Data Mining Applications for Space Mission Operations System Health Monitoring

David L. Iverson*
NASA Ames Research Center, Moffett Field, California, 94035

Recent developments in data mining techniques for anomaly detection make it possible to use the wealth of available archived spacecraft system data to produce advanced system health monitoring applications. These "data driven" applications are capable of characterizing and monitoring interactions between multiple parameters and can complement existing practice to provide valuable decision support for mission controllers. Data driven software tools, including Orca and the Inductive Monitoring System (IMS), have been successfully applied to mission operations for both the Space Shuttle and the International Space Station. Orca uses a nearest-neighbor approach to search for unusual data points in multivariate data sets by calculating the distance of each data point from neighboring points. The IMS tool uses a data mining technique called clustering to analyze archived spacecraft data and characterize nominal interactions between selected parameters. This characterization, or model, is compared with real time or archived system data to detect off nominal behavior. Augmenting traditional mission control software with advanced monitoring tools, such as Orca and IMS, can provide controllers with greater insight into the health and performance of the space systems under their watch. We will describe how such techniques have been applied to NASA mission control operations and discuss plans for future mission control system health monitoring software.

I. Introduction

ARCHIVED spacecraft telemetry data can contain a wealth of information about complex system behavior. Recent developments in data mining techniques for anomaly detection make it possible to examine this archived data and extract embedded information to produce advanced system health monitoring applications. Such applications can aid mission controllers and engineering analysts in their task of ensuring that spacecraft systems under their watch are operating properly. In contrast to common individual parameter monitoring schemes, these "data driven" applications are capable of characterizing and monitoring interactions between multiple spacecraft parameters and can provide additional insight and valuable decision support for controllers and engineers.

Several data driven software tools, including Orca and the Inductive Monitoring System (IMS), have been successfully applied to mission operations for both the Space Shuttle and the International Space Station. Orca[1] is a data mining tool that searches for unusual data points, or outliers, in multivariate data sets by calculating the distance of each data point from neighboring points. The presence of outliers in spacecraft system data is of interest to mission controllers because they may indicate malfunctioning system components. The IMS tool[2] uses a data mining technique called clustering to analyze archived spacecraft data and characterize nominal interactions between selected parameters. This characterization, or model, of normal operation is stored in a knowledge base that can be used for real time system monitoring or analysis of archived events. Spacecraft data is compared with the nominal model built by IMS to produce a measure of how well that data matches normal behavior captured in the training data that was used to build the IMS knowledge base. Significant deviations from the nominal system model can alert the controller to a system malfunction or precursor to a significant failure.

---

* Computer Engineer, Intelligent Systems Division, Mail Stop 269-4

II. Data Mining for Mission Control

NASA maintains years of archived Space Shuttle and International Space Station telemetry data in the Operational Data Reduction Complex (ODRC) at the Johnson Space Center (JSC). Methods from the field of data mining are useful for analyzing the type of data found in the ODRC and extracting information about typical telemetry parameter behavior and how those parameters relate to one another. In particular, recent developments in data driven anomaly detection techniques can process the data to find unusual events, or outliers, in data for a given spacecraft subsystem. These anomaly detection techniques can also automatically analyze archived nominal system data to characterize normal system performance. Comparing incoming real-time data to that nominal model can let the user know if the current system behavior differs from previous system performance.

A. Distance-Based Anomaly Detection

One powerful feature of many data driven anomaly detection techniques is the ability to analyze multiple parameters simultaneously. This feature allows them to discover and model interactions between related parameters that might be difficult to notice when monitoring the parameters individually. A basic data structure used for distance-based analysis is a vector of parameter values (Fig. 1). Vectors containing N values can be treated as points in an N-dimensional vector space. An appropriate distance metric can be used to calculate the distance between points. The familiar Euclidean distance metric has proven effective in several applications, though other metrics may also be useful.

