Integrated Multivariate Health Monitoring System for Helicopters Main Rotor Drives: Development and Validation with In-Service Data

Alberto Bellazzi[1], Giovanni Jacazio[2], Bruno Maino[1], Gueorgui Mihaylov[2], Franco Pellerey[3] and Massimo Sorli[2]

[1] *AgustaWestland, Cascina Costa di Samarate, VA 21017 VA , Italy*
Alberto.Bellazzi@agustawestland.com, Bruno.Maino@agustawestland.com,

[2] *Politecnico di Torino, Department of Mechanical and Aerospace Engineering, Turin, 10129 TO, Italy*
giovanni.jacazio@polito.it, gueorgui.mihaylov@polito.it, massimo.sorli@polito.it

[3] *Politecnico di Torino, Department of Mathematical Sciences, Turin, 10129 TO, Italy*
pellerey@calvino.polito.it

Abstract

The implementation into service of accelerometric health monitoring systems of mechanical power drives on helicopters has shown that the generation of false failure alarms is a critical issue. The paper presents a combined application of several multivariate statistical techniques and shows how a monitoring method which integrates these tools can be successfully exploited in order to improve the reliability of the diagnostic systems. The first phase of the research activity was addressed to exploring the potential advantages of using multivariate classification/discrimination/anomaly detection methods on real world accelerometric condition monitoring data. The second phase consisted of an implementation into actual service of an innovative integrated multivariate health monitoring system.

1. Introduction

Failure diagnostics via condition monitoring on mechanical systems and components is a broad and relevant topic. Different approaches based on the development of specific sensors and data-driven methods have been applied in various contexts. For example in (K. Liu, 2013) is described the construction of a composite health index through the fusion of multiple sensor data. In many cases the calibration of reliable data-driven models is obstructed by the lack of data regarding the failure modes of the mechanical system. In such circumstances sophisticated anomaly detection and decision mechanisms might be required (see for example (Ramasso & Gouriveau, 2010)).

This project has been developed under research contract granted by AgustaWestland. It was focused on monitoring the health conditions of mechanical power drives of helicopters. Accelerometric monitoring systems have been previously installed on helicopters produced by AgustaWestland. The adopted vibration monitoring methods are based on analyzing analog signals provided by a set of accelerometers (we refer the reader to (Randall, 2011) and especially (CAA-PARER-2011, 2012)). A set of accelerometers is arranged in appropriate locations on the power drive. To each component of the power drive is associated an accelerometric analog signal. The accelerometric outputs undergo Fourier spectral decomposition and the description of the local (not global) properties of the energy distribution through the spectrum of vibrational modes leads to a set of scalar health indicators, which are supposed to detect specific damages. For example, actually physical indicators represent the energy of the spectral components corresponding to the main rotational frequency and its multiples, the energy contained in a localised energy bands etc. Other indicators, obtained from the second-level signal analysis in both time and frequency domain are related to local variations, correlations between specific spectral channels, local shape factors, signal standard deviations and signal quality.

The description of the state of each mechanical component is done by a specific set of health indicators, selected by AgustaWestland as appropriate for this scope. The health state monitoring method of each component is based on fixed critical thresholds for the values of each condition indicator. Damage alerts are generated when **any** of the indicators exceeds the threshold for certain number of measures. More in detail "yellow alert" is generated if the value of the indicator ex-

Alberto Bellazzi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ceeds certain threshold and a “red alert” is generated when the value exceeds a higher threshold. In other words the adopted monitoring method concerns a univariate (independent) **interpretation** of the health indicators.

The implementation of this health monitoring system on power drives in actual service has shown that a relatively high number of false alarms is generated, thereby requiring additional troubleshooting workload.

The purpose of this research was to develop an innovative health monitoring method **based on the same accelerometric features**, which is able to reduce to the very minimum the false positives. It is important to underline the fact that our proposal does not require installation of additional sensors in order to obtain further physical information.

The empirical observation during the employment of this monitoring/diagnostic system over a long period of time, highlighted the fact that in certain failure circumstances, groups of health indicators react simultaneously to some anomalies. For this reason, even though (by construction) the condition parameters are processed as univariate indicators, multivariate statistical techniques should be taken into account.

The efficiency of the existing diagnostic system has been improved via **third-level multivariate processing of the condition indicators**. A monitoring method which combines several multivariate statistical techniques has been developed and implemented in an efficient integrated tool. The method is able to distinguish with very high level of statistical confidence true failure situations and false anomaly alerts if these have been previously observed and diagnosed on any other aircraft of the same type.

This article provides a more detailed presentation, with addition of some later results, of the research which was preliminarly introduced in (A.Bellazzi et al., 2014).

2. IN-SERVICE DATA

The research was focused on mechanical power drives of helicopters which consist of an assembly of several gears rotating on shafts supported by ball and roller bearings.

AgustaWestland provided a **very large amount of real data** collected on 115 aircrafts of the same type flying in different conditions. The full available experimental data set consists of huge quantity of measurements of the condition indicators of several mechanical components and was collected over a period of four years and thousands of flight hours. The investigation mainly concerned the following set of power drive components in which true (confirmed by inspection of the power drive) and false alerts were detected:

- TTO Pinion, characterised by twelve relevant condition indicators. A representative calibration data set of 6291 measurements has been extracted. During the monitored period one true failure (confirmed by inspection) was observed and three false alerts were reported by the monitoring system on different helicopters.

- IGB Pin, characterised by twelve relevant condition indicators. The calibration data set is composed by 5496 measurements. Five false alerts of three different types were reported by the monitoring system.

- TGB Gear, characterised by twelve condition indicators. The data set contains 6291 measurements. During the monitored period one confirmed true failure was observed and three false alerts were reported on different aircrafts.

- TRDS, characterised by two health indicators. The calibration data set contains 3925 measurements. One confirmed damage and three false alerts has been generated.

- 2nd Stage Pin RH Brgs, characterised by six relevant condition indicators. The data set contains 6514 measurements. One true failure and two false alerts were generated.

-Oil cooler BRG, characterised by six relevant condition indicators. The calibration set is composed by 3954 measurements. The standard (univariate) control system did not report any anomalous behaviour as none of the alert thresholds has been exceeded.

- Hangar Ball Brg. characterised by nine condition indicators. The calibration set contains 4390 measurements. During the monitored period one true failure was observed and three false alerts were reported by the system.

The TRDS and the Hangar Ball Brg are monitored by the same accelerometer. The other mechanical components are monitored by different single accelerometers.

In some cases (TRDS and the Hangar Ball Brg) the individual thresholds of several health indicators were strongly exceeded (largely over the “red threshold”) in a false alert situation. Unexpectedly a true damage provoked more moderate reaction of the monitoring system (values of the health indicators just above the “yellow” threshold). These cases were considered as particularly critical as the **mono-variate evaluation of the damage appears to be misleading**.

In the rest of the article the set of N health indicators of a mechanical component of a power drive will be interpreted as an element in a real N-dimensional vector space and called the **vector state** of the component.

In order to save space, the results will be illustrated by referring to some relevant examples obtained from the above components. The computations have been done by using R statistical software (for more information see (B. Everitt, 2011) and (Everitt, 2005)).

