IOP Conference Series: Earth and Environmental Science

PAPER • **OPEN ACCESS**

Estimation of gas turbine technical condition using machine learning methods

To cite this article: V l Blinov and G A Deryabin 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1045** 012170

View the article online for updates and enhancements.

You may also like

- Novel method for flue gas flow of gas turbines in offshore oil production facilities
Xianglong Liu, Guang Hu, Zhi Zeng et al.

- Thermodynamic and design considerations of organic Rankine cycles in combined application with a solar thermal gas turbine
R Braun, K Kusterer, T Sugimoto et al.

- Technical Condition Estimation of the Gas Turbine Axial Compressor
V L Blinov and G A Deryabin

This content was downloaded from IP address 194.226.245.255 on 17/10/2022 at 12:45

Estimation of gas turbine technical condition using machine learning methods

V l Blinov and G A Deryabin

Department of "Turbines and Engines", Ural Federal University, 19 Mira Street, Ekaterinburg, 620002, Russian Federation

E-mail: deryabingleb.were@gmail.com

Abstract. This paper considers the method to estimate the technical condition of gas turbine power for natural gas transportation, using machine learning methods. Source data was used to archive gas-dynamic parameters from the automatic control system of the gas turbine. The method is based on changing the enthalpy of the natural gas before and after the centrifugal gas compressor is used for creating a dataset with measured parameters and power from the gas turbine. The actual power is determined from the line of modes for a certain period. The software is implemented using Python and the Scikit-learn library is used to create machine learning models. A mean average percentile error is chosen as the model quality criterion. In this paper, different sets of feature parameters and sample sizes are researched by the quality of the prediction machine learning models. Recommendations on the use of models are given. It has been established that the approach is not applicable for predicting future technical condition without the presence of data on a similar technical condition in the training sample. It is recommended to use the described approach to determine the technical condition in a period of operation in the past.

1. Introduction

Gas turbine engines (GT) are widespread in heat and electricity generation, gas, oil, and chemical industry, in aviation and ship transport [1-3]. At the natural gas compressor stations, GT is used to drive centrifugal gas compressors (GC) [4]. At the same time, GT consumes up to 5% of the volume of transported gas for their own needs. The creation modern maintenance system based on the technical condition of GT will improve the efficiency of their operation.

To implement this system, it is required to know the exact level of the technical state of the GT at each moment. The technical condition coefficient (TCC) by the power used in PJSC Gazprom is it is the most common. Precision estimation of GT power during operation is the main difficulty in TCC definition. The method for determining the power of a GT by the power consumed by the GC is most widely used. The accuracy of measuring the temperature at the inlet and outlet of the GC, gas flow through the GC has a critical effect in this case.

The quantity of data about GT has increased significantly due to the intensive growth of available computing power, reducing the cost of digital sensors, and the introduction of automatic control systems. The volume of accumulated data and the possibility of obtaining new data in real-time open up new ways for increasing the efficiency of the GT operation.

Machine learning is one of the tools that has become available with the development of information technology and the accumulation of big data; it is used for a wide range of problems and tasks [5-7]. The use of machine learning makes it possible to process and analyze large statistical data in a short

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

time. Historically, the most active application of machine learning for predicting the performance of complex technical equipment is developing in the aerospace industry [8]. At the same time, companies operating and serving GT have recently been considering machine learning methods for predicting equipment performance. Nevertheless, all the prerequisites for the introduction of such technology at enterprises are available.

In this paper, various machine learning methods for determining and predicting the GT technical condition using standard collecting data are considered.

2. Materials and methods

The object of research is a gas compressor unit when a gas turbine drives a natural gas centrifugal compressor. The study determines the TCC of a three-shaft converted aircraft gas turbine.

To conduct the research, a software package was created in Python 3 using the Pandas and Scikit-learn [9-10] libraries for working with tables and machine learning models. The research stages are presented in the diagram in figure 1.

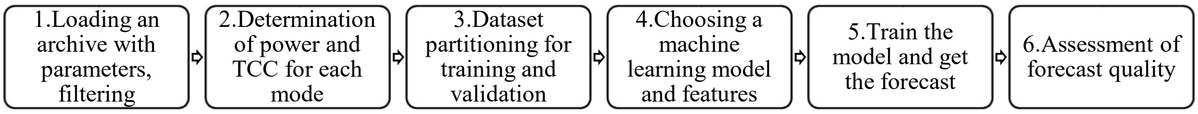

Figure 1. Scheme for assessing the technical state of a GT using machine learning methods.