Pressure A	*Valve 1 Position*	*Pressure B*	*Valve 2 Position*	*Pressure C*	*Temperature 1*	*Temperature 2*
2857.2	86.4%	1218.4	96.2%	1104.1	49.8	37.6

Figure 1. Sample Data Vector

For system health monitoring applications, vector parameters are typically instantiated with concurrent sensor values collected from a time slice of the data stream. Additional computed (derived) or historic parameter values can be included in the vector as well. For instance, increased system insight can often be obtained by incorporating values in the vector such as the rate of change of a pressure value or the difference between two related temperature sensors. Flight controllers and engineers familiar with the monitored system can often suggest useful telemetry and derived parameters to use in the health monitoring vectors.

An archive data set typically covers a number of dissimilar ranges for the vector parameters. For example, in a given data set a pressure parameter may range from 0 to 8000 kPa, while a temperature parameter may only cover 10 to 50 degrees. This would imply that a temperature change of one degree is probably more significant than a one kPa pressure change, since the temperature change covers a larger portion of the expected parameter range than the pressure change. Thus, data normalization within the vector is required to avoid disproportionately weighting parameters with differing ranges when calculating distances between vectors. One useful normalization technique represents values as a percentage of the expected parameter range, so each vector parameter value will fall somewhere between zero and one hundred. With this normalization scheme, a 1 degree change in our example temperature data would translate to a normalized value change of 2.5%, while a 1 kPa pressure change would translate to a normalized value of 0.0125%, reflecting the greater significance of the temperature change. Another useful method is a Z-score normalization using the mean and standard deviation calculated for each parameter from the archived data set. Z-score normalized vector values are calculated by subtracting the parameter mean from the current value then dividing the difference by the standard deviation.

In some cases, it may be advantageous to increase or decrease the significance (weight) given to certain vector parameters. For instance, if maintaining a specific operating pressure is critical to a system, the weight of that pressure value could be increased so a small change in the pressure would manifest as a larger change in the associated vector parameter, increasing monitoring sensitivity to variations in that parameter. Conversely, if the monitored system is not particularly sensitive to a certain parameter, such as ambient temperature, the weight of that vector value could be decreased to reduce the chance of unnecessary alarms when that parameter value changes by an insignificant amount.

Each monitored system will present unique characteristics. Normalization and parameter weighting schemes can be adapted to suit the situation. Often the most effective way to determine which scheme is most appropriate for monitoring a given system is to analyze familiar, well understood data and try candidate normalization and weighting values until the analysis results are acceptable. If the familiar test data is representative of typical system behavior, the normalization and weighting selected by this technique will usually produce reasonable results when analyzing unfamiliar data.

B. Orca Distance-Based Outlier Detection

Orca is a data mining tool that analyzes multivariate data sets.[1] It uses a nearest neighbor approach for outlier detection. For each point in the data set, Orca locates the other points in the data set that are closest to that point, called the point's nearest neighbors. Distance between points is measured with the Euclidean distance measure for continuous parameters and the Hamming distance for discrete parameters. The program outputs a score for each point representing the average distance to the nearest k neighbors in the data set. The value of k is specified by the user. Points that have a larger average distance to their nearest neighbors than most other points in the data set are considered outliers. Orca is able to find outliers within a single data set, or compare one data set with another to determine which points in the first data set are unusual in comparison to the second set.

One approach to using Orca with spacecraft telemetry is to form data vectors, as described previously, using sensor values from a given spacecraft subsystem and search for outliers among those vectors. For instance, vectors could be formed from temperature, pressure, and fuel flow rates in a rocket engine. An Orca analysis of this data set can locate data from time periods during the engine firing that display unusual characteristics compared with the rest of the data. Unusual data points may be symptoms of engine malfunctions, such as a faulty pressure regulator or an incorrect fuel-oxidizer mixture ratio.

C. IMS: Inductive Monitoring System

The Inductive Monitoring System (IMS) is a tool that uses a data mining technique called clustering to extract models of normal system operation from archived data.[2] Like Orca, IMS works with vectors of data values. IMS analyzes data collected during periods of normal system operation to build a system model. It characterizes how the parameters relate to one another during normal operation by finding areas in the vector space where nominal data tends to fall. These areas are called nominal operating regions and correspond to clusters of similar points found by the IMS clustering algorithm. These nominal operating regions are stored in a knowledge base that IMS uses for real-time telemetry monitoring or archived data analysis.

