Condition Monitoring of the Aircraft Structural Fatigue Test Based on PCA Algorithm

C Y Yan¹, Q Y Zhang¹, J F Lin¹ and Q Q Hou¹

¹ AVIC Aircraft Strength Research Institute, Xi’An, China

E-mail: ychenyao@foxmail.com

Abstract. For the frequent occurrence of structural damage and abnormal conditions in the aircraft structural fatigue test, this paper puts forth a condition monitoring method for aircraft structural fatigue test based on PCA algorithm. First, PCA theory is introduced and its principle is deduced. Second, combined with the characteristics of aircraft structural fatigue test, the off-line principal component model of fatigue test under normal condition is established, in which the static measurement data is treated as the training sample set. According to $T^2$ statistic and statistical test theory, the detection index is constructed to realize abnormal state recognition. Finally, the effectiveness of the proposed method is verified by a test case.

Keywords: Aircraft structural fatigue test, condition monitoring, anomaly detection, PCA, data analysis

1. Introduction

Aircraft structural fatigue test is an important part in the process of aircraft development and finalization. Its purpose is to expose the weak parts of aircraft structure as much as possible, verify the fatigue characteristics of aircraft structure, and provide an important basis for determining the service life of aircraft and formulating a reasonable maintenance plan. In the fatigue test of aircraft structure, the initiation and expansion of the damage of aircraft structure will further affect the safety of aircraft structures. Whether or not the abnormal condition in aircraft fatigue test can be monitored as early as possible, is directly related to the success of the test. In order to fully grasp the state parameters of the aircraft structure, engineers oftentimes need to install a lot of sensors on the aircraft parts. The large-scale sensor network brings difficulties to the processing and analysis of test data. In order to solve such a problem in aircraft structural fatigue test, this paper aims to develop a condition monitoring technique based on the multivariate statistical process.

The aircraft structural fatigue test can be regarded as a complex system. The sub-components of aircraft structure are viewed as subsystems, which are not independent of each other, but closely coupled. In other words, there exists a strong correlation between the deformations of each sub-component. Especially the deformation of the internal structure of each sub-system has a strong coupling effect. At present, all methods of condition monitoring of aircraft structure fatigue test merely monitor the parts of aircraft structure individually. Although the existing methods of condition monitoring have made some achievements, the independent condition monitoring may still misinterpret the normal reaction of the adjacent sub-components of the structure as the...
abnormal information and further give an error warning. On the other hand, it may also miss the abnormal running state because the abnormal information is hidden by coupling. The condition monitoring method based on multivariate statistical analysis has the advantages of processing large-scale and high-dimensional data, so it is more suitable to be utilized in the case of numerous variables, strong correlation and large amount of data [7],[8]. In order to conduct real-time and accurate condition monitoring of the entire aircraft structure during the fatigue test operation, while process a large number of multi-related detection variable information efficiently, Principal Component Analysis (PCA) in multivariate statistical process monitoring technique is first introduced in this paper to realize the operation condition monitoring of aircraft structural fatigue test.

PCA is a multivariate statistical method to investigate the correlation between high-dimensional variables. It can be used to build a low dimensional model to describe the system principle based on the statistical principle. Currently, the PCA is mainly used in the fields of pattern recognition, fault diagnosis, feature extraction and image processing [9-12]. The condition monitoring method of the aircraft structural fatigue test based on PCA is based on the multivariate statistical theory. The multivariate statistical theory can not only effectively analyze and interpret the flight process measurement data, but also judge the running state of the process and detect the abnormal situation. Then it makes sense to ensure the safety and reliability of the test and reduce the accident rate.