The archive with routinely measured parameters is a spreadsheet with the operating parameters of the GC and GT. These parameters are recorded by the automatic control system with a step of 2 ... 6 hours for 11 months. The total number of modes (lines) is 2086. The analysis was carried out according to the following 12 parameters: T_3, air temperature at the inlet to the axial compressor; n_{LPC}, low pressure compressor rotor speed; n_{HPC}, high pressure compressor rotor speed; n_{PT}, power turbine rotor speed; T_{PT}, the temperature of combustion products at the inlet to the power turbine; P_4, air pressure behind the axial compressor; P_{FG}, fuel gas pressure; P_{1GC}, the pressure of natural gas at the inlet of the gas compressor; T_{1GC}, the temperature of natural gas at the inlet of the gas compressor; ΔP_R, pressure drop across the GC restriction; P_{2GC}, the pressure of natural gas at the outlet of the gas compressor; T_{2GC}, the temperature of natural gas at the outlet of the gas compressor.

The filtration of the GT operation modes is carried out. This makes it possible to exclude transient unsteady modes of equipment operation from further analysis. The effective power and TCC of the GT are calculated depending on the operating time.

Power and TCC for each data line are determined at the second stage. First, the power for each mode (line) is determined. The effective power is determined by the formula:

$$N_e = (i_{2GC} - i_{1G} \ ) * K\sqrt{\Delta P_R * \rho_R} + N_m \tag{1}$$

where N_m is a mechanical loss in the GC, determined depending on the power of the GC, ρ_R is a gas density at the restrictor, ΔP_R is a pressure drop across the GC restriction, K is a geometric restrictor coefficient, i_{GC} is the enthalpy of natural gas at the inlet, and outlet of the GC. The ΔP_R parameter is measured, the K parameter is determined when checking the confuser, the remaining parameters are calculated according to the measured parameters P_{1GC}, P_{2GC}, T_{1GC}, T_{2GC}.

Further, for each mode, the reduced power is calculated relative to atmospheric conditions using the formula:

$$N_{e\,r} = \frac{P_{a\,0}}{P_a} * \left(\frac{T_{a\,0}}{T_3}\right)^{1/2} * N_e \tag{2}$$

where $P_{a\,0} = 0{,}1013$ MPa and $T_{a\,0} = 288$ K are constants, P_a is a measured atmospheric pressure, T_3 is the axial compressor inlet temperature.

A period is selected to determine the TCC. For this period is plotting the power versus the limiting parameter. This graph forms the mode line as shown in picture 2. These graphs are used to determine the power at the nominal mode in order to determine the actual reduced power of the GT. The duration of the period was chosen for 4 days (figure 2).

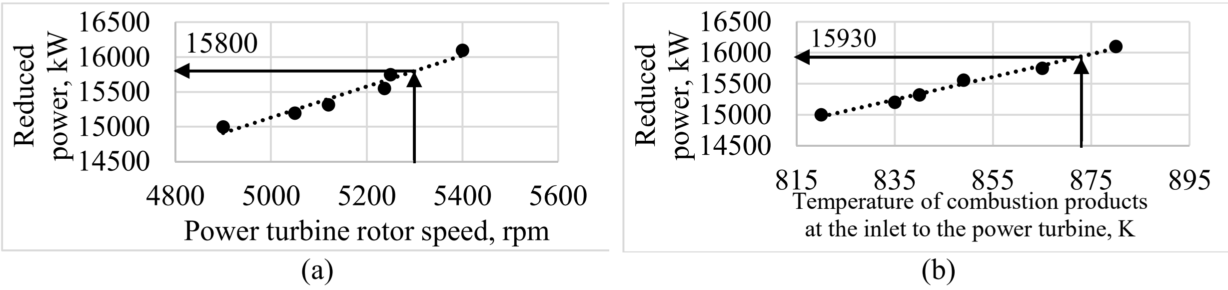

Figure 2. Determination of the reduced power of the GT at the nominal mode depending on the limiting parameter. (a) – by power turbine rotor speed, (b) – by temperature of combustion products at the inlet to the power turbine.

The smallest of the powers, determined by the limiting parameter, is substituted into the formula for calculating the TCC. In the software package, the capacities are determined not according to the schedule, but analytically:

$$K_{Ne} = N^{f}_{e_0 r} \,/\, N_{e_0} \tag{3}$$

where N_{e_0} is a nominal (passport) power of the gas turbine, kW, $N^{f}_{e_0 r}$ is an actual reduced power of GT. The resulting value characterizes the technical condition for the selected 4 days. All dimension rows are assigned the same value. Similar steps are repeated for the entire study period.

