Article

Hybrid data-driven and physics-based modelling for gas-turbine prescriptive analytics

Sergei Belov [1], Sergei Nikolaev [1,*] and Ighor Uzhinsky [1]

[1] Skolkovo Institute of Science and Technology
* s.nikolaev@skoltech.ru

Abstract: Abstract. This paper presents a methodology for predictive and prescriptive analytics of complex engineering systems. The methodology is based on a combination of physics-based and data-driven modeling using machine learning techniques. Combining these approaches results in a set of reliable, fast, and continuously updating models for prescriptive analytics. The methodology is demonstrated with a case study of a jet-engine power plant preventive maintenance and diagnostics of its flame tube.

Keywords: hybrid modelling, prescriptive analytics, gas engine, machine learning

---

1. Introduction

For predictive maintenance of gas turbines, a physics-based approach and a data-driven approach are the two most common methods.

1.1. Physics-based approach

The physics-driven approach uses numerical modeling of a gas turbine and its subsystems. Gas turbines' working processes are described mostly by thermodynamics equations. Modeling includes computational fluid dynamics (CFD) simulation of compressors and turbines [1], simulation of the burning process [2], and high-level modeling [3]. Typically, a physical model is designed using a combination of disciplines, such as hydraulics, pneumatics, mechanics, electromechanics, thermodynamics, and chemistry. The physical model is validated using field data.

Previous papers have presented generalized models for steady-state condition monitoring [4]. Algorithms based on exergetic analysis have shown a good correlation with the field data. This model can be used to determine thermodynamic parameters providing suitable operating conditions.

However, physics-based modeling also has drawbacks:

- Subsystems designed with simplified differential equations are not accurate enough for realistic simulation of a system's dynamics. More complex models will require high computation power.
- Usually, operators and predictive system developers do not have component characteristics (e.g. performance maps of compressors and turbines) required for physics-based model development.

1.2. Data-driven approach

A data-driven approach uses field data to design statistics-based or machine learning-based models. Compared with physics-based modeling, the data-driven approach does not need engineering documentation (performance maps, etc.) to develop a model. However, this approach requires a lot of field data. Algorithms such as support vector machines (SVM) have shown good results in [5,6]. SVM has been used for gas turbine fault detection, where it showed an accuracy greater than 80% for test data and gas generator prognostics. To monitor vibration levels, algorithms such as random forest or gradient boosting are used [7]. The main problem of the implementation of these algorithms is the lack of field data for training.

Artificial Neural Networks (ANNs) are promising algorithms in industrial predictive analytics. An interesting application of neural networks to gas turbines is unsupervised learning [8]. This study examined the problem of gas turbine combustion monitoring. The performance of an unsupervised model was better than that of a model with handcrafted features.

The drawback of fault detection is that it does not determine the reason for the abnormal behavior it detects. The new paradigm in machinery maintenance is prescriptive analytics. Not only should the model identify any abnormal behavior of the gas turbine or other equipment, but also identify the reason for it. Moreover, lack of labeled data, which are a group of samples that have been tagged with one or more labels like different faults, complicates the development of feasible algorithms.

The complexity of the task makes it difficult to develop a model using only one of these approaches; the task requires more complicated hybrid modeling, which uses both physics-based and data-driven approaches.

1.3. Digital Twin

Traditionally, computer-aided engineering (CAE) systems and tools have focused on the development and production stages of the product life cycle, including design, testing, model validation, and manufacturing. Since 1980s, a huge amount of data has been generated, and it is used to develop and manufacture complex products in industries such as aerospace, automotive, and machinery. Nowadays, more and more attention is paid to multi-level simulation of products to support their development process and reduce the number of physical tests needed.

System models built using systems modeling languages (SysML), functional models built in Modelica-oriented environments, finite-element models, and finite-volume models are managed using special modules in product life cycle management (PLM) systems called simulation process and data management (SPDM) modules. All the models mentioned above, along with increasing computing power, make it possible to introduce the concept of digital twins [9]. A digital twin is based on the virtual physics-driven model of a product, system, or process. The digital twin enables real-time monitoring to avoid malfunctions before they occur in its physical counterpart [10]. Also, a digital twin reduces the cost of system testing and verification. Unlike an ordinary virtual model, which simulates a perfect product, a digital twin represents a particular instance of a product at different stages of its life cycle (testing, production, maintenance, and disposal).

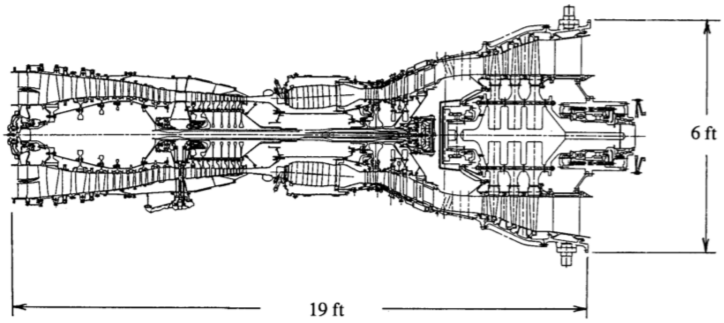
In this study, we aim to utilize the concepts of digital twins and hybrid modeling for prescriptive analytics of a gas-turbine engine. The efficacy of the approach will be demonstrated with a case study.

2. Materials and Methods

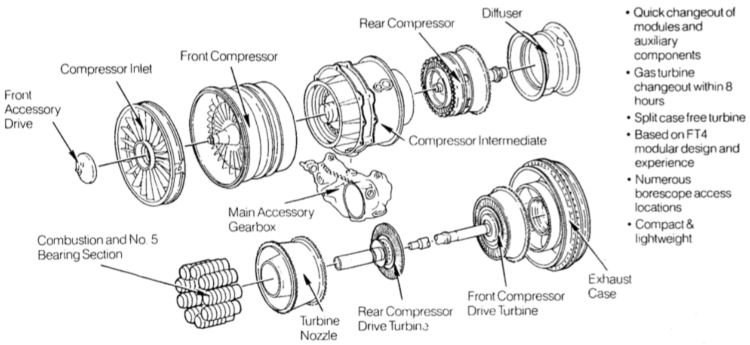
The study considers prescriptive analytics of an FT8 gas generator. It is derived from the Pratt & Whitney JT8D series aircraft engine [11]. The JT8D is a two-spool engine with a fan at the compressor inlet. The configuration of the FT8 includes two new compressor stages instead of the fan. The FT8 has a turbo-annular combustor arrangement. The combustor consists of 9 individual flame tubes [12]. A high-pressure turbine (HPT), a low-pressure turbine (LPT), and a power turbine (PT) are placed after the combustion chamber. Fig. 1 shows the scheme of the FT8 gas generator.

