TOWARDS TRUSTWORTHY DATA-DRIVEN GAS TURBINE PROGNOSTICS

Asteris Apostolidis [1], Solenn Le Dantec [2], Konstantinos P. Stamoulis [1]

[1] Amsterdam University of Applied Sciences, Rhijnspoorplein 2, 1091 GC Amsterdam, The Netherlands
[2] Mechanical Engineering Department, SIGMA Clermont, TSA 62006, CEDEX, 63178 Aubiere, France

Abstract

Trustworthy data-driven prognostics in gas turbine engines are crucial for safety, cost-efficiency, and sustainability. Accurate predictions depend on data quality, model accuracy, uncertainty estimation, and practical implementation. This work discusses data quality attributes to build trust using anonymized real-world engine data, focusing on traceability, completeness, and representativeness. A significant challenge is handling missing data, which introduces bias and affects training and predictions. The study compares the accuracy of predictions using Exhaust Gas Temperature (EGT) margin, a key health indicator, by keeping missing values, using KNN-imputation, and employing a Generalized Additive Model (GAM). Preliminary results indicate that while KNN-imputation can be useful for identifying general trends, it may not be as effective for specific predictions compared to GAM, which considers the context of missing data. The choice of method depends on the study's objective: broad trend forecasting or specific event prediction, each requiring different approaches to manage missing data.

Keywords: Data quality; Missing Data; Prognostics; Model Trustworthiness; Exhaust Gas Temperature

1. Introduction

Aircraft components produce and register sensor data since the introduction of the Flight Data Recorder (FDR) in 1960. While sensors were primarily designed for system control, the recording of Aircraft Operational Data (AOD) [1] started facilitating the introduction of data-driven Diagnostics and Prognostics, which can be subsequently utilized for strategies such as Condition-based Maintenance (CBM) and Predictive Maintenance (PdM). The goal of both CBM and PdM strategies is to minimize downtime, reduce operational costs, and extend the lifespan of gas turbine components. [2] Reliable data-driven Prognostics are essential for the understanding of the physical condition of gas turbine assemblies and parts, and for the support of maintenance operations. However, even for the most recent commercial aircraft types in service, the sensor types and topology are still dictated primarily by the requirements for effective system control, while the operational introduction of CBM and PdM is slow and fragmented. This can be attributed to various technical, operational, and regulatory reasons that must be addressed. [3] Model trustworthiness is one of them as, in general terms, effective data-driven Prognostics can only be developed if they can support real operational needs, while they must be proven superior to present, preventive methods. This last part is essential for the support of the extremely high safety standards of global civil aviation.

Trustworthy data-driven prognostics in gas turbine engines are important for three main reasons: Safety evidently comes first, while cost-efficiency, and sustainability -in terms of both lifing and emissions- can also be delivered with accurate predictions. Model quality is highly dependent on data quality, as well as model accuracy, uncertainty estimation, and practical implementation considerations in operational conditions. As with all data-driven Artificial Intelligence (AI) models, data are extremely crucial for the quality of the results. In this work, different data quality attributes

are discussed, so trust can be built. For this purpose, anonymized real-world engine data during cruise are used and examined with regards to their traceability, completeness, and representativeness.

The used dataset originates from the engines of a now decommissioned fleet of high-bypass turbofans operated between 2011 and 2018. In general, a complete dataset must contain all the possible operational conditions that an engine (or any other component) is likely to encounter when making performance predictions. An example is the difference between the different climate characteristics between Europe, the Middle East, or Southeast Asia, which demonstrate different deterioration patterns. In addition, a dataset is representative of a certain operating profile when the aforementioned conditions are also distributed proportionally to the actual encountered conditions.

2. Data Quality and Attributes

According to SAE International [4], on-wing data need to satisfy a series of different quality characteristics that can render a real-life dataset trustworthy for further use in CBM and PdM, as well as in any other AI-based application. These characteristics include Traceability, Sufficiency, Completeness and Representativeness [5] in order to be able to eliminate bias to an acceptable level.

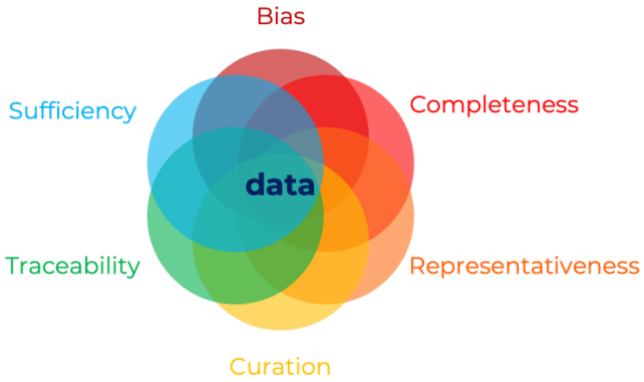

Figure 1 – Fundamental data quality concepts

2.1 Data Traceability

Data traceability refers to the ability of identifying the original data sources and the complete data path from generation to consumption, to enable trustworthy results. One of the ways that traceability can be enabled is via comprehensive data tagging. For every data point, associated timestamps need to be created including the engine serial number. Next to that, any data transformation in a data pipeline should create a similar timestamp, so operators can trace back the data processes that took place and be able to identify potential issues in relation to model behaviour. It is important to underline that as data anonymisation is a common practice, the data tagging must be reversible to identify the real-world engine.