In the aircraft structure fatigue test, the random load spectrum is applied to the cyclic loading of the aircraft structure to simulate the actual use of aircraft. Each load spectrum block contains a variety of load types, and each load type contains different load conditions. The frequency of each load condition in the load spectrum block is, therefore, different. However, each load condition has the characteristic of recurring. In this paper, the off-line PCA model was established based on the repeated characteristics of load conditions, by processing the static measurement data under the same load condition. The condition monitoring method of aircraft structure fatigue test based on PCA model is to use the control diagram of a single variable (T^2 statistic) to monitor the change of multiple monitoring variables (strains, displacements and loads) during the test. The main procedures of the proposed method are depicted in Figure. 1.

![Figure 1. The main procedures of the proposed method.](image)

2. Principle of PCA and the data pre-processing method

2.1. Principle of PCA

PCA is a multivariate statistical method proposed by Pearson [13], its purpose is to convert a group of related feature variables into a few independent principal components (PCs) by using the feature transformation technique. The PCs are a set of comprehensive variables rearranged in order of importance, which can reflect most of the information of the original variables.

The observation sequence composed of m-dimensional variables is $X = \{x_1, x_2, \cdots, x_m\}$, each variable has n observed values, hence the observation matrix is $X(X \in R^{nxm})$. If there is much redundant information between the variables in the observation matrix, it is desired to map the observation matrix to the subspace of $p$ ($p<m$) dimension, in which the mapped data can retain the original data information to the greatest extent. The essence of PCA is feature transformation, and the new variable sequence obtained by the transformation is expressed as $T = \{t_1, t_2, \cdots, t_p\}$, $t_i (1 \leq i \leq p)$ is the linear combination of primitive characters $x_i (1 \leq i \leq m)$, $t_i$ is expressed as
\[ t_i = Xp_i. \]  

To maximize the information of the original observation matrix, we resort to maximize the variance, which can be expressed as

\[ \text{Var}(t_i) = \frac{1}{n} \|t_i\|^2 = \frac{1}{n} p_i^T X^T X p_i = p_i^T \Sigma p_i, \]  

where \( \Sigma = \frac{1}{n} X^T X \) is the covariance matrix of the original observation matrix \( X \). The problem of maximizing variance is transformed into another optimization problem, which is expressed as

\[ J_{PCA}(p_i) = \frac{p_i^T \Sigma p_i}{p_i^T p_i}. \]  

In order to unify the scale of the new variable \( t_i \), the coefficient vector \( p_i \) of the linear group transformation is required to be the unit vector, that is

\[ p_i^T p_i = 1. \]  

The optimization problem is equivalent to solving the extreme value of the Lagrange function as follows

\[ L(p_i, \lambda) = p_i^T \Sigma p_i - \lambda (p_i^T p_i - 1), \]  

where \( \lambda \) is the Lagrange multiplier. The optimal solution \( p_i \) of the equation (5) can be obtained by taking the derivative to satisfy the following equation

\[ \Sigma p_i = \lambda p_i. \]  

From the linear algebra theory, \( p_i \) is the eigenvector of the covariance matrix, \( \lambda_i \) is the corresponding eigenvalue. By substituting equation (6) into the formula (2), the following equation can be obtained:

\[ \text{Var}(t_i) = p_i^T \Sigma p_i = p_i^T (\lambda_i p_i) = \lambda_i p_i^T p_i = \lambda_i. \]  

To maximize the variance of \( t_i \), the eigenvalue \( \lambda_i \) must be maximized, that is, \( p_i \) is the eigenvector corresponding to the maximum eigenvalue of the covariance matrix \( \Sigma \).

