cause-and-Effect Relations in Case of Occurred Damages

External indicators of the engine performance changes	**Main damages causing the engine performance changes**	**Complex accompanying phenomena caused by the engine performance changes**
Engine stall	Destruction of the compressor details: ▪ working blades; (deformation blade airfoil, destruction); ▪ outlet straightening blades; ▪ centre support bearing; ▪ maze sealing; ▪ compressor blow-off.	Pop: ▪ growth of the gases temperature in front of turbine; ▪ sparks and black smoke explosion; ▪ foreign sound appearance; ▪ vibration increase; ▪ rotor velocity fall.

Such approach to determining the diagnostic parameters is built on the base of the analysis of the occurred damage. The analysis is carried on, as a rule, by the specialists-experts using the heuristic methods. The heuristic methods are usually opposed to the formal methods of solution based on the exact mathematical models. The diagnostic value of the parameters is determined by means of the cause-and-effect relations.

This approach to defining the diagnostic parameters has both advantages and disadvantages. The main disadvantage is that before damaging the parameters are not ranged according to the extent of their significance for engine controlling and diagnosing. The regression analysis is suggested for emphasizing the parameters according to their significance using the multiple regression models. The application of the regression analysis is attractive because of its ability to reveal the relationships between engine parameters, taking into account the existing statistical data, not making the complex calculations indicating the influence of one or some parameters on the other ones and to construct the complex diagnostic parameter. The regression analysis is known to be used for forecasting and it means that the selected diagnostic parameters should possess the definite prediction strength. Let us consider the regression analysis application for selecting diagnostic parameters.

2. Statistical Model for Gas Turbine Engine Diagnostics

We can use the following statistical model [2] for solving the tasks of GTE diagnostics:

$$\hat{y} = a_0 + \sum_{i=1}^{k} a_i x_i + \sum_{i<j} a_{ij} x_i x_j + \sum_{i=1}^{k} a_{ii} x_i^2 + \dots , \qquad (1)$$

where $\hat{y}$ is dependent variable estimate (estimated parameter of GTE); k is the number of factors; x_i and x_j are the factors (independent variables) defining the physical meaning of the equation (1). Model coefficients $a_0, a_i, a_{ij}, a_{ii} \dots$ are the statistical estimations calculated according to the experimental data presented as the matrix of the output signals (parameters) of the object $\mathbf{Y}_r$ and the matrix of the input signals (factors) $\mathbf{X}_{rl}$, $r = 1,\dots,n; \, l = 1,\dots,p$, where n is the number of the parameter's measurements (the number of rows), p is the number of the registered parameters (the number of columns). The mathematical model (1) is the model that reflects with some definite exactness the behaviour of GTE or separate systems in the frames of the underlined factors space. The factual data on GTE functioning are used for constructing the diagnostic model. These data can be obtained from the maintenance documentation, where the values of the measured parameters are registered. Therefore, the statistical model is built according to the results of observing the engine maintenance. These results can be presented as the information tables with the emphasized matrix-column of the output signal (response) $\mathbf{Y}_r$ and the matrix of factors $\mathbf{X}_{rl}$.

Taking into consideration the data on the GTE maintenance, it is necessary to investigate the dependence of the temperature of the lubricant at the output of the high pressure turbine centre support bearing T_{oil_output} on the maintenance factors. The speed of the high-pressure compressor n_{HPC} , oil pressure at the input p_{oil_input} and the temperature at the input T_{oil_input} are used as maintenance factors [3]. Initial statistical data for observed engine are presented on Figure 1. The Statistica package is used for the regression analysis.

	1	2	3	4
	T_{oil_output}	n_{HPC}	p_{oil_input}	T_{oil_input}
1	107	7490	4,8	38
2	28	5240	4,4	22
3	36	5240	5	20
4	70	5840	5	32
5	80	7350	5,2	42
6	42	5420	4,8	22
7	74	6040	5	32
8	84	7240	5,4	26
9	84	7240	5	26
10	84	7180	5	26
11	82	7180	5,2	26
12	86	7290	5,2	28
13	86	7290	5,2	28
14	86	7290	5,2	28
15	88	7280	5	28
16	90	7310	5	28
17	60	6410	4,8	22
18	74	6860	5	28
19	74	7360	5,2	30
20	68	6770	5,2	26
21	76	6960	5	26
22	84	7110	5,2	28
23	82	7190	5	26

Figure 1. Initial statistical data

The regression model for the definite engine was obtained according to the initial data using Statistica package. We consider the following three independent (explaining) parameters: compressor speed, oil pressure and temperature at the input. The obtained results of regression model calculations are presented below on Figure 2.

