

IDENTIFICATION OF ABNORMAL EVENT BY USAGE AND FLIGHT DATA MONITORING

Pierre Bect ^{1,2}, Zineb Simeu-Abazi ², Pierre-Loic Maisonneuve ¹ and Patrick Lyonnet ³

¹ Eurocopter, Aéroport International Marseille Provence, 13725 Marignane, France

² G-SCOP Laboratory, 146 avenue Felix Viallet, 38031 Grenoble, France

³ ENISE, 58 rue Jean Parot, 42100 Saint-Etienne, France

Abstract

Many maintenance actions, related to mechanical, electrical, hydraulic systems, are mandatory to maintain a helicopter in operational conditions. A lot of researches have been performed in order to optimize maintenance actions. Each field of expertise assesses the good health of a specific part of the helicopter by using their own experts, their own methods, and sometimes their own data. Nevertheless, all these domains composed the same machine and no interaction between systems is considered. Our study is not based on physical approaches but it uses operational data and mathematical tools to assess, on ground, the current state of the system. The proposed paper concerns a new concept consisting in characterizing the normal usage of the helicopter by using flight data in order to detect any abnormal event. This abnormal event could be assimilated to a global behavioural drift of the aircraft, a behaviour which is different than usual.

Keywords: abnormal event detection, abnormal event characterization, Statistical Data Analysis, Trends monitoring, Machine Learning Customer Feedback.

Introduction

During the last years, maintenance became one of the main priorities of manufacturing industries. A lot of researches have been performed in order to optimize maintenance actions in the aeronautic field:

- Avionic failure diagnostic methods have been improved by [1] and [2].
- Avionic failure prognostic was studied by [3].
- Mechanical degradation diagnosis by vibration analysis is continuously improved.
- Condition based maintenance, a maintenance concept in expansion, has been described by [4]. Monitor the real usage of the aircraft allows the expansion (or reduction) of the time between predictive maintenance operation.
- Mastery of the product ageing is continuously improved [5].

A complex system is composed of different parts, with its own field of expertise, experts, methods and architecture. These sub-systems are monitored in the same way, but their maturity levels are different. Fig 1, presented by [6], describes these different modules which compose the full health monitoring. However, in a complex system, interactions between sub-systems involve maintenance difficulties. Indeed, providing global information, in relation with the system, becomes harder when the system becomes more complex. All detection methods are linked to a specific sub-

system monitoring. A behaviour which is not common to a specific condition attached to another sub-system is not recognized as abnormal.

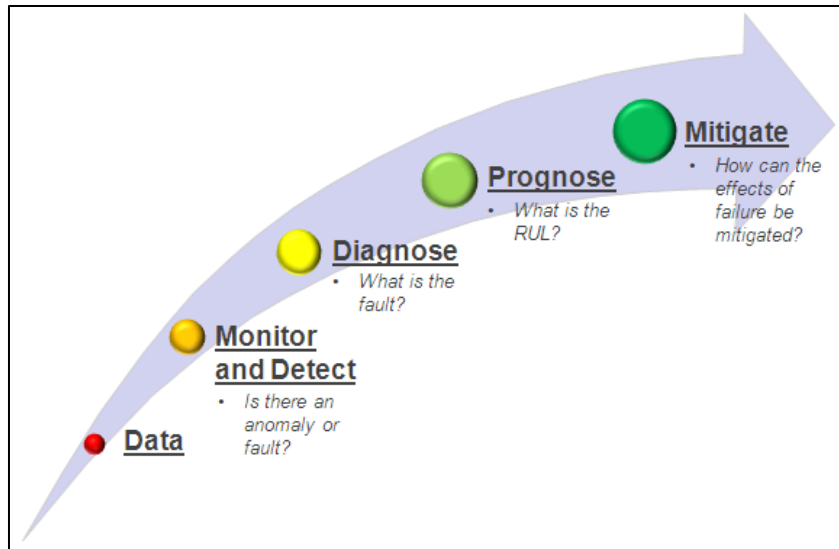


Fig 1: Full health monitoring system

A system is considered as normal when every observable output of the system is seen without default or failure [7]. So, using the normality concept consists in recording every data in a good way of running and characterizing them as “normal” in order to identify the normal behaviour of the system. Each new record, which does not correspond to a “known way of running”, should be considered as “abnormal” and has to be detected and characterized.