2.2 Data Completeness, Representativeness and Sufficiency

Data completeness refers to the coverage of every possible operating condition within the training dataset. In other words, whether a given dataset covers the Operational Design Domain (ODD) of interest [6]. This enables an understanding of how an engine functions across diverse aircraft configurations, as well as operating conditions, ambient conditions, within a variety of different climates and environments [7]. If the data are complete, the model will perform well for the different diverse conditions. However, a model will underperform in conditions that were not part of its training dataset. Conversely, inadequate data completeness confines the model's efficacy to solely the operational conditions covered by the available data, potentially leading to inaccuracies in other operating regions. Data representativeness should not be confused with data completeness. Completeness refers to coverage, whereas representativeness refers to the correct distribution of data points within this range of conditions.

In this instance, the dataset in question comprises of measurements taken during the cruise phase, as the thermal instabilities and the variable environmental factors during take-off can increase uncertainty. Thus, only the cruise environment parameters will be discussed in figure 2, which in this case are for simplicity the flight altitude, the Mach number, and the ambient temperature. It is imperative to ensure that these three parameters cover the full operational spectrum and flight envelope.

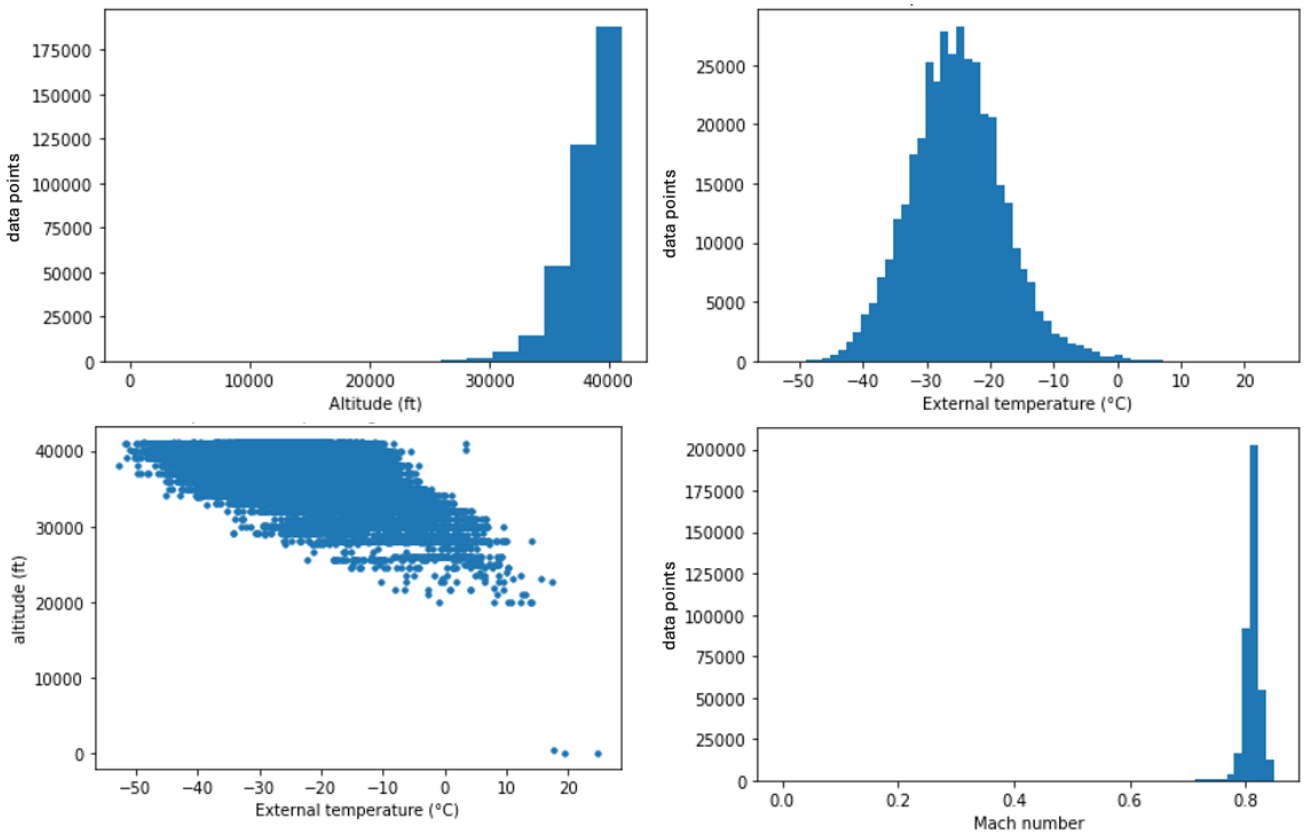

Figure 2 – Visualisation of different dataset parameters

When it comes to flight level, there is a complete coverage of the area between 30,000 ft and 40,000 ft. Nonetheless, as expected there is a lower frequency of flights occurring at lower altitudes, since they are unusual for cruise. In this case, the different altitudes are represented in a realistic way. Next, the external temperature distribution was studied. For a better visualisation of the spectrum coverage of this parameter, a histogram has been plotted. There is a complete coverage of the external temperature of the spectrum, that is, between –50°C and 0°C. It is also important to underline that the external temperature distribution is the same for each engine. Thus, it can be said that the data completeness for external temperature is good. The external temperature distribution seemingly adheres to a normal distribution pattern, centred within the range of -50°C to 10°C. Correlating the ambient temperature with altitude, we clearly see how the two correlate, despite the low frequency of cruise points at low altitudes. Finally, the Mach number distribution was visualized, and it is indicative of the typical Mach numbers during cruise.