2.1. Problem statement

The combustion system is a critical part of any gas engine. Combustion conditions are commonly monitored in the industry [8,13]. A combustion process usually occurs under high pressure, temperature, and gas flow rate conditions that create high thermodynamic loading for the combustor components. Combustion instabilities and injector malfunctions may lead to imbalanced fuel distribution in the flame tubes and cause serious faults, such as fuel nozzle faults, flame tube burnouts, and intensive vibrations. These anomalies may lead to catastrophic failures of gas turbines.



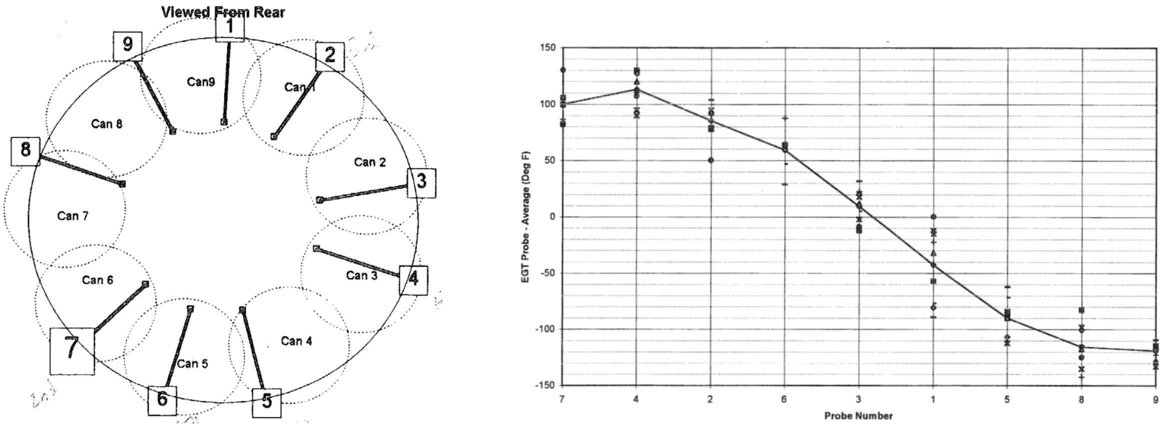
(a) FT8 gas generator scheme



(b) Major assembly sections of FT8 gas generator

Figure 1. FT8 gas generator [12]

Thus, early detection of abnormal behaviors and flame tube malfunctions is important during the gas turbine life cycle. The flame tube malfunction can be detected based on the temperature profile of thermocouples placed at a certain distance from the combustion chamber. The FT8 gas generator is equipped with exhaust gas temperature (EGT) probes. They enable operators to ensure an even temperature distribution. The system for EGT measurements consists of nine thermocouples located after the LPT (Fig. 2a). Because of the EGT probes' locations, a properly performing gas engine has a characteristic curved EGT profile. The profile has a special shape, as shown in Fig. 2b.



(a) Rotated probes location due to flow swirl **(b)** Characteristic EGT profile of gas engine

Figure 2. EGT probes' placement and temperature profile characteristics [8]

In our work, the following data was available:

1. Two years of real FT8 gas generator exploitation data

2. 24 features, such as pressures and temperatures in different parts of the turbine, environment conditions, shafts' rotary speed.
3. Moments of time when flame tubes were simultaneously repaired
4. Numbers of flame tubes which were broken

2.2. Physics-based model of a gas turbine

In this chapter, the development of a gas turbine physics-based model is described. Gas turbine power plants produce mechanical power from the expansion of hot gases in a turbine. The Brayton cycle is a typical thermodynamics cycle upon which all gas turbines operate [14]. A typical industrial gas turbine power plant has two shafts and a detached power turbine, which is connected to a generator [12,15].

The physics-based model of a gas turbine enables the detection of malfunctions even without any field data, using thermodynamic laws and equations [16]. A digital twin of a gas turbine power plant would allow the operators to simulate any malfunction to perform prescriptive maintenance of gas turbines. However, the lack of data on the physical characteristics of gas turbines complicates the task of developing the digital twin of a gas turbine power plant.

Much information about FT8 gas turbine power plants required for developing a digital twin is unknown. Some papers describe an approach for the analysis of FT8 gas turbines [17,18], but this analysis is not enough to develop a digital twin.

A scaling procedure for performance maps is a helpful tool for the analysis of compressors and turbines [19]. The scaling procedure described in present article suggests the following approach:

1. An engineer defines a type of the considered compressor or turbine.
2. A component of the same type with a known performance map (reference map) is taken.
3. The reference map is scaled by mass flow rate, pressure ratio, rotary speed, and isentropic efficiency factors to obtain an actual performance map.

Taking into account reference maps for components of different types [1,20–24] makes it possible to design a physics-based model of the FT8 gas turbine.

2.3. Workflow for physical model development

The physics-based model was developed using the Python programming language. The model considers only a gas engine without a power turbine. The model could also be simplified assuming that the steady-state relationship of the pressure after low-pressure compressor (LPC) vs. the rotary speed of high-pressure compressor (HPC) could be defined, as shown in Fig. 3. The graph was taken from the FT8 gas turbine documentation [25]. Fig. 4 presents a workflow for the physics-based model development.

The approach used in this study consisted of three main steps:

1. The model is represented as a black box. Only the input parameters and output parameters are known. The field data and model documentation are collected.
2. Some subsystems are identified using papers, documentation, and measurements, while subsystems remain as black boxes.
3. Optimization algorithms are used to tune the parameters of unknown subsystems using field data.

2.3.1. Subsystems description

In this study, the physics-based model of a gas turbine was divided into submodels connected in series. The submodels were described by equations determining how they work. The model developed in this study used a gas model with a 6 component gas:

$$\overline{y} = (y_i) = \begin{pmatrix} N_2 & O_2 & Ar & H_2O & CO_2 & C_{10}H_{20} \end{pmatrix}$$

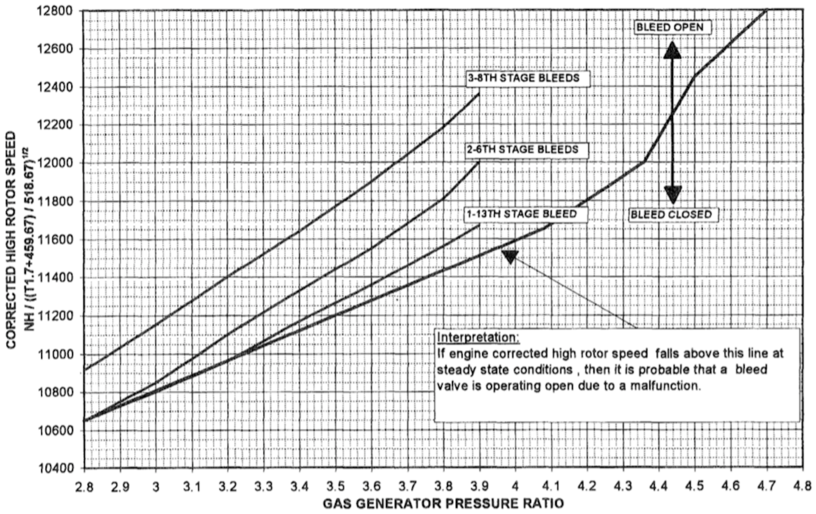


Figure 3. Pressure after LPT vs. on HP system rotary speed [25]