Today, a lot of in-service data are collected from customers. These types of data are heterogeneous: failure code, vibrations data, acoustic data, usage and flight data... The other part of these information is provided by customers operating the helicopter: maintenance repair and overhaul data, flight report... Different utilizations of these different kinds of data are performed:

- Vibration data are commonly used to monitor specific mechanical parts of the aircraft [8] and [9]. Monitor vibration data allows the detection of degradation and in some case the cause of degradation on mechanical parts.
- Failure codes are the base of avionic diagnostic [1], [2] and [10]. These information are filtered, grouped and analysed to detect and localize failure on avionics system.
- Usage and flight data allow the follow-up of the real usage of the aircraft and by this way, the development of Condition Based Maintenance [11].

Focus on usage and flight data. These data are made up of information like: temperature, pressure, command position, rotor speed, altitude, fuel flow/quantity, weight... By the way, these data are representative of environmental context and helicopter behaviours. Consequently, this information could be useful for each field of expertise and represent the global environment system. To improve maintenance performances, taking into account this kind of data is a good way to proceed but it raises many difficulties. Indeed, the objectivity, the non-specificity, the important quantity of data and heterogeneity made flight data difficult to use. In order to ease the flight data utilization, two ways are essential: reducing quantity of data and synthetizing the main information.

Data mining offer a wide range of specific tools to resolve this problem [12] but need important computing capacities [13]. To reduce this substantial amount of data, a sequencing of several statistical tools was set up. It allows the storage of an information representative of historical data and by this way, experts are able to compare easily current information with the evolution trends.

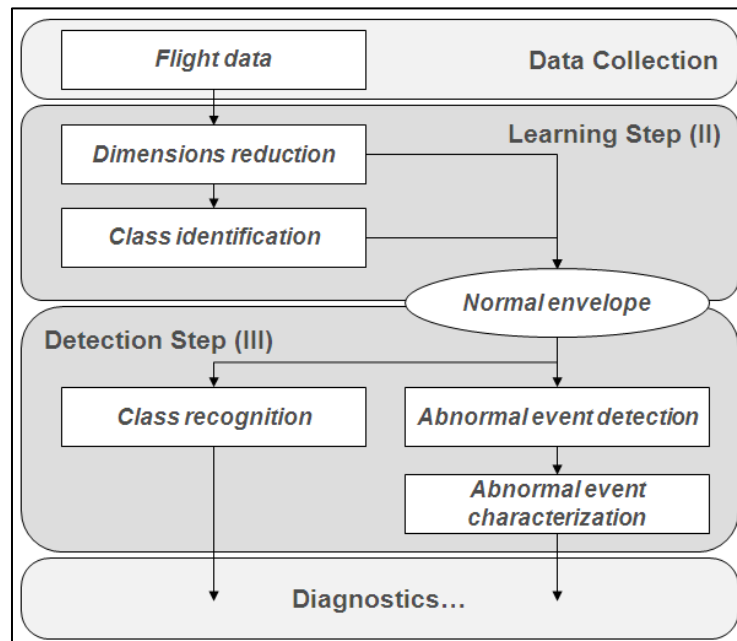


Fig 2. Architecture of the methodology

This paper is divided into two different parts (Fig 2). The first one presents how is made the “life profile of helicopter”. It describes the principle of construction of our referential. This referential is essential; it is where every information records will be projected. The outline of this grouping of projections is called “normal envelope”. It could be assimilated to the normal usage of the helicopter, its life profile. The second part presents how the normal envelope is used as a comparison element. Initially, the normal envelope allows the detection of an unknown event. Then this event is automatically characterized as abnormal and different indicators are defined in order to evaluate its distance and direction from the identified norm.

I. Construction of the normal envelope

The main objective of this normal envelope is to contain every data relative to a normal way of running. The normal envelope must be representative of everything that was seen by helicopter without failure occurrence. Normal envelope could be assimilated as the known “good way of running”. So, every record which is not wrapped in the envelope could be considered as abnormal.

The “grouping of normal data” is very important in term of recording. To facilitate the substantial amount of data analysis and its comprehension, it is necessary to synthesize the information.

An abnormal event could be potentially dangerous. It is difficult to evaluate criticism of this abnormal state of the helicopter via statistics analysis. A partition of the

“grouping of normal data” is performed; it will facilitate the physical and behavioral description of “sub-grouping of normal data”.

