A Multivariate Statistical Approach for Improved Automated Process Control and Anomaly Detection in Mechanical Systems

David Ogden
Southwest Research Institute
San Antonio, Texas, USA
david.ogden@swri.org

Tom Arnold
tom.arnold@swri.org

Walt Downing
walt.downing@swri.org

Abstract— This paper describes an approach for applying multivariate statistical process control techniques to improve and automate anomaly detection in mechanical systems. Implementation examples include turbocharger performance, wind turbine generator failures, transient analysis of aircraft jet engine operation, air compressor performance monitoring, and other projects where the technique has been successfully applied to detect operational anomalies in mechanical systems.

Keywords—multivariate analysis, anomaly detection, condition monitoring, statistical process control

I. Introduction

The multivariate statistical process control methodology detects anomalies in the complex relationships of sensor data gathered during mechanical system operations. The methodology detects changes in the statistical relationships of the parameters being monitored and is useful to detect wear and degradation of components.

The methodology has similarities to conventional Statistical Process Control (SPC), which has been widely adopted by industry. The key difference is the focus on monitoring *n* parameters simultaneously. When properly designed, the *n*-dimensional relationships within the system can be monitored over a wide operating range, which is more powerful than using a static univariate limit.

This capability is an important part of a condition monitoring system to improve mission capability or reduce maintenance costs. Advance knowledge of impending failure or detection of degraded performance is very difficult, but critical to implementing advanced maintenance plans and procedures. The multivariate methodology described herein can provide forewarning of impending failures or degraded performance.

The motivation for a multivariate approach is to assist customers in improving failure detection or wear detection sensitivity and reduce false alarms of Supervisory Control and Data Acquisition (SCADA) systems or other monitoring systems. Additional considerations included the need to present the results in a format that could be easily understood by staff who may have no training in statistical analysis. A frequent constraint in implementing condition monitoring systems is the inability to add new sensors due to cost or other constraints. The multivariate approach has proven to be very effective with limited sensor suites and very low data collection rates.

II. Approach

The multivariate approach presumes the investigator has some familiarity with the operating principles of the asset to be monitored. Feature selection and feature extraction efforts to review and reduce the number of variables under consideration (dimensionality) are employed. Sensor value range checks are also necessary to remove records with faulty sensor readings. In some of the cases described in this paper, sensor validation was performed during the data recording process and therefore additional sensor validation was unnecessary.

Observations for the examples presented were captured over a wide range of operating conditions. Analyses were performed to determine all operating regimes for the systems. When different operating regimes were determined to affect the results, filters were employed to segregate the data for analysis based on each appropriate operating regime.

The performance characteristics of some of the assets were sensitive to the effects of ambient conditions. Conversion of some data to standard day conditions to normalize the data for operating environment changes (29.92 in/Hg at 15° C) was employed as part of the preprocessing. Other data preprocessing or feature extraction techniques were also used to provide the appropriate input to the process. Examples include corrections for noise or non-linearity, and extraction of specific frequency content in vibration signals. Preprocessing the data is automated in cases where continuous monitoring was employed.

To enable monitoring, an asset's baseline performance must be established. The number of observations used in the baseline set is dependent upon the number of parameters being monitored and the target false alarm rate. For 5-parameter

978-1-5090-0793-6/16/$31.00 ©2016 IEEE

system with a 0.1% false alarm rate, a baseline only requires a minimum of 37 samples.

Table 1 summarizes the execution time necessary to compute an initial baseline and calculate the subsequent trends for a sample size of 10,000 observations for a single asset. Benchmarks were established running on a 64-bit desktop computer with I7 2.30 GHz processor and 8 gigabytes of memory. This demonstrates these calculations can be performed very quickly using conventional low-cost computing platforms. The analysis results can be available in near real time. For the types of monitoring suitable for this technology, computational load is not a concern. The software does not need a dedicated or expensive computing hardware and multiple instances of the software could reside on a single system that could support a fleet of equipment.

Table 1. Baseline Formulation Times

# of Parameters	Execution (Seconds)
5	~ 1
10	~ 2
20	~ 6
50	~ 51
60	~ 85

Once the minimum number of required observations for the historical data set has been met the baseline is calculated. Calculations for new observations against the established baseline execute in less than 1 second.

