DETECTION INDICES FOR A MINIATURISED HEALTH AND USAGE MONITORING SYSTEM UNIT

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

Keywords: *Detection Indices; SmartHUMS; Autocorrelation*

Abstract

This paper presents the application of Detection Indices (DI) in a miniature Health and Usage Monitoring System (HUMS) unit developed by Defence Science and Technology Organisation (DSTO) of Australia in co-operation with GPS Online Pty Ltd. The significance of using DI is to process the data onboard without going so far as to prevent more complicated algorithms being used later but far enough to reduced the amount of data storage required. A simulation analysis study was carried out to understand the characteristics of the autocorrelation process. The characteristics were then further verified by experimental procedures. The results indicated that the autocorrelation could be used to identify the occurrence of anomalies within a monitored system. This ability makes the autocorrelation an attractive contender for the DI algorithm.

1 Introduction

When a Mechanical Vehicle (Air, Land or Sea) is in operation, the ability to track and assess its state of health and usage is imperative. In many cases the vehicle might be engaged in a military or civil mission, where human lives may be involved either directly or indirectly.

The question is what needs to be tracked and assessed? A typical Mechanical Vehicle includes rotating components as well as the structural frame support. Many of the rotating components have lives that are limited by fatigue considerations, i.e. the component must be replaced when a pre-defined number of hours has been accumulated, generally referred to as the Component Retirement Time (CRT), and is dependent on fatigue strength, loads and the usage spectrum of the component. The usage spectrum is initially based on assumptions for the proportion of time a vehicle spends in various conditions. In practice, the actual usage of the vehicle may vary markedly from the assumed design usage [1].

The helicopter community has made a vast leap in developing technology for in-service tracking of component accumulated usage hour, usage spectrum, and critical health status. This technology is commonly referred as HUMS. As described in [2], HUMS has the potential to provide significant improvements in such areas as operational improvements, fleet management improvements, reliability and cost benefit improvements. These improvements are commonly described as an improvement for Life Cycle Costing and increase in reliability and maintainability.

For the past twenty years, HUMS technology has advanced very significantly, where the prognostic and diagnostic capability has been greatly improved [3]. However, one of the major drawbacks for HUMS today is the generation of vast amounts of raw data. In most cases raw data can only be converted into meaningful information at a ground-based station, and very often only a small portion of raw data are of any interest. As mentioned in [4], HUMS data is collected with the purpose of recording all-important events and activities for future analysis. However, review and analysis of this data is typically ad hoc, relatively infrequent and requires significant human involvement. As a result, the data accumulates much faster than it can be processed.

Since large portions of the HUMS data are of little significance, an algorithm needs to be developed to isolate the vital data during the recording process. This paper describes such algorithm as Detection Indices (DI). The basic concept of DI is to monitor incoming data, and if an abnormal event occurs, data sampled during this period will be stored. In addition, DI may be able to indicated general information about the incident.

The major difference between the proposed DI algorithm and the algorithm used by the conventional HUMS unit is the diagnosis methodology. While conventional HUMS uses 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. That means a conventional HUMS will only detect a structural crack if an algorithm to detect that crack is included while the SmartHUMS would detect the crack as long as it affected the transfer of any significant signal.

Three DI algorithms will be researched: autocorrelation (sometimes called serial correlation), cross correlation and signal averaging. For the purpose of this paper, only the autocorrelation method will be discussed.

According to [5] time series data sometimes have repetitive behaviour or has other properties, whereby 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 with the previous period, the raw data 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 that means the time series data is being compared to itself at a reduction of 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 when the random data exist within the time series. If the time series contains large amounts of noise, the autocorrelation process will still be able to present the periodic patterns and filter out most of the noise.

The HUMS unit used for this research is called SmartHUMS, which is a miniaturised HUMS co-developed by Defence Science Technology Organisation (DSTO) of Australia and GPS Online Pty. Ltd.