N=23	Regression Summary for Dependent Variable: T_{oil_output} (Yunusov R= ,92560099 R?= ,85673720 Adjusted R?= ,83411675 F(3,19)=37,874 p<,00000 Std.Error of estimate: 7,4445					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
Intercept			-83,4125	40,35485	-2,06698	0,052637
n_{HPC}	0,805051	0,118445	0,0202	0,00297	6,79685	0,000002
$p_{oil\ input}$	-0,003504	0,108279	-0,3111	9,61479	-0,03236	0,974525
$T_{oil\ input}$	0,217779	0,098441	0,8107	0,36644	2,21228	0,039393

Figure 2. Results of regression model calculation

Let us consider the obtained model. Coefficients of the model are the following: $b_0 = -83,4125$; $b_1 = 0.0202$; $b_2 = -0.3111$ and $b_3 = 0.8107$. So, the obtained regression model is

$$\hat{T}^*_{oil\_output} = -83,4125 + 0,0202 n_{HPC} - 0,3111 p_{oil\_input} + 0,8107 T_{oil\_input}, \qquad (2)$$

where $\hat{T}^*_{oil\_out}$ is the variable T_{oil_output} estimate.

The parameters which determine the quality of created model are:

- multiple correlation coefficient R is 0.9256;
- coefficient of determination R^2 is 0.85674;
- Fisher`s criterion is equal to 37,874.

Let us analyse the obtained coefficients for checking up the hypothesis of the regression non-significance. The coefficients of the correlation and multiple determination and the adjusted R^2 have quite high values. Statistical significance of the model can be checked by an F-test [4, 5]. In considered example the value of Fisher`s criterion is equal to 37,874, it means that this is considerably higher than the critical value $F = 3.1$ and accepting the level of significance α=0.05, and we will reject the hypothesis of the regression non-significance. Therefore, the obtained model is significant.

The analysis of *t*-criterion (the sixth column of the table) shows that the hypothesis of the variable p_{oil_output} non-significance is accepted (with p-level 0.975). In this case it will be possible to exclude this variable from consideration and to build the regression model with two other variables.

For multiple regression model according to Gauss-Markov condition the independent (explaining) variables must not correlate with each other. The obtained values of partial correlation coefficients (shown on Figure 3) prove that there exists some correlation between independent variables, but values of partial correlation are not critical to give the opportunity to say that one independent variable is the linear function of the other variables. Let us accept that in considered regression the multicollinearity is absent.

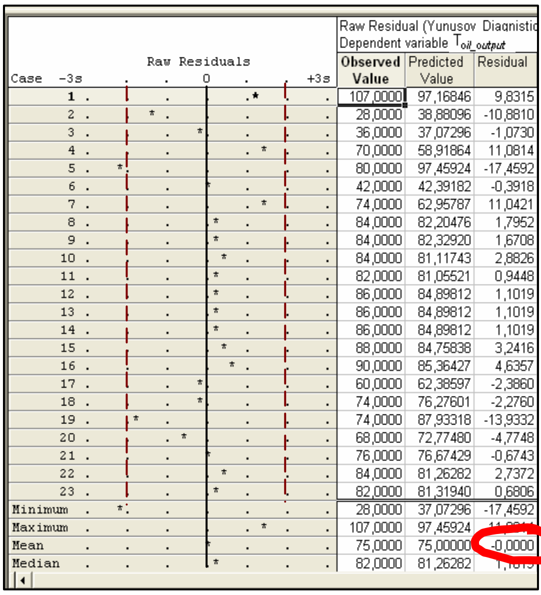
Variable	Correlations (Yunusov_Diagnistics_Stat.sta) n_{HPC}	$p_{oil\ input}$	$T_{oil\ input}$	$T_{oil\ \_output}$
n_{HPC}	1,000000	0,596982	0,470439	0,905410
$p_{oil\ input}$	0,596982	1,000000	0,261225	0,533986
$T_{oil\ input}$	0,470439	0,261225	1,000000	0,595590
$T_{oil\ output}$	0,905410	0,533986	0,595590	1,000000

Figure3. Values of the Correlations between Variables

The adequacy of the obtained model can be also checked up according to residuals analysis (see Figure 4) as the mean of the residuals is equal to zero and the variance is constant, the model is considered to be adequate. Durbin–Watson's test is used for to detect the presence of autocorrelation in the residuals. The criterion is determined by means of the following formula:

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}, \qquad (3)$$

where e_i is the residual associated with the output parameter observation at *i*-th moment of time.