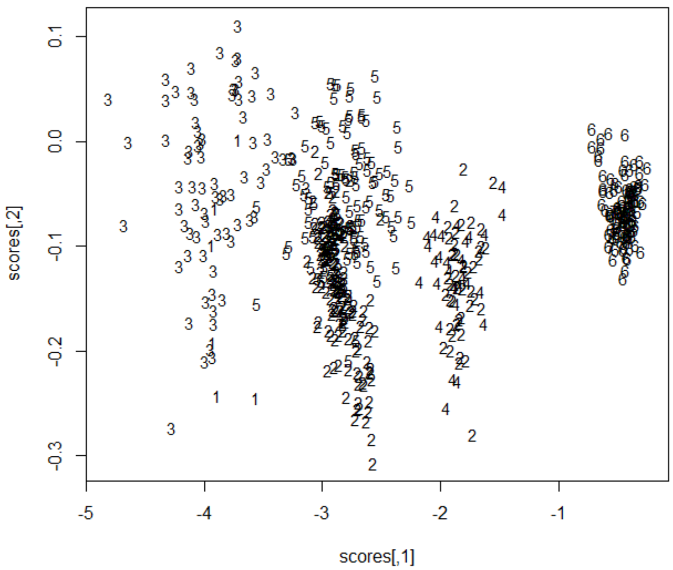

Figure 1. PCA scores of healthy operational states of TRDS component of six helicopters of the same type. Vector states measured on individual helicopters are labelled by different numbers.

3. MULTILINEAR RE-CALIBRATION AND ANOMALY DETECTION

The very first relevant problem we came across in this research program was the fact that values of the standard health indicators, which characterise the healthy operational regime of a mechanical component vary quite consistently between individual aircrafts of the same type. Typically, if compared to each-other, the vector states of the same component in healthy regime on different helicopters form neatly visible clusters inside the vector space of indicators (a striking illustration is given on Fig. 1). Observe that, in this specific case, the individual helicopter clusters spread along the direction determined by the first principal component. This means that by far the most consistent portion of the variance in the data set of healthy operational states can be attributed to differences between individual aircrafts.

The fact that healthy operational states of a power drive installed on different aircrafts cannot be compared, makes **impossible the calibration of any sort of statistical model, based on historical collection of vector states** measured on a fleet of helicopters. Moreover the mechanical components selected for the investigation are typically subject to a very low number of failures. A calibration and a validation of a reliable multivariate model on **each single aircraft** appears therefore as extremely unrealistic.

A solution to these problems is described herein.

Besides the set of component vectors, a historical collection of simultaneous measurements of the following parameters of the operational condition of each aircraft was available: Engine 1 Torque, Engine 2 Torque, Rotor Speed, Roll Angle, Pitch Angle, True Airspeed, Radio Altitude, Vertical Speed, Normal Acceleration, Density Altitude, Tail Rotor Torque, Main Rotor Torque, Roll Rate, Pitch Rate, Yaw Rate, Longitudinal Acceleration.

It has been hypothesised that the accelerometric measurements are in some extent influenced by the environmental state of the aircraft. In order to test that hypothesis, **canonical correlation analysis** has been applied on the available data set.

The canonical correlation method describes the interconnection between two random vector variables by means of a double set of latent variables (directions in the corresponding state vector spaces). Those latent variables reproduce the structure of the correlations between the “physical” observed variables of different groups, minimising in the meanwhile the impact of the correlations between variables in the same group. These latent variables are called canonical components and are ordered according to the magnitude of the common eigenvalues of certain matrices, which has been defined by Hotelling in (Hotelling, 1936). The observable parametrisation of the physical vector states of the variables in the groups can be replaced by a more synthetic one, which is obtained in terms of projections in the directions determined by the canonical components. The linear correlations established between the latent variables, constructed in such a way, are called **canonical correlations** of the model. As an example, the list the canonical correlations obtained by analysing the interconnections between the environmental state vectors and the vector states a TGB gear is displayed below (values of the canonical correlation coefficients close to 0 indicate low correlation, values close to 1 indicate high correlation between canonical variables):

$$\begin{array}{ll}
\rho(a_1b_1) = 0,99999838 & \rho(a_2b_2) = 0,74719544 \\
\rho(a_3b_3) = 0,60608554 & \rho(a_4b_4) = 0,47571818 \\
\rho(a_5b_5) = 0,39483775 & \rho(a_6b_6) = 0,37293685 \\
\rho(a_7b_7) = 0,26062950 & \rho(a_8b_8) = 0,15779505 \\
\rho(a_9b_9) = 0,13292464 & \rho(a_{10}b_{10}) = 0,10704979 \\
\rho(a_{11}b_{11}) = 0,06135586 & \rho(a_{12}b_{12}) = 0,02884099
\end{array}$$

In each of the analysed cases the first canonical correlation is extremely high. This fact, considered the high number of dimensions, can be considered as accidental. More relevantly it has been observed that many components are characterised by three or four canonical correlations with considerably high values (over 0,5). This fact is much more meaningful with respect to the interrelations between the environmental vector state and the component vector state. Unlikely, in some cases (Hangar Ball Brg) the canonical correlation profile is characterised by very low second canonical correlation.

The existence of relevant multi-correlation between the aircraft states and component states led us to the construction

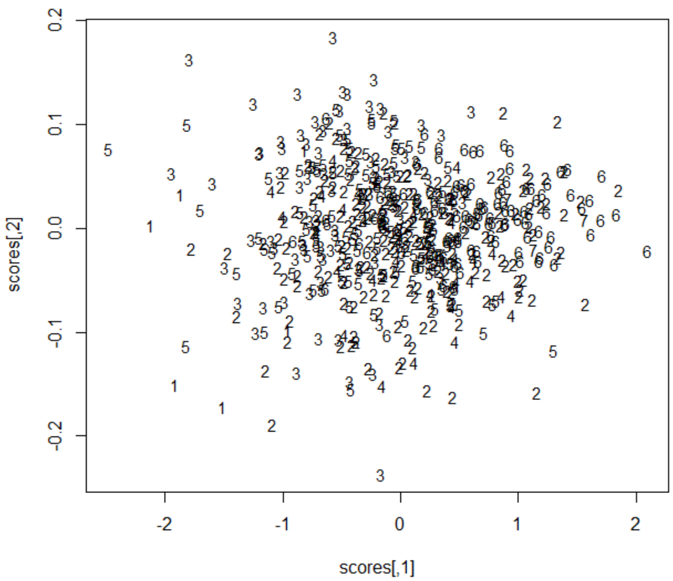

Figure 2. PCA scores of healthy operational states of TRDS of the same six helicopters after linear re-calibration. Again vector states measured on individual helicopters are labelled by different numbers.

of what has been called a **multilinear filter**. A liner map $f : R^{17} \longrightarrow R^N$ (where N is the dimension of the component vector) which provides a "predicted" component vector state in correspondence to each environmental state has been calibrated. The k-th row of the matrix associated to this linear map (with respect to the canonical basis of physical variables of the state vector space) represents the coefficients of a multiple liner regression of the k-th component of a state vector over the set of environmental parameters. The calibration is done in healthy conditions and the analysis is then performed in terms of residuals with respect to the predicted value.

