USING DETECTION INDICES (DI) TO DETECT AIR VEHICLE CHARACTERISTIC CHANGES FOR HUMS APPLICATION

Eric C.J. Lee  Graham F. Forsyth
Defence Science and Technology Organisation
Cees Bil
RMIT University

Keywords: Detection Indices; Autocorrelation; Cross-Correlation; and SmartHUMS Unit

Abstract

This paper presents the application of ‘Detection Indices’ (DI) to monitor the health and usage aspects of air vehicles. The two DI described in this paper utilise ‘Autocorrelation’ and ‘Cross-Correlation’ algorithms. The primary function of the autocorrelation process in this context is to precondition the raw data segments obtained from the monitored data stream into a format that allows accurate comparison between two consecutive data segments. The second type of DI is the cross-correlation. The main emphasis of the cross-correlation analysis is to verify if differences exist between the two compared autocorrelated data sets, thus indicating whether changes have occurred in the characteristics of the vehicle being monitored. The described DI will eventually be imbedded in a miniaturised HUMS unit, called SmartHUMS, currently under development by the Defence Science and Technology Organisation (DSTO) in cooperation with GPS Online Pty Ltd.

A number of experimental results obtained by the preproduction SmartHUMS unit are presented in this paper. The experimental test setups used for the experiments consist of a bench top electric motor driven test rig and a two-stroke model helicopter engine driven experimental test rig. During the bench top electric motor and model helicopter engine experiments, artificial disturbance was introduced to demonstrate the DI algorithm’s ability to detect the disturbance. This paper presents the result for each of these experiments and show that the proposed DI can be used to create a low-cost HUMS solution.

1 Introduction

The concept of ‘Health and Usage Monitoring Systems’ (HUMS) is relatively new in the field of aerospace engineering. It is quite common for many aerospace engineers to misunderstand or to have not even heard of the terminology ‘HUMS’, especially in the field of fixed-wing aircraft. The application of HUMS technologies is generally agreed to have started with the rotorcraft community because helicopters have a higher rate of mechanical failure accidents, and are much more vulnerable to catastrophic mechanical failures than public transport or fixed wing aircraft [1]. The higher helicopter accident rate is simply because of the higher number of single load path critical parts within the rotor and transmission systems and the reduced redundancy within the helicopter design [2]. In order to decrease the failure instance rate, equipment capable of detailed monitoring of different critical helicopter functions is routinely fitted to medium and larger sized helicopters used by civil and military operators. The combination of this equipment forms a system that is generally referred to as ‘Health and Usage Monitoring System’.

Currently, health (vibration) monitoring systems have been made mandatory in UK on large helicopters certified or validated since certification requirements were tightened by the CAA following the HARP report [3]. An additional airworthiness directive in 1999 also
made health monitoring systems mandatory in the UK on older types of helicopter carrying more than 9 passengers [3]. The main reason for only large helicopters being fitted with HUMS is mainly due to the cost issue. Larger helicopter generally cost more and show more scope for financial benefits of improved reliability. Most helicopter operators will only consider to installing HUMS in their fleet only if the system would provide significant economic benefits that would outweigh the costs in the short term. Also majority of benefits for HUMS implementation is distributed over the remaining life of the aircraft, which is why older helicopters are less likely to be considered for HUMS installation than helicopters that are about to enter service.

The physical size and cost are other reasons why HUMS is rarely considered for small helicopters and small fixed-wing aircrafts. In a small aircraft the payload dimension and weight are critical factors. Unfortunately HUMS generally have noticeable size and weight, as well as being generally too expensive to be justified to fit into small aircraft, which might cost less or equivalent to the cost of the HUMS itself. Take a medium or small UAV as an example, to install HUMS in these types of UAV is quite often physically impossible and financially impractical. As a result, the main emphasis of this paper is to investigate a novel approach that will allow the realisation of HUMS benefits with significantly lower cost and physical dimensions.