Once the individual asset's unique baseline has been computed, new observations are automatically preprocessed and compared to the Upper Control Limit (UCL) established for the asset. To assist users in visualizing the results of the analysis, a line chart is used. The observations are ordered by date and time along the X-axis and the weighted value of the multivariate calculation forms the Y-axis. One line representing the UCL and a second line for the calculated statistic are plotted on the graph. If the calculated statistic for a given point exceeds the UCL the asset has signaled. Observations that have signaled are then automatically decomposed to determine which parameter, or combination of parameters, are contributing to the signal. This decomposition can assist a knowledgeable individual with diagnosing the maintenance issue with the asset or can be tied to troubleshooting trees for the asset in some cases.

SwRI has integrated the multivariate approach into a software tool called nSPCT™ (n-variable Statistical Process Control Tool) that consumes Comma Separated Value (CSV) files as it data source and as a Microsoft.Net Dynamic Link Library (.dll) component consuming SQL Server data as the observation data source. The tool can compute up to 9th order relationships within large data sets. Work is currently in progress to include LabView TDMS files as a data source. Providing the analysis engine as part of a Microsoft Azure marketplace web service component is also being considered.

III. APPLICATION

All examples described below are derived from data provided by equipment operators for post-failure analysis to evaluate the value of the nSPCT tool.

A. Turbocharger

The initial data set included 52 time stamped parameters. Data points were collected every 3 hours over the course of a year. The turbocharger was not operated continuously over the period during which data was collected. The data set was preprocessed to remove observations when the engine was not running. Temperatures and rotational speeds were corrected to standard day conditions using the ambient temperature provided in the dataset.

Multivariate analysis identified an anomalous shift, which continued until the asset's on-board monitoring system shut the asset down months later. Figure 1 is the multivariate control chart for the asset. Each point on the chart represents a weighted system score of all parameters for a given observation. If the system score exceeds the Upper Control Limit (UCL), the physical system has left its multivariate control region. The X-axis represents each observation taken when the engine was running over the past year. The point at which performance shifted above the UCL is highlighted in Figure 1.

Decomposition of the signal revealed two parameters had left their individual (1st order) control regions and a second order relationship between two other parameters was also detected. None of the parameters exceeded the manufacturer's specified operating limits. The root cause was determined to be an issue with the turbocharger wastegate.

B. Wind Turbine Generator

The original data set included observations collected over a 1.5-year time frame. Approximately 73,000 observations were provided for wind turbines generators on two different towers. The same ten parameters were provided for each wind turbine generator. Standard day corrections were applied to the data. All data was from the same operating regime. One of the wind turbine generators was considered healthy by the customer, and the other unhealthy. SwRI was not informed in advance which data set was from which generator.

Figure 2 is a plot of the multivariate performance for the healthy wind turbine generator. All observations are below the UCL. The fluctuations depicted by the red oval in the figure could be reduced further by optimizing standard day corrections and additional sensor validation.