If the reader compares Fig. 1 to Fig. 2, will observe that as a consequence of re-calibration, scores of healthy operational states measured on different helicopters slightly concentrate (compare the scales of the diagrams) and mix together quite uniformly. Similar effects are observed for all the mechanical components, for which the canonical correlation analysis reveals considerable level of linear correlation. Linear re-calibration makes vector states measured on individual helicopters of the same type **comparable**. A specific situation on an aircraft can be compared to analogous situation on another aircraft.

One of the standard anomaly detection tools in multivariate statistics, based on the statistically relevant Mahalanobis distance, is the so called multidimensional Shewhart control chart (we refer the reader to (Shewhart, 1931) and (Shewhart, 1986)). Control charts are based on an evaluation of the likelihood on a single event in the context of a random process. Consider a vector space endowed with a probability distribution f and a sample (a process) of random vectors $(X)_i \in V$. As long as the sample vectors belong to regions where the probability density is judged sufficiently high, the process is considered **under control**, or **out of control** otherwise. Under certain symmetry assumptions on the probability distribution density f, control charts can be implemented as distance based statistical methods. A state X is considered out of control if it is "far enough" from the expectation value of the distribution of the ordinary regime of the process.

In a population characterised by a **multidimensional Gauss distribution**, the Mahalanobis distances from the mean value follow the $T_k^2(n)$ distribution. Moreover there is an exact correspondence between the $T_k^2(n)$ distribution and the Snedekor-Fisher variable F:

$$\frac{n-k+1}{nk} T_k^2(n) \cong F_{k,n-k+1},$$

which is exploited for inference purposes. This means that plausibility of a state is compared to a statistical significance level imposed on the values of the $F_{k,n-k+1}$ distribution. Distances which exceed the one corresponding to the significance level indicate a phenomenon which is very improbable under the hypothesis of being a manifestation of the ordinary regime of the process. For this reason such a state is judged as a modification of the process due to not accidental causes.

The normality of the distribution of the healthy states of the mechanical component is a necessary condition for the application of a Shewhart control chart. On Fig. 3 are displayed the scores of the unfiltered healthy operational states of a TGB gear of one helicopter with respect to the first two principal components. The reader can observe that the cluster of PCA scores is characterised by an asymmetric "tail" in the direction determined by the second principal component. The PCA scores of the healthy states of the same component after linear re-calibration procedure are displayed on Fig. 4. The first obvious consequence of linear filtering is that the shape of the cluster of PCA score components becomes more ellipsoidal (recall that level sets of the Gaussian distribution are ellipsoids).

The extent to which the filtered healthy operational states of each component of the power drive fit with a multidimensional Gauss distribution has been tested. This fact was verified by various multivariate normality tests like Kolmogorov-Smirnoff, Jarque-Bera etc. (see (Kolmogorov, 1936; A. Justel, 1997; C. M. Jarque, 1987)). It has been observed that the distribution of filtered healthy operational states of a component of a single helicopter can be considered as Normal with very high level of statistical confidence (p-value around 2×10^{-15}). Analogous behaviour was observed in all the analysed mechanical components.

The above results can be interpreted by saying that the linear re-calibration procedure **filters the deterministic impact**

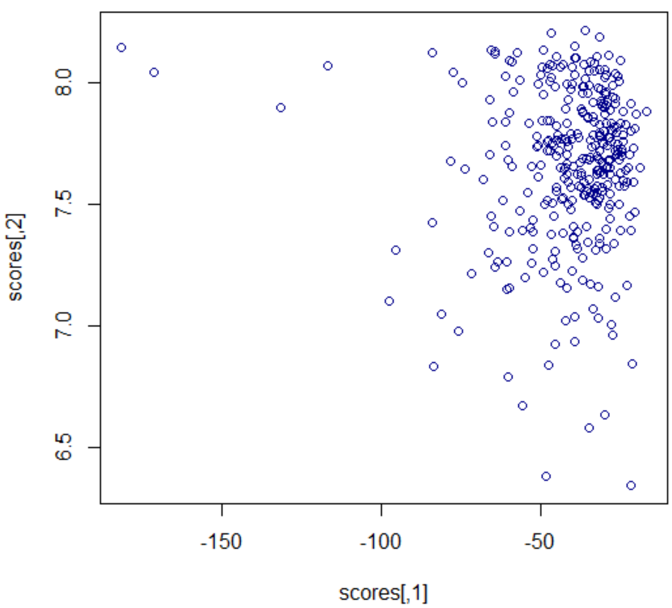

Figure 3. PCA scores of healthy operational states of the TGB gear of a single helicopter before linear re-calibration.

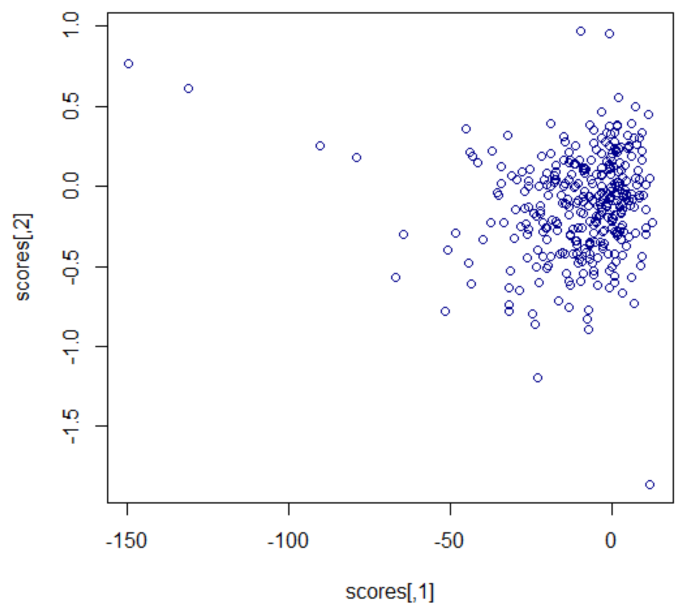

Figure 4. PCA scores of healthy operational states of the TGB gear of a single helicopter after linear re-calibration.

of the general state of the aircraft onto the accelerometric measurements. Once filtered the influence of the specific exploiting regime of the aircraft, the intrinsic variability of the healthy operational states of each mechanical component can be modelled over a random (white) noise process.

Fig. 2 illustrates the fact that analogous remark regards the set of filtered healthy operational states of the same component installed on different helicopters of the same type. They are normally distributed with roughly the same statistical confidence but with slightly higher variability.

Shewhart control charts have been calibrated on the set of healthy operational states of each mechanical component on a single helicopter. A small portion (less than 2%) of healthy vector states exceed the control limit. The same control chart was applied to healthy operational states of the same power drive, installed on other "twin" helicopters and bigger portion of states was judged out of control (15% for the Hangar Ball Brg). This means that even though linearly filtered data are used, there are still residual differences between the healthy regimes of components of different aircrafts. The same control chart has been also validated in the context of anomalous situations occurred on the same helicopter with very good results. In the case of Hangar Ball Brg roughly 73% of the anomalous states were judged out of control.