To achieve the research aims, this paper demonstrates the application of utilising SmartHUMS unit (miniaturised HUMS system) and DI algorithms for size and cost consideration. Although the experimental result discussions are based on test rig systems that were designed to mimic a small UAV propulsion system, it does not necessary mean the final SmartHUMS system cannot be applied to other mechanical systems or platforms. The main reason why small UAV designs were targeted is because the successful application of SmartHUMS unit in a small UAV will help the realisation of the practicality of HUMS technology in smaller and less expensive fields of mechanical systems. Additionally because UAVs are pilot-less, HUMS technologies are a way to ensure the continuous safety of a UAV over a populated area.

The discussion of this paper will be focusing on the DI aspects of the research. The SmartHUMS hardware design and development are mainly performed by the GPS Online under the guidance of DSTO. A preproduction SmartHUMS hardware unit has been produced. This preproduction unit is currently being utilised to help the investigation of the intended DI algorithms.

As mentioned in [4], HUMS data are collected with the purpose of recording all important events and activities for future analysis. However, review and analysis of these data are typically ad hoc, relatively infrequent and require significant human involvement. As a result, data may accumulate much faster than they can be processed. As large portions of the HUMS data are of little significance, the proposed DI algorithms need to be able to isolate the vital data during the monitoring process.

The major difference between the proposed DI algorithms and the algorithms used by the conventional HUMS unit is the diagnosis methodology. While conventional HUMS use algorithms that specifically look for individual faults (or faults in individual gears, bearings, etc.), the DI techniques described in this paper will look for faults in terms of changes in transfer functions. Which means, for example, a conventional HUMS will only detect a structural crack if an algorithm to detect that crack is included, while the SmartHUMS, with the imbedded DI, would detect the crack as long as it affected the transfer of any significant signal. The development of the SmartHUMS unit is not intended to replace any existing HUMS system. In the contrary the SmartHUMS research is aiming to extend or assist current HUMS technology, in order to introduce HUMS benefits into disciplines which are previously thought to be financially impossible or physically impractical to be applied.
2 DI Background

The two DI algorithms been investigated in this paper are: Autocorrelation (sometimes called serial correlation), and Cross-Correlation.

2.1 Autocorrelation

According to [5], time series data sometimes show repetitive behaviour or other properties where current values have some relation to the earlier values. Autocorrelation is a statistic that measures the degree of this affiliation. The ability of autocorrelation to determine changes to otherwise regular patterns sets an excellent backdrop for the DI application. If, during the monitoring of a mechanical vehicle, a difference is detected between the behaviour of the current data from that relating to the previous period, the raw data during both period is stored and compressed for further analysis. The autocorrelation technique has two most significant parameters, which are the time series data length and the lag amount. Essentially the lag amount is the parameter that allows the comparison of the time series to itself. If the lag amount is equal to 1, the time series data is being compared to itself shifted by one data point at a time.

The other advantage of using autocorrelation as a DI is that it has the capacity of detecting periodic patterns even in the presence of random data (noise). If the time series contain large amount of noise, the autocorrelation process will still be able to present the periodic patterns by filtering out most of the noise.

The general mathematical expression for autocorrelation function is commonly described as [6, 7]:

\[ R_x(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt \]  

(1)

where T is the record length, \( R_x(\tau) \) represents the value of the autocorrelation function at the time delay \( \tau \), \( x(t) \) represents the value of the signal \( x \) at time \( t \), and \( x(t+\tau) \) is the value of the signal \( x \) at delayed time \( t+\tau \). In terms of discrete time, Eq. 1 becomes:

\[ R_m = \frac{1}{N-m} \sum_{t=1}^{N-m} (x_t - \bar{x})(x_{t+m} - \bar{x}) \]  

(2)

where \( N \) (sample size) is the approximation of \( N-m \) (the difference between \( N-m \) and \( N \) is in fact negligible in most cases), and \( m \) is the delay value called lag. Introducing \( \bar{x} \) (mean of entire time series) into Eq. 2 gives:

\[ \text{Autocovariance} = \frac{1}{N} \sum_{t=1}^{N-m} (x_t - \bar{x})(x_{t+m} - \bar{x}) \]  

(3)

Autocovariance is one of the two major components in the formulation of the autocorrelation coefficient function for a given lag value. According to [5], autocovariance literally means, “How something varies with itself”, where a time series gets compared to itself and the main tool in the system is the lag. It is a quick way of evaluating deviations between the one unaltered time series and one that is lagged, as shown in Figure 1. When generating autocovariance there are two rules of thumb [8]. The first rule is that the data set should contain more than 50 values. The second rule is the largest lag for the autocovariance calculation is equal to one quarter of the total number of values in the data set.