The result of stage 2 is a table with the measured parameters of the gas turbine at each time point and the corresponding TCC. This table is the original dataset in this study.

In this study, the possibility of determining the TCC of a GT under operating conditions by standard-measured parameters using machine learning methods is investigated. For these purposes, different machine learning models were selected that made predictions for different training sample sizes and features. As a result, we identified the error in the assessment of TCC and formed recommendations for engineering practice.

At the fourth stage, a model and characteristic parameters were selected, according to which the target value will be determined. The study used the supervised machine learning models provided in the Scikit-learn library.

At the fifth stage, the model was trained using a certain amount of data. At the sixth stage of the algorithm, the forecast quality was assessed. To assess the quality of the TCC forecast, the mean absolute percentage error (MAPE) was used:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \tag{4}$$

where n is a number of predictions, $\hat{y}_i$ is a predicted value, y_i is a value for validation.

3. Results

Initially, the same features were considered, according to which the effective power of a gas turbine was calculated when determining the coefficient of technical condition. Training is carried out on 75% of the sample, the model is tested on 25% of the sample. The values of errors in determining the TCC when using various models are presented in table 1.

Further analysis is carried out only for the entire dataset. The analysis of the change in the error in determining the TCC when using different amounts of data in the sample for training the considered models is carried out. The results are shown in figure 3.

Based on the data obtained, several models with the smallest error in determining the TCC were noted. Further research is related to the use of various sets of features for training these models. Before the formation of the sets of features, a correlation analysis was carried out, showing the closeness of the relationship of all considered variables with the coefficient of the technical state of the GT.

Table 1. The value of the error in determining the TCC by various models (features: P_{1GC}, P_{2GC}, T_{1GC}, T_{2GC}, ΔP_R).

No	ML model	Error in determining the TCC for all data	Error in determining the TCC for half of the data	Error in the forecast of the TCC for the half of the data that was not used during training
1	AdaBoost	0.2055	0.1736	14.9435
2	Extremely Randomized Trees	0.3380	0.2674	13.2767
3	Decision Tree regressor	0.3462	0.2514	14.5109
4	Random Forest regressor	0.3924	0.3056	13.8178
5	Bagging regressor	0.4189	0.2923	14.5426
6	K-neighbors regressor	0.5243	0.5808	12.3288
7	Histogram-Based Gradient Boosting	0.5538	0.4360	13.9362
8	Gradient Tree Boosting	0.5769	0.4532	16.2103
9	Polynomial regression	1.4496	1.1597	14.2861
10	RANdom SAmple Consensus	2.3888	1.5460	8.8177
11	Bayesian Regression	2.3948	1.5020	11.0620
12	Linear regression	2.3949	1.4959	10.8028
13	Least Angle Regression	2.3949	1.4940	10.0504
14	Orthogonal Matching Pursuit	2.3949	1.4959	10.8028
15	Partial Least Squares	2.3949	1.4959	10.8028
16	Ridge regression	2.4053	1.4998	10.9675
17	Kernel ridge regression	2.4730	1.5342	9.2060
18	Stochastic Gradient Descent	2.5030	1.5846	12.6139
19	Generalized Linear Regression	2.6993	1.7906	10.6639
20	Lasso	2.7784	2.3784	12.1213
21	LARS Lasso	2.8978	2.4913	12.2404
22	Gaussian Processes	3.0149	3.0673	90.8093
23	Multi-layer Perceptron	5.7697	5.7426	46.1593
24	Support Vector Regression	6.3520	2.7994	10.5413
25	Elastic-Net	6.3822	2.4913	12.2404
26	Passive Aggressive Algorithms	6.4845	5.5739	7.6462