2 Theoretical background

The basic mathematical equation for autocorrelation is commonly described[6, 7] as:

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

where T is the record length, Rx(τ) represents the value of the autocorrelation function at the time delay τ, x(t) represents the value of the signal x at time t, and finally the x(t+τ) is the value of the signal x at time delay t+τ. Eq. 1 can be approximated by:

$$R = \frac{1}{N}\sum_{t=1}^{N-m}(x_t)(x_{t+m}) \qquad (2)$$

where N is the number of segments, and m is the delay value called lag. Introducing $\bar{x}$ (mean of entire series) into Eq. 2 gives:

$$\text{autocovariance} = \frac{1}{N}\sum_{t=1}^{N-m}(x_t - \bar{x})(x_{t+m} - \bar{x}) \qquad (3)$$

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

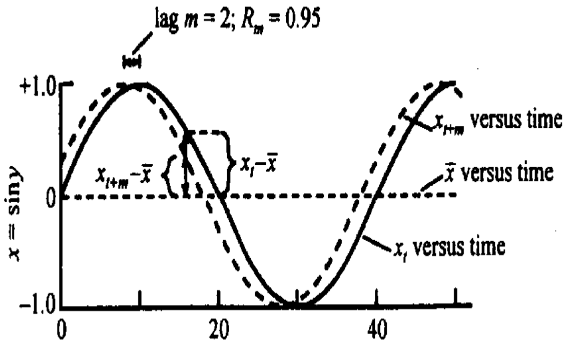

Fig. 1 Time Series (solid), lags (dashed)[5]

thumb [8]. The first rule is that the data set should contain more than 50 values. The second rule is that the largest lag for the autocovariance calculation is equal to one quarter of the total number of values in the data set.

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

$$\text{variance} = \frac{1}{N}\sum_{t=1}^{N}(x_t - \bar{x})^2 \qquad (4)$$

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:

$$\text{autocorrelation}\,(R_m) = \frac{\text{autocovariance}}{\text{variance}} = \frac{\frac{1}{N}\sum_{t=1}^{N-m}(x_t - \bar{x})(x_{t+m} - \bar{x})}{\frac{1}{N}\sum_{t=1}^{N}(x_t - \bar{x})^2} = \frac{\sum_{t=1}^{N-m}(x_t - \bar{x})(x_{t+m} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} \qquad (5)$$

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. A common way of analysing the autocorrelation coefficients and their respective lag values is by plotting the autocorrelation coefficient against the lags. The plot is called correlogram and is a comprehensive way to indicate the relationship between time series data. In the case where the time series have no relationship to each other, the correlogram will present an irregular pattern with amplitude close to zero, except when the lag is equal to zero, as shown in figure 2. In contrast, when the time series have a strong relationship, the correlogram will show high coefficient values and a regular pattern as shown in figure 3.

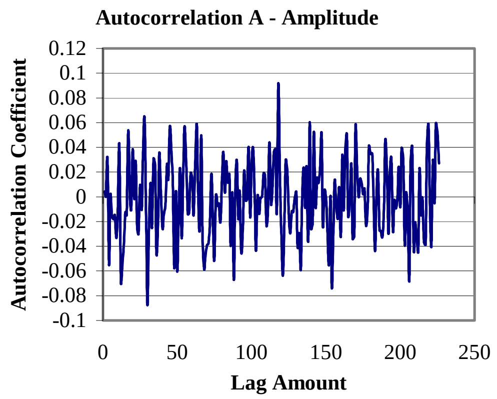

Fig. 2 Uncorrelated correlogram (random time series)

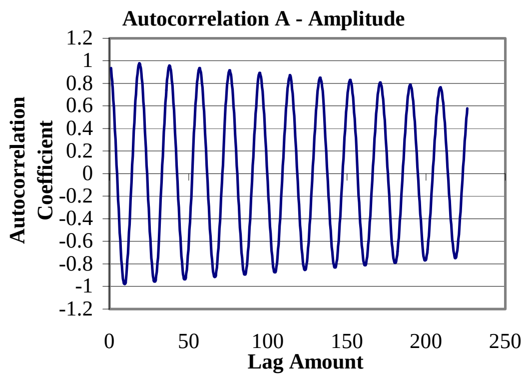

Fig. 3 Correlated correlogram

When a correlogram has been generated, the confidence of the result needs to be addressed. According to [9], for data with no trend and no correlated relationship, 95% of the coefficients theoretically fall within:

$$\text{95\% confidence limits} \approx \pm \frac{2}{\sqrt{N}} \qquad (6)$$

where N is the total number of values in the time series data set. It is important to note that there are still about 5 per cent of the coefficients that could exceed the confidence limits, and hence be uncorrelated. When plotting the 95% confidence limits on a correlogram, the confidence bands are two horizontal lines at constant value above and below zero.