In aeronautics, the flight envelope defines the operational limits for an aircraft with respect to maximum speed and load factor given a particular atmospheric density, given a particular altitude [8]. The flight envelope is the region within which an aircraft has been tested and can operate safely. To ensure the data completeness, the flight envelope has been plotted (figure 3), to see if the measurements covered all the relevant spectrum. In this plot, *Mmo* and *Vmo* represent the maximum operating limit speeds for the turbine that may not be deliberately exceeded in any regime of flight. *Vs* is the stalling speed and *CAS* is the calibrated airspeed, the indicated airspeed corrected for instrument and position error.

With regards to data sufficiency, a model designer needs to assess whether the available data are enough to train a model with the desired accuracy. Insufficient data might still be complete and

representative in relation to the ODD, but the reduced frequencies can affect the desired predictive accuracy, especially when a high granularity is expected.

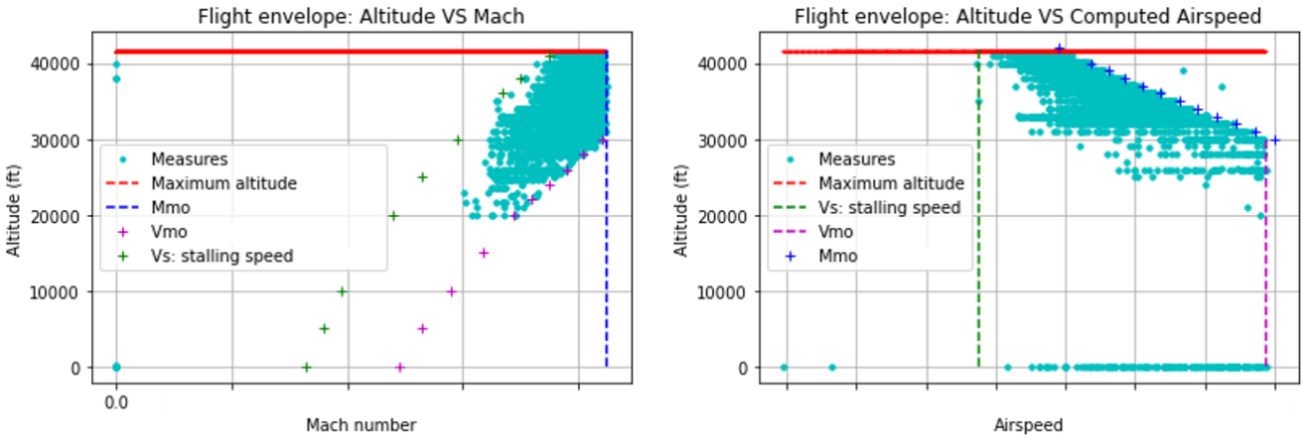

Figure 3 – Dataset visualisation with regards to the Flight Envelope

3. Missing Data

3.1 Categories of Missing Data

A major challenge in real-world datasets, such as the one utilised in this case, is the treatment of missing data points. Missing points are common, they introduce bias, and they deteriorate the quality of both the training data and the prediction.

It is imperative to consider how the missing values are distributed. There are four categories of a missing data occurrence [9]. The first category is known as missing completely at random. This is a case when the probability of a missing variable is the same for all observations, the missingness of data is unrelated to any study variable. Essentially, this implies that the causes behind the missing data have no connection to the data itself. Many complexities arise because missing data can be consequently ignored, apart from the obvious loss of information. Second, there is the case of data missing at random. In this case, the missing ratio varies for the other variables. So, the likelihood of a data point being missing is not directly tied to the nature of the missing data, but rather connected to some of the observed data. Third, there is the not missing at random case, which can be subdivided into two subcases. The first subcase involves missing data that depends on unobserved predictors. In this situation, the missing values are not randomly distributed, they are related to an unobserved variable. The second subcase is the missing that depend on the missing values themselves. This implies that the probability of a missing value is directly correlated with the missing values itself.