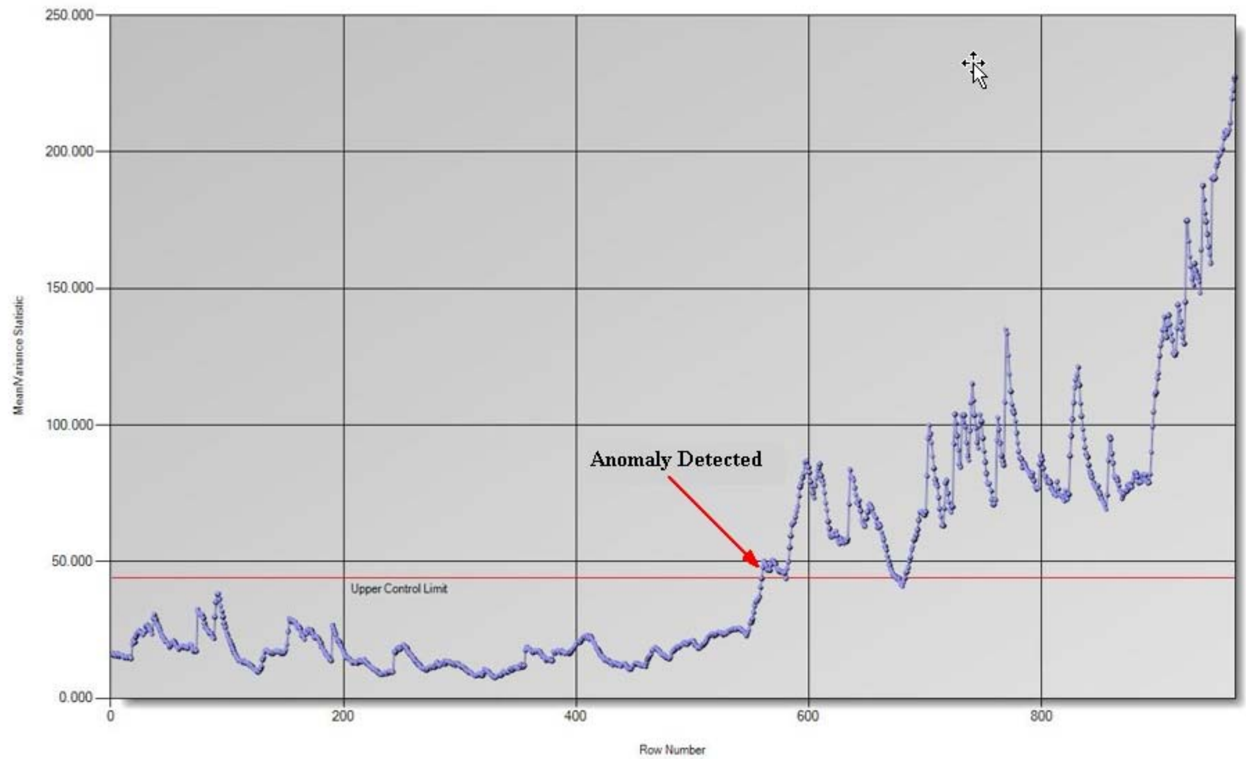

Figure 1. Turbocharger Multivariate Control Chart

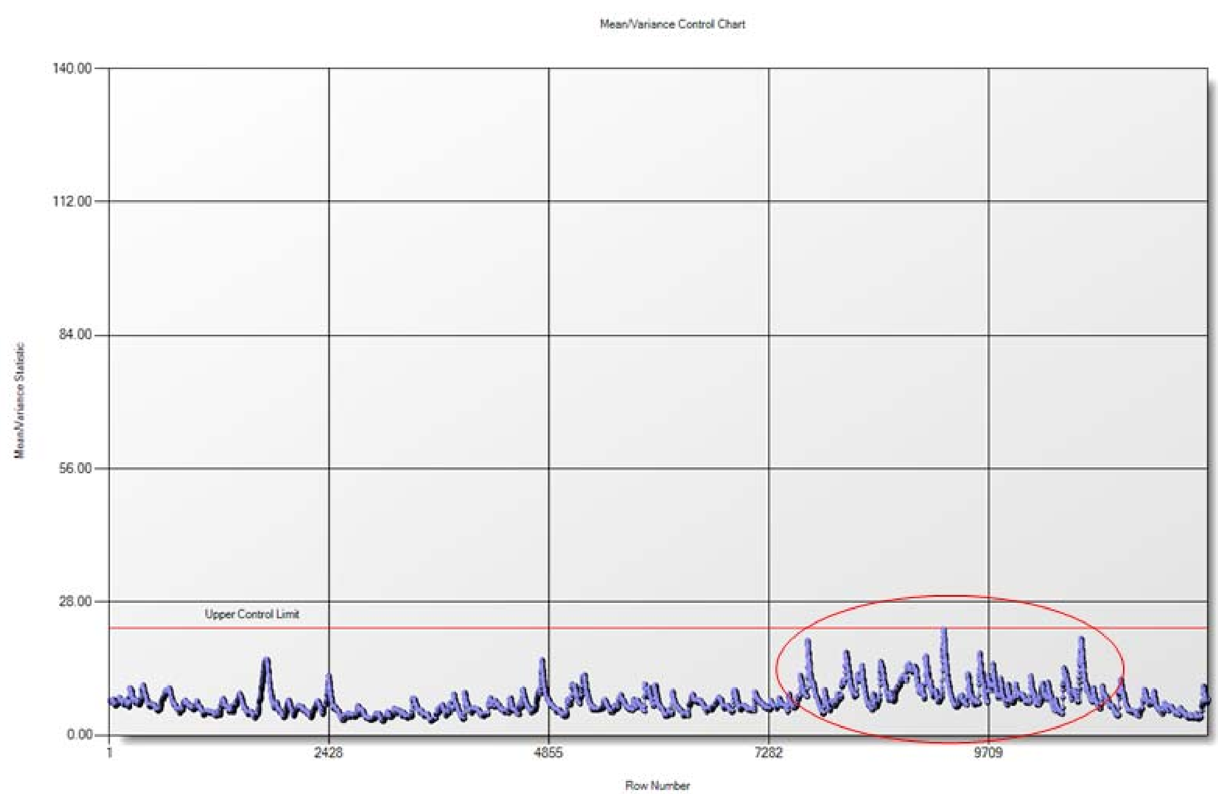

Figure 2. Healthy Wind Turbine Generator Control Chart

Figure 3 shows the unhealthy wind turbine generator data. The unhealthy generator seems to be operating well until about observation 10,200. At this observation and continuing until about observation 11,680, the control chart shows a distinct increase in the frequency of anomalous signals, indicating a rapid degradation of the wind turbine generator condition. The red oval in Figure 3 indicates this region of degraded condition. At point 11,680 the high speed bearing in the gearbox failed, setting off an alarm in the existing OEM SCADA system, and taking the turbine out of service. After weeks of troubleshooting, the gearbox was replaced at a cost of nearly $500,000 for parts, labor and crane expenses. Once the maintenance was performed on the tower, the trend line returned to the original control region.

Decomposition analysis indicated one bearing temperature parameter, corrected to standard day, was the principal contributor to the anomalous report. Interestingly, the bearing's temperature rarely exceeded its univariate control region during the monitoring period. However, for almost every anomalous observation in the nSPCT control chart, the bearing temperatures relationship with multiple parameters (2nd order relationship) has shifted and left its control region. Further decomposition analysis revealed the same bearing temperature anomalies can be credited for all of the previous spikes seen in as early as observation 8,400. The failure could have been detected months earlier allowing for planned corrective maintenance. Coordinated crane usage or taking assets off-line during periods of lower average demand are examples of savings opportunities.

C. Transient Analysis of Aircraft Engines

Much of the anomaly detection performed on legacy USAF engines is based on snapshot data taken when predetermined conditions are achieved during flight, takeoff or cruise are examples of such conditions. Transient data is also captured for some engines but not generally used for day-to-day trending purposes.