During the monitoring operation, IMS reads real-time or archived data values, formats them into the predefined vector structure, and searches the knowledge base of nominal operating regions to see how well the new data fits the nominal system characterization. For each input vector, IMS returns the distance that vector falls from the nearest nominal operating region. Data that matches the normal training data well will have a deviation distance of zero. If one or more of the data parameters is slightly outside of expected values, a small non-zero result is returned. As incoming data deviates further from the normal system data, indicating a possible malfunction, IMS will return a higher deviation value to alert users of the anomaly. IMS also calculates the contribution of each individual parameter to the overall deviation, which can help isolate the cause of the anomaly.

III. Mission Control Applications

The Orca and IMS software tools have both been applied in NASA mission control to support real-time telemetry monitoring and engineering analysis of mission data. In support of the Johnson Space Center (JSC) Mission Evaluation Room (MER) engineering analysis activity, the tools were applied to data from the Space Shuttle Wing Leading Edge Impact Detection System (WLEIDS) to find potential impact signatures. In the International Space Station (ISS) flight control room they have been used to build real-time health monitoring applications for the ISS Control Moment Gyroscopes. Applications for real-time monitoring of ISS thermal control systems are currently under development.

A. Space Shuttle Wing Leading Edge Impact Detection System

The Space Shuttle WLEID system was developed in response to the loss of the Columbia orbiter on the STS-107 mission. During the launch of STS-107 a piece of foam shed from the Shuttle external fuel tank

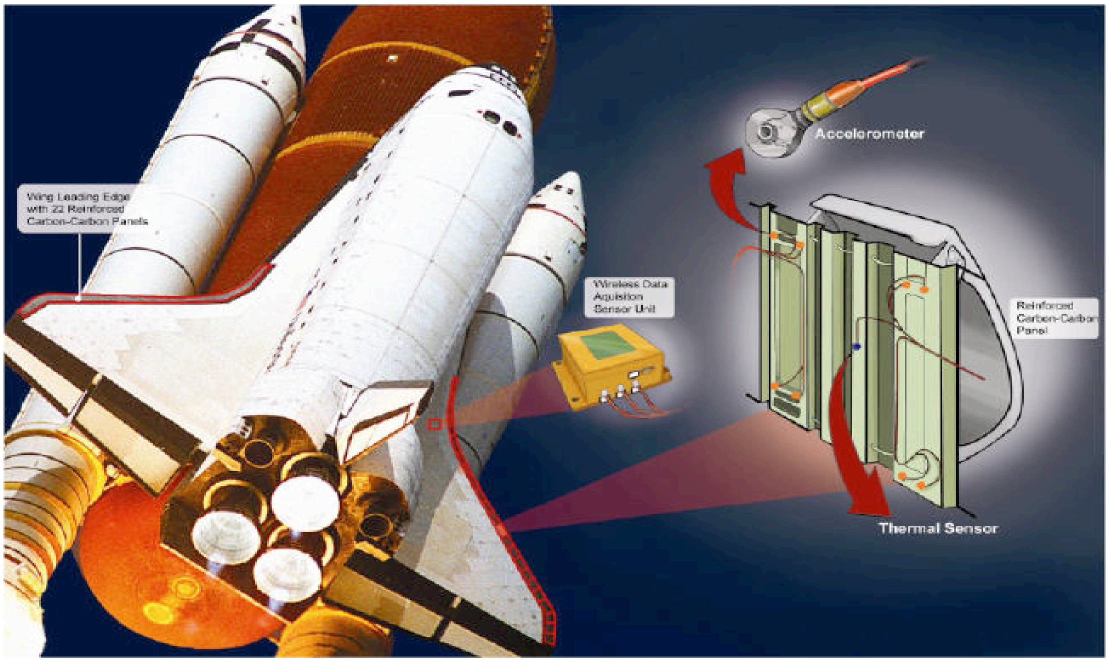

Figure 2. Space Shuttle Wing Leading Edge Impact Detection System

struck the leading edge of the orbiter's left wing, compromising the thermal protection system. This damage resulted in the tragic loss of vehicle and crew during reentry due to overheating and failure of the internal wing structure.[3]

