UDC 629.73:658.114(669.1)(045)

O.C Okoro, M. Zaliskyi, S. Dmytriiev
(National Aviation University, Ukraine)

Statistical simulation regression models for efficient aircraft operations

Regression analysis is a simple predictive tool that examines the relationship between dependent and independent variables. Regression models can be incorporated into an existing framework to improve prediction accuracy. Segmented regression models are an alternative variant of approximating empirical curves, and their use allows for increased prediction accuracy. Therefore, this study proposes its application for improving the prediction accuracy of failures of aircraft components, subsystems, systems, and structures. Three segmented regression models are developed and analysed using pre-processed daily operations aircraft data; the model with the least standard deviation value for the switching point m is considered the most precise. The proposed models can provide valuable insights for aircraft maintenance planning and serve as part of a data-driven predictive aircraft maintenance framework. Furthermore, realism is needed in how the aircraft maintenance optimization problem is formulated during the design and manufacturing phase of the aircraft lifecycle; the proposed model can provide insights.

Introduction

Traditional maintenance actions i.e., corrective, and preventive maintenance actions are no longer able to address increased complexity of systems therefore a shift towards complex maintenance approaches can be implemented to ensure quality and reliability. Unscheduled maintenance results in costly delays and inconvenience to passengers. A significant reduction in the number of unscheduled maintenance task would transform the aviation industry. This may be enabled by an intelligent aircraft system that identifies and communicates component/system anomalies and self-diagnoses faults/failures, wear, software glitches, low fluids, etc.

Recent research highlights that statistical data processing algorithms can be related to intelligence-based information technologies that use adaptivity, system and process approach, and robustness to improve efficiency. The operations phase of the aircraft lifecycle generates most of the statistical data in the aircraft life that can be used to develop statistical data processing algorithms for estimating the time of possible failure with the aim of preventing it based on correct and timely operational actions [1]. The statistical data generated by the aircraft contains various data processing algorithms which can be used for developing mathematical models, regression analysis, correlation analysis, detection and estimation algorithms, heteroscedasticity analysis and diagnostics; this paper will only be focused on regression analysis.

<u>Overview of regression analysis</u>

Regression analysis can be defined as a set of data analysis methods that provide an understanding of the interrelationships among variables. The relationship is expressed as a model or an equation that connects the response (dependent) variable and one or more predictor (independent) variables [2]. The response variable is denoted by *y* and is of particular interest. The independent, explanatory, or regressor variables are used to predict the behaviors of ***Y*** and are denoted by $X_1, X_2, \ldots, X_k$ [3]. The relationship between *y* and x_i*'s* can be expressed via a function ***f***

$$Y \approx f(X_1, X_2, \ldots, X_k)$$

The relationship between the response variable *Y* and predictor variable *X* is given as a linear model

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

where β_0 and β_1 are referred to as the model regression, unknown coefficients, and ε is a random disturbance or error.

Equation (1) gives an acceptable approximation of the true relation between *Y* and *X*, i.e., *Y* is an approximate linear function of *X*, and ε measures the difference in that approximation. According to the observed sample, equation (1) can be written as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i=\{1, 2, \ldots n\} \quad (2)$$

where y_i represents the *ith* value of the response variable Y, x_i represents the ith value of the predictor variable X, and ε_i represents the error in the approximation of y_i

Segmented regression models are models where two or more lines are connected at the switching points which represents the threshold [4]. It separates the data into different regions, and a regression function is fitted to each one [2]. Segmented regression is a variant of approximating empirical curves; its use in aircraft operations will allow for increased correctness for calculating extreme probability values of occurrence of failures. In the context of this study, the term 'failure" refers to faults and failures of aircraft components, subsystems, systems, or structures.

Methodology

Considering the increased precision accuracy that segmented regression models give, three models are developed in this study. The models developed are the quadratic-linear segmented regression model, linear-linear segmented regression model, and quadratic-quadratic segmented regression model; the functional dependence of each model is given respectively in (3), (4), and (5).