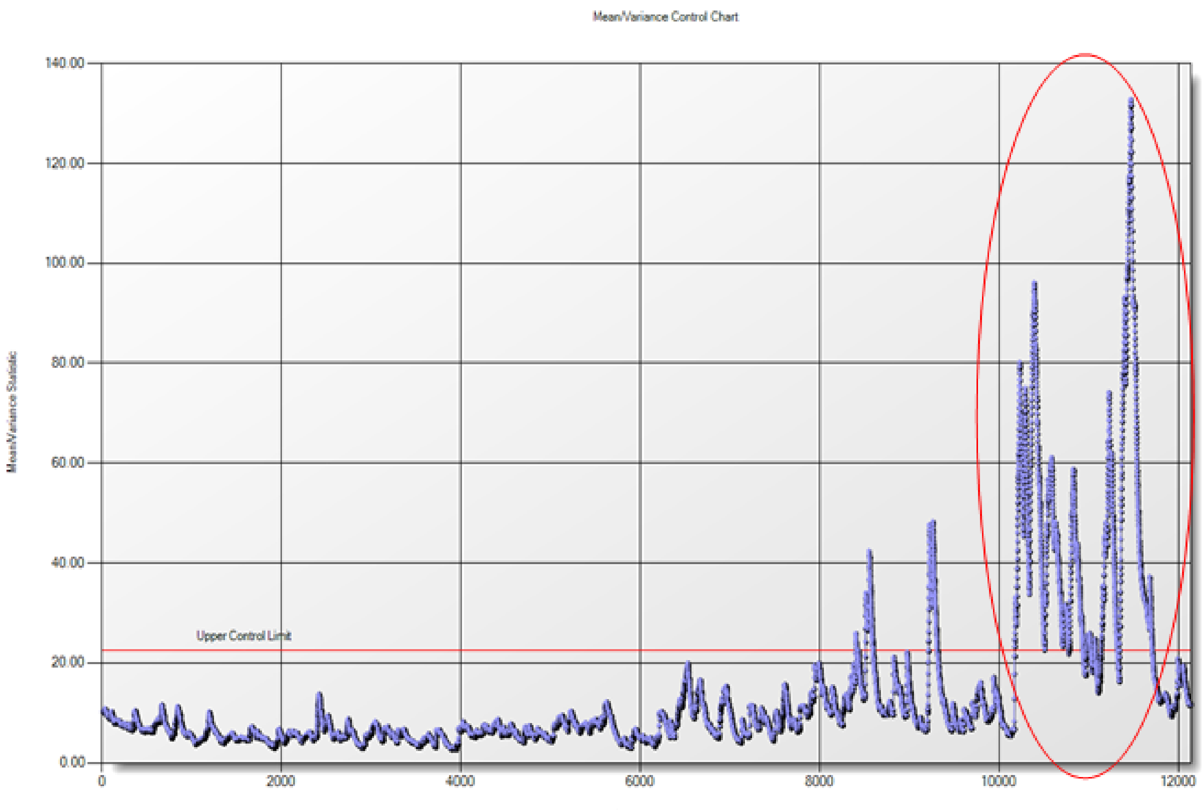

Figure 3. Unhealthy Wind Turbine Generator Control Chart

Mission aborts can occur for many reasons. Only mission aborts attributed to engine issues were considered in the analysis. Data for five aircraft recording an engine related mission abort were analyzed to see if multivariate analysis could be used to detect the condition prior to the flight on which the mission abort occurred. Early detection can allow for maintenance or changes in flight schedules and prevent mission aborts. Of the five data sets analyzed, anomalies were detected with sufficient time to prevent the mission abort in each case.

In the case of the mission abort described in this paper, the crew of the aircraft aborted their mission due to fluctuations in an engine's fan speed and oil pressure. The existing univariate tools used to trend the engine, detected no engine events prior to the date of the mission abort. Multivariate analysis of the data taken during the cruise snapshots for multiple days leading up to the mission abort also did not identify any issues based on the cruise snapshots. Figure 4 is the multivariate plot for the cruise data. The X axis is the date and time ordered cruise snapshots recorded by the system. However when multivariate analysis was performed on the transient data, anomalies were apparent (Figure 5).

Decomposition of the signal indicated the relationship between corrected exhaust gas temperature and corrected fan speed. If transient analysis had been in place, maintainers would have been alerted to the deteriorating condition weeks before the mission abort.

The aircraft type analyzed captures data for multiple parameters at a sub-second rate from the time power is applied to the aircraft systems until the aircraft is powered down after flight. For the purpose of the analysis, SwRI first had to establish conditions under which the engine was undergoing transient stress and limit the records to those conditions. The transient conditions were defined for a specific change in power level angle used during takeoff. The final data sets analyzed were comprised of 10 seconds worth of data in 1 second increments when the takeoff condition was achieved. Data was corrected to standard day. Five performance parameters were used for the analysis.

Subsequent to this demonstration, the USAF has begun to implement multivariate statistical analysis on engines for the F-15/F-16 and A-10 aircraft engines.