Each successive autocorrelation coefficients, R_m, can be highly interdependent, where an autocorrelation coefficient is large simply because its previous lag value of autocorrelation coefficient R_{m-1} is large. This interdependency presents a difficulty in assessing how many of the lag values are actually significant within the correlogram. To overcome this difficulty, the large-lag standard error confidence bands [10] were used, where the confidence bands on the correlogram appear to be most narrow at lag 1 and slowly widen at higher lags. Formulation of the large-lag standard error described in [10] is as follows:

$$\text{standard error } (r_m) = \pm \sqrt{\frac{1}{\mathrm{N}}(1 + 2\sum_{t=1}^{M} R_t^{\,2})} \qquad (7)$$

where M < m. The term within the square root is basically the variance equation with an adjustment due to the summation term. The summation term in Eq. 7 is related to the sample size as well as the estimated autocorrelation coefficients at shorter lags. For example, with $r_{m=3}$ the summation term depends on the autocorrelation coefficients at lags 1 and 2 and the summation is over lags 1 to M, with M = 2 in this case. Figure 4 shows an example of a correlogram with large-lag standard error bands.

3 Simulation analysis

To understand the characteristics of the autocorrelation analysis and what kind of information can be extracted from it, this section presents a number of controlled time series data sets that were analysed by the autocorrelation process.

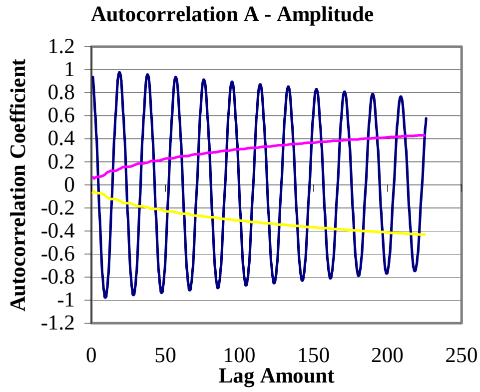

Fig. 4 Correlogram with large-lag confidence bands

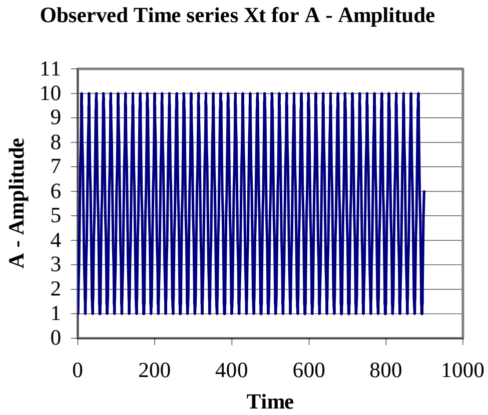

Fig. 5 Table 1 in time series

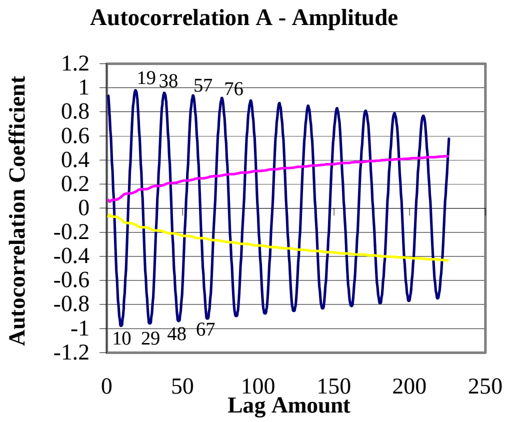

Fig. 6 Correlogram for Table 1

Assume the time series source is generated by a rotating component, e.g. rotating turbine blade, bearing, etc, at a constant speed, where a segment within the rotating component has a fault. If the fault remains unchanged, the time series data can be represented as shown in

Table 1 Time series data for an unchanged fault

Time	Data	Time	Data	Time	Data		Time	Data	Time	Data
1	*1*	20	*1*	39	*1*		875	*1*	894	*1*
2	*2*	21	*2*	40	*2*		876	*2*	895	*2*
3	*3*	22	*3*	41	*3*		877	*3*	896	*3*
4	*4*	23	*4*	42	*4*		878	*4*	897	*4*
5	*5*	24	*5*	43	*5*		879	*5*	898	*5*
6	*6*	25	*6*	44	*6*		880	*6*	899	*6*
7	*7*	26	*7*	45	*7*		881	*7*	900	*7*
8	*8*	27	*8*	46	*8*		882	*8*		
9	*9*	28	*9*	47	*9*		883	*9*		
10	*10*	29	*10*	48	*10*	....	884	*10*		
11	*9*	30	*9*	49	*9*		885	*9*		
12	*8*	31	*8*	50	*8*		886	*8*		
13	*7*	32	*7*	51	*7*		887	*7*		
14	*6*	33	*6*	52	*6*		888	*6*		
15	*5*	34	*5*	53	*5*		889	*5*		
16	*4*	35	*4*	54	*4*		890	*4*		
17	*3*	36	*3*	55	*3*		891	*3*		
18	*2*	37	*2*	56	*2*		892	*2*		
19	*1*	38	*1*	57	*1*		893	*1*		