Raw Residual (Yunusov Diagnistic
Dependent variable T_{oil_output}

Case	Observed Value	Predicted Value	Residual
1	107,0000	97,16846	9,8315
2	28,0000	38,88096	-10,8810
3	36,0000	37,07296	-1,0730
4	70,0000	58,91864	11,0814
5	80,0000	97,45924	-17,4592
6	42,0000	42,39182	-0,3918
7	74,0000	62,95787	11,0421
8	84,0000	82,20476	1,7952
9	84,0000	82,32920	1,6708
10	84,0000	81,11743	2,8826
11	82,0000	81,05521	0,9448
12	86,0000	84,89812	1,1019
13	86,0000	84,89812	1,1019
14	86,0000	84,89812	1,1019
15	88,0000	84,75838	3,2416
16	90,0000	85,36427	4,6357
17	60,0000	62,38597	-2,3860
18	74,0000	76,27601	-2,2760
19	74,0000	87,93318	-13,9332
20	68,0000	72,77480	-4,7748
21	76,0000	76,67429	-0,6743
22	84,0000	81,26282	2,7372
23	82,0000	81,31940	0,6806
Minimum	28,0000	37,07296	-17,4592
Maximum	107,0000	97,45924	[illegible]
Mean	75,0000	75,00000	-0,0000
Median	82,0000	81,26282	[illegible]

Figure 4. Residuals analysis

The critical values D_1 and D_2 are determined on the base of special tables for the demanded level of significance α, the number of observations n and the quantity of explaining variables m. If we are not addressing the tables, it is possible to use the approximate rules and to consider the residuals autocorrelations to be absent, if $1{,}5 < DW < 2.5$, $D_1 < DW$ and $d_2 < DW < 4\text{-}d_2$.

According to the table $\alpha = 0{,}05$ $m = 3$ and $n = 23$, $d_1 = 1{,}26$, $d_2 = 1{,}65$; and we have: $1{,}65 < 2{,}1926$; $1{,}65 < 2{,}1926 < 2{,}35$. Therefore the residuals autocorrelation is absent.

Let us construct the following charts for analysing the obtained diagnosing statistical model. The first chart (Figure 5) shows the comparison of the observed and calculated values T_{oil_output}, the second chart (Figure 6) presents the confidence intervals for the estimated parameter $\hat{T}^*_{oil\_output}$ as well as for the observed variable T_{oil_output}.

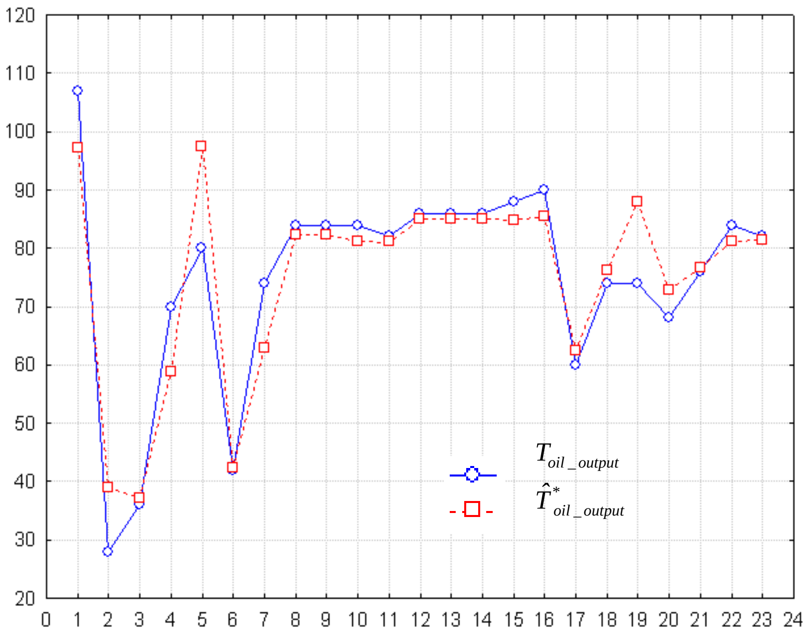


Figure 5. Values of observed T_{oil_out} and estimated $\hat{T}^*_{oil\_output}$ dependent variables