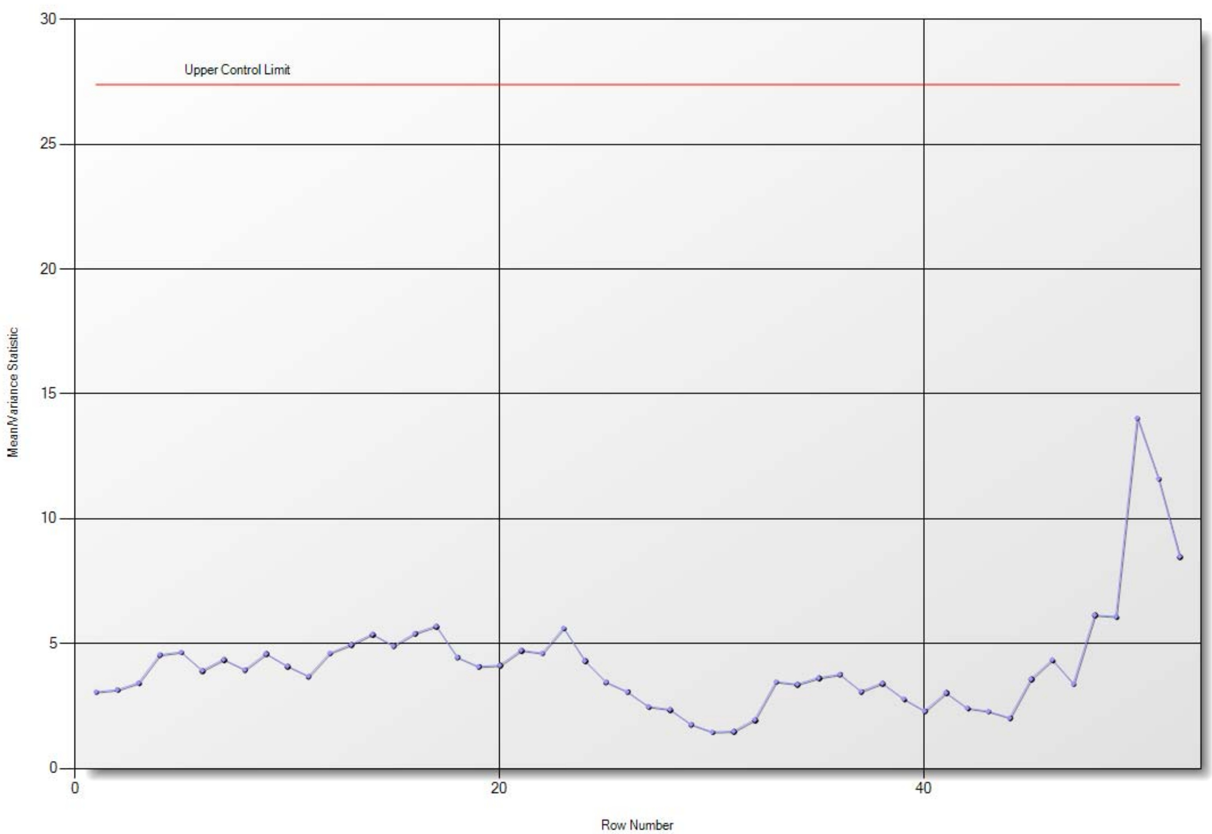

Figure 4. Cruise Data Control Chart

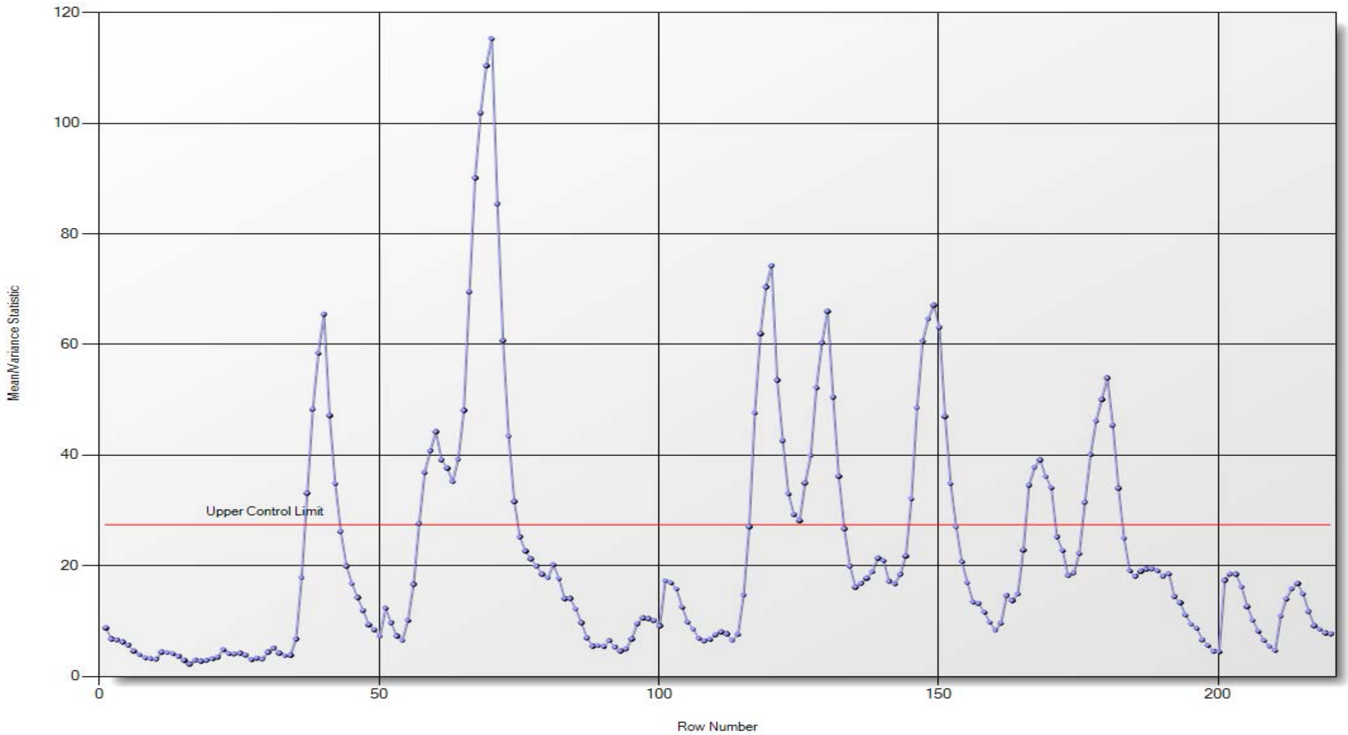

Figure 5. Transient Data Control Chart

D. Air Compressor

A customer that maintains compressor equipment experienced a failure of an Inlet Guide Vane (IGV) on an axial compressor which is part of one of the world's largest nitrogen generation systems. The system is used to increase the production of oil and gas wells by injecting the gas into the shale formation.

The failed IGV reduced system production by about 6%, negatively impacting the oil and gas field profitability. The failure was not detected when it occurred, and was only discovered during the next regularly scheduled maintenance activity. The customer tasked SwRI to redesign the failed part, and requested an analysis of the data that was collected by the monitoring system.

The monitoring system captures 22 parameters of data in 1 minute samples. The monitoring system data spanned approximately 6 months of operations (around 200,000 observations). Standard day corrections were applied to the data prior to analysis. Incomplete and erroneous observations were removed from the data set. Eight of the initial 22 parameters were deemed to be of no use in the analysis and were removed. Figure 6 is the multivariate plot of the air compressor's performance after sub-sampling the data a 1 sample per three hours.

A distinct shift in performance was detected approximately 6 weeks after the beginning of the data set. This shift is the moment the IGV failure occurred. The pressure on the pad bearings distinctly increased, but not to a level that caused a system alarm.

This is a good example of an important characteristic of the nSPCT tool – all failure modes do not have to be predefined for detection to occur. SwRI has seen other examples of critical failures due to a monitoring system that did not include alarm conditions for the particular failure mode that ultimately occurred.

