

# DATA-DRIVEN APPROACH TO OPTIMAL AIRCRAFT MAINTENANCE

Onyedikachi Chioma Okoro<sup>1</sup>, Maksym Zaliskyi<sup>2</sup>, Serhii Dmytriiev<sup>1</sup>, Qudus Sadiq<sup>3</sup>

<sup>1</sup>National Aviation University, Department of Continuing Airworthiness, Kyiv, Ukraine
<sup>2</sup>National Aviation University, Department of Telecommunication and Radioelectronic Systems, Kyiv, Ukraine
<sup>3</sup>Tamara Niger Aviation, Quality & Safety Department, Niamey, Niger

### Abstract

An aircraft typically comprises millions of parts collected globally and assembled in a highly complex process. Its' life cycle consists of four phases, the longest being operations. This phase generates most of the statistical data in the aircraft life cycle, which can be used to develop statistical data processing algorithms to improve aircraft operations' efficiency. Aimed at efficient and cost-effective aircraft maintenance, this study presents analytical, numerical, and simulation methods. The models are based on probability theory and statistics, reliability theory, and segmented regression. In addition, the models were tested for goodness-of-fit to prove their accuracy. Input data was extracted from maintenance and pilot reports of a fleet of aircraft. An advantage of the proposed model is its simplicity, making it easy to use by personnel for planning aircraft maintenance activities. Furthermore, based on information provided by the models, aircraft operators can adjust their aircraft maintenance programs to reflect their actual requirements. The models can also be used for improving the design of new and existing aircraft systems, subsystems, and components

Keywords: aircraft maintenance, aircraft systems, operations, data processing, regression analysis

## 1. Introduction

An aircraft typically comprises millions of parts collected globally and assembled in a highly complex manufacturing process, resulting in vast multi-modal data. Its operational phase generates a wealth of real-time data, which is collected, transferred, and processed with 70 miles of wire and 18 million lines of code for the avionics and flight control systems alone. These real-time operational data can be transformed into statistical data for developing models and algorithms for data-driven predictive aircraft maintenance approaches, which result in reduced maintenance costs, avoiding unnecessary preventive maintenance actions, and reducing unexpected failures. Operational data such as past aircraft faults/failures and maintenance actions can be used to estimate the probability of aircraft component failure and plan maintenance actions accordingly.

The proposed data-driven methodology for optimal aircraft maintenance is carried out in two steps: a) reliability analysis of aircraft components, subsystems, and systems to determine reliability parameters and b) regression models for predicting the time moment of the occurrence of a fault/failure. The models and algorithms in this paper are based on probability theory and statistics, reliability theory, and regression models. The methodology described is applied to real-life aircraft operational data to validate the proposed models and prove their applicability. Historical datasets of pilot and maintenance records of faults/failures from aircraft operating in Nigeria are utilized. The results of the analysis described in this paper can provide insights into future faults/failures of aircraft components, sub-systems, and systems – it can supplement an existing aircraft maintenance strategy. This results in reduced waste which arises due to early maintenance and failure costs connected with late maintenance actions [1].

# 2. Methodology

The proposed methodology for data-driven optimal aircraft maintenance is in two steps: 1) Reliability analysis [2] and 2) Regression analysis.

## 2.1 Statistical Simulation Model for Reliability Analysis

Maintenance costs, to a large extent, depend on the reliability of aircraft components, subsystems, and systems. A failure probability generally measures reliability, and optimization ensures that the latter remains lower than the given threshold [3]. Over the last two decades, reliability analysis methods have been developed – these have stimulated interest for the probabilistic treatment of structures [4]. Reliability analysis involves the evaluation of the level of safety of a system. Given a probabilistic model (an n-dimensional random vector **X** with probability density function  $f_X$ ) and a performance model (a function **g**), it uses mathematical techniques to estimate the system safety level in the form of a failure probability [5]. Failure is generally defined as an event **F**= {**g**(**X**)≤**0**} and the probability of failure is defined as:

$$p_f \equiv \mathbb{P}(\{g(\mathbf{X}) \le 0\}) = \int_{D_{f=\{\mathbf{X} \in \mathbb{R}^n: g(\mathbf{X}) \le 0\}}} f_{\mathbf{X}}(\mathbf{X}) \, \mathrm{d}\mathbf{X}$$