In conclusion, anomaly detection method based on a Shewhart control chart must be calibrated on each single helicopter. A software tool implementing a multivariate **self-learning** Shewhart control chart, which calibrates itself automatically on the healthy regime of a single mechanical component and highlights anomalous states, has been produced. The program computes automatically the upper control limit by means of a Gaussian approximation of the Fisher-Snedecor distribution.

In many cases (especially TRDS and Hangar Ball Brg) the Mahalanobis distance between states corresponding to false alerts and the mean value of the healthy regime exceeds the distance of the true damage states. For this reason the multivariate self-learning Shewhart control chart is an excellent tool for the detection of anomalous situations, but it is not sufficient for the discrimination of true failure states and anomaly alerts which do not correspond to a failure. Thus, additional discrimination statistical tools have been applied, as described later on. In the following sections of this article statistical models are calibrated and validated on filtered data.

The re-calibration filter can be made even more powerful by applying higher order regression of the health indicators over the set of environmental parameters of the aircraft. As an example, the reader can compare the previous canonical correlations of the linear filter of the TGB gear with the following canonical correlations of a quadratic multiple regression on the same component:

$$
\begin{array}{ll}
\rho(a_1b_1) = 0,9999989 & \rho(a_2b_2) = 0,8211778 \\
\rho(a_3b_3) = 0,7046168 & \rho(a_4b_4) = 0,6313677 \\
\rho(a_5b_5) = 0,5325972 & \rho(a_6b_6) = 0,4903959 \\
\rho(a_7b_7) = 0,4192670 & \rho(a_8b_8) = 0,4000518 \\
\rho(a_9b_9) = 0,3749749 & \rho(a_{10}b_{10}) = 0,3463869 \\
\rho(a_{11}b_{11}) = 0,2774180 & \rho(a_{12}b_{12}) = 0,2452053
\end{array}
$$

In conclusion the quite encouraging results obtained by linear re-calibration procedure can be even further improved.

4. MULTIVARIATE DISCRIMINATION METHODS

The linear re-calibration strongly reduces the differences between the healthy operational regime of power drives installed on aircrafts of the same type. This fact enables us to apply a set of standard multivariate statistical methods on a historical database of a fleet of helicopters. For a detailed description of those techniques the reader can refer to the following texts (Ferrell, 1979; Rencher, 2002; Timm, 2002; W. K. Härdle, 2012; Izenman, 2008).

In this study a particular geometric viewpoint on multivariate statistics has been adopted, as long as an Euclidean approach (or a more general metric geometry) provides some very useful intuitions on multivariate methods (see (Wickens, 1995) and (Epps, 1993)). We also refer the reader to (Tyurin, 2009), where a more intrinsic (coordinate free) geometric prospective on multivariate statistics is presented. In this context the analysis has been developed in terms of directions (random variables) and projections (magnitudes) onto relevant subspaces of the space of state vectors. Analogously but independently on the work exposed in (Gniazdowski, 2013) our approach interprets correlations as angles, but further radicalises this viewpoint by identifying statistical variables (both observable and latent-ones) in terms of real projective classes in a space of random vectors.

4.1. Structure of variance

The complete set of available states (healthy, true failures, false alerts) of each mechanical component was processed by Principal Component Analysis (PCA). This method is a direct implication of the Spectral Theorem in linear algebra. Principal components are the directions in the vector space of random variables, which maximise the variability of the data set. This technique highlights existing **spontaneous clusterings** in the variance structure of the data set. On Fig. 5 is displayed an example of scores of complete data sets on the subspace generated by the first three principal components.

A remarkable fact is that, after filtering, healthy operational states measured of many helicopters of the same type form a well-defined (green) cluster (see Fig. 5). Furthermore, there is an evident spontaneous clustering of the healthy and the anomalous true/false anomalous states. PCA leads to a consistent dimensional reduction in the space of states. Equations of linear and quadratic separation surfaces between the projections of the group clusters have been easily worked out and simple control methods have been based on the spontaneous clustering for each of the analysed mechanical components.

In the “critical case” of Hangar Ball Brg the projections on the subspace generated by the first and the second principal component do not reveal a significant clustering of the vector states. Nevertheless there is a relevant spontaneous clustering of the scores with respect to the second and the third principal components which was exploited in order to define discrimination conditions (see Fig. 6).

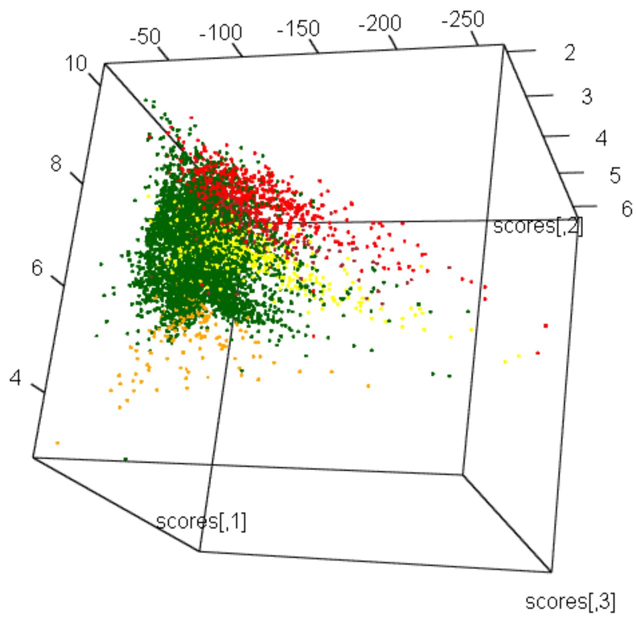

Figure 5. PCA scores of the states of a TGB Gear. Green dots represent scores of healthy operational states measured on 18 helicopters, red dots - true failure states measured on one of those helicopters, yellow dots - false alert on one of those helicopters.

On Fig. 7 are displayed PCA scores of a 2nd Stage Pin RH Brgs measured on a number of twin helicopters. The ordinary healthy operational states arrange in a very compact cluster. The set of blue dots represents a false alert occurred on one helicopter of the fleet. The yellow and the red dots represent anomalous states of the component measured on another helicopter of the fleet. In this case the chronological analysis of the data set led us to the following interpretation. An early fault (cluster of yellow dots) evolves towards a failure (cluster of red dots). The distinction between false and true anomalies is extremely sharp in this case and the direction in which the projections of true anomalous states spread in the space generated by the first three principal components is **indicative regarding the type of failure even before the definitive failure occurs**.