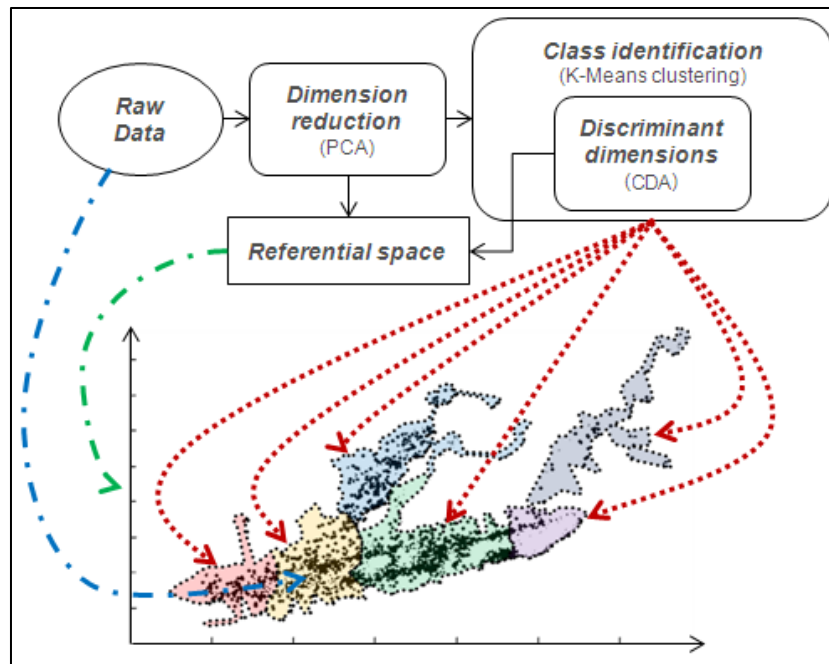


Fig 3: Principe of normal envelope construction

In our studying case, a 2 Hertz sampling frequency allows the record of 136 parameters. We dispose of roughly 980 000 records data per flight hour. Each recording session is divided flight by flight but in order to compare similar recording sessions, a preliminary segmentation work is performed. A decision tree designed by experts allows the flight phases recognition. Our studying case is performed on stabilized flight. The construction of the normal envelope has been performed on a learning database which is composed of roughly 170 record hours on stabilized flight (1224 thousands records).

I.1. Dimensions reduction step

Many mathematical tools could tackle the reduction problem. One of the most pertinent is the Principal Component Analysis (PCA) [14]. The objective of PCA is to transcribe the maximum of information in a reduced number of dimensions. PCA constructs new dimensions named vector. These vectors are based on a linear combination of every input parameters. By this way, each parameter (which corresponds to initial dimension) has a contribution in the vector. Then, it is possible to seeing the relation between the principal contributors of each dimension.

The main difficulty that this methodology has encountered is the construction of the PCA. The objective is to define dimensions that could be improved, readjusted in function of helicopter's records and also insensible to flight peculiarities. It could be possible to construct the PCA on the entire learning database at a stretch. However, in real operational environment, this procedure has to be performed after each flight and it raises many problems. Consequently, a new approach has been designed: an iterative PCA. It impacted the calculation process of the PCA. The proposed solution consists in performing a PCA after each flight, store this result and merge PCA of different flights in order to have a generic PCA. To merge two vectors from various

flights efficiently, the duration of the flight has to be taken into account. The equation [1] shows the generalization to n flights of this method. Where PCA_{gen} is the matrix which is composed of the PCA vectors for all flight, $NFra_x$ the number of frame for flight x and PCA_x the matrix of PCA vectors for flight x .

$$PCA_{gen} = \frac{\sum_{x=1}^n PCA_x \cdot NFra_x}{\sum_{x=1}^n NFra_x} \quad (1)$$

The same method could be iterative (equation (2)).

$$PCA_{gen}(t+1) = \frac{PCA_{gen}(t) \cdot NFra_{tot}(t) + PCA_x * NFra_x}{NFra_{tot}(t) + NFra_x} \quad (2)$$

$$NFra_{tot}(t+1) = NFra_{tot}(t) + NFra_x$$

When the general PCA vectors are defined, these new dimensions compose a new referential. This referential will be used to compare and project the recorded data. A matrix multiplication is performed to obtained projection coordinates of each recorded data in the referential.