Many mathematical definitions and probability distributions are used to perform different types of maintenance and reliability studies. However, the exponential distribution is the most used probability distribution because it is easily applied in various types of analysis of failure rates of components, subsystems, and systems during their useful life [6]. The exponential distribution probability density function is defined by

$$f(t) = \lambda e^{-\lambda t} \text{ for } t \ge 0 \ \lambda > 0 \tag{1}$$

where *t* is time, f(t) is the probability density function (PDF) and  $\lambda$  is the distribution parameter, which in reliability studies refers to the constant failure rate [6]. Based on the exponential distribution law of mean time to failure, a statistical simulation model for reliability analysis of aircraft systems and structures in view of effective maintenance is proposed in [2], and the algorithm is shown in Figure 1.

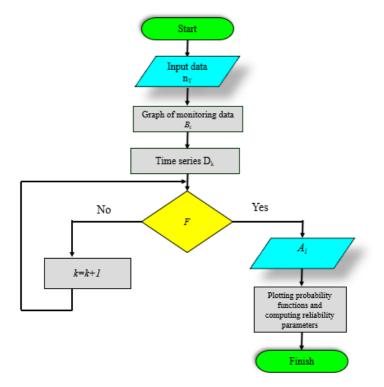


Figure 1 – Flowchart for reliability analysis based on exponential distribution.

For the proposed model, the aircraft systems, structures, and powerplant were categorized per the ATA numbering system. Input data  $n_T$  is extracted from pilot and maintenance reports of a fleet of aircraft. The graph of monitoring data (Figure 2) gives a visual analysis of how the faults/failures occur. The output *F* refers to the time at which the *i*-th fault/failure occurs, but on its own, *F* cannot be used to plot the PDFs because it is a two-dimensional array; hence  $A_i$  is formulated to extract a one-dimensional array for plotting the PDFs. The following reliability indices are calculated using the resulting PDFs:

– Fault/failure rate  $\lambda$ ,

- Mean time between fault/failure (MTBF),
- Number of faults/failures per 1000 flight hours (K<sub>1000</sub>).

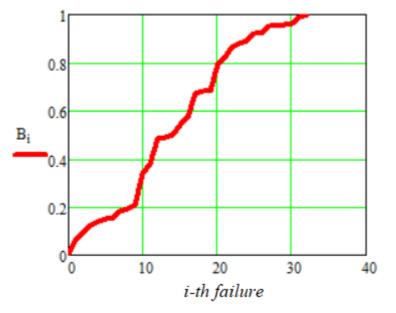


Figure 2 – Graph of monitoring data

After the reliability analysis is carried out to determine the least reliable aircraft components and systems, the next phase will be the regression analysis to predict the time at which faults/failures will occur.

# 2.2 Regression Analysis for Aircraft Maintenance

Regression analysis is a simple predictive tool that investigates the relationship between independent and dependent variables [7]. Regression models are statistical models where we make a regression assumption [8]. They can be integrated to improve the prediction accuracy of faults/failures of aircraft components, subsystems, and systems, thereby providing valuable insights for maintenance planning. For the scope of this study, regression models are employed to predict the occurrence of faults/failures during aircraft operations in view of determining optimal maintenance interval for aircraft systems. Regression analysis can also be viewed as a set of data analytic techniques that help understand the interrelationships among variables. The relationship is expressed in the form of a model or an equation that connects the dependent or response variable and one or more explanatory or predictor variables [9]. For this study, there are no predetermined coefficients for the regression analysis because calculations are based on aircraft operational data. The nomenclature for the parameter used in the regression analysis are given in Figure 3.