In this work we chose to use the Exhaust Gas Temperature (EGT) Margin of the engines as the parameter to be predicted. The EGT is the single most important health indicator of gas turbines, being a metric for the cumulative deterioration of all gas path engine assemblies [5]. Previous work [10,11] has shown that the EGT timeseries can be predicted with the use of Machine Learning models making use of certain features that resemble the actual instrumentation in aero engines. Considering that we are dealing with a time series context, it is crucial to examine the distribution of missing data depending on the time. Indeed, a lack of a lot of value in the same period might lead to an impossibility of studying this specific period as there is no information about what happened.

3.2 Prediction of Dataset General Trend

In the examined case, we wish to determine the general behaviour of the missing parameters of the dataset. Consequently, our aim is to identify a general pattern, which could manifest as either a cyclic behaviour or a trend. It is essential to remember that our focus is not on individual data points, but rather on groups of points that collectively exhibit these patterns.

The method to be used depends on the occurrence of missing data as well as the distribution of missing data depending over time, since this is a case with time series data. As a result, if the missing data points are randomly distributed over time, then the missing values could be predicted. However,

if the missing values of parameters are not randomly distributed over time, an alternative approach may be necessary to estimate those missing values.

In this use case, it appears that the parameters are missing at random over time indeed. To illustrate an example of this behaviour, we consider the missing data for specific parameters of Engine A. As demonstrated in previous work [5] we choose to use specific parameters for the prediction of EGT, as these parameters are common among real-world engine datasets. These parameters are:

1. High-Pressure compressor (HPC) intel total temperature
2. Total Temperature at HPC outlet
3. Static Pressure at HPC outlet
4. Engine core speed and
5. Fuel flow

Table 1 – Example of missing values for Engine A

Datetime	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	EGT margin
Missing Data point 3778	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 5173	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 5833	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 5992	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 6237	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 6313	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 6894	NaN	NaN	NaN	NaN	NaN	To be predicted
Missing Data point 7011	NaN	NaN	NaN	NaN	NaN	To be predicted

As observed, the missing values for these parameters are randomly distributed over time. Nevertheless, except for a single missing value for the fuel flow, there are data points which miss values for all five parameters. Given the randomness of the effect and the few occurrences, these data points can be omitted with no significant effects in the prediction of the EGT Margin. However, when the missing data are numerous, but still randomly distributed over time, estimate values might be needed as substitutes. Given our objective of predicting the general behaviour of the engine, employing K-Nearest Neighbour Imputation (KNN-Imputation) to estimate the missing values might stand as a suitable solution [12].

The KNN-Imputation is one of the proposed methods for estimation of the missing data. This method suggests the replacement of any missing values with the mean value of a number K of nearest neighbouring parameters. The two main advantages of this method are the easy implementation, and the presence of the correlation notion of the data. The main disadvantage of this method is the lengthy times of implementation, as well as the right selection of parameter K. Opting for higher K-values incorporates attributes that may differ from the real value, while lower K-values risk missing out on crucial attributes. Therefore, given the temporal nature of the dataset, it is important to estimate the values of missing data during the period of interest. This involves identifying an appropriate time window, which depends on the sampling frequency, the aircraft utilisation and the diversity of conditions.

To prove the efficiency of the KNN-Imputation, we chose to explore the behaviour of the EGT Margin over time. Our initial exploration involved the visualisation of the EGT Margin for Engine A, illustrated in figure 4.

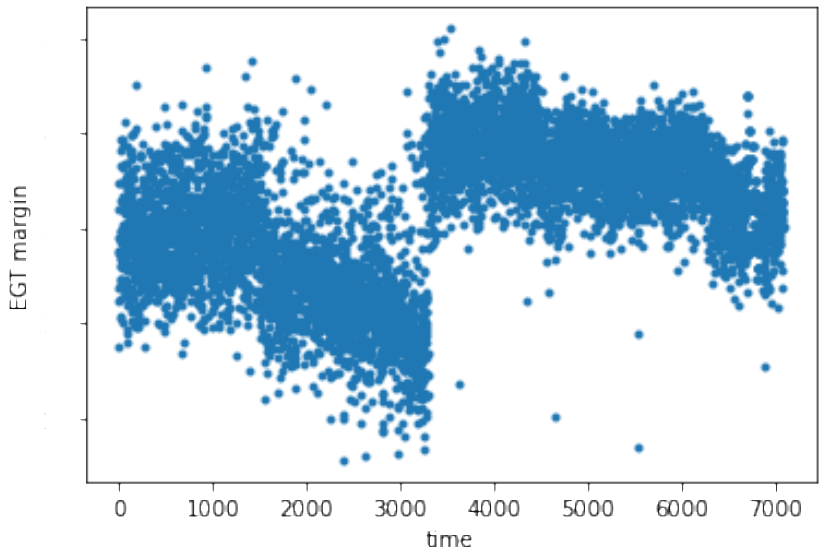

Figure 4 – Temporal evolution of EGT Margin for Engine A

The scatter plot reveals two distinct time periods. In the first one, the EGT Margin experiences a continuous decrease until after the 3,000th data point, at which there is an abrupt discontinuity, causing the EGT Margin to suddenly rise. In the second one, the EGT Margin decreases again. This increase in the EGT Margin can be attributed to an engine washing procedure.