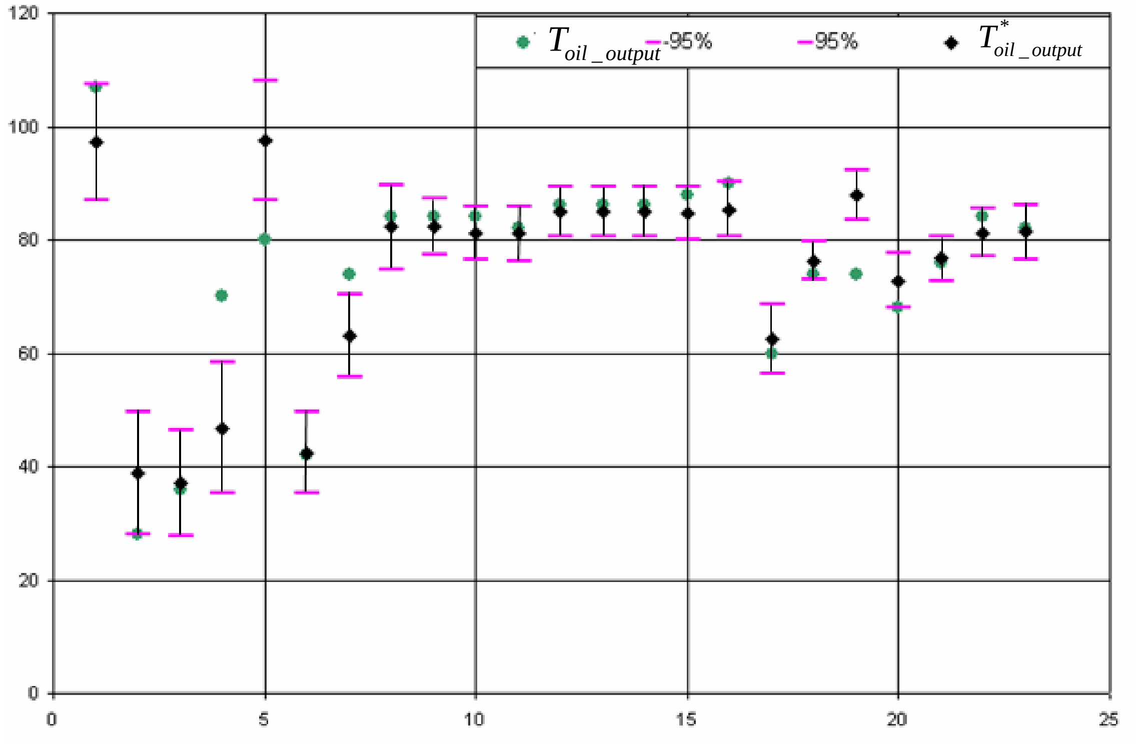


Figure 6. Confidence interval for observed T_{oil_output} and estimated $\hat{T}^{*}_{oil\_output}$ dependent variables

3. Comparative Analysis of the Models with Different Number of Variables

Testing the quality of obtained multiple regression model using different criteria we have demonstrated that the model is significant and can be used for predicting the technical status condition of the aircraft engine. In the case of three independent variables using one of the variables p_{oil_input} (oil pressure at the input) is non-significant and has to be excluded from the model. The presented below model was built without variable p_{oil_input} and has two independent variables n_{HPC} (speed of the high-pressure compressor) and T_{oil_input} (temperature at the input T_{oil_input}):

$$\hat{T}^{**}_{oil\_output} = -84{,}6237 + 0{,}0201 \cdot n_{HPC} + 0{,}8107 \cdot T_{oil\_input}\,, \tag{4}$$

where $\hat{T}^{**}_{oil\_out}$ is the variable T_{oil_output} estimate.

The obtained characteristics of the model (4) are shown on Figure 7.

	Regression Summary for Dependent Variable: T_{oil_output} (Yunusc R= ,92559673 R?= ,85672930 Adjusted R?= ,84240223 F(2,20)=59,798 p<,00000 Std.Error of estimate: 7,2562					
N=23	**Beta**	Std.Err. of Beta	B	Std.Err. of B	t(20)	p-level
Intercept			-84,6237	14,69914	-5,75705	0,000012
n_{HPC}	0,802917	0,095914	0,0201	0,00241	8,37122	0,000000
T_{oil_input}	0,217867	0,095914	0,8110	0,35703	2,27148	0,034312

Figure 7. Characteristics of the Model with Two Variables

The comparative analysis of the first and the second models showed that the second model was not worsened, but on the contrary, was even improved due to some indicators, thus Fisher`s criterion being bigger (59,789), but the standard error of the estimate being less.

The chart presented below on Figure 8 demonstrates, that the estimates $\hat{T}^{*}_{oil\_output}$ and $\hat{T}^{**}_{oil\_output}$ of dependent variable T_{oil_output} are practically equal in the considered models (2) and (4).