Nomenclature for the parameters and variables used in the regression analysis

- a: matrix of n
- n: sample size
- m: switching point
- $\phi$ : Heaviside step function which is equal to 0 before the switching point and 1 after the switching point
- T<sub>i</sub>: time moment of malfunction/failure
- Y: predicted value i.e. optimal maintenance flight hour
- X: *ith* number of failures, i

After the parameters are defined, the next step is determining which regression model is optimal for predicting the time moment of the next failure – for this purpose, three segmented regression models are tested. Segmented regression models are models where two or more lines are joined at unknown points called the switching points representing the threshold [10]. It partitions the data into different regions, and a regression function is fitted to each one [9]. Segmented regression is an alternative variant of approximating empirical curves. Its use in aircraft operation allows for increased correctness for calculating extreme probability values of occurrence of faults/failures in aircraft components, subsystems, and systems. The segmented regression models considered are:

- Quadratic-linear segmented regression model,
- Linear-linear segmented regression model,
- Quadratic-quadratic segmented regression model.

The matrix of unknown coefficients in the segmented regression models is estimated using the ordinary least squares method. The vertical distances represent errors in the response, and the ordinary least squares method gives the line that minimizes the sum of squares of vertical distances from each point to the regression model function. These errors can be obtained by writing equation (2) for example, for linear regression model as

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i,$$
 *i*=1, 2, ...n (2)

The sum of squares of these distances can be written as

$$S(\beta_{0}, \beta_{1}) = \sum_{i=1}^{n} \epsilon_{i}^{2} = \sum_{i=1}^{n} (\gamma_{i} - \beta_{0} - \beta_{1} x_{i})^{2}$$

The values of  $\widehat{\beta_0}$  and  $\widehat{\beta_1}$  that minimize  $S(\beta_0, \beta_1)$  are given by

$$\widehat{\beta_1} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$
$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1} \bar{x}$$

and

The estimates  $\hat{\beta}_0$  and  $\hat{\beta}_1$  are called the ordinary least squares estimates of  $\beta_0$  and  $\beta_1$  because they are the solution to the ordinary least squares method [9].

### 2.2.1 Quadratic-linear segmented regression model

The quadratic-linear segmented regression model for predicting the time moment of the subsequent fault/failure of the aircraft system has the following form

$$Y_1(X) = \beta_{0,1} + \beta_{1,1}X + \beta_{2,1}X^2 - \beta_{2,1}(X - m)^2\phi(X - m)$$
(3)

This model uses two segments joined together at the switching point *m* and three unknown coefficients  $\beta_{0,1}$ ,  $\beta_{1,1}$  and  $\beta_{2,1}$ . Using the ordinary least square method unknown coefficients are calculated as follows:

$$\begin{bmatrix} \beta_{0,1} \\ \beta_{1,1} \\ \beta_{2,1} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 - \sum_{i=m}^{n} (i-m)^2 \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 - \sum_{i=m}^{n} [(i-m)^2i] \\ \sum_{i=1}^{n} i^2 - \sum_{i=m}^{n} (i-m)^2 & \sum_{i=1}^{n} i^3 - \sum_{i=m}^{n} [(i-m)^2i] + \sum_{i=m}^{n} (i-m)^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T$$

#### 2.2.2 Linear-linear segmented regression model

The functional dependence (4) for the linear-linear segmented regression model also uses two

segments joined together at the switching point *m*.

$$Y_2(X) = \beta_{0,2} + \beta_{1,2}(X) + \beta_{2,2}(X - m)\phi(X - m)$$
(4)

The unknown coefficients  $\beta_{0,2}$ ,  $\beta_{1,2}$  and  $\beta_{2,2}$ . are calculated as follows

$$\begin{bmatrix} \beta_{0,2} \\ \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=m}^{n} (i-m) \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=m}^{n} [i(i-m)] \\ \sum_{i=m}^{n} (i-m) & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} (i-m)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} (T_i) \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T_i \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \\ \sum_{i=m}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} [T_i(i-m)] \end{bmatrix}^{-1} \end{bmatrix}^{-1$$