![Fig. 1 Time Series (solid), lags (dashed)][5]

The second ingredient for the autocorrelation coefficient for a given lag is called variance and it is obtained by standardising Eq. 3 the autocovariance equation, therefore it can then be compared directly to other standardised autocovariances [5]. The equation for variance is basically the sum of the square term \( (x_t - \bar{x})^2 \) for each observation in the original time series, divided by \( N \):

\[ \text{Variance} = \frac{1}{N} \sum_{t=1}^{N} (x_t - \bar{x})^2 \]  

(4)

The second ingredient for the autocorrelation coefficient for a given lag is called variance and it is obtained by standardising Eq. 3 the autocovariance equation, therefore it can then be compared directly to other standardised autocovariances [5].
With the equation for both components known, the description for the autocorrelation coefficient for a given lag is basically the autocovariance divided by the variance as presented in Eq. 5:

$$R_{xx} = \frac{\sum_{t=1}^{N-m} (x_t - \bar{x})(x_{t+m} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}$$

Eq. 5 is one of the many forms that describe the autocorrelation coefficient approximation, also called the lag autocorrelation coefficient or the lag serial correlation coefficient. The autocorrelation coefficient values range between +1 to –1, with +1 meaning the time series compared are exact duplicates of each other, which also means the lag value is equal to zero, and –1 meaning the time series compared are mirror images of each other. Zero means the compared time series have no relation to each other, which basically means they are random.

2.2 Cross-Correlation

The cross correlation algorithm is a measure of the similarities and shared properties between data series. The arithmetic aspect of cross correlation is very similar to that of the autocorrelation. The only difference is the variable composition. In autocorrelation there is only one series to deal with, but in cross correlation there are usually two data series. The two data series can be any type of series for example related, non-related, or even identical (in such a case, it becomes an autocorrelation analysis). Once the cross correlation has been performed the association between the two data series will be revealed. Similar to the autocorrelation, the cross correlation results are often being described as a non-dimensional format. With the non-dimensional property, it is easier to compare the cross correlated results to other results obtained from different data sources. The non-dimensional cross correlation result is also known as the cross correlation coefficient. Like autocorrelation coefficients, the cross correlation coefficient values always lie between -1 and +1. +1 means 100% correlation in the same sense as autocorrelation analysis, -1 means 100% correlation in the reverse order (anti-phase), and 0 signifies zero correlation (means the series are completely independent of each other) or two completely randomised series.
Cross correlation method is the second DI technique investigated in this paper also in this research. The application of this DI is to assess the amount of the similarities between two autocorrelated data series, and use these information to decide whether a characteristic change has occurred for the platform in question.

Cross correlation is also a type of statistical analysis. The common mathematical expression for the continuous time cross correlation function is generally defined as [9, 10]:

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)y(t+\tau)dt$$  \hspace{1cm} (6)

As cross correlation is used to examine the common properties between two sequences of data series, it is required to move the sequences past one another entirely. This prerequisite is different from that of the autocorrelation, where the calculation only compute positive lags from 0 to +T to obtain all possible comparisons between the time series and itself. In the case of cross correlation if two different series are being considered as shown in Eq. 6, the negative lags of the correlation must be considered as well (i.e. incorporated all data from –T to +T). This process will ensure the entire length of one series to move pass the other series, hence all possible match positions are being scrutinised. From Eq. 6, \(R_{xy}(\tau)\) represents the value of the cross correlation function at the time delay (or lag) \(\tau\), \(x(t)\) represents the value of the series x at time t, and \(y(t+\tau)\) is the value of the series y at lagged time \(t+\tau\).