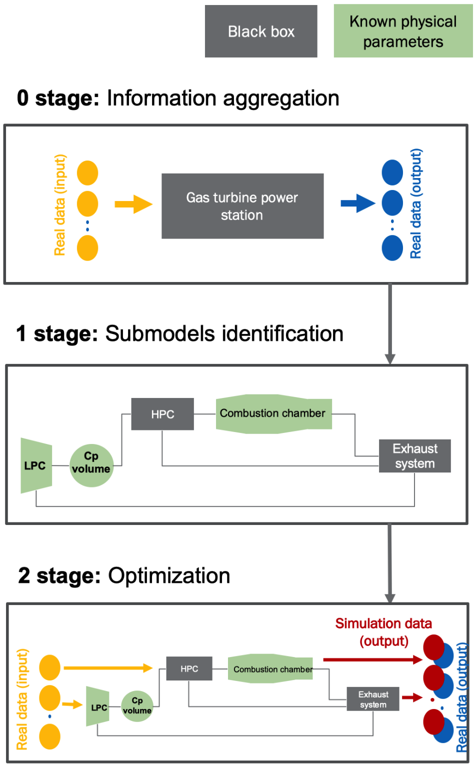


Figure 4. Workflow for physics-based model of a gas turbine power plant development

Enthalpy of formation of species h_{f0_i} and specific heat c_{p_i}, c_{v_i} vs. temperature in the range of 200 K to 6000 K were taken from NASA documentation. Fig. 5 represents the composition of the physical model.

The model uses field data to define initial conditions, such as the pressure in chambers, environmental temperature, pressure, and humidity. The submodels are elaborated below.

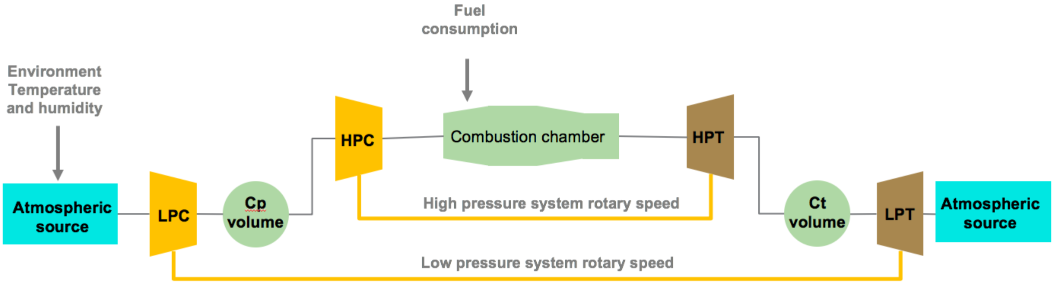


Figure 5. Architecture of the gas turbine model

2.3.2. Atmospheric source

The atmospheric source represents a submodel with no input and constant thermodynamic values. To be initialized, the model required pressure P, temperature T, species fraction y, and molar mass μ. The submodel calculates a specific enthalpy h_{out} and a heat capacity ratio γ:

$$c_{p_{mix}}(T) = \sum_{i=1}^{6} y_i c_{p_i}(T),\ c_{v_{mix}}(T) = \sum_{i=1}^{6} y_i c_{v_i}(T)$$

$$\gamma = \frac{c_{p_{mix}}}{c_{v_{mix}}}$$

$$h_{out} = c_{p_{mix}} T$$

2.3.3. Chamber with constant volume

The chamber model solves the variation of internal energy using the first law of thermodynamics for an open system. This study assumed that thermal losses within the environment are negligible. To be initialized, the model requires the values of pressure P, a temperature T, species fraction y, molar mass μ, and volume V. The submodel calculates the specific universal gas constant for the mixture.

$$r_{mix} = \sum_{i=1}^{n} \frac{y_i}{\mu_i} R$$

$$\rho_{mix_{init}} = \frac{P_{init}}{r_{mix} T_{init}}$$

$$\rho_i = y_i \rho_{mix}$$

The following equations are used to calculate a steady-state condition during simulation:

$$\frac{d\rho_i}{dt} = \frac{1}{V}\left(y_{i_{in}} \dot{m}_{in} - y_{i_{out}} \dot{m}_{out}\right)$$

$$\frac{dT}{dt} = \frac{\dot{m}_{in} h_{in} - \dot{m}_{out} h_{out} - m \sum_{i=1}^{6} \frac{dy_i}{dt} u_i - \frac{dm}{dt} \int c_{v_{mix}} dT}{m c_{v_{mix}}}$$

$$P = \rho_{mix} r_{mix} T$$

$$h_{out} = c_{p_{mix}}(T) T$$

2.3.4. Compressor

A compressor performance map is the most important data for submodel initialization. The performance map defines several steady-state compressor operating regimes with interpolation for other regimes. The model also requires the values input and output pressure P_{in} and P_{out}, input temperature T_{in}, and rotary speed ω to be initialized. The performance map is represented as two functions $PM_m(\omega, \frac{P_{out}}{P_{in}})$ and $PM_\eta(\omega, \frac{P_{out}}{P_{in}})$, which define mass flow rate $\dot{m}$ and isentropic efficiency η with a correction to a real compressor's working conditions:

$$\omega_{corrected} = \omega_{actual}\sqrt{\frac{290\ K}{T_{in}}}$$

$$\dot{m}_{corrected} = PM_m(\omega_{corrected}, \frac{P_{out}}{P_{in}}),\ \eta = PM_\eta(\omega_{corrected}, \frac{P_{out}}{P_{in}})$$

$$\dot{m}_{actual} = \dot{m}_{corrected}\sqrt{\frac{290\ K}{T_{in}}}\frac{P_{in}}{1\ atm}$$