### 2.2.3 Quadratic-quadratic segmented regression model

The quadratic-quadratic segmented regression model for predicting the time moment of the next failure of the aircraft system has the following form

$$Y_{3}(X) = \beta_{0,3} + \beta_{1,3}X + \beta_{2,3}X^{2} + \beta_{3,3}(X-m)\phi(X-m) + \beta_{4,3}(X-m)^{2}\phi(X-m)$$
The five unknown coefficients  $\beta_{0,3}$ ,  $\beta_{1,3}$ ,  $\beta_{2,3}$ ,  $\beta_{3,3}$  and  $\beta_{4,3}$ . are calculated as follows
(5)

$$\begin{bmatrix} \beta_{0,3} \\ \beta_{1,3} \\ \beta_{2,3} \\ \beta_{3,3} \\ \beta_{4,3} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 & \sum_{i=m}^{n} (i-m) & \sum_{i=m}^{n} (i-m)^2 \\ \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 & \sum_{i=1}^{n} i^4 & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} [i^2(i-m)^2] \\ \sum_{i=1}^{n} (i-m) & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} [i^2(i-m)] & \sum_{i=m}^{n} (i-m)^2 & \sum_{i=m}^{n} (i-m)^3 \\ \sum_{i=m}^{n} (i-m)^2 & \sum_{i=m}^{n} [i(i-m)^2] & \sum_{i=m}^{n} [i^2(i-m)^2] & \sum_{i=m}^{n} (i-m)^3 & \sum_{i=m}^{n} (i-m)^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} (T_i i) \\ \sum_{i=1}^{n} (T_i i) \\ \sum_{i=1}^{n} (i-m) T_i \end{bmatrix}$$

In step three, the value of the optimal switching point *m* is selected for each model based on the corresponding least value of standard deviation  $\sigma$ .

$$\sigma = \sqrt{\frac{1}{n-I} \sum_{i=1}^{n} (T_i - \widehat{Y})^2}$$

where *I* is the degree of freedom for each of each selected model,  $\hat{Y}$  corresponds to the response variable for each of the segmented regression models i.e.,  $Y_1(X)$ ,  $Y_2(X)$  and  $Y_3(X)$ .

From the observations of  $\mathbf{m}$  and  $\boldsymbol{\sigma}$ , the optimal values of the switching point for the three segmented regression models are calculated. The segmented regression model with the least value of  $\mathbf{m}$  is considered the optimal model for predicting the flight hour at which a fault/failure is likely to occur in the observed system.

#### 3. Analysis and Results

To test the models, input data for the simulation was extracted from pilot and maintenance reports of a fleet of aircraft in Nigeria. A basic sample of statistical data was generated from a fleet of aircraft over an operational period of four years [11]. As shown in Table 1, the fault/failure information of each system, structure, and powerplant was grouped according to the ATA numbering system. In the table,  $n_T$  refers to the total number of faults/failures observed by pilots and maintenance personnel for the time interval.

ATA №	n⊤	ATA №	nT	ATA №	n⊤	ATA №	nT
21	11	31	31	53	165	73	48
22	104	32	211	55	13	74	1
23	39	33	76	56	4	75	54
24	57	34	173	65	192	76	5
25	27	39	9	66	37	77	8
26	15	45	17	67	76	78	4
28	9	51	70	71	24	79	48
29	46	52	53	72	20	80	15

Table 1 – Faults/failures information of aircraft systems and structures

The simulation for the reliability analysis is performed using the algorithm in Figure 1 for 10000 iterations, and the probability density functions (PDFs) are plotted based on the output. A sample PDF (ATA 34) is shown in Figure 4 – The blue line proves that the simulation results coincide with the theoretical distribution. For this study, the reliability parameters of the top-most failing ATA chapters are analyzed i.e., ATAs chapter 22, 24, 32, 33, 34, 53, 65, 67 and 75

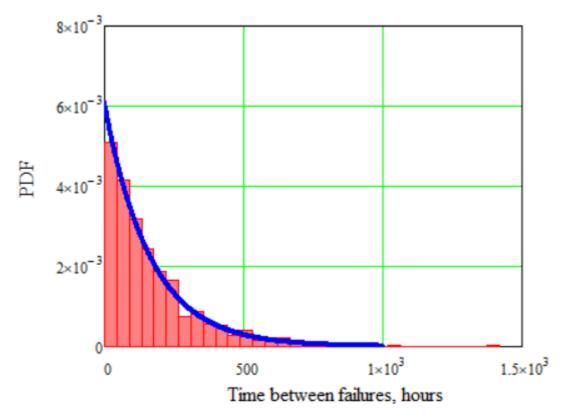


Figure 4 – Probability density function for observed time between failures of the navigation system (ATA 34)