Naturally, the normal behaviour of the EGT Margin is to decrease over time as an engine degrades and his general behaviour could be predicted by linear or polynomial regression. In order to validate the effectiveness of KNN-Imputation, the prediction of the general behaviour of the EGT Margin was carried out under three different scenarios: First, without any missing values; second, with randomly distributed missing values; and third, with estimated values obtained through KNN-Imputation for the missing data.

To accomplish this, a period with no missing values was selected. This choice enables the comparison of the estimated values with the actual values. Initially, random missing values were introduced to replace 50% of the data within the chosen period. Following, the missing values were predicted, and the general behaviour of each situation could be compared, as illustrated in figure 5.

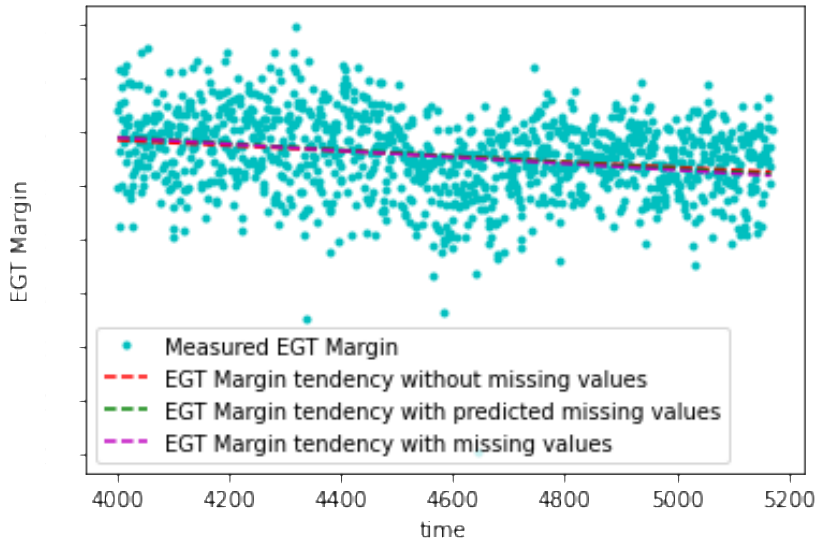

Figure 5 – EGT Margin trends for different substitution methods

To understand better the error between the true trend and the estimate ones, the relative errors between a. the predicted trend with the true value and b. the predicted trend with the estimate missing data were calculated and visualized. Same for a. the relative error between the predicted trend with the true value and b. the predicted trend with the missing data. This approach allows for quantification and visualisation of the discrepancies between these scenarios.

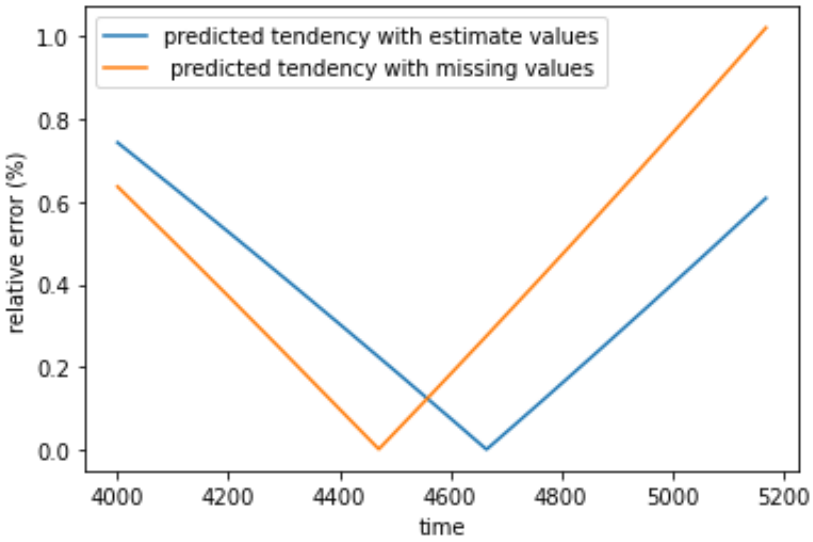

Figure 6 – Relative error for estimated and missing values

It can be observed that, whether using the KNN-Imputation or the dataset with missing values, the general quality prediction does not improve significantly. However, there are other cases where the use of KNN-Imputation can be challenging. For instance, if the missing values do not occur randomly over time and a time-dependant parameter is to be predicted, KNN-Imputation may struggle to predict the missing values, as the nearest neighbour could be located far away.