The outlet temperature and enthalpy are then computed:

$$T_{out} = T_{in}\left[1 + \frac{(\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}} - 1}{\eta}\right]$$

$$h_{out} = c_{p_{mix}}(T)T$$

2.3.5. Combustion chamber

The combustion chamber was described using the same equations as the *Chamber* with some additional heat generated due to chemical reaction:

$$15O_2 + C_{10}H_{20} \Rightarrow 10H_2O + 10CO_2$$

The generated heat flow was calculated using the enthalpy of formation of species i:

$$Q_{react} = \sum -dm_i h_{f0_i}$$

To calculate a steady-state condition, the thermodynamic equations from *Chamber* had to be changed:

$$\frac{d\rho_i}{dt} = \frac{1}{V}\left(y_{i_{in}}\dot{m}_{in} + y_{fuel}\dot{m}_{fuel} - y_{i_{out}}\dot{m}_{out}\right)$$

$$\frac{dT}{dt} = \frac{\dot{m}_{in}h_{in} + \dot{m}_{fuel}h_{fuel} - \dot{m}_{out}h_{out} + \frac{dQ_{react}}{dt} - m\sum_{i=1}^{6}\frac{dy_i}{dt}u_i - \frac{dm}{dt}\int c_{v_{mix}}dT}{mc_{v_{mix}}}$$

$$P = \rho_{mix}r_{mix}T$$

$$h_{out} = c_{p_{mix}}(T)T$$

2.3.6. Turbine

The turbine submodel had the same initialization requirements, inputs, and outputs as the *Compressor* submodel, but the outlet temperature was calculated in a different way:

$$T_{out} = T_{in}\left[1 + ((\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}} - 1)\eta\right]$$

where η is the isentropic efficiency of a turbine. Mass flow rate $\dot{m}$ and isentropic efficiency η were determined by the performance map of a turbine. Mass flow rate and rotary speed were corrected as described in the *Compressor* model.

doi:10.20944/preprints202009.0460.v1

2.3.7. Exhaust gas temperature distribution

We assumed a Gaussian temperature distribution at the end of flame tubes and simulated each flame tube as a heat source with the heat flow and temperature distribution in a radial direction, as presented in Fig. 6. According to our assumption, a breakage in a flame tube leads to a widening of the tube's temperature profile (in a simplified model), but in fact it is actually more complex. In fact, the mean position of the temperature distribution peak changes too, because burnouts are not symmetrical.

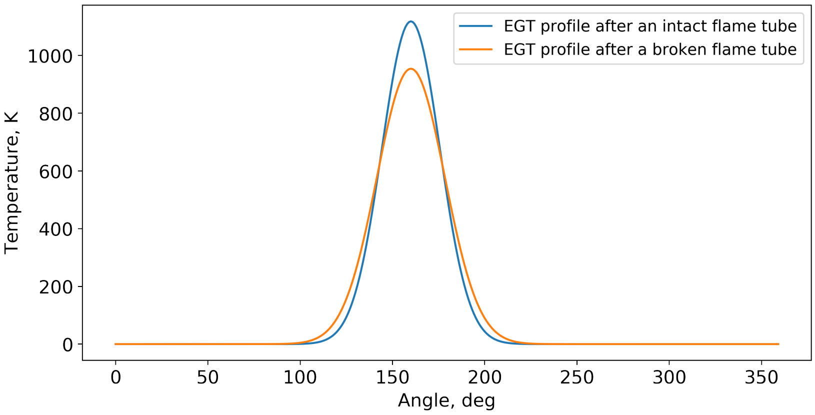


Figure 6. Exhaust gas temperature distribution after one flame tube

Temperature distributions at the end of each k flame tube were determined as:

$$T(\alpha, \alpha_k) = A * exp(\frac{(\alpha - \alpha_k)^2}{2\sigma^2})$$

And

$$\int_0^{2\pi} \sum_{k=1}^{9} T(\alpha, \alpha_k) d\alpha = 2\pi T_{EGT}$$

where α_k is the angle of the k-th thermocouple placement.

2.4. Tuning of the gas turbine physical model

Having developed the physics-based model architecture, we tuned its parameters to fit the real operation data of the gas turbine. We used the **Nelder-Mead** method to minimize the difference between the mass flow rate of the gas turbine from field data and the mass flow rate of the components of the physical model.

As was mentioned above, performance maps determine the operating conditions of turbines and compressors. In the study, the LPC performance map was obtained from the turbine operator, but the the maps for other turbine components had to be acquired. We decided to use an optimization algorithm to find the scaling parameters for fitting the performance maps. We used 10 points of the operation data from a different period of the gas turbine's performance. These points were obtained from the steady-state working conditions of the gas turbine because, in steady state, we can say that the mass flow rates of all compressors and turbines are equal. The optimization algorithm was applied to minimize the following functions:

$$F_j^N(RP_j) = \sum_{i=1}^{N} (\dot{m}_{iLPC} - \dot{m}_{ij}(RP_j))^2$$

where RP_j represents resizing parameters for the F_j function and $j = \overline{1,3}$ corresponds to HPC, HPT, and LPT respectively. N is the number of points used for optimization, which equals 10.

To determine the exact position of the EGT probes, we used the **Nelder-Mead** method. The algorithm varied the distribution of the temperature at the end of flame tubes determined by the variance and positional angle of the EGT probes to minimize the following goal function:

$$F^N(\sigma, \alpha_1, \alpha_2, ...\alpha_9) = \sum_{j=1}^{9} \left[\sum_{i=1}^{N} \frac{\widetilde{TC}_{ij}(\sigma, \alpha_j) - TC_{ij}}{N} \right]^2$$

where $\widetilde{TC}_{ij}(\sigma, \alpha_j)$ and TC_{ij} are j simulated and measured temperature on the j-th thermocouple in the i-th experiment.

2.5. Model validation

To validate the tuned model, we simulated the operation of the turbine for 2000 minutes. Fig. 7 compares the field data values and the simulation results of the physical model.

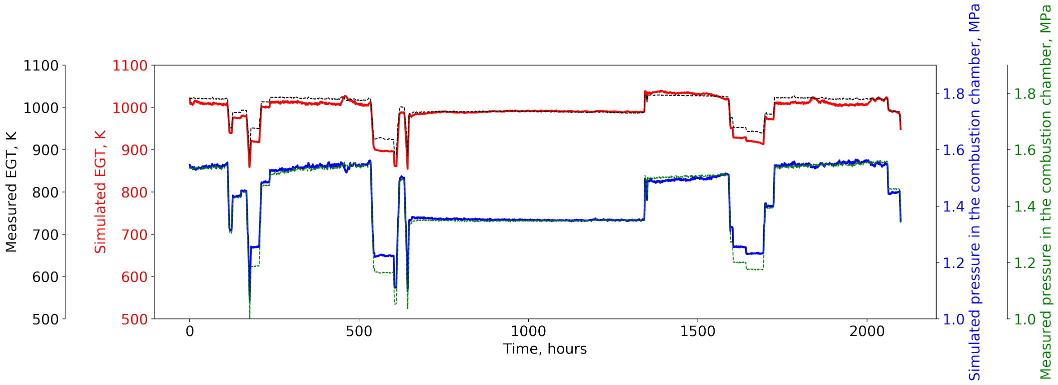


Figure 7. Simulated and measured combustion pressure and EGT histories

The physics-based model showed good accuracy in the gas turbine's steady-state regime of work, where the error was less than 2%.