ATA №	ATA Chapter Name		$\lambda_{calc}$	K <sub>1000 calc</sub>	
22	Auto flight	285.20	0.004	4	
24	Electrical power	580.45	0.002	2	
32	Landing gear	136.75	0.007	7	
33	Lights	361.91	0.003	3	
34	Navigation	164.52	0.006	6	
53	Fuselage	173.14	0.006	6	
65	Tail rotor drives	151.76	0.007	7	
67	Rotors flight control	409.80	0.002	2	
75	Engine air	506.60	0.002	2	

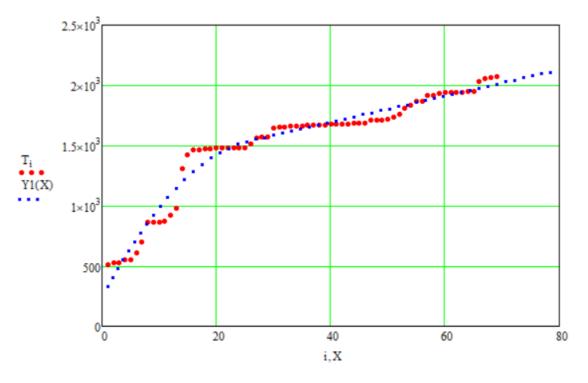
Table 2 – Reliability indices based on the PDFs

To predict the time moment at which an aircraft fault/failure will occur, three segmented models were previously developed in this study. To determine which of the segmented regression models gives the most prediction accuracy, all models will be tested using real-life aircraft operational data [11]. The selected aircraft system (ATA 34) is further transformed for the analysis (Table 3.)

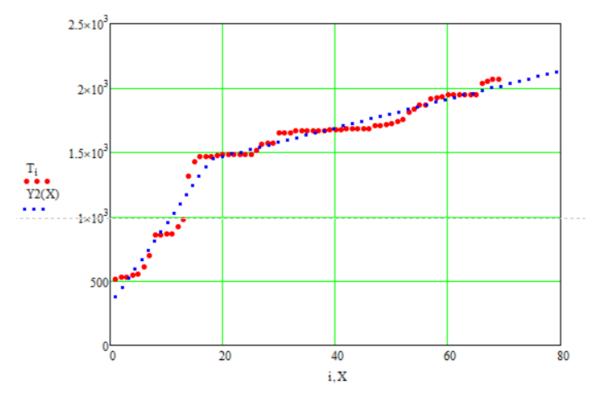
Table 3. Statistical data generated from aircraft operations

	Time between		Time		Time between		Time between
Failure <i>i</i>	failures	Failure <i>i</i>	between failures	Failure <i>i</i>	failures	Failure <i>i</i>	failures
1	0	19	1.5000	37	2.5000	55	24.9501
2	510.9672	20	3.3333	38	0.1000	56	32.6334
3	17.0833	21	6.3833	39	2.5000	57	2.4500
4	0.0833	22	0.4000	40	2.5000	58	44.7332
5	20.2667	23	0.4000	41	4.3000	59	5.0333
6	4.336	24	0.4000	42	1.8333	60	10.3833
7	54.6334	25	0.0833	43	1.8333	61	13.3600
8	90.8332	26	0.0833	44	1.8333	62	0.1333
9	161.7500	27	33.2168	45	1.8333	63	0.1333
10	0.5000	28	48.7167	46	1.8333	64	0.6167
11	4.1667	29	5.6667	47	1.8333	65	0.3334
12	4.1667	30	2.4833	48	20.9167	66	2.6667
13	56.0999	31	78.2099	49	2.4167	67	83.8634
14	56.0999	32	4.3333	50	1.5667	68	22.4334
15	330.5002	33	1.5000	51	8.2334	69	11.3999
16	111.7334	34	8.8166	52	18.9332	70	3.0666
17	42.7768	35	0.1000	53	18.9332		
18	1.5000	36	0.1000	54	53.1834		

The transformed statistical data (Table 3) is a matrix **A** and is the input data for the simulation. Each proposed segmented model is tested using the input data, and the graphs plotted to check the fitting (Figure 5, 6, 7). Each model's matrix of unknown coefficients is calculated using the ordinary least square method.