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

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

$$Y_3(X) = \beta_{0,3} + \beta_{1,3}X + \beta_{2,3}X^2 + \beta_{3,3}(X-m)\phi(X-m) + \beta_{4,3}(X-m)^2\phi(X-m) \quad (5)$$

where *Y* is the predicted value i.e., optimal maintenance flight hour, *X* is the *ith* number of failures, *m* is the switching point, ϕ is the Heaviside step function which is equal to 0 before the switching point and 1 after the switching point.

The unknown coefficients β are calculated respectively as follows using the ordinary least square method:

$$
\begin{bmatrix} \beta_{0,1} \\ \beta_{1,1} \\ \beta_{2,1} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 - \sum_{i=m}^{n} (i-m)^2 \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 - \sum_{i=m}^{n} \left[(i-m)^2 i \right] \\ \sum_{i=1}^{n} i^2 - \sum_{i=m}^{n} (i-m)^2 & \sum_{i=1}^{n} i^3 - \sum_{i=m}^{n} \left[(i-m)^2 i \right] & \sum_{i=1}^{n} i^4 - 2\sum_{i=m}^{n} \left[(i-m)^2 i^2 \right] + \sum_{i=m}^{n} (i-m)^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} T_i i \\ \sum_{i=1}^{n} \left(i^2 T_i \right) - \sum_{i=m}^{n} \left[(i-m)^2 T_i \right] \end{bmatrix} \tag{6}
$$

$$
\begin{bmatrix} \beta_{0,2} \\ \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=m}^{n} (i-m) \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=m}^{n} [i(i-m)] \\ \sum_{i=m}^{n} (i-m) & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} (i-m)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} \left(T_i i \right) \\ \sum_{i=m}^{n} \left[T_i (i-m) \right] \end{bmatrix} \tag{7}
$$

$$
\begin{bmatrix} \beta_{0,3} \\ \beta_{1,3} \\ \beta_{2,3} \\ \beta_{3,3} \\ \beta_{4,3} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=m}^{n} (i-m) & \sum_{i=m}^{n} (i-m)^2 \\ \sum_{i=1}^{n} i & \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} [i(i-m)^2] \\ \sum_{i=1}^{n} i^2 & \sum_{i=1}^{n} i^3 & \sum_{i=1}^{n} i^4 & \sum_{i=m}^{n} [i^2(i-m)] & \sum_{i=m}^{n} [i^2(i-m)^2] \\ \sum_{i=m}^{n} (i-m) & \sum_{i=m}^{n} [i(i-m)] & \sum_{i=m}^{n} [i^2(i-m)] & \sum_{i=m}^{n} (i-m)^2 & \sum_{i=m}^{n} (i-m)^3 \\ \sum_{i=m}^{n} (i-m)^2 & \sum_{i=m}^{n} [i(i-m)^2] & \sum_{i=m}^{n} [i^2(i-m)^2] & \sum_{i=m}^{n} (i-m)^3 & \sum_{i=m}^{n} (i-m)^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} T_i \\ \sum_{i=1}^{n} (T_i i) \\ \sum_{i=1}^{n} (T_i i^2) \\ \sum_{i=m}^{n} [(i-m) T_i] \\ \sum_{i=m}^{n} [(i-m)^2 T_i] \end{bmatrix} \tag{8}
$$

After determining the coefficient of all three segmented regression models, the value of the optimal switching point m is selected for each model based on the corresponding least value of standard deviation σ.

$$
\sigma = \sqrt{\frac{1}{n-l} \sum_{i=1}^{n} \left(T_i - \widehat{Y} \right)^2} \tag{9}
$$

where l is the degree of freedom for each of each selected model, where T_i is the time moment of the failure, $\widehat{Y}$ corresponds to $Y_1(X)$, $Y_2(X)$, and $Y_3(X)$.