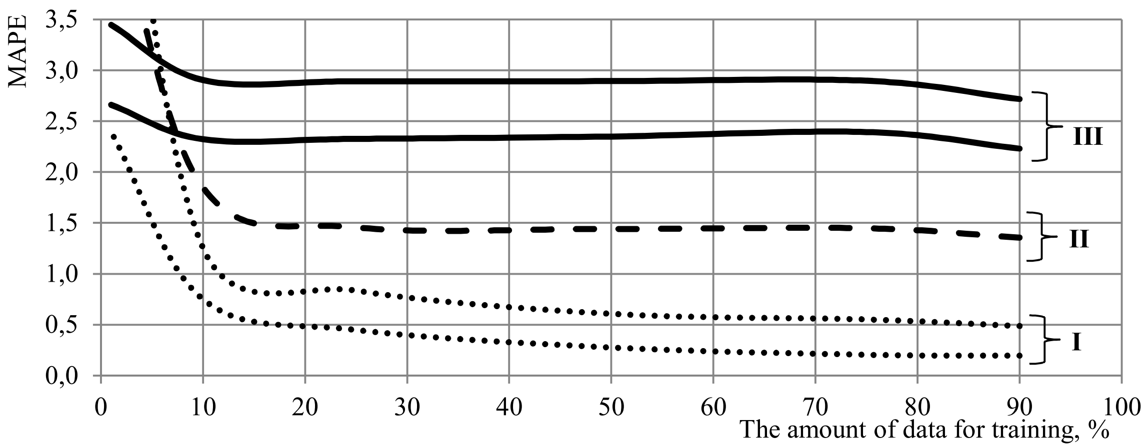

Figure 3. The level of error in determining the TCC by different models when changing the sample size for training (features: P_{1GC}, P_{2GC}, T_{1GC}, T_{2GC}, ΔP_R). I is area of ensemble models, II - Polynomial regression, III – linear models.

The values of errors in determining the TCC by the considered models during training on various sets of features are presented in table 2. Training was carried out on 50% of the sample, the accuracy of the model was assessed on the remaining 50% of the sample. Taking into account the results of the correlation analysis, the following sets of features were selected (in the order of the columns in table 2):

- all GT parameters in the sample (T_3, n_{LPC}, n_{HPC}, n_{PT}, T_{PT}, P_4, P_{FG}, T_{1GC}, P_{2GC}, P_{1GC}, T_{2GC}, ΔP_R);
- parameters used to determine the effective power of the gas turbine (P_{2GC}, P_{1GC}, T_{1GC}, T_{2GC}, ΔP_R);
- only GT parameters (T_3, n_{LPC}, n_{HPC}, n_{PT}, T_{PT}, P_4, P_{FG});
- parameters of the engine with a high tightness of connection with the TCC (P_{FG}, ΔP_R);
- parameters of GT with high and medium tightness of connections with the TCC (n_{LPC}, n_{HPC}, n_{PT}, P_{FG}, T_{1GC}, T_{2GC}, ΔP_R);
- rotational speed of all GT shafts (n_{LPC}, n_{HPC}, n_{PT});
- parameters of GPU with low tightness of connections with the TCC (T_3, T_{PT}, P_4, P_{2GC}, P_{1GC});
- parameters of the GPU with an average tightness of connections with the TCC (n_{LPC}, n_{HPC}, n_{PT}, T_{1GC}, T_{2GC});
- gas turbine operating hours only;
- all parameters of the GT and operating time (Operating time, T_3, n_{LPC}, n_{HPC}, n_{PT}, T_{PT}, P_4, P_{FG}, T_{1GC}, P_{2GC}, P_{1GC}, T_{2GC}, ΔP_R)

Table 2. Values of errors in determining the TCC during training on various sets of features (superscript numbers 1,2,3 indicate the smallest errors for a specific model).

No	Model	Set of features									
		1	2	3	4	5	6	7	8	9	10
1	Linear regression	1.36^2	2.35	2.10	2.60	1.42^3	5.24	5.25	4.90	2.19	1.28^1
2	Polynomial regression	1.13^2	1.44^3	1.61	2.38	1.15^2	3.96	3.08	3.06	2.16	1.06^1
3	Random Forest regressor	0.44^3	0.47	0.76	1.95	0.52	2.68	1.77	0.72	0.08^1	0.21^2
4	K-neighbors regressor	1.27	0.58	1.96	2.21	1.44	2.61	4.05	2.43	0.08^1	0.15^2
5	Decision Tree regressor	0.42^3	0.42^3	0.80	2.18	0.48	2.83	1.70	0.58	0.05^1	0.13^2
6	Bagging regressor	0.46^3	0.51	0.80	1.96	0.52	2.74	1.85	0.79	0.08^1	0.25^2
7	Extremely Randomized Trees	0.40	0.36^3	0.70	1.99	0.45	2.49	1.54	0.61	0.09^1	0.26^2
8	AdaBoost	0.21^3	0.28	0.50	1.96	0.29	2.17	0.98	0.39	0.05^1	0.09^2
9	Gradient Tree Boosting	0.56^3	0.60	0.88	2.04	0.63	3.16	2.28	1.08	0.05^1	0.25^2
10	Histogram-Based Gradient Boosting	0.52^3	0.61	0.82	1.98	0.62	2.95	1.97	0.92	0.28^1	0.34^2

4. Discussion

Based on the analysis of the received errors in determining the TCC by various models throughout the data set (column 3 of Table 1), the following models were identified with the best accuracy: AdaBoost, Extremely Randomized Trees, Decision Tree regressor, Random Forest regressor, Bagging regressor, K-neighbors regressor, Histogram-Based Gradient Boosting, Gradient Tree Boosting. The error in determining the TCC does not exceed 0.6% for these models. The Polynomial regression model has an error of 1.4%. For the rest of the models, the error in estimating the TCC is in the range from 2.4 to 6.5%. In this work, a relatively high error in determining the TCC was obtained using the MLP neural network, however, standard parameters were used, no additional settings were made, the model must be investigated separately.