Table 1. The data in Table 1 shows a sequence of numbers, increasing to a maximum value of 10 and then gradually reducing to 1. The assumption is that when the number reaches 10, a fault occurs. Figure 5 is the plot of Table 1 in the time domain and figure 6 is the correlogram plot of Table 1. When plotted in time domain, it is very difficult to determine whether the time series data are random or correlated, but when plotted as a correlogram, it is much easier to determine.

In figure 6 the lag value associated with the first four peaks and troughs is shown. Where the lag value of 19, 38, 57 and 76 corresponds to a minimum data value of 1 in table 1. Lag values of 10, 29, 40 and 67 correspond to the maximum data value of 10 (assumed fault data). In this particular case the troughs of the correlogram represent the fault of the rotating component. The lag period between each fault is 19, and in this example each lag value corresponds to 1 second. The period for each fault to occur is 19 seconds. If the starting position of the rotating component is known, the fault location can be immediately identified.

There are two types of fault, Non-critical and critical. Non-critical faults can be faults that do not cause immediate failure to the system. Critical faults are faults that will lead to immediate catastrophic failure, such as severe wear of the bearing balls or rapid crack growth on the inner or outer race of the bearing.

The controlled time series example in Table 1 is shown to be a non-progression fault (non-critical fault). The next controlled example will simulate the progression of a fault for a rotating system. The amplitude of faults in table 1 is assumed to be 10, but for this example the fault is assumed to increase in sequence with an increment of plus 0.2 (i.e. 10, 10.2, 10.4, 10.6 etc).

The resulting plot of the progressing faults is shown in figure 7 in the time domain, and figure 8 as a correlogram. When comparing figure 7 with figure 5, figure 7 shows a linear increase of peak values with a steady increment of 0.2. By overlapping figure 6 and figure 8 into figure 9, the trough's amplitude of the progressing fault (red line) is less than that of the non-progression fault (blue line), but the peak amplitude remains the same. Obviously for this example, the troughs represent the simulated faults, as expected, because the only change made in this case is the increment of fault values and the only differences between figure 6 and figure 8 are the amplitudes of throughs. The conclusion can be drawn that as faults are changing in size, the corresponding peak or trough amplitudes of the correlogram will vary.

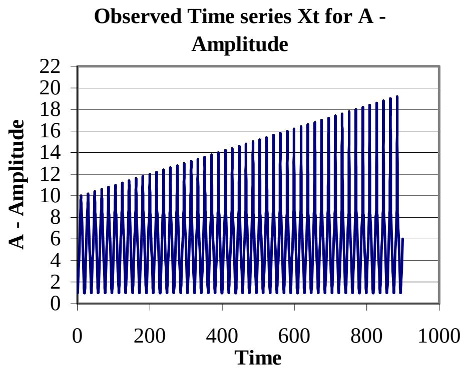

Fig. 7 Fault progression time domain

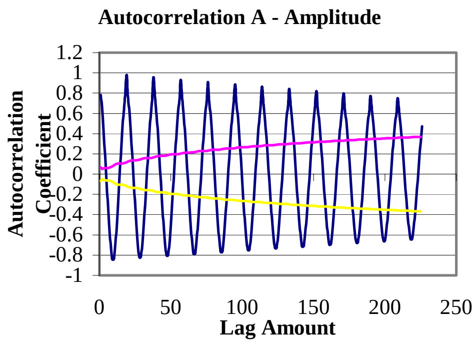

Fig. 8 Correlogram for fault progression

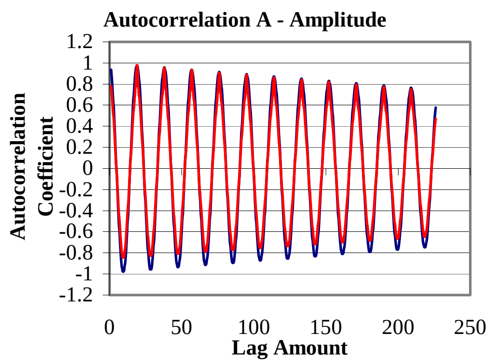

Fig. 9 Overlapping of figure 6 & 8

So far, both examples are simulated under the assumption of constant rotational speed. The following examples will demonstrate the effect of different rotational speeds and compare them to the original speed as shown in the first two examples.