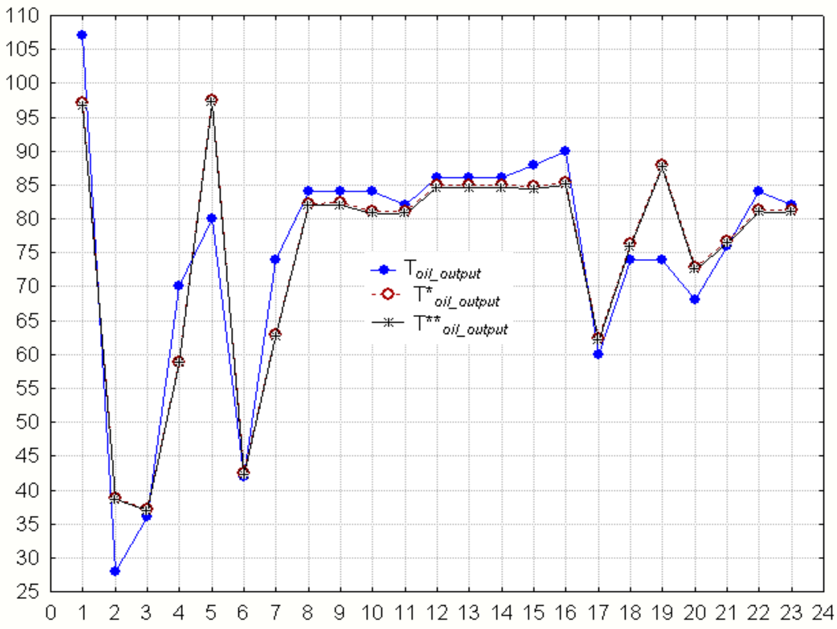


Figure 8. Comparison graphs of models

The deviations between the values of the first and the second models are not considerable. The model with one variable was analysed for the frequency of the high-pressure compressor (HPC) rotor revolving and the results happened to be worse in comparison with the second model. The main diagnostic parameter is the frequency of the HPC revolving, but such parameter as the oil temperature at the input should be considered too, because its registration improves the comprehension of statistical model of diagnosing the aircraft engine.

4. Conclusions

The multiple regression models of GTE diagnosing enable forecasting the aircraft engine state. The Statistica package enables the construction of the statistical models with different number of independent variables (factors) and their comparative analysis accomplishment.

The article acquaints us with the method of determining the model statistical significance and its adequacy as well as with the new approach to checking up its accordance with definite criteria. Only the significant and adequate models can be compared and analysed. The results of the analysis reveal the best model for the aircraft engine diagnosing and this model will posses the parameters of the highest diagnostic value.

Such approach allows to determine the most important parameters for engine controlling and diagnosing as well as to range these parameters according to the extent of their diagnostic value. The selection of the diagnostic parameters with the help of regression analysis has a lot of advantages in comparison with the heuristic approach, because there the distinguished choice of parameters exists that should be analysed in performance before occurred damages, as these parameters possess the essential diagnostic information on the engine state. Emphasizing the diagnostic parameters in accordance with the extent of their diagnostic significance, it is possible to determine the frequency of measurements of these parameters values, i.e. the most important parameters should be registered more often.

The article is written the financial assistance of European Social Fund (Project Nr.2009/0159/1DP/1.1.2.1.2/09/IPIA/VIAA/006 - The Support in Realization of the Doctoral Programme “Telematics and Logistics” of the Transport and Telecommunication Institute) and Latvian Council of Science (Project Nr. 10.0009).

References

1. Novikov, A.S., Paikin, A.G., Sirotin, N.N. *Control and Diagnostics of the Gas-Turbine Engine Performance.* Moscow: Nauka, 2007. 469 p. (In Russian)
2. Voznesensky, B.A. *Statistical Method of Experiment Planning in the Technical-Economic Investigations*. Moscow: Finance and Statistics, 1981. 263 p. (In Russian)
3. Lozitsky, L.P., Stepanenko, V.P. a. o. *Practical Diagnostics of the Gas-Turbine Engines.* Moscow: Transport, 1985. 103 p. (In Russian)
4. Greenglaz, A. Kopytov, E. *Mathematical Statistical with the Examples of Solution with the Computers: Text-book,* Riga: Higher School of Economics, 2002. 236 p. (In Russian)
5. Yatskiv, I.V. *Multi-dimension Statistical Analysis: classification and decrease of dimension: Text-book of the course "Computing Statistics"*. Riga: Transport and Telecommunication Institute, 2004. 207 p. (In Russian)