During the cross correlation analysis, if two data series are identical, the analysis procedure actually becomes very similar to that of autocorrelation analysis. The corresponding results in a cross correlation plot (i.e. cross correlogram) will be a mirror image of itself around lag 0, and with the highest amplitude (i.e. value of 1) at this point. The interpretation of the result in this situation should not be treated as the same as in the autocorrelation. Because in cross correlation one sequence is being ‘moved past’ the other rather than being lagged behind from a position of initial equivalence, it is therefore common to describe the successive comparisons as matched positions rather than lags.

Since the data series examined by the cross correlation are usually in discrete time domain, it is therefore much more convenient to describe Eq. 6 in discrete time as well. The discrete time domain expression for Eq. 6 is very similar to the discrete time domain of the autocorrelation function. The expression is shown in Eq. 7.

$$R_{xy}(m) = \frac{1}{N-m} \sum_{t} (x_{i} \cdot y_{i+m}) = \frac{1}{N} \sum_{t} (x_{i} \cdot y_{i+m})$$  \hspace{1cm} (7)

N is the series size which is also the approximation of N-m (the difference between N-m and N is in fact small and can be ignore), and m similar to the application of lag value in autocorrelation analysis, but in cross correlation, it is referred to as match position. In the summation term, the variable i is the representation of the time limit –T and +T, as mentioned in order to compare all possible position of the two series, cross correlation computation will started with the negative lags (the match positions that are less than zero or in the negative region) during the analysis.

$$r_{xy}(m) = \frac{\text{CrossCorrelationFunction}}{S_1 \cdot S_2}$$

$$= \frac{1}{N} \sum_{i} (x_{i} - \bar{x})(y_{i+m} - \bar{y})$$

$$= \frac{1}{N} \sum_{i} (x_{i} - \bar{x})^2 \cdot \frac{1}{N} \sum_{i} (y_{i+m} - \bar{y})^2$$

$$= \frac{1}{N} \sum_{i} (x_{i} - \bar{x})(y_{i+m} - \bar{y})$$

$$= \frac{1}{N} \sum_{i} (x_{i} - \bar{x})^2 \cdot \frac{1}{N} \sum_{i} (y_{i+m} - \bar{y})^2$$  \hspace{1cm} (8)

In order to allow the cross correlation solutions to be able to evaluate with other cross correlation results, cross correlation function in Eq. 7 needs to be normalised. The normalisation of Eq. 7 produced the cross correlation coefficient equation which is as shown in Eq. 8. Different to the autocorrelation standardisation procedure, in cross correlation the standardisation is done using the standard deviation (\(S_1, S_2\)) from both compared autocorrelated data sources as shown in Eq. 8. Also different to the autocorrelation
standardisation procedure, the cross correlation mean values \((\bar{x}, \bar{y})\) of both data sources are included in the equation to minimise the data calibration requirement for the comparison purposes.

Similar to the autocorrelation DI the easiest way to understand the characteristics of cross correlation coefficients is to plot them. The cross correlation coefficient plot is usually referred to as ‘Cross Correlogram’, which has the same amplitude range between +1 and -1 as the correlogram from autocorrelation. However, with the cross correlogram the horizontal axis contain parameters which are match positions rather than lag values. If the majority of the match positions shown high amplitude of coefficient and the cross correlogram shows high degree of organised cyclic patterns, which basically means the two compared autocorrelated data series have high correlation to each other.

High correlation in cross correlation analysis actually means both series shared large numbers of common properties and characteristics. If the maximum cross coefficient amplitude of 1 (in an ideal case) is achieved at the match position 0, and couple with both cross correlogram from – and + region are mirror image of each other. The two autocorrelated series in this case are very likely to be identical. In real life, however, cases of minor discrepancies for the properties mentioned are always to be expected. Figure 4 contains a cross correlogram representing the cross correlation analysis of two identical (ideal case) autocorrelated data series. Figure 5 presents the superimposed plot of – region and + region of the cross correlogram of Figure 4. As the plot shown in Figure 5, the – and + region of the plot is exactly identical, which means they are mirror image of one another.

In the case where two autocorrelated data series have no relation or shared properties (i.e. random) with each other, the corresponding cross correlogram plot is presented in Figure 6, where the amplitudes of the plot are almost equal to zero. Most importantly, the maximum amplitude does not occur at match position 0.