In example 1 and example 2, the simulated time series is in sequence number 1 to 10 and then from 10 back to 1. For the first example the value 10 is assumed to be the fault and it remains constant. However, in example 2, the variation is the sequential increment of 0.2 for all the fault values. Now in the third example an increase in rotational speed is assumed, where the time series data is assumed to be in sequence of 1 to 9 and 9 to 1. For this example, value 9 is the assumed fault. The plot of the third example is shown in figure 10, where the red dashed line is the current example representation and the solid line is the original time series data.

The fourth example involves a slower rotational speed for the rotating component. In this case, the sequence number is increased to a value of 11, such as 1 to 11 and then 11 to 1. The number 11 is the assumed fault value in this time series set. The correlogram of example 4 is plotted in figure 10 as a green dashed line. By observing figure 10 and comparing the correlograms a conclusion can be drawn, which if the phase of each plot from different time periods are not in phase, then the rotating system behaviour has also changed. So far, the simulation analysis has covered the changes in amplitude and changes in constant rotational speed. In the next example changes in speed are simulated, but this time in a constant accelerated progression. In this fifth example, the time series data set will have different assumed fault values. The data sequence is 1 to 10, 10 to 1, 1 to 11, 11 to 1, 1 to 12, 12 to 1, 1 to 13, 13 to 1 and so on, such that the faults are now 10, 11, 12, 13 etc. The total number of data is the same as in Table 1, 900 data points.

Figure 11 shows the correlogram of example 5, where the pink line represents the constant accelerated progression time series and the blue line is the original rotational speed correlogram. The pink line shows a rapid reduction in amplitude and large shift of phases to the right when compared to blue line. As shown in example 4, rotational speed reduction will cause phase shift to the right, therefore example 5 is actually simulating constant deceleration.

Fault values in example 5 are different in each sequence and do not occur at the same interval due to the deceleration. The amplitude difference in figure 11 can only signify that there is variation of fault values in the system