Applying the optimization algorithm to determine the thermocouples' position allowed us to obtain the optimized function's value:

$$F^{10}(\sigma, \alpha_1, \alpha_2, ...\alpha_9) = 12.15K^2$$

$$F^{100}(\sigma, \alpha_1, \alpha_2, ...\alpha_9) = 10.24K^2$$

2.6. Hybrid modelling

After the physics-based model of a gas turbine was created, it was combined with a machine learning model.

Taking into account the lack of labeling (only one label was known when the flame tubes were repaired), we decided to use a machine learning model trained by a physics-based model. In this configuration, we produced much-labeled data, which were then fitted to the machine learning–based model.

The problem and assumptions can be stated as follows:

1. Nine flame tubes were placed in the combustion chamber. Each could be broken in the same period of time (in the present case study, seven flame tubes were broken simultaneously). The number of broken flame tubes at a given moment may range from zero to nine with unknown probability.
2. Malfunction of the flame tubes led to changes in the temperature profile of the EGT probes.
3. Malfunction of the different flame tubes led to unique differences in the temperature profile of the EGT probes.

We used neural networks as the architecture for the machine learning model, as they are well suited for multiclass classification problems [26,27]. It was also interesting to check different neural network structures and observe the trainability of the model.

We tried different architectures for the neural networks with different numbers of hidden layers and neurons. Although different architectures produced almost equal results, the neural network with three hidden layers performed the best.

We used two different activation functions in the hidden layers (Rectified Linear Unit (ReLU) and Exponential Linear Unit (ELU)) to compare their performance to possibly improve the results [28,29]:

$$ReLU(x) = max(0, x)$$

$$ELU(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

For the output layer, the sigmoid activation function was used:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \; prediction = \begin{cases} 1 & \sigma(x) > 0.5 \\ 0 & \sigma(x) \leq 0.5 \end{cases}$$

The inputs to the model were the EGT probes' temperatures T_i from the physics-based model, and the outputs were nine possible classes a_j corresponding to flame tube malfunctions:

$$input = \begin{pmatrix} T_1 \\ T_2 \\ \ldots \\ T_9 \end{pmatrix}, \; output = \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_9 \end{pmatrix}, \; where \; a_j = \begin{cases} 1 & j \text{ flame tube is broken} \\ 0 & j \text{ flame tube is whole} \end{cases}$$

Fig. 8 shows the neural network structure.

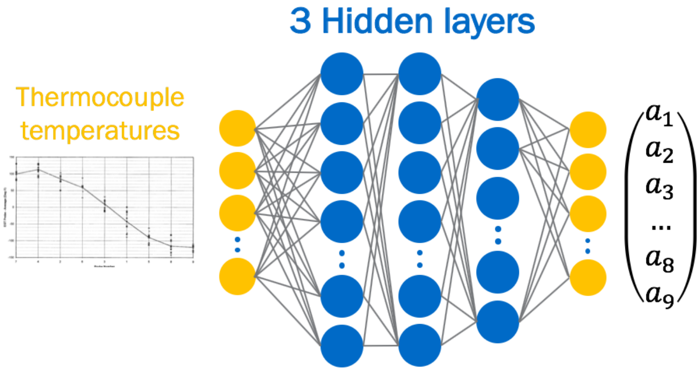


Figure 8. Structure of the neural network used for flame tube malfunction detection

We used the *Adam* optimizer [30] and *binary crossentropy* loss function [31] for the model. Let [input, hidden layer 1, hidden layer 2, hidden layer 3, output] represent the number of neurons in a network. In this study, we used the following architectures for the model:

1. $[9 - 50 - 50 - 40 - 9]$, dropout = 0.1, ReLU activation function in hidden layers
2. $[9 - 50 - 50 - 40 - 9]$, dropout = 0.1, ELU activation function in hidden layers
3. $[9 - 70 - 70 - 50 - 9]$, dropout = 0.1, ELU activation function in hidden layers

2.6.1. Combination of approaches for flame tubes health monitoring

Physics-based and machine learning models were combined in the way in which the physics-based model simulated flame tube breakages. The following algorithm describes the simulation process:

1. The physics-based model chose a random number N from zero to nine, which determined the amount of broken flame tubes in the data sample.
2. Random flame tubes $n_j,\ j = \overline{1, N}$ were defined as broken.
3. Malfunction in each n_j flame tube was simulated as the increased variance of the temperature distribution. Fig. 9 shows an example of the simulated malfunction in the third flame tube.
4. The physics-based model simulated the steady-state working conditions of the turbine and calculated the EGT distribution.

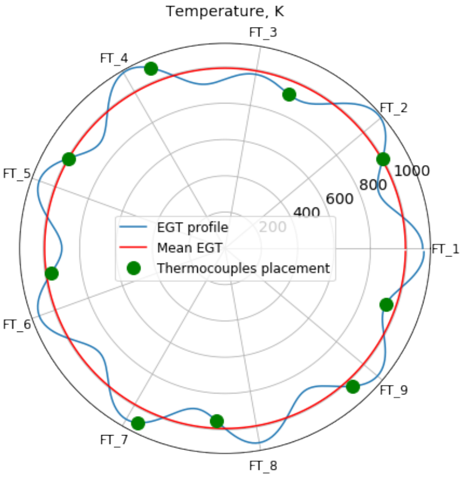


Figure 9. Temperature distribution after LPC turbine with a broken third flame tube

In the combined model, the physics-based model generated new labeled data each iteration, which prevented the neural network from overfitting. Fig. 10 shows how the physics-based model was connected with the machine learning model.

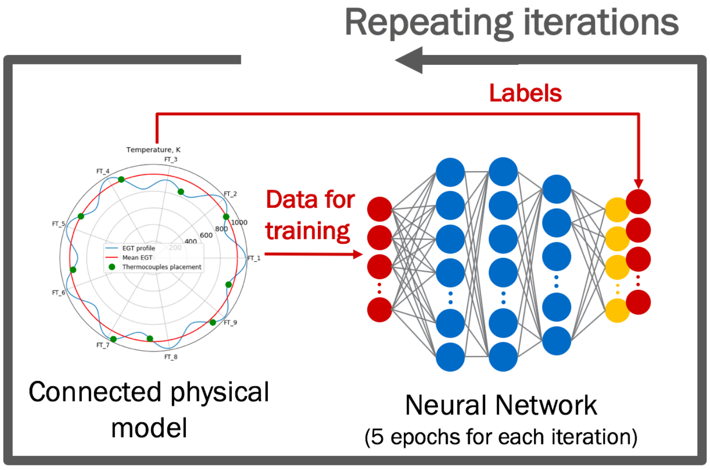


Figure 10. Architecture of a combined model

The model was trained in M iterations. The following algorithm was used for the training:

```
for $i = 0$ to $M$ do
    The physics-based model generated 900 data samples and labels;
    The accuracy of the neural network was estimated based on the generated data;
    The neural network was trained for 8–16 epochs with the data;
end
```