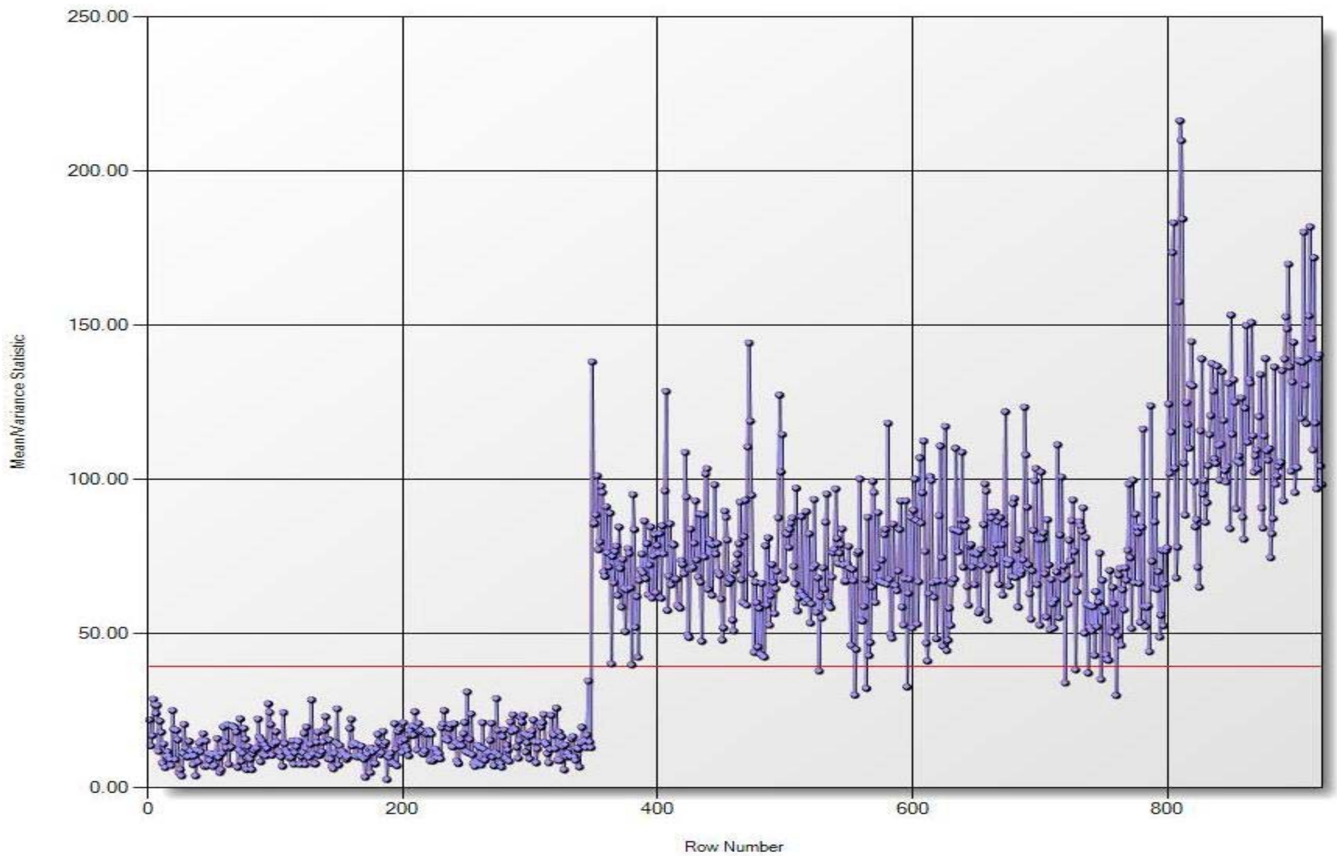

Figure 6. Compressor Control Chart

IV. Conclusion

The multivariate statistical analysis technique is a powerful tool for detecting mechanical system degradation. It has advantages over previously applied techniques in that it has high sensitivity, a low false alarm rate, and provides good detection with limited sensor sets. The desired false alarm rate can be adjusted and failure modes do not have to be defined in advance.

SwRI has applied this technique to diverse mechanical systems ranging from aircraft turbine engines to process water cooling towers. The potential for other applications is immense.

The nSPCT tool for multivariate statistical analysis is highly automated. This automation, combined with the high sensitivity and low false alarm rate, relieves maintenance staff from the burden of manually checking system performance. Degradation indicating mechanical wear or signs of impending failure can be detected well in advance, providing the maintainer an opportunity to perform a cost effective maintenance action. Early notification of performance changes may allow options for extending system life that were not previously available. In addition, better information on system health can reduce excessive preventative maintenance that is unwarranted and costly.