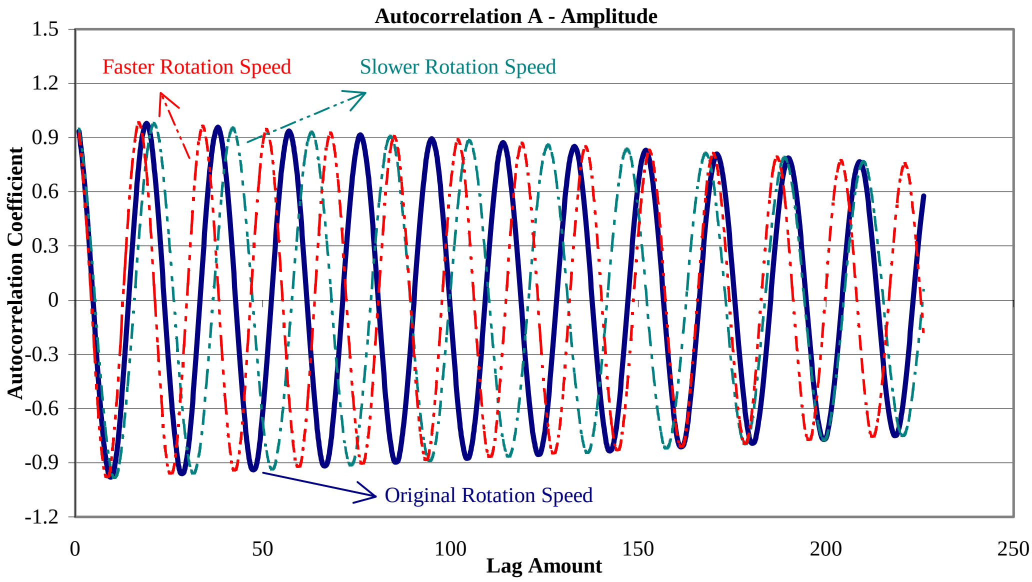

Fig. 10 Correlogram plots for different rotational speed

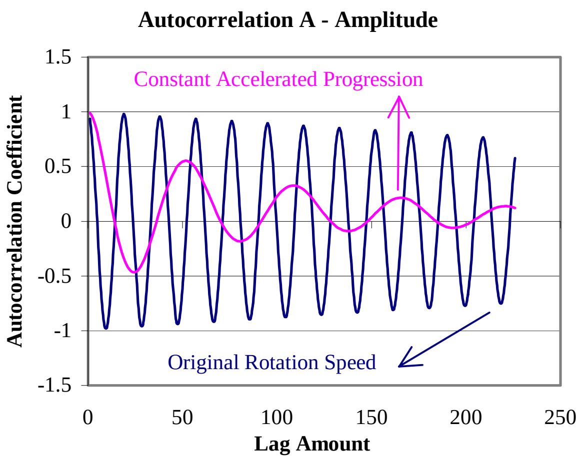

Fig. 11 Constant accelerated progression

and that they are appearing at different periodic intervals when compared to the original rotational speed plot.

4 Experimental analysis

The experimental set-up for the examination of autocorrelation as a DI algorithm consists of a prototype SmartHUMS unit and an electric motor driven test rig. The SmartHUMS unit has two internal sensors, which are a triaxial accelerometer and a microphone to measure the vibration and sound generated by the test rig. The test rig itself consists of an electric motor, a driving shaft supported by three bearings, and one end of the shaft is connected to a gearbox. Attached to the shaft are two circular discs, where screw fasteners can be screwed in to create unbalance in the system. The set-up of the test rig is shown in figure 12. Note that the autocorrelation DI algorithm has not yet been implemented into the SmartHUMS unit, therefore the unit is purely acting as a sensor at moment.

During the experimental process the electric motor runs at 770 RPM (12.8 Hz) for each test. There were five tests conducted, starting from no fastener on the circular disc, up to all 4 fasteners used. Each fastener weighs 2.7g. Each test lasted 30 seconds and after the raw data is obtained, it is then analysed by the autocorrelation process with lag amount of 1. Figure 13 shows the correlograms for sound and XYZ vibration data of all the tests. By examining figure 13, the periodic characteristic of the data can be easily observed, which also indicated that the raw data gathered is not random.

Figure 13 also shows a change in amplitude and phase shift. The trend of the phase shift is to the right as the weight on the disc is increased. As expected when the weight on the disc is increased, the unbalance in the system also increases. This then has a carry on effect on the rotational speed of the drive shaft.