1.2. Class Identification step

A segmentation of the « grouping of normal data » is performed to ease physical and behavioral description. This identification is not mandatory to qualify a new record as “abnormal” but it could allow to be more accurate and it provides another kind of information which could be very helpful. Indeed, having good knowledge on aircraft physical phenomenon and on behavioral specificities could be benefit for maintenance performance, design improvement. Data mining methods offer a lot of clustering tools [14]. K-means clustering solved efficiently this partition problem [15] and the combination of K-means clustering and Principal Component Analysis has been well-tried [16].

To assess efficiently the current state of the system by using operational data, it is necessary to be totally objective. K-means clustering is a non-parametric method and consequently, it corresponds exactly to this constraint of objectivity.

The only input that needs the K-means clustering is the number of clusters to create. The most difficult problems in data clustering is the automatical determination of the optimal number. To provide this information, [15] proposed different methods that were used in literature. By using a mathematics criterion, the process is totally objective.

When the clusters are stable, the first question that is coming is “How it is possible to distinguish efficiently the clusters?”. The problem involved by this fact, could be solved by the identification of discriminant dimensions.

Canonical Discriminant Analysis (CDA) could untangle this problem. Canonical discriminant analysis is a dimension-reduction technique related to principal component analysis and canonical correlation. An interesting comparison between CDA and PCA is done by [17]. The objective of the CDA is to define dimensions in order to separate the different clusters while PCA tries to maximize the information on

a minimum dimensions. At the opposite of the PCA construction, CDA process must be performed on the entire knowledge database. Indeed, during a flight, a specific class could not appear, so the construction on the CDA flight by flight is impossible. On Figure 4, the first graph presents 3 clusters projected on PCA vectors while the second graph presents the same clusters projected on CDA vectors:

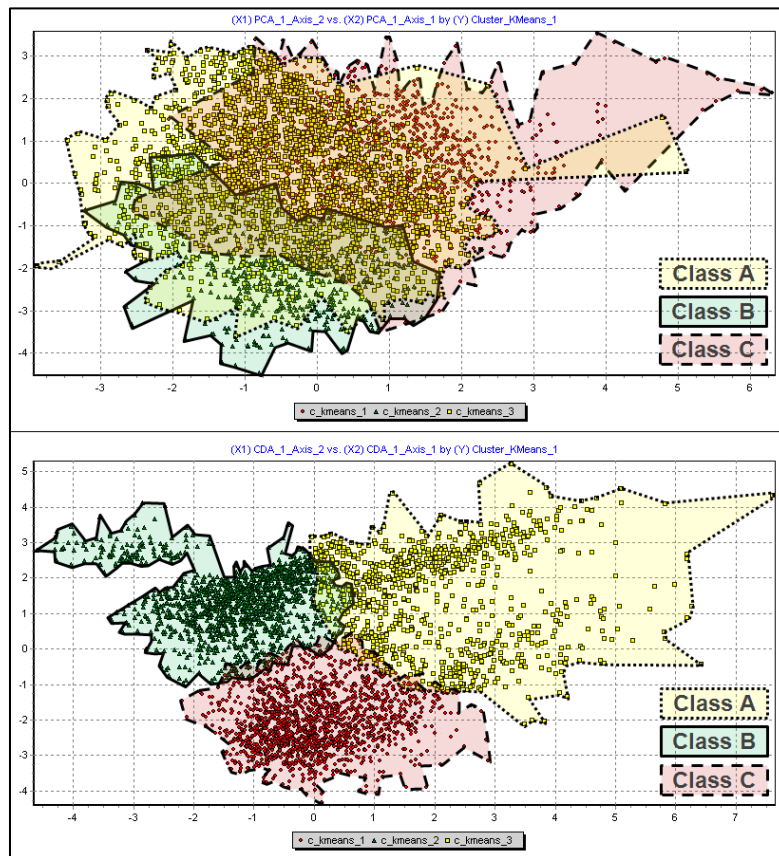


Fig 4. 3 clusters projected on different dimension (PCA vs CDA)

After the CDA constructions step, the defined dimensions could be added to PCA dimensions in order to improve the referential. These new dimensions facilitate the segmentation of clusters.

This section has set up a referential by using different mathematics tools while remaining as much objective as possible. This referential offers the possibility to compare different flights. By this way, it becomes possible to dispose of an element of comparison. The comparison element proposed is a substantial amount of data records in “good way of running”. The projection of this comparison element is the normal grouping of data, and its outline designs represents the normal envelope.