The structure of variance in the data sets has been further explored by applying multivariate discrimination methods like Liner Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) (see (W. K. Härdle, 2012)). The standard Fisher’s linear discriminant model is based on a linear transformation of the vector space, which maximises the differences between the transformed sample mean values of the distinguished groups. In other words LDA defines a new basis (a set of latent variables) such that the impact of the **between** component of the covariance matrix gets maximised at the

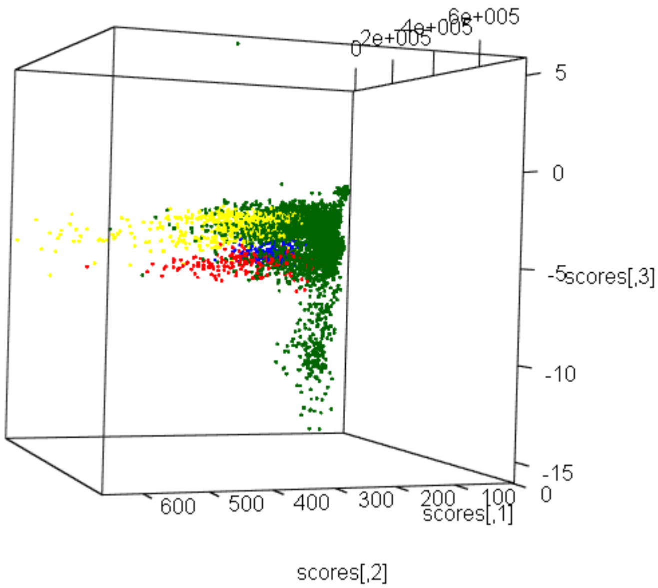

Figure 6. PCA scores of the states of a Hangar Ball Brg. Green dots are scores of healthy operational states from 18 helicopters, red dots - true failure states of one of those helicopters, yellow and blue dots are different false alerts on two helicopters.

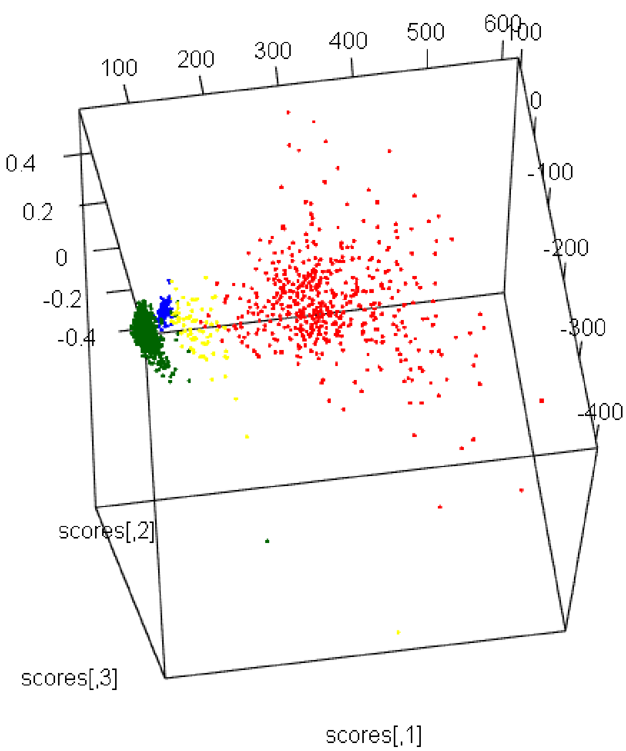

Figure 7. A 2nd Stage Pin RH Brgs fault and failure detection by means of PCA. Green dots represent healthy states measured on a fleet of helicopters, red dots - true failure occurred on one of those helicopters, yellow dots - fault states, blue dots - false alert occurred on one helicopter of the fleet.

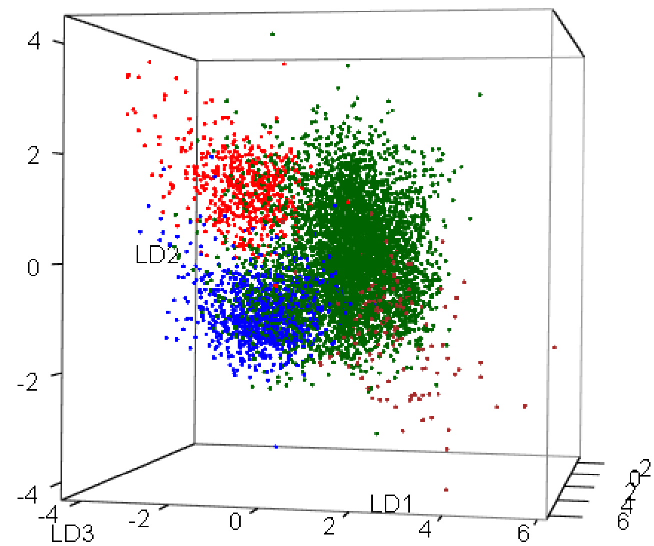

Figure 8. LDA scores of TGB Gear. Green dots are scores of healthy operational states measured on 18 helicopters, red dots - true failure states measured on one of those helicopters, blue dots - false alert on one of those helicopters.

Table 1. Leave-one-out LDA re-classification of 2nd Stage Pin RH Brg vector states

real \ classified as	false alert	healthy	true failure
false alert	74	0	0
healthy	1	1869	0
true failure	8	67	495

expense of the **within** component. The decision boundaries of LDA are linear affine subvarieties of the space of states.

The set of component state vectors has been divided into three groups, healthy operational states, false alerts and true failures. On Fig. 8 are displayed projections of TGB Gear states onto the subspace generated by the first three linear discriminant functions.

The calibrated linear discriminant models were validated by standard leave-one-out procedure using the complete data set of the fleet. On Table 1, and Table 2 are displayed some examples of LDA re-classification results.

There is a well-known quadratic classifier which exploits the minimisation of the Mahalanobis distance (with some corrections) from the mean vectors of the pre-assigned groups (see (Rencher, 2002)). In general QDA is a more flexible and precise method than LDA. Its decision boundaries are determined by the equality condition (equal probability) of the quadratic discriminant functions and are therefore (portions of) quadric hyper-surfaces in the space of states, typically ellipsoids or paraboloids. On Table 3 and Table 5 ere displayed some examples leave-one-out quadratic discriminant validation results.

Table 2. Leave-one-out LDA re-classification of Hangar Ball Brg vector states

real \ classified as	false alert	healthy	true failure
false alert	54	6	4
healthy operat.	29	1513	20
true damage	5	49	117

Table 3. Leave-one-out QDA re-classification of 2nd Stage Pin RH Brg vector states

real \ classified as	false alert	healthy	true failure
false alert	74	0	0
healthy	0	1860	10
true failure	0	0	570

The results obtained by both LDA and QDA leave-one-out cross validation are quite encouraging, especially because of the small portion of miss-classified true failure states. In the "critical" case of the Hangar Ball Brg both methods provide statistically significative number of correctly classified true failure states. This means that **true failure can be unambiguously detected**.

Several validation procedures based on splitting of the huge initial data set into calibration and validation data subsets have been applied in order to compare different helicopters of the same type. The results provided by the alternative validation methods are basically analogous to the leave-one-out and are therefore quite satisfying.

Both discrimination methods provide excellent results in the case of the fault and failure detection of the 2nd Stage Pin RH Brgs. The component states were divided into four groups (healthy/false alert/true fault/true failure) and the results of a QDA re-classification is displayed on Table 5.