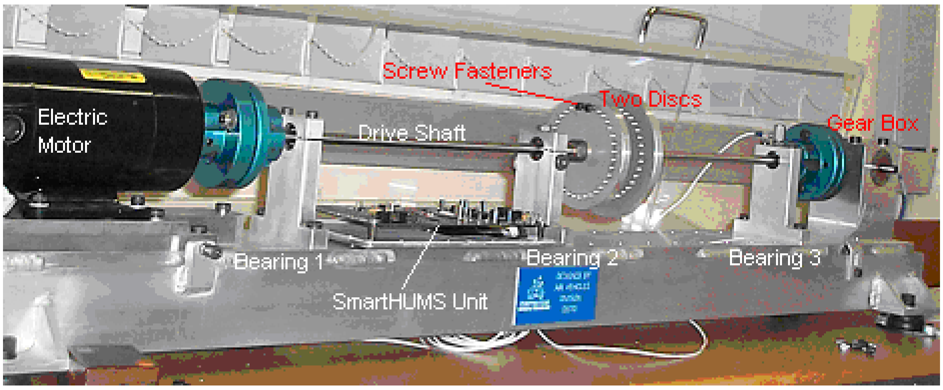

Fig. 12 Experiment set-up

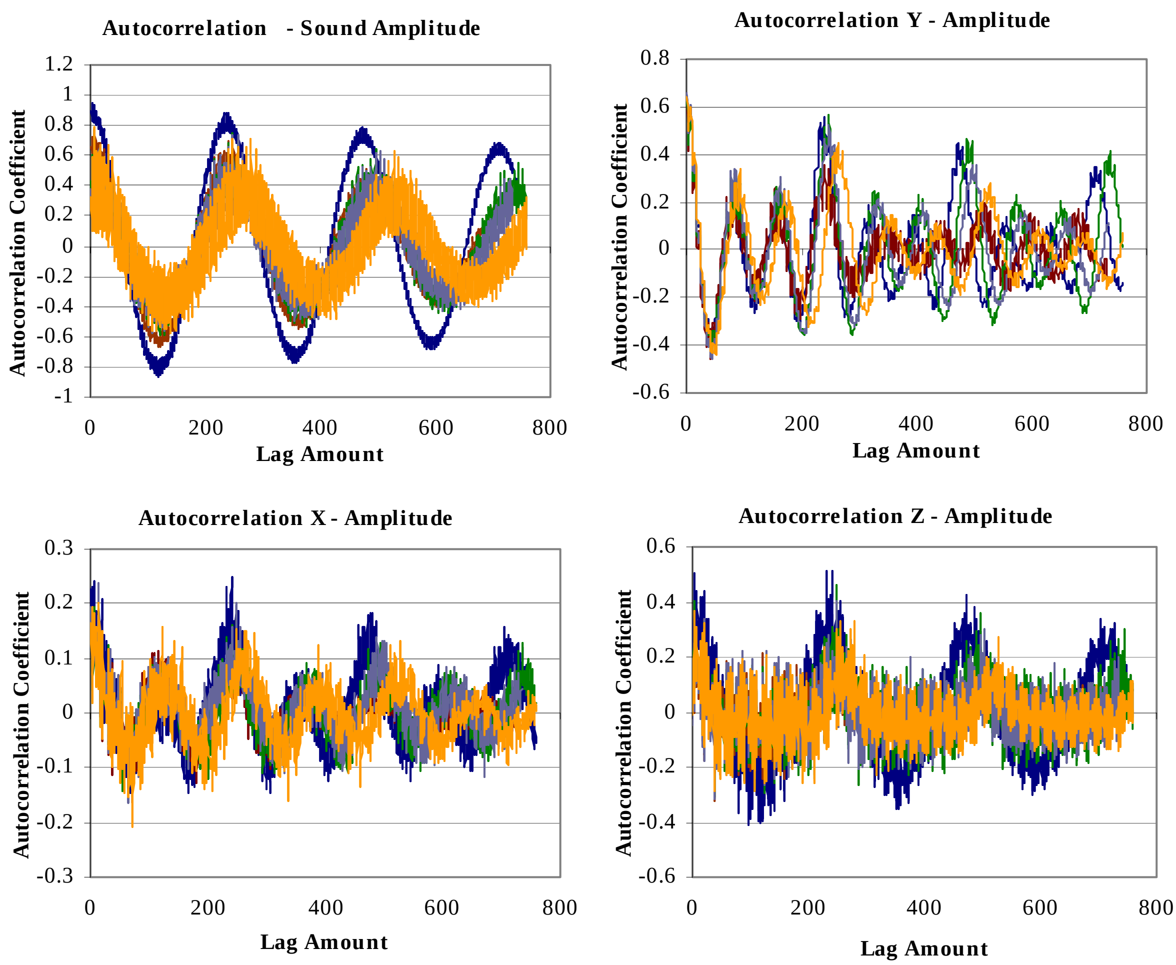

Fig. 13 Comparison of the experimental tests with lag value equal to 1

5 Discussion

From the simulation analysis, three major characteristics have being established. When comparing the correlograms of different periods from the same system, where amplitude and/or phase change is detected, the system experiences a change in behaviour. The simulated analyses illustrate that the autocorrelation method is a potential candidate for the Detection Indices (DI) algorithm generation.