Calculating the eigenvalues of the covariance matrix \( \Sigma \), and sorting them in a descending form as \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \). Its corresponding eigenvectors are \( p_1, p_2, \ldots, p_m \). \( t_i (1 \leq i \leq p) \) satisfy \( \text{Var}(t_1) \geq \text{Var}(t_2) \geq \cdots \geq \text{Var}(t_p) \), the covariance of \( t_i \) is \( \lambda_i \), so \( t_i \) is the first principal component, it contains the largest of the largest of the raw data, \( t_2 \) is the second principal component, the information of which contains the original data comes next, and so on. \( p \) number of principal components \( t_i (1 \leq i \leq p) \) can be constructed from \( p \) eigenvectors \( p_i (1 \leq i \leq p) \). Cumulative Percent of Variance (CPV) is usually used to determine the number of principal components. The CPV can be calculated as:

\[ CPV_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{m} \lambda_i}, \quad k = 1, 2, \ldots, m. \]  

Generally speaking, as the cumulative contribution rate of the current \( k \) principal components is more than 85%, it is considered that the principal component space contains enough information in the original data.

2.2. Data pre-processing in PCA

In practical application, the variables in the observation matrix may have different dimensions, and the units of same variable may be different, and the principal component may change due to the
different dimensions and orders of magnitude of the observation variables. In order to eliminate such effects, the first step in PCA application is to standardize the observation matrix $X$, which is

$$\hat{x}_{i,j} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, m),$$  \quad (9)$$

where $\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$ is the mean of the observed variables $x_j$, $s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}$ is the standard deviation of $x_j$.

After normalization, each variable in the observation matrix is processed into an observation sequence with a mean value of 0 and a variance of 1. For illustration purposes, the preprocessed data is still represented as $X(X \in \mathbb{R}^{n \times m})$.

3. The condition monitoring of aircraft structural fatigue test based on PCA

3.1. The basic principle

In the aircraft structure fatigue test, the measurement data of test include the information of strain, displacement and load of each test part of aircraft structure. The PCA was introduced to condition monitoring aircraft structure fatigue test, its basic idea is to construct the PCA model by using the history measurement data under the normal operation. During the running state of test, the new testing data is detected continuously to calculate the deviation degree of the principal component model, and its purpose is to determine whether there is any abnormal condition. The degree of deviation is quantified by constructing the anomaly detection index based on the principal component model.

3.2. The condition monitoring model

The condition monitoring model of aircraft structural fatigue test is a principal component model, which describes the running state of the system in the form of vector and matrix through statistical analysis of the historical data under normal conditions. Corresponding to the principal component model is a set of statistical information, obtained after PCA treatment of the training samples, which mainly includes: principal component matrix $T$, load matrix (eigenvector matrix) $P$, eigenvalue matrix $\lambda$, and principal component subspace $\hat{X}$.

The principal component matrix is $T = \{t_1, t_2, \ldots, t_m\} = XP$. The $k$th principal component is the element $t_k$, that is

$$t_k = Xp_k = \left[ \sum_{j=1}^{m} p_{jk} x_{1j}, \ldots, \sum_{j=1}^{m} p_{jk} x_{nj} \right]^T,$$  \quad (10)$$

where $p_{jk}$ is the $j$th component product of $p_k$, the load matrix is $P = [p_1, p_2, \ldots, p_m]$.

The eigenvalue matrix is $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$.

The expression of the original observation data $X$ is

$$X = \sum_{i=1}^{k} t_i p_i^T + \sum_{i=k+1}^{m} t_i p_i^T = T_k P_k^T + E,$$  \quad (11)$$

where $T_k = \{t_1, t_2, \ldots, t_k\}$, $P_k = [p_1, p_2, \ldots, p_k]$. The principal component subspace only reserve the first $k$ principal components to reconstruct the process data, which describes almost all the change information of systems during the
monitoring of aircraft structural fatigue test. The expression of the principal component subspace $\hat{X}$ is

$$\hat{X} = \sum_{i=1}^{k} t_i p_i^T = T_k P_k^T = X P_k P_k^T.$$  

(12)

3.3. Detection indicators of anomaly state

According to the theory of multivariate statistics, in this paper, the Hotelling $T^2$ statistic constructed in the principal component subspace is taken as the anomaly detection index. The Hotelling $T^2$ statistic reflects the change of multivariate by the fluctuation of the module of the principal component vector inside the principal component model.