We determined the absolute accuracy with the following formula:

$$acc = \frac{\sum_{j=1}^{M} \sum_{i=1}^{9} p_{ij}}{9 * M}, \ where \ p_{ij} = \begin{cases} 1 & \tilde{a}_{ij} = a_{ij} \\ 0 & \tilde{a}_{ij} \neq a_{ij} \end{cases}$$

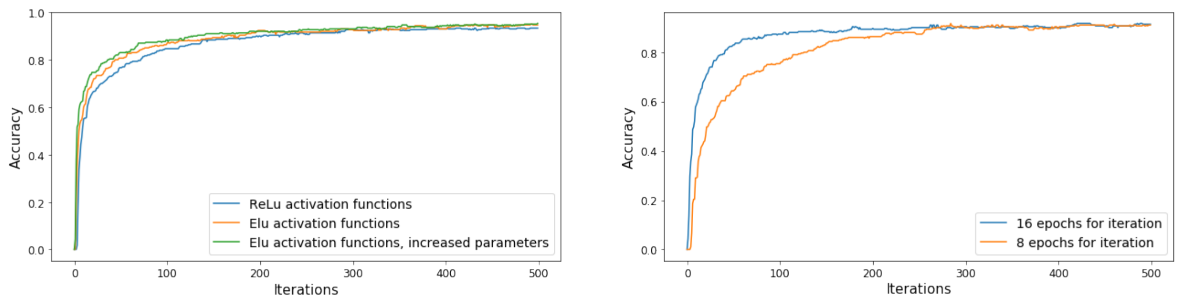
a_{ij} - label for i flame tube in j experiment out of M

$\tilde{a}_{ij}$ - prediction for i flame tube in j experiment out of M

2.6.2. Model training and architectures comparison

The most important task was to choose the right neural network architecture. We compared the neural networks with different numbers of trainable parameters and hidden layers. Each architecture was trained for 30 iterations, and its accuracy was checked on the same test dataset. The ELU activation function was used in hidden layers. Each architecture was also trained with 15 epochs in each iteration and was fitted with 100 batch sizes.

We investigated the hyperparameters of the neural network with three hidden layers. Each architecture for the model was trained with 500 iterations. The results show good trainability for the model, as presented in Fig. 11. The architecture $[9 - 70 - 70 - 50 - 9]$ with the ELU activation function in hidden layers produced the most accurate results (-0.95).



(a) Model accuracy for different NN architecture **(b)** Model accuracy for different epochs number

Figure 11. Neural network accuracy

We have chosen the following hyperparameters for training:

- Epochs number - 16
- Batch size - 40

The mean timing for each iteration shows that the physics-based model, which generates datasets, limits the speed of training:

- Mean time of the dataset generation = 9.6 s for 900 labeled data points
- Mean time of the neural network training = 1.22 s for 16 epochs

We considered other metrics as well. In the designed model, we checked the accuracy of the newly generated datasets. However, we considered a more common way to validate the model that generates just one test bench. On this test bench, we applied absolute accuracy and other metrics, such as F1 score, Area Under the Receiver Operating Characteristic Curve (AUC ROC) score, precision score, and recall score [32,33]. Fig. 12 shows the results on the accuracy for different metrics. Absolute accuracy showed the same trend with the validation set as with the generated datasets.

The accuracy of the number of broken flame tubes was also investigated. The increase in the number of broken flame tubes complicates the EGT profile, so the model should have a lower accuracy value.

We designed an experiment where the pre-trained model was applied to eight different validation sets consisting of 600 data samples each. In each validation set, the constant number of broken flame

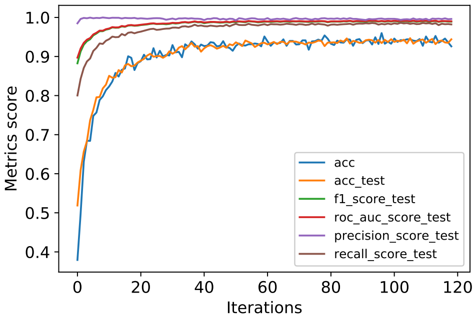


Figure 12. Accuracy results for different metrics

242 tubes was simulated (0, 1, 2, . . . , 7 broken flame tubes). Fig. 13 shows our hypothesis that the accuracy
243 would decrease with an increased number of simultaneously broken flame tubes was correct.

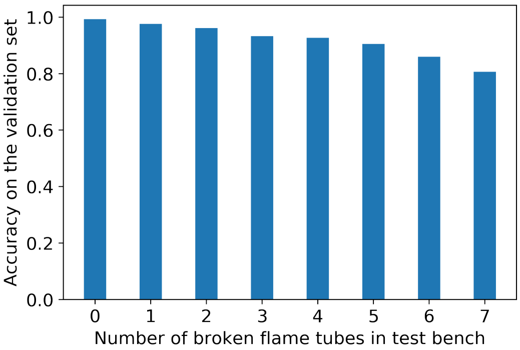


Figure 13. Absolute accuracy for different number of broken flame tubes

244 3. Results and discussion

245 The physics-based model was used to train the neural network under various conditions of
246 malfunctioning flame tube, and it achieved an accuracy of more than 0.95. The next step was to
247 validate the model using field data.

248 *3.1. Prescriptive results*

249 We applied the model to predict the probability of flame tube malfunction for each field data point.
250 As the model worked with low accuracy on the region without steady-state gas turbine performance,
251 we observed the cumulative error for 15 minutes of gas generator work (Fig. 14). We defined the 15 minutes interval as time index.