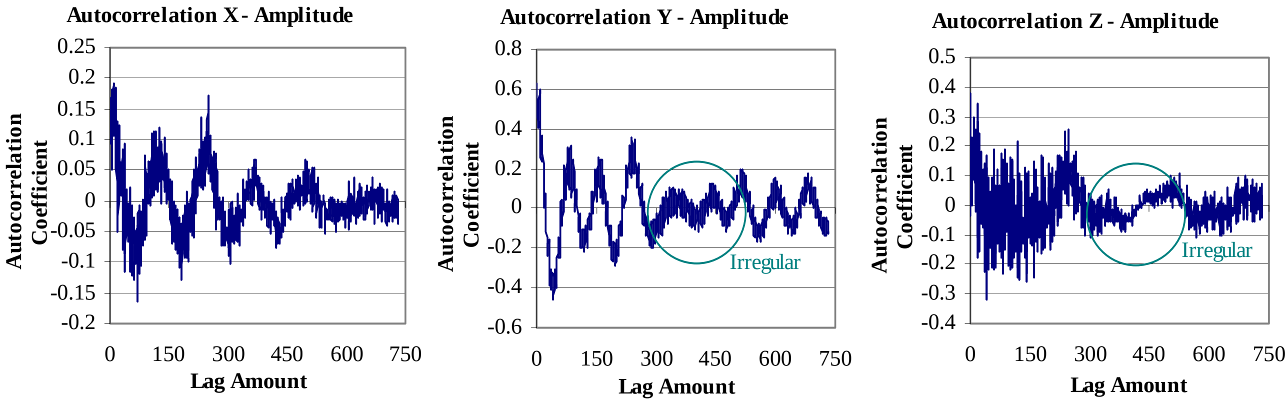

Fig. 14 XYZ-axis vibration correlogram plots for fastener weight equal to 2.7g

To further attest autocorrelation as a potential DI, an experimental set-up was used to demonstrate the capability of the autocorrelation. Compared to the simulated analysis, the experimental results indicated the existence of multiple periodicities with a greater amount of noise. As described in the experimental analysis, the drive shaft is supported by number of bearings and connected to a gearbox, and also that the entire arrangement is mounted on a beam structure as shown in figure 12. Because of this experimental design set-up, it is prone to create multiple sound and vibration sources. The large number of periodicities and noise seen in the correlogram plots demonstrated this point, but also demonstrated the ability of the autocorrelation process to identify these sound and vibration sources. Refer to figure 13 for the correlogram plots.

Figure 13 shows a fairly constant shift of phases as the weight on the disc is increased, but when referring to figure 14 where the vibration correlogram plots are for the weight on the disc equal to 2.7g, a section of irregular pattern can be seen from the Y and Z axis plot. It was found that the inner race of each of the bearings that were attached to the driving shaft were not rigidly fixed, which produced slippages during the experimental process. It is strongly speculated that the slippages resulted in a disruption of the vibration transfer from the shaft to the bearings and subsequently into the whole system. As a result, figure 14 shows that for a lag amount around 400, there is an irregular section in both of the Y and Z vibration correlograms. Again this demonstrated the ability of the autocorrelation process to identify anomalies when they occur. The X-axis vibration correlogram shows very few irregular patterns, because when the electric motor is rotating most of the forces and displacements are occurring in the transverse and vertical directions. There are very few axial direction movements, which is evident from the low amplitude values of the autocorrelation coefficient for the X-axis correlogram plot in figure 14.

6 Conclusions

An examination of the autocorrelation process as a potential DI algorithm has been studied. Simulation and experimental analysis has proven that the autocorrelation method has the ability to identify changes in a monitored mechanical system by comparing the correlogram at different time periods. Three basic characteristics have been observed, which are change in amplitude, phase shift or rapid phase shift with rapid reduction in successive waveform. If any of these characteristics occurred that means the system has a fault.

The simulation has also demonstrated that the autocorrelation method is capable of representing a fault as troughs in the correlogram for the simulated time series data set. Furthermore, the experimental section shows that the irregular pattern of the correlogram has lead to the speculation of slippage between the bearings and the drive shaft of the experiment set-up. Although more investigation is needed, the autocorrelation method has demonstrated the potential of identify occurring faults and possibly the location of those faults within the monitored system.

References

[1] Forsyth G F. *Structural Usage Monitoring in Black Hawk and Seahawk Helicopters: Data System Options*. Defence Science & Technology Organisation, Melbourne Australia, 2000.

[2] Fraser K F. *Effectiveness of Helicopter Health and Usage Monitoring Systems (HUMS) in the military Environment*. Defence Science & Technology Organisation, Melbourne Australia, March 1996.

[3] Larder C T B, Vossler G and Azzam H. *Smith Industries HUMS: Changing the M from Monitoring to Management*. IEEE, 2000.

[4] Mathur A. *Data Mining of Aviation Data for Advancing Health Management*. SPIE's 16th International Symposium on AeroSense, 2002.

[5] Williams G P. *Chaos Theory Tamed*. London: Taylor & Francis, 1997.

[6] James E B and Robinson C. *Description of peakvue and illustration of its wide array of applications in fault detection and problem severity assessment*. Emerson Process Management Reliability Conference, 2001.

[7] Zubaydi H A and Swamidas A S J. *On the use of the autocorrelation function to identify the damage in the side shell of a ship's hull, Marine Structures*. Vol. 13, pp. 537-551, 2000.

[8] Davis J C. *Statistics and data analysis in geology*. New York, Wiley, 1986.

[9] Makridakis S G, Wheelwright S C and McGee V E. *Forecasting, methods and applications*. New York, John Wiley, 1983.

[10] Duncan A O. *Time series analysis and forecasting: the Box-Jenkins approach*. London, Butterworth, 1976.