It is important to underline that a more sophisticated data-driven predictive model may not be a suitable solution in this case either. Indeed, in some cases the estimation of a parameter can be challenging when depending on other parameters. Therefore, these predicted values, which have a low accuracy, could have a negative impact on the pattern prediction.

For example, let us consider a scenario where the 70% of a dataset comprises of missing values, and we wish to compare a. the general trend of the EGT Margin predicted on true values (without missing data), b. on a dataset using KNN-Imputation, and c. on a dataset using a data-driven predictive model. In this case, a Generalized Additive Model (GAM) [13] was used to predict the EGT Margin based of Engine B on the five parameters mentioned in section 3.2. Figure 7 illustrates the results of the general prediction of the EGT Margin.

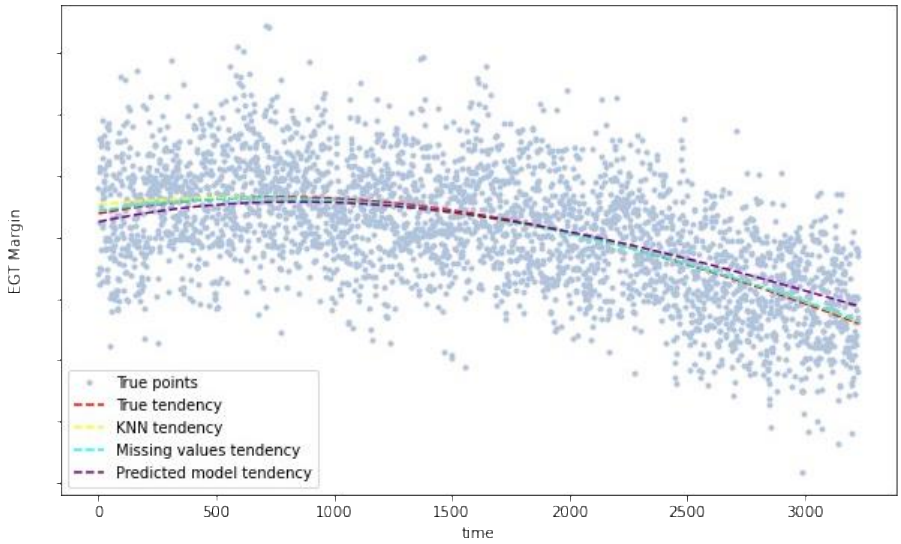

Figure 7 – EGT Margin tendency of Engine B for different missing data treatments

It can be concluded that using KNN-Imputation to predict missing values or not predicting missing values at all are both good solutions for predicting the general pattern, as the curves are nearly superimposed. However, there is a bigger error between the true general pattern and the pattern predicted on the dataset that contains predicted values generated by the GAM predictive model.

Thus, KNN-Imputation is a viable solution, but its effectiveness is comparable to not predicting the missing values at all. Moreover, KNN-Imputation should not be used if the missing values are concentrated on a specific period. The GAM predictive model is not a recommended solution in this case, as there is a risk that if the prediction accuracy.

3.3 Prediction of Missing Points

The processing of data when our goal is to find specific missing points requires a different approach. The initial step entails the visualisation of the data and subsequently the identification of any erroneous values, with no physical meaning. The importance of this step is significant, as it allows the proper curation of the dataset. Only values which have no physical meaning are affected here. In other words, it is imperative to exclusively remove outliers that are believed to be not real outliers.

Once again, the preferable scenario is when the presence of missing values is limited, and there is no need to predict the missing ones. However, if there is a high number of missing values, their distribution must be checked. In such instances, relying on KNN-Imputation might not be an optimal solution as employing methods based on mean or median values could lead to dataset flattening, so the true outliers may not be distinctly identifiable as intended.

Therefore, employing a data-driven predictive model is more suitable in this context. As a first step, the dataset is split into two subsets: one without missing values, and one with missing values. The first dataset is used for training, while the second dataset for test. Next, a model is created to predict the missing, target variable based on the other attributes present in the training dataset, and these predictions are used to populate the missing values in the test dataset. A detail of key importance here is that the missing data should be missing at random, since some correlated parameters must be present.

To illustrate the efficiency of this method, a use case was prepared: the HPC Inlet Total Temperature (T25) based was predicted, based on the ambient temperature and fan speed. An engine with no missing data was selected, so data could be selected and removed randomly, resulting in the training and the test subsets of the original dataset. In this example, there is 20% of simulated missing data, in a dataset of total 2,909 data points.

To predict the missing values, the Generalized Additive Model (GAM) was used once again. The choices of the hyperparameter values are summarised in the table below.