The WLEIDS consists of 132 single axis accelerometers mounted along the length of the orbiter's leading edge wing spars (Fig. 2). During launch, the accelerometers collect data at a rate of 20 kHz and store that data onboard for subsequent downlink to Mission Control. Within 6 to 8 hours of launch, summary files containing periodic sub-samples of the data collected by each accelerometer are down linked to the MER for analysis to find potential impact signatures. This analysis must be completed within 24 to 48 hours of the launch so the results can be used to schedule detailed on-orbit wing leading edge inspections using cameras mounted on the Shuttle robotic arm.

The WLEIDS analysis is performed by MER engineers by visually examining three dimensional graphs of summary data that show accelerometer location and vibration magnitude along a time axis (Fig. 3). The analysts search the graphs for localized peaks among the normal vibration signals caused by the Shuttle engines and aerodynamic forces, looking for unusual peaks that may have been caused by an impact on the wing leading edge. When potential impact events are identified, a half second of raw data collected by the affected accelerometer during that time period is downloaded for more thorough analysis to determine the likelihood that an impact occurred.

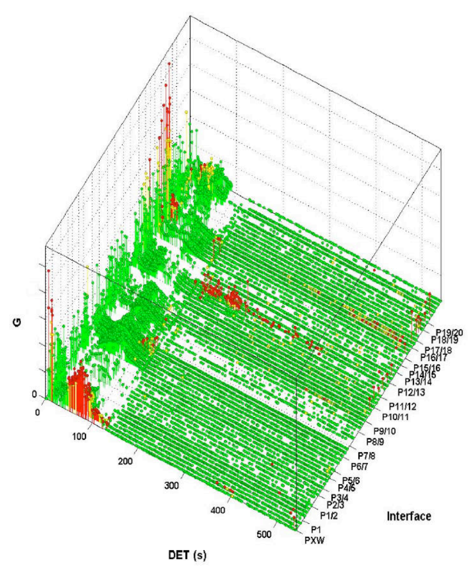

Figure 3. Sample WLEIDS Summary File Graph

The Orca and IMS tools have been used to support the WLEIDS analysis on three Shuttle launches. The goal was to provide a quick, automated initial scan of the WLEIDS summary files to locate unusual points and help focus the MER analysts' efforts. For each accelerometer sensor, vectors were formed from concurrent values collected from that sensor and seven nearby sensors that might pick up radiating impact energy (Fig. 4). Prior to the launch, IMS was used to analyze normal data from previous launches to characterize typical vibration patterns for each group of accelerometers. Data from the current launch is compared to this characterization to identify unusual vibration patterns that might have been caused by impact events. Orca is used to search for outliers within the data collected during the current launch.