II. Exploitation of the normal envelope

The normal envelope allows capitalization of operational data in a normal way of running. In consequence, it could be assimilated as to a comparison element which is representative of the life profile of the threaten system. This learning step established, different questions link to its utilization, are asked:

- What is the signification of a new operational data projection?

- Is this projection contained in the normal envelope?
- If this projection is considered as abnormal, how far is it from the normal envelope?

This section will try to answer to these questions.

II.1. Class Recognition step

In order to describe easily physical and behavioural phenomenon included in the normal envelope, a subdivision of the grouping of data have been performed. However, this subdivision was performed with the K-means clustering. K-means clustering is an objective technique which runs on an important quantity of data. In consequence, this method could not be applied on a new record. It is necessary to find a methodology which is able to recognize the membership classes of a new record and the multivariate normal law could easily tackle this problem.

The multivariate normal law is a Gaussian which is implemented on N dimensions. The multivariate normal law is constructed for each class and on all dimensions of the referential that we have construct in the first part. This method provide for the new record a membership probability to a define class. In the case of helicopter, six classes have been identified on the entire knowledge database (for the stabilized flight phase). This method has been tested and provides interesting results (by crossed validation, we obtained 96.8% of good recognition).

This step allows class attribution to a new record. It is important to specify that it is only a mathematical way and that a physical approach has to be performed to provides an accurate explication on the behavioural phenomenon.

II.2. Abnormal event detection

In order to detect an abnormal phenomenon, it is necessary to identify a new record which is not contained in the normal envelope. The first idea consists in defining the outline. There are two possibilities:

- Convex envelope. A convex polygon is a polygon that has all interior angles less than 180° [18].
- Concave envelope. At the opposite of a convex polygon, a concave polygon is a polygon that has one or more interior angles greater than 180° [18].

The concave envelope takes in consideration all specificities of projected data but there is not a defined method to build them. In the studying case, the important number of dimensions increases the complexity of the algorithm. In order to simplify this difficulty, the problem has to be taken in another way.

The second idea consists in identifying if a new record is contained in the normal envelope. In this case, we consider that a new record is abnormal if it is isolated. It could be assimilated as "the new record is contained in the concave envelope". By using the previous projected data, it determines if a previous record has been projected at proximity of the new record. By this way, this method is able to provide a Boolean which indicates if the new record is normal (so a record exists at its proximity) or abnormal (this new record is singular). This method has to be configured on two different ways: the learning database (which contain the normal projections coordinates) and the proximity parameter.

In the previous section, a class has been attributed to every new record. It could be very interesting to know if this new point is contained in the normal envelope of its attributed class. In consequence, detecting if the new record is contained in their recognized class could be the solution. By introducing this processing on the global

envelope and on the recognized class outline it becomes possible to detect wrong class recognition. This procedure allows detection of abnormal event in a specific flight condition.

To illustrate, a frame recorded during a stabilized flight at high altitude is attached to a class A. This class is relative to a high altitude. If when this record is projected on the referential, the point is contained in the normal envelope but not in the class A envelope, there is a problem. It could be a wrong recognition or it could be a real defect because of the non-resemblance with the normal data. By this way, the detection process is improved.

After detection of an abnormal phenomenon, it is necessary to assess its estrangement.

II.3. Abnormal event characterization

Usage data represent the helicopter behaviours and outside events encountered by the machine. Consequently, a single abnormal record is not very representative and so it is better to take into consideration more than one record. Records grouping utilization allows to observe a drift of the points cloud. By this way, it becomes possible to dispose of a trend monitoring and to provide interesting indicators.

In order to easily work on these indicators, it is imperative for us to normalize them. So, all of these indicators provide a value between 0 and 1. In this paper, we propose 3 different indicators:

- Detection indicator which provide the rate of abnormal records during a time period.
- Distance indicator which provide an idea of the estrangement from the normal envelope during a time period. By a trend monitoring it could be assimilated to a speed drift from the normal envelope.
- Direction vector. This information presents the area in which appears abnormal records. This information should become the best indication in order to diagnose an eventual degradation or failure.