4.2. Failure detection via canonical correlation

Canonical correlation analysis can be employed for detecting anomalies. Suppose that the healthy operative regime of a process is characterised by a strong correlation between vector variables X and Y. In such case one estimates the values of Y starting from known values of X by a suitable linear model. If Y assumes "unexpected" values i.e. its behaviour contrasts with the established correlation, this fact can be considered as a manifestation of some anomaly.

The reader can notice the analogy with the so called **consistency based anomaly detection** methods in which the deviations or inconsistences with a fixed functional model are considered as anomalies. In this study a multilinear model, which returns a state of a mechanical component as a function of the environmental parameters of the helicopter has been calibrated. The hypothesis that anomalous behaviour of a mechanical component is uncorrelated with the environmental data, i.e. is a manifestation of an inconsistency with

Table 4. Leave-one-out QDA re-classification of Hangar Ball Brg vector states

real \ classified as	false alert	healthy	true failure
false alert	60	2	2
healthy operat.	63	1430	69
true damage	2	33	136

Table 5. Leave-one-out QDA re-classification, fault and failure detection of a 2nd Stage Pin RH Brgs

real \ classified as	normal	false alert	fault	true failure
normal	1860	0	10	0
false alert	0	74	0	0
fault	0	0	75	0
true failure	0	0	0	495

the linear model was then tested. One would expect that the linear correlations between the environmental parameters and the components health indicators should decrease in presence of anomalous behaviour of the component. Therefore the data sets of healthy states and data sets containing anomalous states have been compared in order to establish whether the relevant (high) linear correlation coefficients decrease.

The situation which emerges from this procedure appears a bit chaotic. For the TRDS the linear correlation is very strong and the values of the coefficients drastically drop in mixed regime which contains true failure states. For the IGB pin the linear correlation is strong, the correlations in mixed regime get certainly worse, but monitoring of that component did not give evidence for real failures, so the measured anomalies correspond to false alerts. The TGB gear is characterised by relatively high values of the significant correlation coefficients and its mixed regime contains a true failure, but it seems that the second canonical correlation slightly improves in mixed regime.

In conclusion, for components for which the linear correlation with the environmental states is particularly high, the theoretical hypothesis is confirmed. This means that for those components the canonical correlation method can be considered as a supplementary anomaly detection resource. We expect that higher order filtering models as the one previously mentioned, will provide more unambiguous results.

4.3. Structure of covariance

In this study, a particular modification of the covariance matrix of the vector states of some mechanical components in case of anomalous measurements has been observed. The states of true damage are often characterised by increased correlation of certain vector components. The behaviour of the correlation matrix appeared slightly different in the case of false anomaly reports.

A possible explanation of this phenomenon could be given

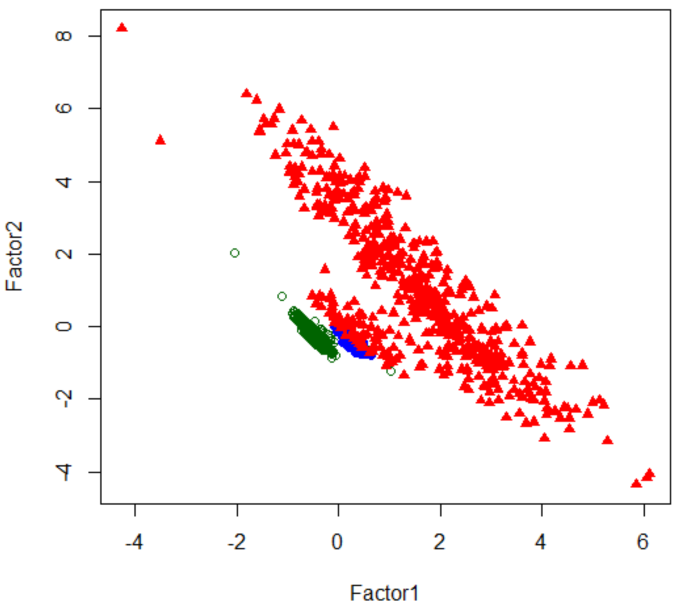

Figure 9. Bartlett factor scores of the 2nd Stage Pin RH Brgs. Green circles represent scores of healthy operational states measured on several helicopters, red triangles - true failure states measured on one of those helicopters, blue dots - false alert on one of those helicopters.

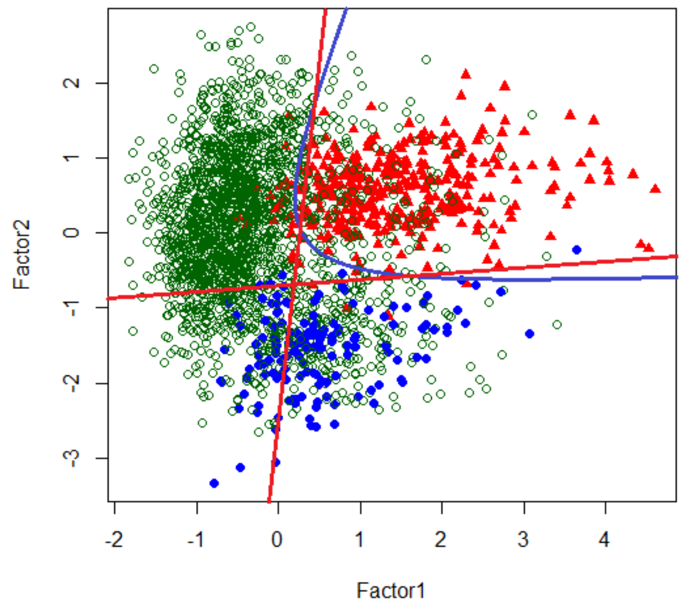

Figure 10. Bartlett factor scores of TGB Gear. Green circles represent scores of healthy operational states measured on 18 helicopters, red triangles - true failure states measured on one of those helicopters, blue dots - false alert on one of those helicopters.

if, in the case of true failure, different health indicators react simultaneously in a consistent and correlated way (failure states provoke an enhancement of certain elements of the correlation matrix). On the contrary false alerts can be interpreted as anomalous measurements not necessarily induced by a consistent reaction of the monitoring system.

The main purpose of the so called *factor analysis* consists of describing the structure of the correlations of a set of random variables by means of a small number of underlying uncorrelated latent variables called **factors**. In such sense it is analogous to the methods of principal component analysis, in which the structure of the variance in the sample is described by dimensional reduction. In this case the aim is obtaining a significant description of the structure of the covariance in the multivariate statistical sample in suitable subspace.

A compact multidimensional version of the defining equation of a factor model is:

$$X = \mu + \Lambda F + U,$$

where X denotes a k-dimensional vector random variable, Λ is a $k \times m$ matrix and U is the vector of specific factors. The matrix Λ is called the loadings matrix of the model.

The columns of Λ have an immediate geometric interpretation, they represent vectors which detect the directions of the latent factor multivariate variables. The vector variable F is nothing else but the k-uple of the projections (components) of the physical vector state X along those directions. In other words X is decomposed in certain relevant directions and its projections represent magnitudes of new variables. The vector F can be itself considered as a random m-dimensional vector variable.

The above expression only apparently resembles a multivariate linear model, in fact care must be taken as the whole expression in the second term of this equation is based on latent i.e. unobservable variables.