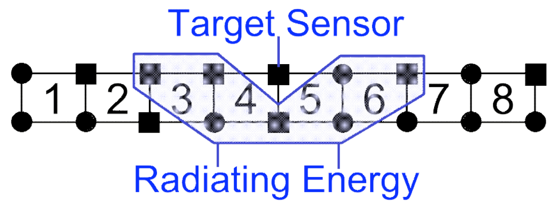

Figure 4. WLEIDS Analysis Sensor Selection on Shuttle Wing Leading Edge Panels

To gauge the effectiveness of the data mining tools, we compared Orca and IMS results to visual WLEIDS summary file analysis performed by MER engineers on the STS-115 launch of Space Shuttle Atlantis. The analysts classified the events they identified as critical, probable, or questionable based on the likelihood that the data signature was caused by an impact. Their analysis of the STS-115 launch WLEIDS summary data produced 6 critical events, 23 probable events, and 2 questionable events. The Orca analysis placed all critical events in the top 50 outliers. IMS identified 334 interesting events, divided nearly evenly between the two wings. Those events included all 6 critical events, 18 of 23 probable events, and all of the questionable events found by the MER analysts. Most of the anomalies identified by Orca and IMS that were not noted by analysts could be eliminated as normal global vibrations that shook the entire vehicle, leaving a small subset that included the events of interest. A technique to automatically identify and remove those global vibration events was later developed. Additionally, during all launches where Orca and IMS have been used, the tools identified several lower energy vibration signatures that did not stand out in the visual data inspection. These events were investigated with raw WLEIDS data downloads from the affected sensors. Fortunately all of the potential impact events identified in the WLEIDS data were shown to be the result of non-damaging phenomena, such as aerodynamic events, sensor data spikes, or minor impacts, and all missions concluded with safe and uneventful reentry and landing.

B. ISS Control Moment Gyroscopes

The International Space Station (ISS) Control Moment Gyroscope (CMG) attitude control system consists of four large gyroscopes, each mounted in a gimbal system that can rotate the CMG about the two axes perpendicular to the gyroscope spin axis (Fig. 5). The CMGs operate as non-propulsive attitude control devices that exchange momentum with the ISS through induced gyroscopic torques.

As they have aged, some of the CMGs have degraded enough to malfunction and require replacement. A failed CMG1 was replaced with a new unit in July 2005, and a faulty CMG3 was replaced in August 2007. Given their history, the ISS Attitude Determination and Control Officer (ADCO) flight controllers are interested in detecting early symptoms of degradation in the CMGs. A deployment of data driven system health monitoring applications in the ISS flight control room is assisting with that task.

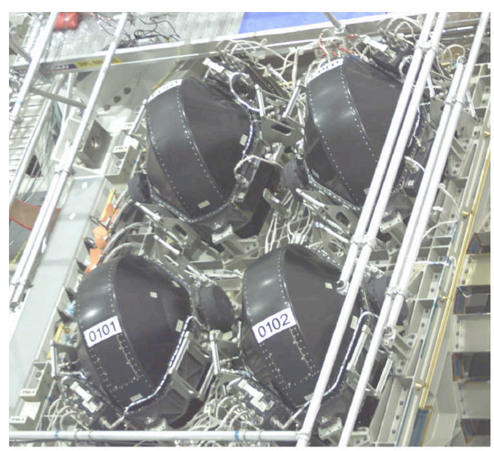

Figure 5. ISS Control Moment Gyroscopes

Working with the ADCO flight controllers, 13 CMG parameters were selected for real time monitoring. These parameters include CMG vibration, bearing temperatures, rotation speed, gimbal rates, electrical current, and ISS rotation rates, along with derived parameters for rates of change of temperatures and electrical current. Archived data collected over a period of 10 months for CMG1, 2, and 4 was analyzed. Seven months of data was analyzed for the recently installed CMG3. The data was sampled at a 1 Hz rate and formed into vectors of 13 values. The vectors were normalized using a variation of the Z-score method described previously. Each CMG was analyzed individually to capture its unique characteristics.

Because IMS was trained strictly on nominal data, the first operation with the CMG data was removal of any anomalies from the archived data. This was accomplished by searching for outliers within each data set using the Orca tool. Data records with significant deviations relative to the remainder of the data for that