Suppose that the new detection sample at time $t$ is $x_t = (x_{i1}, x_{i2}, \cdots, x_{im})$, and the result after pre-processing is $\bar{x}_t$. The score vector $t_i$ of the principal component of the new test sample and the estimated value $\hat{x}_i$ of reconstruction are calculated as follows

$$t_i = \bar{x}_i P_k$$
$$\hat{x}_i = t_i P_k^T = x P_k P_k^T$$

(13)

The $T^2$ statistic is defined as follows

$$T^2 = n(\bar{X} - \mu_0)^T S^{-1}(\bar{X} - \mu_0),$$

(14)

where $\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$, $S = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T$, and $\mu_0 = [\mu_{01}, \mu_{02}, \cdots, \mu_{0m}]^T$. The value of $T^2$ statistic represents the distance between $\bar{X}$ and $\mu_0$, the value of $T^2$ statistic is calculated under the new test sample, and the process data is judged by using the statistical test principle. If $T^2$ is even large, which means $\bar{X}$ is far away from $\mu_0$, then reject the hypothesis $H_0: \mu = \mu_0$.

The $T^2$ statistic corresponding to the new test sample $x_t = (x_{i1}, x_{i2}, \cdots, x_{im})$ at time $t$ is [14]

$$T^2 = t_i D_k^{-1} t_i^T = x P_k D_k^{-1} P_k^T x_t^T,$$

(15)

where $D = diag(\lambda_1, \lambda_2, \cdots, \lambda_k)$ is the eigenvalue matrix of the first $k$ principal components. Since the measurement data $X$ of the test under normal operation state was subject to normal distribution, the $T^2$ statistic were subject to $F$ distribution with degrees of freedom of $n$ and $n-k$, $n$ is the number of samples of the observation matrix, and $k$ is the number of principal components.

If the significance level is $\alpha$, the control threshold of $T^2$ statistic is [15]

$$\delta_{T^2, \alpha} = \frac{k(n^2 - 1)}{n(n-k)} F_{\alpha}(k, n-k),$$

(16)

where $F_{\alpha}(k, n-k)$ is the critical value of $F$ distribution with degrees of freedom of $n$ and $n-k$ at test level $\alpha$. To determine whether the abnormal conditions occurred or not during the test, the control threshold of $T^2$ statistic is usually defined at test level $\alpha = 0.05$ as the warning control threshold.

The essence of $T^2$ statistic is the length of the projection vector of the new test sample under the principal component space, which mainly explains the comprehensive fluctuation degree of the first $k$ principal components. When the $T^2$ statistic exceeds its control threshold, abnormality can be determined during operation.

The value of $T^2$ statistic at time $t$ is a scalar sum accumulated by multiple monitoring variables during the test. The contribution plot of variables to the comprehensive statistic variables can reflect
the influence of changes of each variable on the stability of the statistical model. The contribution \( C_{jh} \) of the \( j \)th process variable at time \( t \) to the \( h \)th principal component is calculated as follows

\[
C_{jh} = x_j P_{hj}
\]  
(17)

The contribution plot can be used to analyse the contribution of each variable to the statistics, and to further determine which variables are responsible for the abnormal monitoring state.

### 3.4. Implementation steps of condition monitoring and abnormal detection

The condition monitoring of aircraft structure fatigue test is based on PCA, which implementation steps are as follows.

- **Step 1:** Collecting and pre-processing historical sample data to establish off-line principal component model.
- **Step 2:** Calculating the control threshold of abnormal detection indicator \( T^2 \) statistic.
- **Step 3:** Detecting and pre-processing new sample data.
- **Step 4:** Calculating the value of \( T^2 \) statistic, and comparing it with the threshold. If the threshold is exceeded, an abnormal alarm of the test will be announced.
- **Step 5:** Calculating the contribution rate of process variables to \( T^2 \) statistic, the variable with the largest contribution rate is the one that causes the abnormal state alarm.