In this case standard recursive methods for the calibration of factor models have been applied and canonical factor models have been defined on the set of state vectors. Typically the calibration of factor model based on two factors was possible (the calibration procedure converges), but in some cases as the one of the Hangar Ball Brg, the iterative procedure does not converge with two but with three factors.

In terms of projections onto the space generated by the principal factors, the theoretical hypothesis translates in the following way. One could expect that the projections of the healthy operational cluster (near by the origin) and true failure cluster (away from the origin) onto the subspace generated by the principal factors show different characteristic profiles. The direction in which failure states projections spread away from the origin is indicative regarding the correlation modifications introduced by the simultaneous reaction to a damage. The shape of the cluster of healthy operational states characterises the intrinsic covariance structure of the component. In this context we expect that anomalous or false alerts should reveal some sort of irregular behaviour.

On Fig. 9 and Fig. 10 are shown the projections of the states of the 2nd Stage Pin RH Brgs and the TGB gear. Relevant clustering is rather visible in both cases. Projections (factor scores) of true failure states spread away from the origin in a direction, which is characteristic for the modified covariance structure. The investigation based on in-service data substantially confirmed the theoretical hypothesis. It is easy to work out linear or quadratic decision boundaries on factor scores as those displayed on Fig. 10.

In the case of Hangar Ball Brg the factor scores of the healthy operational states concentrate again near by the origin and the anomalous states spread far from it. Nevertheless these projections do not reveal a striking separation between true and false alert states.

In conclusion, for some mechanical components the covariance structure of the vector data set provides further useful resources for defining discriminant procedures.

5. PROJECTIVE STRUCTURE OF DATA SETS

Random variables has been interpreted as real projective classes in a vector space. From this viewpoint it is natural to hypothesise that the correlation structure of the data set can be better understood in terms of directions of the state vectors. In this context the module of a vector state plays a minor role and a direction in a vector space can be identified by a unit vector. In order to test this hypothesis, an original ”experiment” has been performed. Normalised state vectors has been considered, the set of N-dimensional vector states arranges over an $(N-1)$-dimensional sphere and factor models on the set of unit vector states have been calibrated.

An obvious effect of the spherical re-definition is a sort of compactification of the operational state clusters (Fig. 11). The hypothesis on the characteristic variations of the covariance structure appears rather plausible. In fact points representing healthy operational states and true damage situations form well-defined compact clusters.

Remarkably, as a consequence of this original procedure, the discrimination between true and false alerts becomes much more striking (compare Fig. 11 to Fig. 9). In this new situation the definition of the linear discriminant conditions appears even easier and more precise with respect to the previous factor models.

The typical behaviour of the unit states of a power drive component is that true damage states condense in a compact region inside the scatter-plot cluster of states. It is often easy to work-out a discriminant condition based on the affinity to that specific compact region. On Fig. 12 is shown the case of a TGB Gear.

Another considerable advantage of the normalisation of the vector states is the elimination of the large spreading of false anomalous alerts far from the mean value of the healthy operational regime. In this context LDA leads to precisely the same classification results, but remarkably QDA of the unit vector states of the “critical case” Hangar Ball Brg produces a slight improvement (compare Table 6 to Table 5).

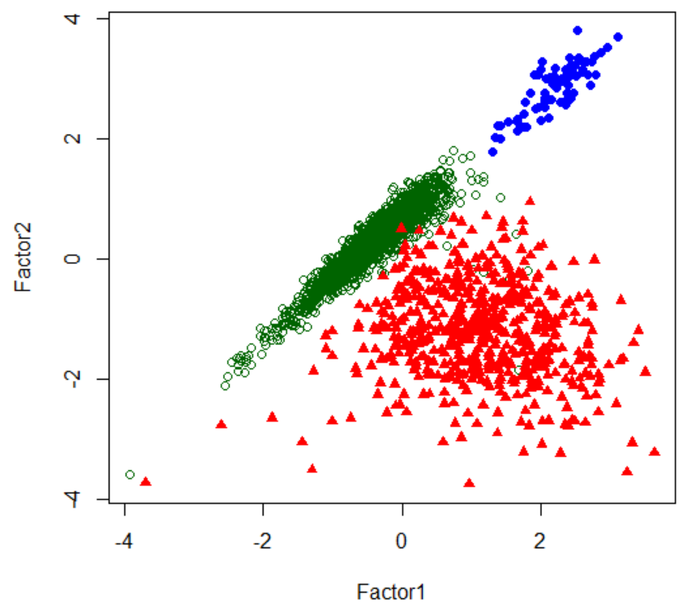

Figure 11. Bartlett type scores of unit states of a 2nd Stage Pin RH Brgs. Green circles are scores of healthy operational states measured on several helicopters, red triangles - true failure states measured on one of those helicopters and blue dots - false alert on one of those helicopters.

Table 6. Leave-one-out QDA re-classification of Hangar Ball Brg unit vector states

real \ classified as	false alert	healthy	true failure
false alert	60	2	2
healthy operat.	63	1432	67
true damage	2	33	136

In conclusion, this peculiar mathematical experiment led to interesting and in some cases unexpected, potentially useful results. The principal factor analysis on unit states gives further, often relevant, information on the anomalous behaviour of some mechanical components, and can be therefore integrated in a control procedure.

6. IMPLEMENTATION

The statistical techniques tested over the available vector data set are based on different mathematical constructions. They provide different and therefore not overabundant results. For this reason the above techniques have been combined in a software implementation of an integrated control process in the following way:

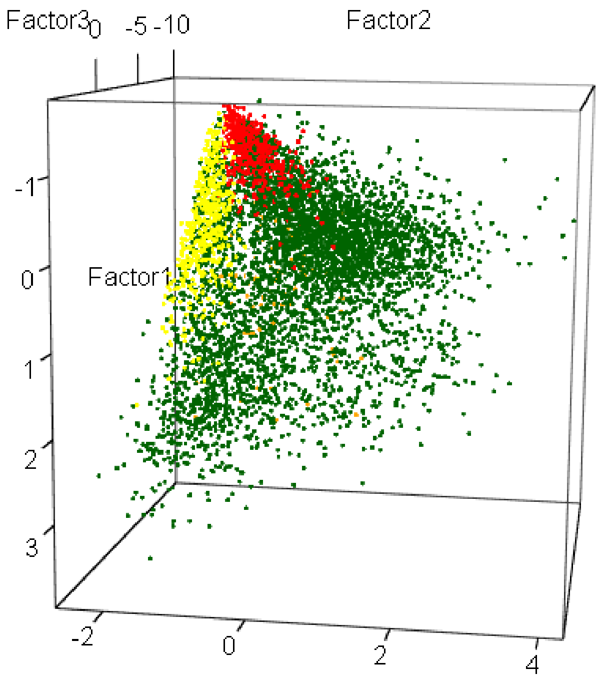

Figure 12. Bartlett type scores of unit states of a TGB Gear. Green dots represent scores of healthy operational states measured on 18 helicopters, red dots - true failure states measured on one of those helicopters, blue dots - false alert on one of those helicopters.