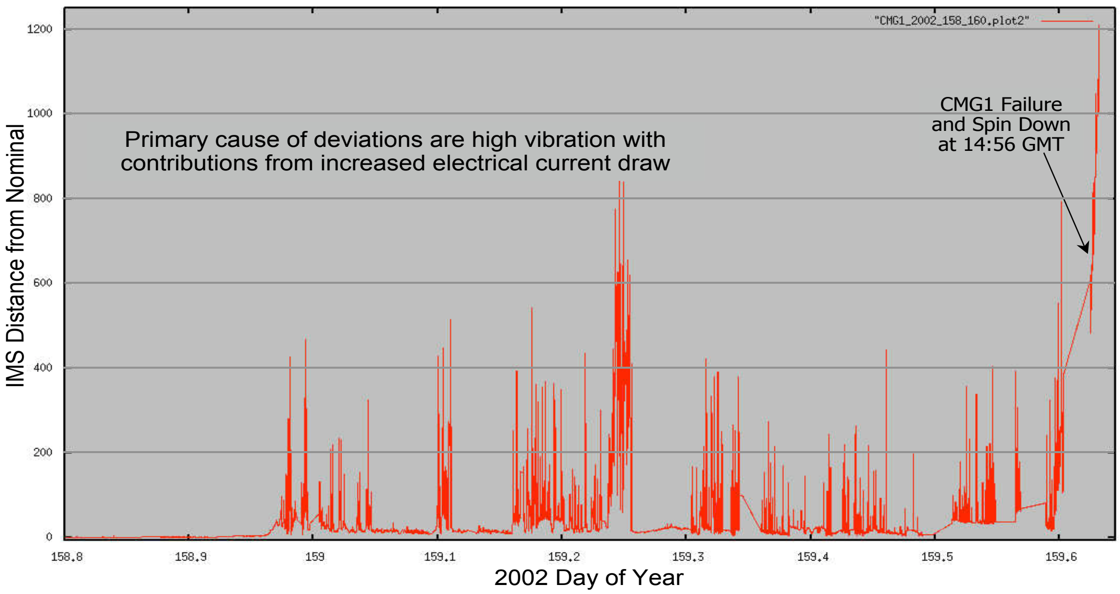

Figure 6: IMS monitoring results prior to CMG1 failure

CMG were removed. These deviations were typically caused by data corruption or minor anomalies in CMG operation. Once the archived CMG data had gone through this cleaning process, the remaining nominal data was used by IMS to build a monitoring knowledge base for each CMG.

To test the IMS CMG monitoring on a known anomaly, a similar process was performed using archived data from 2002 when CMG1 experienced a major failure. CMG1 exhibited increasing vibration levels that damaged a gyroscope spin bearing, prompting controllers to shut it down for safety. The redundant CMGs were able to maintain ISS attitude control without any issues. To test IMS effectiveness on this event, an IMS monitoring knowledge base was constructed from a month of archived nominal data collected prior to the CMG1 failure. The experiment used the same parameters, normalization, and weighting as the deployed IMS CMG monitoring systems. The results are shown in Fig. 6, plotting time on the horizontal axis and IMS results on the vertical axis. Recall that IMS outputs a measure of the distance from expected normal system behavior. Lower values indicate the system is behaving as expected. Increasing IMS values indicate the monitored system is deviating from expected behavior, possibly due to a system fault. In this case, IMS began indicating anomalous behavior more than 14 hours in advance of the eventual CMG1 failure. These and other similar results showed the value of using data driven anomaly detection methods to provide system health awareness and decision support for flight controllers.

The IMS monitoring application was integrated with the NASA Mission Control data server software to access real-time telemetry in the ISS flight control room. Four IMS processes, one per CMG, are run on the ADCO flight control console to provide continuous monitoring. Once per second, when data is available, each IMS process will query the appropriate CMG knowledge base and return the amount of overall deviation, if any, from the nominal training data. It will also return the contribution of each individual parameter to any deviation to aid in isolating the source of any deviation. These IMS results are published back to the data stream for access and monitoring by other Mission Control software applications.

C. ISS Early External Thermal Control System

A study similar to the CMG1 analysis was performed using data from the ISS Early External Thermal Control System (EETCS). The EETCS was used to dissipate heat onboard ISS. Excess thermal energy from inside the ISS was transferred to liquid ammonia cooling loops in the EETCS. The heated ammonia was then circulated to radiators and cooled as thermal energy was released into space.