II.3.1. Quantification

In order to quantify properly the drift, it is important to integrate temporal information. So, it is necessary to have a referential of time to build the speed of drift. The idea is to make indicators by time period, and so compare two indicators becomes possible because of their same referential (the same number of records).

First of all, it is essential to validate and consolidate the abnormal event detection. It could be interesting to quantify the number of abnormal records during a time period. So, to have a global indication, the first indicator that is created is the rate of abnormal records:

$$Ind_{\text{detection}} = \frac{N_{\text{abnormal}}}{N_{\text{total}}}$$

(3)

In this equation, N_{total} represents the number of records of the defined period. By this way, it becomes possible to dispose of information which indicates the abnormal degrees over this period. The same indicator could be created on a flight. In this case N_{total} represents the number of records of the flight.

The second indicator that is created is a distance between two points. The main difficulty is the definition of these points. There is not a single solution to tackle this problem and many distances could be calculated (Figure 5). The interpretation of these distances is also a problem. Scales are designed in the previous section, in consequence their significations are not relevant and it is complicated to assess the dangerousness in function of these distances. By this way, provide an indicator which is representative to a difference between the records and the norm is difficult.

The proposed method consists in the calculation of the membership probabilities of each point in the normal records grouping. The created indicator represents the gap between the records and the norm; this indicator cannot have any physical representation. To perform this calculation, the multivariate law could be used again. The membership probability is not computed on a single record but on the average of records grouping which represent a time period. In this solution, the centre of the new records cloud and their variance/covariance matrix are used to compute this calculation.

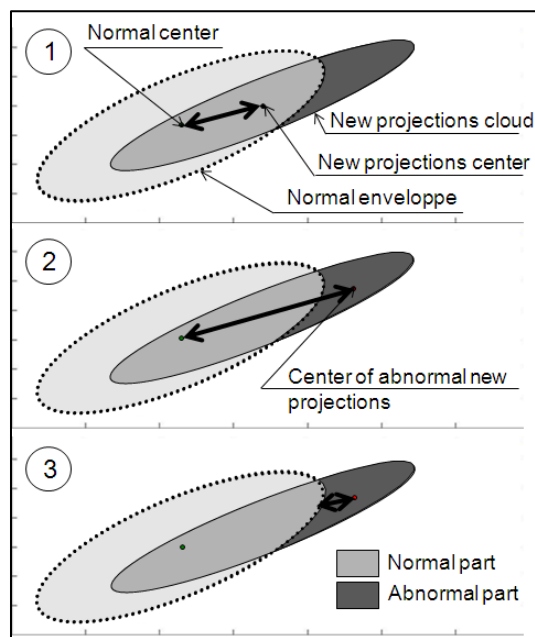


Fig 5: Distances could be calculated to characterize abnormal event

This step allows having an idea of the difference between the records and the norm and on the speed of drift. Disposing of an information relative to the area where appears abnormal records could be very useful in order to understand degradations and to diagnose efficiently a failure or an abnormal wear of part of helicopter.

II.3.2. Qualification

The normal envelope is defined on N dimensions, so the identification of a specific direction is not an easy step. Moreover, to dispose of a direction which could be exploitable, the cloud of records must be compact and properly localized.

The main goal is to identify a drift direction of the normal envelope and if the landmark is the origin of the referential, there will be a problem. The solution consists in fixing the landmark at the centre of the normal envelope. By this way, it becomes possible to identify a specific direction relative to a drift of the normal envelope.

This direction is defined on the N dimensions and it represents the step to perform to pass from the landmark to the centre of the records cloud. This information contains also a notion of distance, but as it was said in the previous part, because of the scales, this distance is not relevant.

The direction is obtained by using the Chasles relations and it is presented as a vector of size N:

$$\overrightarrow{RX} = \overrightarrow{OX} - \overrightarrow{OR}$$

(4)

In this equation, O is the origin of the referential; R is the centre of the normal envelope and X the centre of the cloud of records that is currently analysed.

This direction allows identification of known “abnormalities”. In this goal, an important work of listing has to be performed. The objective is to be able to recognize a specific distortion of the envelope and to link them to a known degradation or failure. To reach this goal, physical interpretation of the drift and a good knowledge of the machine could be an interesting way of investigation.

III. Conclusion and development perspectives

The objective of this method is to identify an abnormal event by flight data monitoring. This approach presents a lot of advantages because of the non-specificity of the data and their global impact. The global impact of these data allows detection at a higher level, at the system level and it could be the link between all other data and so, their computing results.