1. Anomaly detection by means of a self-learning Shewhart control chart. A problem highlighted by the experts of AgustaWestland consists of the fact that the healthy operational regime of some power drives on certain helicopters is characterised by very high values of the health indicators. Such values would be considered as anomalous if compared to other helicopters or to some a priori fixed threshold values. This ambiguity is completely removed by the self learning individual calibration of the control chart. Any vector state judged in control contributes to the real time re-calibration of the control chart i.e. the control chart keeps learning.

2. Anomaly classification based on discriminant methods calibrated and validated over the entire fleet. A vector state judged as anomalous undergoes evaluation based on a set of different discriminant techniques which can regard both the variance and the covariance structure of the calibration data sets (PCA, LDA, QDA, factor scores). A state classified as false alert does not generate an alert.

3. Evaluation. For different power drives, different discriminant methods appear as more efficient. An integrated parallel application of all the calibrated discriminant method is more powerful discrimination tool than the individual application of any single technique. A pre-alert status is produced by a suitable combination of discriminant outputs. Such a combination is chosen in order to maximise the efficiency of the integrated control system.

The integrated control process was tested on a series of real cases contained in the historical database of AgustaWestland. In the cases of the TGB gear and 2nd Stage Pin RH Brgs the integrated discriminant method judges a state as true failure i.e. generates a pre-alert if **each** discriminant method classifies it as a true failure. With this requirement only 3% of the measured states were miss-classified. In the most difficult case of Hangar Ball Brg a pre-alert is produced in 13% of the healthy states, in 28% of the previous false alerts and in 65% of the true failure states. The current univariate version of the control system generates an alert if the values of the health indicators exceed the alarm thresholds in a fixed proportion (usually 2/3) in a number of consecutive measurements. In the integrated method **the density of true failure outputs** required for a failure alarm can be rigorously deduced directly from these last overall validation results. For example, in the case of Hangar Ball Brg 1/2 appears as a suitable proportion.

An engineering software tool, which implements both the control process and the calibration of the parameters of the control routine for each component of the monitored power drives, has been produced.

7. CONCLUSIONS AND FUTURE DEVELOPMENT

The study has highlighted the advantages of this third-level multivariate approach. An efficient control process is based on an integration of several classification techniques. Even in those cases in which true failures and false alerts show misleading univariate profiles, multivariate techniques are able to distinguish them with **very high level** of statistical confidence.

In view of the results obtained by this research, an integrated multivariate health monitoring system is **currently in phase of implementation into actual service** on two models of helicopters produced by AgustaWestland.

The **elimination of the deterministic influence of the environmental states** of the helicopter determines two huge advantages:

1. After filtering the individual behaviour of each power drive can be very faithfully modelled over a random noise process. **The a priori threshold-based anomaly detection was therefore completely replaced by self-learning Shewhart control charts which operate individually on each mechanical component on each aircraft**.

2. Filtering gives the possibility to compare rigorously vector states measured on individual helicopters in different flight conditions. Once guaranteed the homogeneity of the measured data, **powerful classification and discriminant models can be calibrated on historical data obtained from many helicopters**. These models are applied in the context of a

control and diagnostics process over the entire fleet of helicopters. When relevant new data are collected, the statistical models should be updated and improved by re-calibration on a larger and more detailed data set. **Once a precise anomaly gets observed and diagnosed on one aircraft of the fleet, it can be diagnosed elsewhere** by means of its specific multivariate health condition profile.

The analysis of the results of this research from the viewpoint of the a posteriori **prognostics and health monitoring validation** of a diagnostic system will be a very interesting task. This is an extremely relevant topic which concerns the evaluation of the efficiency of the constructed health indicators i.e. how exhaustively they describe the state of the mechanical component (an observability problem). The investigation shows that the univariate processing of the health indicators could provoke a loss of relevant information. The results of this work show that besides the obvious advantages of direct multivariate processing, there is an interesting possibility to define and apply multivariate health monitoring validation protocols which aim to improve the efficiency of each individual health indicator by minimising the overall loss of information.

Another possibility, quite worthy to be explored in future, consists of by-passing the phase of construction of specific condition indicators by adopting a completely **multivariate spectroscopic approach** to the processing of the accelerometric signals. The calibration/monitoring software engineering tool which has been built in the context of this work can be directly applied without any modification in the context of such an alternative approach.

REFERENCES

A.Bellazzi, et al. (2014). A multivariate statistical approach to the implementation of a health monitoring system of mechanical power drives. In *Proceedings of the European Conference of the prognostics and Health Management Society ISBN-978-1-936263-16-5.*

A. Justel, R. Z., D. Pena. (1997). A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters*, *35*(3), 251259.

B. Everitt, T. H. (2011). *An introduction to applied multivariate analysis with R*. Springer.

CAA-PARER-2011. (2012). *Intelligent management of helicopter vibration health monitoring data* (Vol. 01 Based on a report prepared for the CAA by GE Aviation Systems Limited; Tech. Rep. No. ISBN 978 0 11792 403 1). Civil Aviation Authority - Safety Regulation Group.

C. M. Jarque, A. K. B. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, *55*(2), 163172.

Epps, T. W. (1993). Characteristic functions and their empirical counterparts: Geometrical interpretations and applications to statistical inference. *The American Statistician, DOI:10.1080/00031305.1993.10475930*, *47*(1), 33-38.

Everitt, B. (2005). *An R and S-plus companion to multivariate analysis*. Springer.

Ferrell, B. L. (1979). *Multivariate analysis*. Academic Press (Probability and Mathematical Statistics).

Gniazdowski, Z. (2013). Geometric interpretation of a correlation. *Zeszyty Naukowe Warszawskiej Wyzszej Szkoly Informatyki*(9, Rok 7), 27-35.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*(28 (3-4)), 321-377.

Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer (Springer Texts in Statistics).

K. Liu, J. S., N. Gebraeel. (2013). A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE T. Automation Science and Engineering*, *3*(10), 652-664.

Kolmogorov, A. (1936). Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari*(4), 83-91.

Ramasso, E., & Gouriveau, R. (2010). Prognostics in switching systems: Evidential Markovian classification of real-time neuro-fuzzy predictions. In *Ieee international conference on prognostics and system health management, macau, hong-kong, jan. 12-14.*

Randall, R. B. (2011). *Vibration-based condition monitoring - industrial, automotive and aerospace applications*. John Wiley and Sons, Ltd. Publications.

Rencher, A. C. (2002). *Methods of multivariate analysis*. Wiley Interscience (Wiley Series in Probability and Statistics).

Shewhart, W. A. (1931). *Economic control of quality of manufactured product.* David Van Nostrand.

Shewhart, W. A. (1986). *Statistical method from the viewpoint of quality control*. Dover Publications.

Timm, N. H. (2002). *Applied multivariate analysis*. Springer (Springer Texts in Statistics).

Tyurin, Y. N. (2009). Multivariate statistical analysis: A geometric perspective. *arXiv:0902.0408*.

Wickens, T. D. (Ed.). (1995). *The geometry of multivariate statistics*. Lawrence Erlbaum Associates, Incorporated.

W. K. Härdle, L. S. (2012). *Applied multivariate statistical analysis 3rd ed.* Springer.