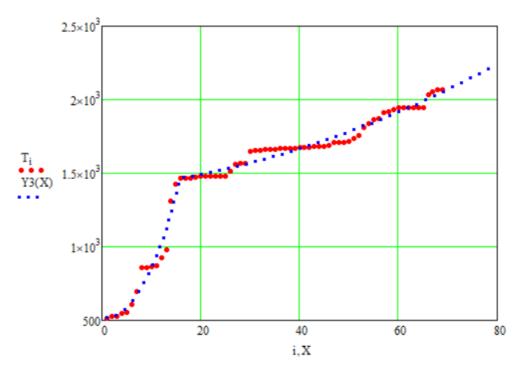


Figure 7 – Quadratic-quadratic segmented model

Designing an accurate predictive model involves fitting it to a set of training data and then adjusting its parameters such that this model will be able to make reliable predictions on new untrained data. Overfitting or underfitting is a common concern when designing a predictive model and it is possible to create a complex structure when fitting the regression model which results in poor performance [12]. Figure 8 illustrates this problem, and we compare it to Figure 5-7 to confirm that the three proposed regression models can be used to forecast time moments of faults/failures of aircraft component, subsystems, or system.

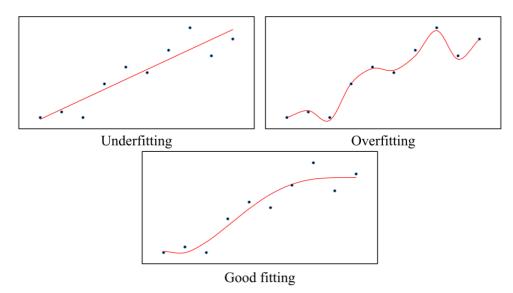


Figure 8 – Pictorial representation of over- and underfitting in regression [12]

To determine which of the three models gives the most precise prediction and at which optimal switching point *m*, an analysis of the values of standard deviation  $\sigma$  for each value of *m*=15...30 is carried out. The results are given in Table 4 – Y1, Y2 and Y3 respectively refer to quadratic-linear segmented model, linear-linear segmented model, and quadratic-quadratic segmented model.

		σ			σ		
m	Y1	Y2	Y3	m	Y1	Y2	Y3
15	115.084	74.426	51.332	23	80.537	84.681	72.787
16	108.777	67.825	47.939	24	79.201	89.517	73.757
17	102.827	64.736	50.674	25	78.425	94.120	74.199
18	97.404	64.643	55.727	26	78.127	98.415	74.243
19	92.615	66.781	60.859	27	78.232	102.525	74.104
20	88.523	70.386	65.227	28	78.683	106.613	73.949
21	85.154	74.862	68.655	29	79.434	110.618	73.780
22	82.503	79.730	71.137	30	80.435	114.485	73.599

Table 4. Values of standard deviation $\sigma$ fo	or each value of <i>m</i> =15-30
---	----------------------------------

According to Table 4, the least value of standard deviation  $\sigma$  is observed with the quadratic-quadratic segmented model when m=16. Therefore, this model is considered the most precise of the three proposed regression models for predicting the fault/failure of an aircraft component, subsystem, or system. The matrix of coefficients of the quadratic-quadratic segmented model are

523.85
- 8.307
4.201
- 119.969
- 4.107
L = 4.107

# 4. Conclusion

In this study, a two-step methodology for data-driven optimal aircraft maintenance was developed. In the first step, a reliability analysis is carried out using an algorithm based on the exponential distribution law of mean time to failure. In the second stage, the flight hour at which a fault/failure will occur is predicted using segmented regression models. The simulation results using real-life operational aircraft data show that the quadratic-quadratic segmented regression model is the most optimal for the prediction.