Based on the values of $\boldsymbol{m}$ and $\boldsymbol{\sigma}$, the optimal switching point for the three segmented regression models is determined; the model with the corresponding least value of $\boldsymbol{m}$ is considered the most accurate for the prediction.

Analysis and results

To analyze all three proposed segmented regression models and to determine which of them allows for the most precise prediction, daily aircraft operations data of an aircraft component was processed to generate statistical data (Table 1). For the simulation, the input data is a matrix K of the transformed statistical data.

The values of the standard deviation σ for each value of *m=5-15* are analysed. The results of this analysis are given in Table 2; *Y1, Y2*, and *Y3* respectively refer to quadratic-linear segmented regression model, linear-linear segmented regression model, and quadratic-quadratic segmented regression model.

Table 1

Statistical data generated from aircraft operations

Failure *i*	Time between failures	Failure *i*	Time between failures	Failure *i*	Time between failures	Failure *i*	Time between failures
1	0	11	0.1	21	6.9834	31	4.8833
2	371.1506	12	0.1	22	18.9332	32	8.4767
3	41.7693	13	2.5	23	18.9332	33	0.1333
4	311.8833	14	6.1333	24	53.1834	34	0.1333
5	311.8833	15	0.4167	25	24.9501	35	0.6167
6	239.45	16	0.2167	26	32.6334	36	3.0001
7	85.9168	17	20.9167	27	2.45	37	28.6966
8	308.377	18	2.4167	28	28.5832	38	55.1668
9	85.0265	19	1.5667	29	16.15	39	28.0667
10	10.3166	20	1.25	30	15.4166	40	8.8332

Table 2

Values of standard deviation σ for each value of m=5-15

m	σ		
	Y1	Y2	Y3
5	145.062	97.239	76.122
6	117.221	66.952	60.519
7	91.796	**43.69**	**40.762**
8	70.36	54.138	42.662
9	56.93	80.249	50.536
10	**54.474**	105.13	52.947
11	61.165	126.575	52.218
12	72.302	144.76	50.689
13	84.673	160.203	50.279
14	96.855	173.416	51.98
15	108.296	184.731	56.239

According to Table 2, the least value of standard deviation σ is observed with the quadratic-quadratic segmented regression model when m=7. Therefore, this model is considered the most precise for predicting failure; based on the resulting matrix of coefficients, the formula for the prediction is given as:

$$Y(X) = 106.572 + 198.881X + 4.284X^2 - 249.694(X-7)\phi(X-m) - 4.137(X-7)^2\phi(X-7)$$

Conclusion

The operations phase of the aircraft lifecycle generates statistical data that can be used for regression analysis. Considering that segmented regression models allow for improved prediction, three models were developed in this study for predicting the failures of aircraft components, subsystems, systems, or structures. All three models were analysed using daily aircraft operations data, and the results show that the quadratic-quadratic segmented regression model allows for the most accurate prediction. The proposed model can provide valuable insights for aircraft maintenance planning decisions and be part of a data-driven predictive aircraft maintenance framework. Furthermore, realism is needed in how the aircraft maintenance optimization problem is formulated during the design and manufacturing phase of the aircraft lifecycle; the proposed models can provide insights.

References

1. Okoro, O. C., Zaliskyi, M., Dmytriiev, S., Solomentsev, O., & Sribna, O. (2022). Optimization of Maintenance Task Interval of Aircraft Systems. International Journal of Computer Network & Information Security, Vol.14, No.2, pp.77-89, 2022. DOI: 10.5815/ijcnis.2022.02.07
2. Chatterje .S. Hadi, A.S., Regression Analysis by Example, 5th Edition, Wiley Series in Probability and Statistics, 2013.
3. Seber G. A. F., Wild C. J., 'Nonlinear Regression', Wiley Series in Probability and Statistics, 2003.
4. J.D. Toms, M. L. Lesperance, 'Piecewise Regression: A tool for Identifying Ecological Thresholds', Ecological Society of America, 2003, pp. 2034–2041
5. 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.