Table 2 – Hyperparameter values for the GAM prediction

Hyperparameter	Value
n_splines	10
λi	0.6
Training subset	70%
Test subset	30%

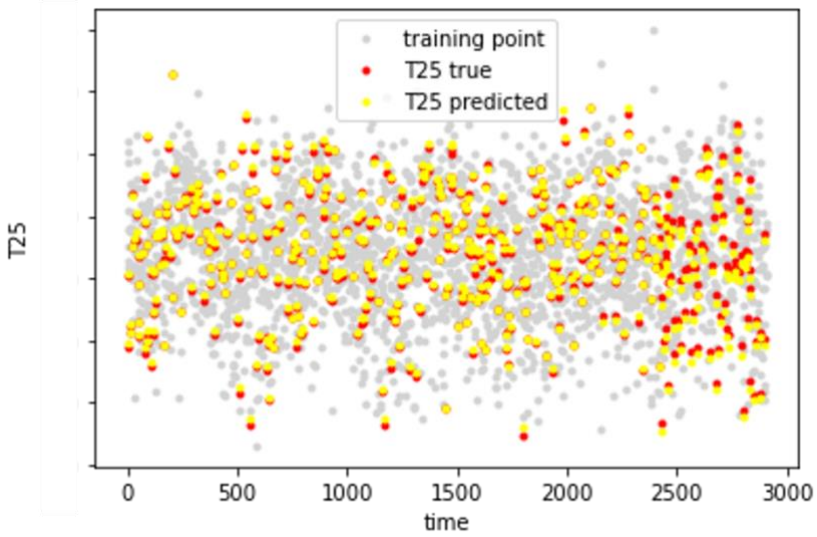

Figure 8 – GAM prediction of T25

Table 3 – Quality of predicted T25 values

RMSE Training	RMSE Test	R2 Training	R2 Test
0.6988	0.7076	0.9949	0.9949

As illustrated in figure 8, the model performs well in predicting the missing values, supported by an R^2 score close to one and an RMSE close to zero, indicated in table 3. However, as T25 was originally used to predict the EGT Margin, the next exploration was a repetition of the original one. This time the original T25 values were replaced by the predicted ones, so a comparison between the case with the missing T25 values and the one with the GAM-predicted T25 values can be performed.

Table 4 – EGT prediction for four different scenarios

	RMSE Training	RMSE Test	R2 Training	R2 Test
True Values	2.5608	3.1938	0.9922	0.9905
GAM-predicted values	2.5592	3.1231	0.9922	0.9910
Missing values	2.5814	3.2200	0.9920	0.9894
KNN-predicted values	4.3464	5.2545	0.9776	0.9742

As indicated in table 4, The disparity in RMSE and R^2 values among the models trained on the dataset containing only actual data, the dataset with estimated data, and the dataset with missing values is minimal. Nonetheless, the models trained on true data and the GAM-predicted values marginally outperform the other one. This implies that, first, the predicted values align closely the real ones, and second, employing a predictive model to estimate missing values proves to be a superior approach compared to not estimating the missing values at all. To check if the predictive model is a better solution than the KNN-Imputation in this situation, the EGT predictions were carried out under the same conditions as mentioned earlier, but instead of having a complete dataset with estimated HPC Inlet Total Temperature thanks to the predictive model, we will have a complete dataset thanks to the KNN-Imputation. The prediction of the Exhaust Gas Temperature is less precise with the KNN-Imputation because this model fit less well and create a bigger total error. As

a result, it is evident that utilizing KNN-Imputation for predicting precise points is less appropriate compared to a predictive model.

4. Conclusions

There are some very important concepts when our objective is to create a data-based predictive model in aviation. The first one is data quality. Since the accuracy of predictions relies on the quality of the data they are based on, it is essential that the data are, first, representative of reality, and second, that the preprocessing applied to it does not influence the predictions. In our study, the dataset is representative of real-life conditions for a commercial aircraft because it covers all possible conditions in which the aircraft operated. If the dataset did not cover all these possibilities, the model based on this data would be limited to specific environments and conditions, potentially excluding some flights, thus making the model less useful or prone to errors.

The preprocessing of data is a crucial aspect of building a data-based model. The first step is to clean the data by identifying the non-natural outliers. Then, working with the data can be challenging due to missing data points, requiring decisions on how to manage this problem by introducing the lowest possible bias. To make these decisions, it is necessary to first determine the context and purpose of the study. If the objective is to identify general trends, the KNN-Imputation could be a suitable solution. However, this method does not always outperform a case where missing values are not replaced. Therefore, KNN-Imputation might not be as effective as expected and may introduce unnecessary bias sometimes. On the other hand, if the goal is to predict specific data points, a data-driven predictive model such as GAM is a suitable solution, since it considers the context in which the missing data points were measured. Nevertheless, this method can become challenging to use when dealing with numerous missing data points.