### 4. Test application and result analysis

#### 4.1. Test application

The structural fatigue test of a certain type of aircraft has been carried out to the stage of triple fatigue life, and is now in the period of concentrated fatigue damage outbreak. Before the fatigue test of this type of aircraft, seven conditions with high severity were selected for static debugging. The structure of aircraft is normal and without damage in the static debugging process. In this paper, the data of 16 static measurements of a typical working condition is taken as the historical data set under the normal fatigue test state, the historical data set makes up the training sample set to establish a PCA model under norm state of fatigue test. The historical data set is an 8-dimensional observation variable matrix containing both strain and displacement information. The strain information is selected from the monitoring point where the wing stress is relatively concentrated, and the displacement information is selected from the monitoring point which can reflect the large deformation of the wing and the posture of the aircraft. The information of the history data set under the static measurement is shown in Table 1.

| NO. | Process Variables   | Units | Mean Value | Standard Deviation |
|-----|---------------------|-------|------------|--------------------|
| 1   | Displacement sensor 1 | mm    | 526.1      | 8.8                |
| 2   | Displacement sensor 2 | mm    | 535.9      | 8.6                |
| 3   | Displacement sensor 3 | mm    | -134.6     | 9.2                |
| 4   | Displacement sensor 4 | mm    | 196.3      | 3.9                |
| 5   | Displacement sensor 5 | mm    | 193.2      | 4.0                |
| 6   | Strain gage 1        | με    | 1133.2     | 21.9               |
| 7   | Strain gage 2        | με    | 1082.6     | 25.8               |
| 8   | Strain gage 3        | με    | -497.0     | 45.9               |

Table 1. The information of the history data set under the static measurement.

After the pre-processing of the historical data set, the contribution rate of each principal component after PCA transformation is shown in Figure 2. As can be seen from Figure 2, the cumulative contribution rate of the first three principal components reached 85.4%. Therefore, the three principal components were finally retained to establish a condition monitoring model fatigue test.
In order to verify the effectiveness of the proposed method for monitoring the fatigue test state of the aircraft structure, the measured data under the same typical condition during the fatigue test were selected as the test sample set, with a sample size of 27 points. The test sample set includes two kinds of abnormal conditions during the test process: (1) the change of wing posture, which is caused by the increase of deformation in the right outer wing. (2) The fracture of the connecting bolt, which is caused by the increase of strain in one part of the wing.

4.2 Result analysis

The principal component scores of the testing sample set in PCA model is shown in Figure 3. From Figure 3, we can see that the distribution of the testing sample set under the principal component model. Most of the data in the test sample set is clustered in the same area, but a few sample points diverge from where most of the points converge, which indicates some anomalies may appear in the process of fatigue test. The monitoring process of $T^2$ statistic of test sample data set is shown in Figure 4, the dotted line in Figure 4 represents the control threshold of $T^2$ statistic as the confidence level is 0.95. From Figure 4, we can see that the value of the $T^2$ statistic at the 9th and 16th sample points exceeds the control threshold. It indicates that the process variable deviates from the principal component model at the corresponding time. Put another way, some abnormal conditions occurred at the detection time corresponding to the 9th and 16th sample points in the fatigue test.
The contribution plot of each variable to the comprehensive statistics can reflect the influence of the changes of each variable on the stability of the statistical model. Therefore, this paper uses the contribution plot of the statistics to determine which monitoring variable is most significant to the abnormal monitoring process. The contribution diagrams of each monitoring variable, at the detection time of the 9th and 16th sample points in the fatigue test, are shown in Figure 5 and Figure 6 respectively.