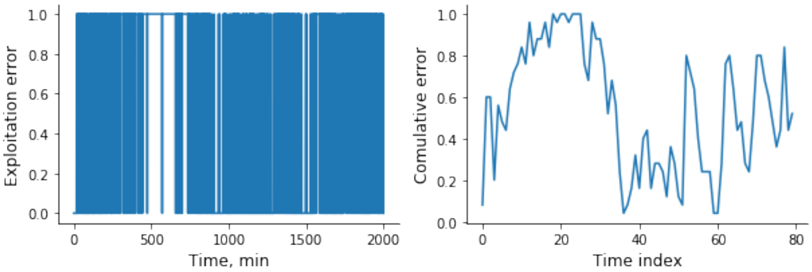
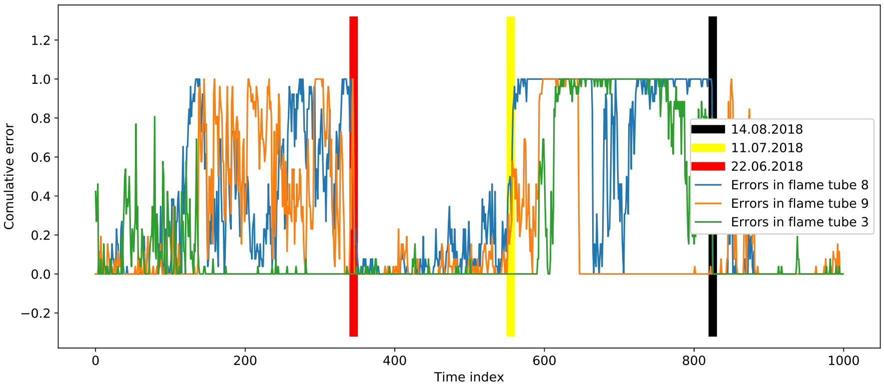


Figure 14. Error on field data and cumulative errors for the 4th EGT probe

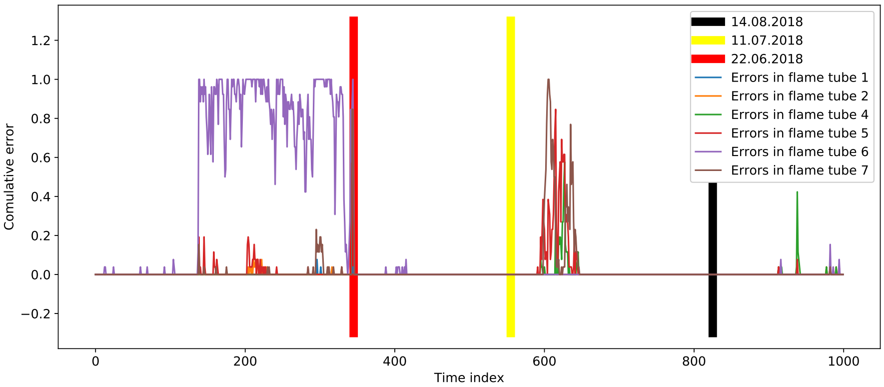
252

253 Fig. 15 shows the cumulative error over six months for all nine flame tubes. As shown in the
254 figure, we identified three regions for discussion of the predictions. The regions are drawn in Fig. 15:

1. Black color - 14.08.2018. The day when flame tubes 1, 2, 3, 4, 6, 8, 9 were repaired
2. Yellow color - 11.07.2018. The day when flame tubes have broken, according to the model prediction.
3. Red color - 22.06.2018. A previous repair of flame tubes, according to the model prediction.



(a) Breakage determination in flame tubes



(b) Intact flame tubes determination

Figure 15. Application of the model to predict exact number of broken flame tubes

The application of the model to the FT8 gas generator dataset correctly determined the breakage time for three broken flame tubes out of seven. False negative results may be due to small defects in the replaced flame tubes.

The model predicted the flame tubes' breakage a month before their repair. This means that the operating company used gas engines with broken flame tubes for about a month.

Between 06-22-2018 and 07-11-2018, some increase in error was seen. This increase presumably is linked with the breakage. It is probably connected to injector lag, so the subsequent improvement of the model will allow us to observe the causes in more detail.

The proposed methodology can be applied not only to prescriptive systems for the development of gas turbines but also to any other complex systems, such as pumps, compressors, and turbochargers. The methodology presented in this study may be integrated into maintenance, repair, and overhaul software and may be used during the machinery life cycle for better performance.

doi:10.20944/preprints202009.0460.v1

3.2. Advantages of combined approach

As this work shows, the combined approach makes it possible to predict malfunction for even a very small amount of labeled data. The following advantages of the combined approach may be of research interest in this field:

- Only a small amount of the field data is needed to adjust the model to a real turbine. In the case of the FT8 gas turbine, we worked with only 10 data points to identify the performance maps. This means that the model could be tuned to new turbines after a day of operation, which indicates the scalability of the approach.
- Compared with a purely data-driven approach, the combined model allows operators to detect not only the abnormal behavior of gas turbines but also the reasons for it.
- Many malfunctions can be simulated. As soon as the physical model is validated, we can simulate many types of faults.

4. Conclusions

This paper presented the workflow of prescriptive analytics of the gas turbine engine based on its hybrid model. The model utilizes a data-driven and a physics-driven approach. The developed model can accurately identify different malfunctions of the engine based on its performance. However, the presented hybrid approach cannot be implemented without developing a physics-based model that is tuned to fit the performance of a gas turbine. So, the developed model has some limitations:

1. Lack of engineering data is the main limitation in the development of a physics-based model. It is not possible to develop a model without performance maps for the compressors and turbines.
2. Unknown malfunctions or other malfunctions not taken into during machine learning model training, could be identified incorrectly.
3. If the construction of a gas turbine is changed (for instance, if the EGT probes are placed differently), the model has to be adjusted.

Acknowledgements

The authors would like to express their gratitude to their colleagues Artem Semenov and Nikita Losyakov for their help with numerical calculations.

Funding

This work was supported by the federal program “Research and development in priority areas for the development of the scientific and technological complex of Russia for 2014–2020” via grant RFMEFI60619X0008.

Conflicts of Interest: Authors declare no conflict of interest

Nomenclature

Latin Symbols

acc	prediction accuracy	$-$
$\tilde{a}_{ij}$	prediction for i flame tube in j experiment out of M	$-$
a_{ij}	label for i flame tube in j experiment out of M	$-$
c_{p_i}	specific heat at constant pressure of i-th gas component	$\frac{J}{kg \cdot K}$
$c_{p_{mix}}$	gas mixture specific heat at constant pressure	$\frac{J}{kg \cdot K}$
$c_{v_{mix}}$	gas mixture specific heat at constant volume	$\frac{J}{kg \cdot K}$
c_{v_i}	specific heat at constant volume of i-th gas component	$\frac{J}{kg \cdot K}$
h_{f0i}	enthalpy of formation of i-th gas component	J
h_{out}	output enthalpy	J
$\dot{m}$	mass flow rate	$\frac{kg}{s}$
M	number of experiments	$-$
P_{in}	inlet pressure	Pa
P_{out}	outlet pressure	Pa
Q	heat flow	Pa
r_{mix}	gas mixture individual gas constant	$\frac{m^3 \cdot Pa}{K \cdot mole}$
T_{in}	inlet temperature	K
T_{out}	outlet temperature	K
V	volume	m^3
y_i	i-th component of the gas mixture	$-$

Greek symbols

α_k	angular placement of the k-th thermocouple	rad
γ	heat capacity ratio	$-$
η	isentropic efficiency	$-$
$\rho_{mix_{init}}$	initial mix density	kg/m^3
ρ_i	pressure of the i-th mix component	kg/m^3
σ	standard deviation of the exhaust gas temperature distribution	K
ω	compressor rotary speed	rad/s