3 Experiments

To verify the HUMS capability of the proposed DI, two bench top experiments were conducted. Each experiment setup is driven by different types of motor. The first experiment setup was driven by an electric motor and the second experiment setup was driven by a two-stroke model plane engine (similar to many used by small size UAVs) as shown in Figure 7. In order to simulate a fault generation an off-balance CPU fan was attached right on top of SmartHUMS housing. At the 10th second of the
Fig. 7 Electric motor (left) and two-stroke motor (right) test setup

Experiment this fan was turned on, the vibration caused by the fan creates an independent vibration signal to the experiment setup. If the proposed DI algorithms are proficient enough they should pick up the disturbances caused by the fan, provided that the fan vibration is strong enough to interfere with transfer function of the test setup.

3.1 Electric Motor Experiment Setup

During the experiment, the test setup was running at a constant rotating speed of 600 RPM. 10 seconds into the experiment the off-balance CPU fan was turned on for about eight seconds. As the only disturbance for the entire experiment was the fan, the DI should only pick up two events, which are fan start up at around 10th second and fan shut down at about 18th second of the experiment.

Figure 8 presents the overlap of Z-axis (data source obtained from Z-axis vibration signal) autocorrelation correlogram plots for the first 10 seconds of the experiment. The first 9 seconds of correlograms are plotted in blue colour and the 10th second correlogram is plotted in red colour. Obviously the red correlogram is very different to the first 9 seconds of blue correlograms. This actually indicates that the fan was turned on during the 10th second. Figure 9 is the cross correlogram plot between the 10th and 9th seconds of the autocorrelated data sets. From the figure, the maximum amplitude at match position 0 is almost equal to 0, and from the Figure 10 zoom in overlap plot, it can be seen that the Figure 9 cross correlogram does not have a mirror image property about the vertical axis at match position 0. To conclude the finding, both DI have identified a characteristic change (start up of off-balance CPU fan) at 10th second of the experiment.
Figure 11 presents the overlap of Z-axis autocorrelation correlogram plots for the 18, 19, 20, and 21 second of the experiment. The correlogram for 18th second is plotted in red colour and the rest of the correlograms is in blue colour. Obviously the correlogram representing 18th second is different to the rest of the correlograms. 18th second is actually the time when fan was shut down.

Autocorrelation Z-Amplitude

-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0

Lag Amount

Fig. 11 Correlated correlogram

Cross Correlogram

-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0

Match Position

Fig. 12 Cross correlogram for 19th and 18th second

Overlap of – and + region

-0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0

Fig. 13 Overlap plot for – and + regions of Figure 12

Figure 12 presents the cross correlogram between 19th and 18th second autocorrelated data set. In this correlogram the evidence of fan disruption is even much clearer, where the maximum coefficient amplitude do not happen at match position 0. The superimposed plot of Figure 13 shows the cross correlogram definitely not mirror image of itself at the vertical axis of match position 0. Therefore, DI algorithms have detected the second event which was the shutting down of the fan.

3.2 Two-Stroke Motor Experiment Setup

In the case of two-stroke motor experiment, the experimental procedures were similar to the electric motor driven experiment. During the experiment the total data log time is 28 second, where the off-balance CPU fan was turn on around 10th second of the experiment and shut off after around 8 seconds. As in the case of the electric motor, it is expected that two events will be detected by the DI. The first event is the starting up of the fan, and the second event is the shutting down of the fan.

It is important to note that during the two-stroke motor experiment there were number of issues arose which caused significant impact to the vibration signals generated by the test setup. As the engine was designed for a radio-control helicopter, by stripping the engine from the helicopter and incorporated into the test setup the loads experienced by the engine are completely different. As a result, there was great difficulty in starting the engine, and once the engine was running, there were problems with fuel tank pressurisation and carburettor tuning due to different load requirements. All these issues caused the two-stroke motor to run in an inconsistent manner. In fact, the two stroke motor was running at gradually changing speed during the 28 seconds of data logging. At the time of this publication a significant amount of effort was still dedicated to calibration of the engine in order to achieve constant rotation for the experiment.