If the entire dataset is divided into two equal parts according to the operating time of the GT, then it is possible to train the model only in one part, and check on the other. Since with an increase in operating hours, the TCC of GT decreases, both samples may contain similar values of the attributes at different values of the TCC. Thus, an assessment was made of the applicability of the machine learning models under consideration for predicting the TCC in the future with the degradation of the technical state of the GT (column 5 of table 1).

The forecast of the TCC values for the second part of the data significantly worsens if we are using a model which trained by the first part of the dataset. Therefore, the approach is not applicable for predicting future TCC without the presence of data on a similar technical condition in the training sample. Linear models demonstrated relatively more accurate forecasting results, but are still not sufficient for the application.

Based on the analysis of the influence of the sample size for training the model, it is assumed that for the problem under consideration in the future, it is optimal to use 50% of the sample during training. The models of groups I and II have sufficient training accuracy even for 10% of the entire sample.

To determine the TCC, it is worth using the entire set of the considered features. Based on the analysis of the use of various sets of features (table 2), it was noted that sets 10, 9, and 1 have the highest accuracy. Also, relatively high accuracy was obtained when training using a set of features 5, which corresponds to the results of correlation analysis. Linear models (Linear regression and Polynomial regression) show the best results when using all GT parameters. In general, by all features, it can be noted that the best results were obtained using the AdaBoost, Extremely Randomized Trees, Decision Tree regressor, and Random Forest regressor models.

5. Conclusion

In this research the following main results were obtained:

- A software package for studying the possibility of using machine learning methods for assessing the technical state of a GT has been implemented. The GT TCC was determined for different sets of feature parameters and machine learning models.
- It has been found that a satisfactory level of error of 1% can be achieved using even 10% of the data for training in some models. For best accuracy, it is recommended to use 50% of the training data.
- The highest accuracy in determining the TCC is achieved when using all 12 measured parameters and operating time assigns. Sufficient accuracy (above 99%) is also achieved when using separately the parameters of the GC or GT.
- To achieve the highest accuracy in determining the TCC, it is recommended to use the AdaBoost and Decision Tree regressor models
- It has been established that the approach is not applicable for predicting future TCC without the presence of data on a similar technical condition in the training sample. It is recommended to use the described approach to determine the TCC in an arbitrary period of operation in the past if there are reliable values of the TCC obtained as a result of tests.
- A promising direction for the development of the work is associated with the study of the method on other types of gas turbine plants, including in the analysis of such parameters of the gas turbine plant as the vibration level, oil quality, and the operation of auxiliary equipment.

6. References

[1] Saravanamuttoo H I H and Rogers G F C 2001 *Gas turbine theory* (Harlow: Pearson Education)
[2] Islas J 1999 The Gas Turbine: A New Technological Paradigm in Electricity Generation. *Technological Forecasting and Social Change* **60(2)** 129-148
[3] Benito A and Alonso G 2018 *Energy Efficiency in Air Transportation* (Oxford: Butterworth-Heinemann)
[4] Rengui C 2000 Gas Turbines Used in Natural Gas Pipelines. *Petroleum planning & engineering* **4**
[5] Michelassi V, Allegorico C, Cioncolini S, Graziano A, Tognarelli L, Sepe M 2018 Machine Learning in Gas Turbines. *ASME Mechanical Engineering* **140(09)** 54–55
[6] Cuccu G, Danafar S, Cudré-Mauroux P, Gassner M, Bernero S, Kryszczuk K 2017 A data-driven approach to predict NOx-emissions of gas turbines. *IEEE International Conference on Big Data* 1283-1288
[7] Goyal V, Xu M, Kapat J, Vesely L 2021 Prediction Enhancement of Machine Learning Using

Time Series Modeling in Gas Turbines. *ASME Turbo Expo* **4**

[8] Lary D J 2010 Artificial intelligence in Aerospace. *Aerospace technologies advancements* 492-516

[9] Pedregosa F et al. 2011 Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12** 2825-2830

[10] McKinney W 2010 Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference* **445** 51-56