Data mining tools offer a lot of mathematic methods to explore a substantial amount of data and provide interesting information.

One of the most difficult problems is to characterize the normal envelope, to represent the life profile of the system. We have considered that the normality is composed of all phenomena that encounter the system during their learning phases without failure or observed degradation. It could be assimilated to the “good way of running”. This hypothesis allows the definition of the normal envelope. This envelope becomes the normal referential and everything which is not contained in this envelope could be qualified as “abnormal”.

In order to validate and characterize an abnormal event, 3 different indicators have been designed. They allow an expert to have an opinion on the global health of the system. This method is still under testing and their first results are very encouraging.

The method which is presented opens the way to a future usage based diagnostics only based on operational data. This method is objective and essentially based on mathematics. By this way, it could be easily adapted to another system where a lot of operational data are collected.

References

- [1] Lefebvre, Arnaud, Zineb Simeu-Abazi, and Jean-Pierre Derain. «A methodology of alarm filtering using dynamic fault tree» Reliability Engineering and System Safety, Volume 96, Issue 2, 2011: 257-266.

15th Australian International Aerospace Congress (AIAC15)

- [2] Simeu-Abazi, Zineb, Maria Di Mascolo, and Michal Knotek. «Fault diagnosis for discrete event systems: modelling and verification» Reliability Engineering and System Safety, Volume 95, Issue 4, April 2010: 369-378.
- [3] Ghelam, Samir. Implémentation d'une fonction de maintenance prédictive appliquée aux systèmes avioniques. 2006.
- [4] Maisonneuve, Pierre-Loic, Patrick Lyonnet, Mathieu Glade, Samir Ghelam, and Emmanuel Mermoz. «An introduction to Condition Based Maintenance using Reliability assessment methods» AIAC-13 Thirteen Australian Aerospace Congress. Melbourne (Australia), 2009.
- [5] Plaut, R. H., R. H. Andruet, and S. Suherman. «Behavior of a Cracked Rotating Shaft During Passage Through a Critical Speed» Journal of Sound and Vibration, Volume 173, Issue 5, June 1994: 577-589.
- [6] Jamie Coble, Matt Humberstone, and J. Wes Hines, "Adaptive Monitoring, Fault Detection and Diagnostics, and Prognostics System for the IRIS Nuclear Plant", Annual Conference of the Prognostics and Health Management Society, 2010
- [7] Bect, Pierre, Zineb Simeu-Abazi, Pierre-Loic Maisonneuve, Marc Pero, and Emmanuel Mermoz. «Normal signature characterization for system health assessment» PHM11. Shenzhen (China), 2011.
- [8] Mermoz, Emmanuel. Automatisation du diagnostic vibratoire des transmissions de puissance d'hélicoptère. 2002.
- [9] Wiig, Johann. Optimization of fault diagnosis in helicopter health and usage monitoring system. 2006.
- [10] Lefebvre, Arnaud, Zineb Simeu-Abazi, Jean-Pierre Derain, and Mathieu Glade. «Diagnostic of the avionic equipment based on dynamic fault tree» IFAC07. 2007.
- [11] Maisonneuve, Pierre-Loic. Développement d'un modèle de maintenance dynamique pour système complexe demandant une haute fiabilité. 2009.
- [12] Gan, Guojun, Chaoqun Ma, and Jianhong Wu. Data Clustering: Theory, Algorithms and Applications. 2007.
- [13] Cios, Krzysztof, Witold Pedrycz, and Roman Swiniarski. Data mining: Methods for knowledge discovery. Kluwer Academic Publishers, 1998.
- [14] Jolliffe, I. T. Principal Component Analysis. 2002.
- [15] Jain, Anil K. «Data clustering: 50 years beyond K-means» Pattern Recognition Letters, Volume 31, Issues 8, June 2010: 651-666.
- [16] Ding, Christophe, and He Xiaofeng. "K-means Clustering via Principal Component Analysis." (ACM Press) 2004: 225-232.
- [17] Zhao, Guang, and Ann L. Maclean. «A Comparison of Canonical Discriminant Analysis and Principal Component Analysis for Spectral Transformation» Photogrammetric Engineering & Remote Sensing, 2000: 841-847.
- [18] Definition and properties of convex/concave polygons with interactive animation, Math Open References website 2009.