In order to make sense of the data logged, which definitely contain certain amount of inconsistent rotations, shorter comparison intervals had been used. For example, to detect the fan starting, the experiment results from 9th
second to 11th second were presented and compared, instead of analysing all 11 seconds of results which most likely included a number of data variations caused by the fluctuation of engine rotation.

![Autocorrelation correlogram for 9th (red) and 10th (blue) second](image1)

**Fig. 14 Autocorrelation correlogram for 9th (red) and 10th (blue) second**

![Autocorrelation correlogram for 10th (red) and 11th (blue) second](image2)

**Fig. 15 Autocorrelation correlogram for 10th (red) and 11th (blue) second**

![Cross Correlation](image3)

**Fig. 16 Cross correlogram for 11th and 10th second**

Figure 14 represents the overlap comparison between 9th (red) second and 10th (blue) second Z-axis experimental data in terms of autocorrelation correlogram. Generally, the phase plot between 9th and 10th second matched well except the amplitude variations. As the two-stroke motor experiment was a real life experiment, some random phenomena would be expected during the test. There were also some variations due to the carburettor calibration problems. As a result, the comparison in Figure 14 was treated as similar, which means that no variation was detected. Figure 15 shows the comparison between 10th and 11th second of Z-axis experiment data. In this plot, the phase misalignment is visible, additionally the amplitude differences are also observed. These characteristics described so far indicated discrepancies between 10th and 11th second data.

To prove 10th and 11th second experimental data are different the cross correlogram plot is presented in Figure 16.

The cross correlogram in Figure 16 shows low amplitudes, which basically means low correlation between 11th and 10th second data. Also the maximum coefficient amplitude did not occur at match position zero and the – and + regions of correlogram plot are not symmetric to each other. Thus the conclusion can be drawn that an event had occurred between 10th and 11th second of the experiment, caused by the fan start up procedure.

Figure 17 represents the comparison of autocorrelation correlogram between 16th and 17th second data. In this figure the amplitude differences are visible, which suggested random interruptions, but from around lag value 100 onwards the phases started to get misaligned.

To further verify the comparison between 16th and 17th second a cross correlogram is produced in Figure 18, where maximum amplitude did not occur at match position 0 and the cross correlogram plot did not have a mirror image about the vertical axis with respect to match position 0, as a result an event had been detected between 16th and 17th second. Figure 19 is the autocorrelation correlogram comparison for 17th and 18th second experimental data, where the two correlograms overlapped relatively well except some minor amplitude variations. Therefore, the CPU fan was shutting down at 16th second of the
experiment, which basically means the fan was not turn on for full 8 seconds as expected.

![Autocorrelation correlogram for 16th (red) and 17th (blue) second](image1)

**Fig. 17 Autocorrelation correlogram for 16th (red) and 17th (blue) second**

![Cross correlation](image2)

**Fig. 18 Cross correlogram for 17th and 16th second**

![Autocorrelation correlogram for 17th (red) and 18th (blue) second](image3)

**Fig. 19 Autocorrelation correlogram for 17th (red) and 18th (blue) second**

### 4 Conclusion

Two experimental test setup results have been demonstrated in this paper. In each case the DI algorithms were able to detect the off balance fan starting up and shutting down process. The ability to detect fan operation demonstrated the DI fundamental concept, where any event that causes significant impact to the system transfer function will be flagged by the DI algorithms. By only analysing these flagged data, significant diagnostic time can be reduced hence the whole system can be put back to operation sooner. As a result, the availability and reliability is increased, therefore, the operational cost will be reduced in certain degree and indirectly the safety aspects will also increased. The improvements mentioned are some of the typical HUMS benefits regularly seen from a HUMS implementation. Therefore, by applying SmartHUMS technology in the field such as UAV will likely to see these benefits being introduced as well.

### References


[3] Civil Aviation Authority. *What is HUMS, and what are its benefits*. Civil Aviation Authority, [http://www.caa.co.uk/docs/389/srg_acp_00012-01-300103.pdf](http://www.caa.co.uk/docs/389/srg_acp_00012-01-300103.pdf)