The findings presented in this paper can be used for determining and optimizing maintenance during the first three phases of the aircraft life cycle. It can also supplement an existing aircraft maintenance strategy and decrease waste due to early maintenance and failure costs connected with late maintenance actions. An advantage of the proposed method is its simplicity, making it easy to use by airline personnel for planning maintenance tasks.

# References

- [1] van Staden Heletjé E., Laurens Deprez, and Robert N. Boute. 2022 "A dynamic "predict, then optimize" preventive maintenance approach using operational intervention data." European Journal of Operational Research. DOI: https://doi.org/10.1016/j.ejor.2022.01.037.
- [2] O. Onyedikachi Chioma and S. Dmytriiev, "A Model for Reliability Analysis of Aircraft Systems and Structures," 2022 XIX Technical Scientific Conference on Aviation Dedicated to the Memory of N.E. Zhukovsky (TSCZh), 2022, pp. 10-14, doi: 10.1109/TSCZh55469.2022.9802497.
- [3] V. Dubourg, "Adaptive surrogate models for reliability analysis and reliability-based design optimization",. Université Blaise Pascal - Clermont-Ferrand II, 2011.
- [4] V. Papadopoulos, D. Giovanis, N. Lagaros, P. Manolis., 'Accelerated subset simulation with neural networks for reliability analysis, Computer Methods in Applied Mechanics and Engineering, 2012, pp. 70-80.
- [5] V. Dubourg, B. Sudret, F. Deheeger, 'Metamodel-based importance sampling for structural reliability analysis, Probabilistic Engineering Mechanics, 2013, pp. 47-57.
- [6] Dhillon, B. S., Maintainability, Maintenance, and Reliability for Engineers. New York, Taylor & Francis Group, 2006, 214 pp 12-17.
- [7] I. Gkioulekas, L.G. Papageorgiou, "Piecewise Regression Analysis through Information Criteria using

Mathematical Programming", Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, 2019.

- [8] M. Evans, J. Rosenthal, 'Probability and Statistics, The Science of Uncertainty Second Edition', University of Toronto, 2009
- [9] S. Chatterjee, A.S. Hadi, Regression Analysis by Example, 5th Edition, Wiley Series in Probability and Statistics, 2013.
- [10] J.D. Toms, M. L. Lesperance, 'Piecewise Regression: A tool for Identifying Ecological Thresholds', Ecological Society of America, 2003, pp. 2034–2041.
- [11] O. Okoro, "Reliability Analysis of Aircraft Fleet in Nigeria," Proceedings of National Aviation University, Vol. 83 (2), 2020, pp. 49–53.
- [12] I. Gkioulekas, L.G. Papageorgiou, "Piecewise Regression Analysis through Information Criteria using Mathematical Programming", Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, 2019.
- [13] O.C. Okoro, M. Zaliskyi, S. Dmytriiev, O. Solomentsev, O. Sribna, "Optimization of Maintenance Task Interval of Aircraft Systems", International Journal of Computer Network and Information Security (IJCNIS), Vol.14, No.2, pp.77-89, 2022. DOI: 10.5815/ijcnis.2022.02.07.
- [14] V.M. Kuzmin, M.Yu. Zaliskyi, R.S. Odarchenko, Yu.V. Petrova, New Approach to Switching Points Optimization for Segmented Regression during Mathematical Model Building, CEUR Workshop Proceedings, 2022, Vol. 3077, pp. 106-122.
- [15] O.V. Solomentsev, M.U. Zaliskyi, O.V. Zuiev, M.M. Asanov, Data Processing in Exploitation System of Unmanned Aerial Vehicles Radioelectronic Equipment, IEEE 2nd International Conference Actual Problems of Unmanned Air Vehicles Developments Proceedings (APUAVD), 2013, pp. 77–80, doi: 10.1109/APUAVD.2013.6705288.
- [16] I.G. Prokopenko, Robust Methods and Algorithms of Signal Processing, 2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS), 2017, pp. 71–74, doi: 10.1109/MRRS.2017.8075029.

#### **Copyright Statement**

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.