As shown in Figure 5, at the detection time of the 9th sample, the contribution rate of the second process variable to both the first principal component and the second principal component was the largest, and the contribution value of the other variables was relatively small. Such a phenomenon indicates that there was an abnormal situation near the measurement site of displacement 2 sensor. Combined with the inspection results, it was finally found that the reason for the abnormal displacement 2 data was that the increased deformation of the aircraft’s outer wing and the changed posture of wing.
Figure 6. The contribution diagrams of each monitoring variable at the detection time of the 16th sample points in the fatigue test.

As given in Figure 6, at the detection time of the 16th sample, the contribution rate of the second process variable to both the first principal component and the second principal component was the largest, and the contribution value of the other variables was relatively small. It indicates that an abnormal situation occurs near the measurement site of strain 1. Combined with the inspection results, it was finally found that the reason for the abnormal strain 1 data was that the increased deformation of the aircraft's right outer wing and the fracture of a connecting bolt of the wing.

In summary, the condition monitoring of the aircraft structure fatigue test based on PCA model is effective in identifying and locating abnormal conditions.

5. Conclusion
In this paper, the correlation between various monitoring variables in aircraft structural fatigue test was studied, and a novel condition monitoring method based on PCA for aircraft structural fatigue test was proposed. This method took the static measurement data of typical fatigue test conditions as the training sample set, and established the off-line principal component model under the normal state. Then the new test samples in the fatigue test was detected and judged to realize the recognition of abnormal state. The experimental results showed that the proposed method can effectively reduce the dimension of monitoring variables, and the T² statistic comprehensively reflected the information of state changes in the process of test, and it is capable of identifying abnormal conditions. The new method provided a new way of thinking for the damage monitoring of aircraft structure fatigue test, which is of great significance to reduce the test risk and has a certain promotion value.

References
[1] WANG Y J, DONG J, LIU X D and et al 2015 Identification and standardization of maneuvers based upon the operational flight data J. Chinese Journal of Aeronautics 29(1) 133-40
[2] ASRI. 2013 Aero structural strength technology (Beijing: Aviation Industry Press) 45-9
[3] ZHONG G Y 2014 Automatic monitoring of aircraft structural damage based on measured strain Proceedings of the 6th China Aviation Society Youth Science and Technology Forum Shenyang China 1589-94
[4] PAN S Z, ZHONG G Y and LIU X D 2017 Structural damage prediction method in aircraft full scale fatigue test Proceedings of the 2017 China aviation science and technology conference Beijing China 69-73
[5] WANG X G, YANG H L, HOU Q Q and et al 2018 Study on structural damage monitoring in aircraft fatigue test based on strain data Proceedings of SAMPE China 2018 International Conference Shanghai China 193-7
[6] AN G, WANG X X and DU Z 2012 The strain monitor method for structure crack inspection on full-scale aircraft fatigue test J. Structure & Environment Engineering 39(02) 14-9
[7] HE X Q 2008 multivariate statistical analysis (Beijing: China Renmin University Press)
[8] HU J 2015 Fault diagnosis and quality monitoring based on multivariate statistical analysis D. Zhejiang University
[9] Yang J, Zhang D, Frangi AF and et al 2004 Two-dimensional PCA: A new approach to appearance-based face representation and recognition J. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(26) 131-7
[10] ZHU S Q and SHI J F 2007 PCA approach to condition monitoring and fault diagnosis J. Machine Tool & Hydraulics 2007(01) 241-3
[11] Jombart T, Devillard S and Balloux F 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations J. BMC Genetics 11(94) 1-15
[12] FU R H 2019 Simulation of hyperspectral image dimensionality reduction algorithm based on PCA transform J. Computer Simulation 11(36) 234-8
[13] PEARSON K 1901 On lines and planes of closest fit to systems of points in space J. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2(11) 559-72
[14] ZHANG J and YANG X H 2000 Multivariate statistical control process (Beijing: Chemical Industry Press)
[15] JACKSON J E and MUDHOLKAR G S 1979 Control procedures for residuals associated with principal component analysis J. Technometrics 21 341-9