Here, we compare the accuracy of the predictions when a. we keep the missing values, or b. when using a KNN-imputation model to predict the missing values, or c. when using a GAM to predict the same missing values. As a result, it is imperative to thoroughly discuss the diverse situations one may encounter and determine the optimal solutions for each scenario. To facilitate this process, a use case has been crafted to aid in making well-informed decisions. Given the inevitability of introducing bias into the dataset regardless of the chosen corrective approach, the selection of the most suitable solution becomes crucial based on the intended use of the data. Opting to retain missing data may lead to an information deficiency, potentially resulting in inaccurate predictions. Conversely, utilising a predictive model to estimate missing data could result in information loss, as the model might homogenize the dataset, causing technical issues to go unnoticed. Hence, it is vital to distinguish between two application cases for the dataset. In the initial scenario, the data is utilized for forecasting broad trends, such as the overall degradation of an engine. Here, the main objective is to discern patterns within the dataset. The second scenario emerges when the goal is to unearth or anticipate particular data points that signal rare events, where components may malfunction with minimal advance notice.

5. Discussion and Future Work

In summary, the preprocessing of data is a crucial aspect of building a data-based model. The first step is to clean the data by identifying the non-natural outliers. Then, working with the data can be challenging due to missing data points, requiring decisions on how to manage this problem by introducing the lowest possible bias. To make these decisions, it is necessary to determine the context and purpose of the study first. If our objective is to find the general behaviour of EGT evolution, the KNN-imputation could be a suitable solution. However, this method does not outperform the studying of the data pattern if the missing values are still missing. Therefore, KNN-imputation might not be as effective as expected and may introduce unnecessary bias. On the other hand, if our goal is to predict specific data points, a GAM predictive model is a suitable solution since it considers the context in which the missing data points were measured. Nevertheless, this method can become challenging to use when we deal with numerous missing data points.

As future work, it would be interesting to examine different data-driven models for the same applications. In addition, the results of this work could be tested for other parameters and applications

within the aeronautical domain. Last, as the available dataset contained low-frequency time series data, the investigation of datasets with various sampling rates could provide additional insights, especially as newer aircraft generations produce data containing more parameters, at very high frequencies.

6. Contact Author Email Address

mailto: a.apostolidis@hva.nl

7. Copyright Statement

References

[1] International Air Transport Association (IATA). *Aircraft Operational Availability*, 2nd edition, IATA, 2022.

[2] Stamoulis KP. *Innovations in the Aviation MRO: Adaptive, Digital, and Sustainable Tools for Smarter Engineering and Maintenance*. 1st edition, Eburon Academic Publishers, 2022.

[3] Stamoulis KP. *Advanced Data Analytics and Digital Technologies for Smart and Sustainable Maintenance*. Proceedings of ISATECH 2021. Sustainable Aviation. 1st edition, Springer, 2024.

[4] SAE International. *Artificial Intelligence in Aeronautical Systems: Statement of Concerns*. AIR6988. 1st edition, SAE International, 2021.

[5] Apostolidis A, Bouriquet N and Stamoulis KP. AI-Based Exhaust Gas Temperature Prediction for Trustworthy Safety-Critical Applications. *Aerospace*, Vol. 9, 722, 2022.

[6] Kaakai F, Adibhatla S, Pai G and Escorihuela E. Data-centric Operational Design Domain Characterization for Machine Learning-based Aeronautical Products, *Proceedings of SAFECOMP 2023*, Toulouse, France.

[7] Apostolidis A. *Turbine cooling and heat transfer modelling for gas turbine performance simulation*, PhD Thesis, Cranfield University, 2015.

[8] Gratton G. *Initial Airworthiness: Determining the Acceptability of New Airborne Systems*. 1st edition, Springer, 2015.

[9] Kang H. The prevention and handling of the missing data. Korean Journal of Anaesthesiology. Vol. 64, No. 5, pp 402-406, 2013.

[10] Kefalas M, de Santiago Rojo J Jr, Apostolidis A, van den Herik D, van Stein B and Bäck T. Explainable Artificial Intelligence for Exhaust Gas Temperature of Turbofan Engines. *Journal of Aerospace Information Systems*, Vol. 19, pp 447–454, 2022.

[11] Protopapadakis G, Apostolidis A and Kalfas AI. Explainable and Interpretable AI-Assisted Remaining Useful Life Estimation for Aeroengines. *Proceedings of the ASME Turbo Expo 2022*, GT2022-80777, V002T05A002, Rotterdam, The Netherlands.

[12] Taunk K, De S, Verma S and Swetapadma A, A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. *Proceedings of 2019 International Conference on Intelligent Computing and Control* Systems (ICCS), Madurai, India, pp. 1255-1260, 2019.

[13] Hastie T and Tibshirani R. Generalized Additive Models. *Statistical Science*, Vol. 1, No. 3, pp 297 – 31