Abbreviations

AUC ROC	Area under receiver operating characteristic
CAE	Computer-Aided Engineering
CFD	Computational Fluid Dynamics
EGT	Exhaust gas temperature
ELU	Exponential Linear Unit
HPC	High pressure compressor
HPT	High pressure turbine
LPC	Low pressure compressor
LPT	Low pressure turbine
NN	Neural Network
PLM	Product lifecycle management
PT	Pressure turbine
ReLU	Rectified Linear Unit
RP	Resizing parameters
SPDM	Simulation Process and Data Management
SVM	Support Vector Machine
SysML	Systems Modeling Language
TC	Temperature at a thermocouple

Subscripts

in	inlet
init	initial
mix	gas mixture
out	outlet

References

1. Martins, D.R. Off-Design Performance Prediction of the CFM56-3 Aircraft Engine. *Tecnico Lisboa* **2015**.
2. Ghose, P.; Patra, J.; Datta, A.; Mukhopadhyay, A. Prediction of soot and thermal radiation in a model gas turbine combustor burning kerosene fuel spray at different swirl levels. *Combustion Theory and Modelling* **2016**, *20*, 457–485. doi:10.1080/13647830.2016.1147607.
3. Durante, A.; Pena-Vergara, G.; Curto-Risso, P.; Medina, A.; Hernández, A.C. Thermodynamic simulation of a multi-step externally fired gas turbine powered by biomass. *Energy Conversion and Management* **2017**, *140*, 182 – 191.
4. Bilgen, E. Exergetic and engineering analyses of gas turbine based cogeneration systems. *Energy* **2000**, *25*, 1215 – 1229.
5. Allen, C.W.; Holcomb, C.M.; de Oliveira, M. Gas Turbine Machinery Diagnostics: A Brief Review and a Sample Application. Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 2017, Vol. 50916, p. V006T05A028.
6. Wong, P.K.; Yang, Z.; Vong, C.M.; Zhong, J. Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. *Neurocomputing* **2014**, *128*, 249–257.
7. Mulewicz, B.; Marzec, M.; Morkisz, P.; Oprocha, P. Failures prediction based on performance monitoring of a gas turbine: a binary classification approach. *Schedae Informaticae* **2017**, *26*, 21.
8. Yan, W.; Yu, L. On Accurate and Reliable Anomaly Detection for Gas Turbine Combustors: A Deep Learning Approach. Proceedings of the Annual Conference of the Prognostics and Health Management Society. PHM Society, 2015, Vol. 6, pp. 59–71. doi:10.1145/1218063.1217942.
9. Grieves, M. Digital Twin: Manufacturing Excellence through Virtual Factory Replication. *A White Paper* **2002**.
10. Madni, A.M.; Madni, C.C.; Lucero, S.D. Leveraging digital twin technology in model-based systems engineering. *Systems* **2019**, *7*, 7.

11. Prario, A.; Voss, H. FT8A, a New High Performance 25 MW Mechanical Drive Aero Derivative Gas Turbine. Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 1990, Vol. 79061, p. V003T07A004.
12. Day, W.H. FT8: A High Performance Industrial and Marine Gas Turbine Derived From the JT8D Aircraft Engine. Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 1987, Vol. 79245, p. V002T03A005.
13. Allegorico, C.; Mantini, V. A data-driven approach for on-line gas turbine combustion monitoring using classification models. Proceedings of the Second European Conference of the Prognostics and Health Management Society, 2014, pp. 92–100.
14. Gas Turbine Power Plants. https://www.isisvarese.edu.it/wp-content/uploads/2016/03/first-law-of-thermodynamics-for-an-open-system-.pdf. Accessed: 11-08-2020.
15. Brooks, F.J. GE gas turbine performance characteristics. *GE Power Systems* **2000**.
16. Xuewu Dai.; Breikin, T.; Zhiwei Gao.; Hong Wang. Dynamic modelling and Robust Fault Detection of a gas turbine engine. 2008 American Control Conference, 2008, pp. 2160–2165.
17. Turan, O.; Aydin, H. Exergy-based Sustainability Analysis of a Low-bypass Turbofan Engine: A Case Study for JT8D. *Energy Procedia* **2016**, *95*, 499 – 506. International Scientific Conference "Environmental and Climate Technologies", CONECT 2015.
18. Dan-Teodor Bălănescu, Sorinel-Gicu Talif, E.H. Aeroderivative Pratt & Whitney FT8-3 gas turbine – an interesting solution for power generation. *INCAS BULLETIN* **2011**, *3*, 9 – 14.
19. Kurzke, J.; Riegler, C. A new compressor map scaling procedure for preliminary conceptional design of gas turbines. Turbo Expo: Power for Land, Sea, and Air. American Society of Mechanical Engineers, 2000, Vol. 78545, p. V001T01A006.
20. Cumpsty, N. *Jet Propulsion: A Simple Guide to the Aerodynamics and Thermodynamic Design and Performance of Jet Engines*; Cambridge University Press, 2003. doi:10.1017/CBO9781316223116.
21. Energy Efficient Engine High Pressure Compressor - Component Performance Report. https://ntrs.nasa.gov/citations/19850025825. Accessed: 11-08-2020.
22. Broichhausen, K. Aerodynamic Design of Turbomachinery Components - CFD in Complex Systems. *AGARD Lecture Series 195* **1994**.
23. Extended parametric representation of compressor fans and turbines. Volume 3: MODFAN user's manual (parametric modulating flow fan). https://ntrs.nasa.gov/citations/19860014467. Accessed: 11-08-2020.
24. Performance of a High-Work Low Aspect Ratio Turbine Tested with a Realistic Inlet Radial Temperature Profile. https://ntrs.nasa.gov/citations/19910010813. Accessed: 11-08-2020.
25. FT8 EGT Spread Troubleshooting. *Turbo Power and Marine Systems* **1996**.
26. Ding, C.H.; Dubchak, I. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* **2001**, *17*, 349–358.
27. Bhardwaj, A.; Tiwari, A.; Bhardwaj, H.; Bhardwaj, A. A genetically optimized neural network model for multi-class classification. *Expert Systems with Applications* **2016**, *60*, 211–221.
28. Clevert, D.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings; Bengio, Y.; LeCun, Y., Eds., 2016.
29. Pedamonti, D. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. *arXiv preprint arXiv:1804.02763* **2018**.
30. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
31. Understanding binary cross-entropy / log loss: a visual explanation. https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a. Accessed: 11-08-2020.
32. Shung, K.P. Accuracy, Precision, Recall or F1? *https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9* **2018**.
33. McClish, D.K. Analyzing a portion of the ROC curve. *Medical Decision Making* **1989**, *9*, 190–195.