The EETCS included accumulators, which are containers that compensate for the expansion and contraction of ammonia due to temperature variation and also keep the ammonia in a liquid state via pressure regulation.[4] As documented in a January 2007 ISS anomaly report, the EETCS experienced conditions that resulted in increasing accumulator quantity sensor values for approximately 9 hours, followed by a sudden drop in accumulator quantities. The anomaly report states that symptoms of this event were noticed by mission controllers on the morning of January 9, 2007. After the fact, it was determined that the root cause was the formation of a bubble of gaseous ammonia within the normally liquid EETCS ammonia fluid loop. As the bubble grew, it appeared that accumulator quantities were increasing. The sudden drop in accumulator quantities occurred when the gaseous ammonia bubble essentially popped and dissipated back into the liquid ammonia. Although this ammonia bubble incident fell outside of normal EETCS operating conditions, there was no significant impact to ISS thermal control capabilities.

To see how a data driven monitoring system would react to the ammonia bubble event, archived data for 23 EETCS parameters was obtained from the time period surrounding the event. These parameters included EETCS pressures, temperatures, accumulator quantities, and pump speeds. 185 days of data collected between June and December 2006 were normalized with the Z-score technique and used to build an IMS monitoring knowledge base. This knowledge base was used to analyze data from January 1 through January 9, 2007, the day the anomaly was reported. It can be seen in the results graph (Fig. 7) that the IMS analysis detected the first signs of the anomaly near the end of day 2. As the ammonia bubble grew, the IMS deviation value steadily increased, until the bubble popped on the afternoon of day 9. (The large IMS spike on day 6 was caused by commanded temperature set point changes that briefly perturbed the EETCS.)

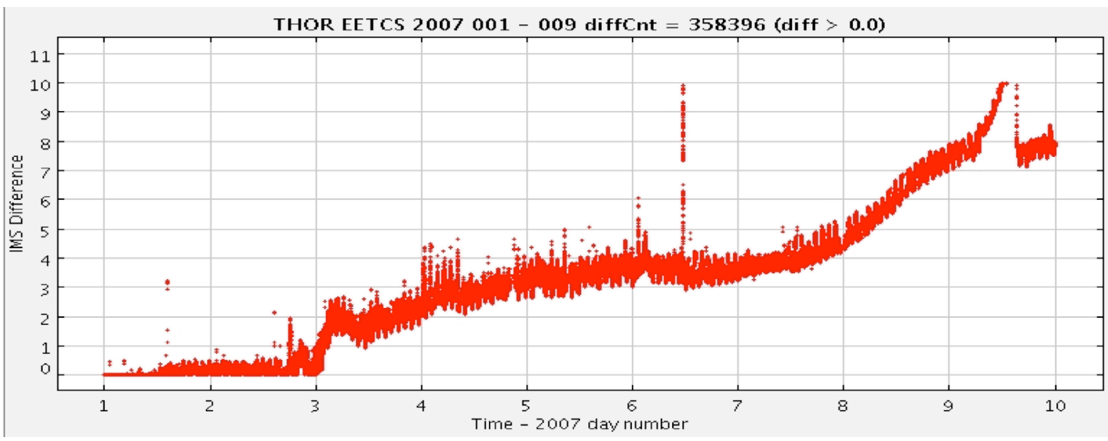

Figure 7. IMS monitoring results for EETCS ammonia bubble event

This EETCS study is another demonstration that data driven anomaly detection can be an effective tool for space operations. In this case, as in the CMG1 example, the data driven monitoring process was able to detect unusual parameter interactions early in the life of the anomalous event. Such timely information on unusual system behavior can be a useful decision aid for mission operations personnel as they monitor the health of their spacecraft systems.

IV. Summary and Future Work

Through practical application, it has been demonstrated that data driven system health monitoring can be useful in a space mission operations setting. Many spacecraft have extensive archives of telemetry data available that can be advantageously exploited by data mining methods. Two data mining tools, Orca and the Inductive Monitoring System (IMS), have been used to analyze data from the Space Shuttle and International Space Station to search for anomalous data points that could be indications of a system fault or damage to the spacecraft. Providing information on possible system anomalies in a timely manner provides controllers and mission support engineers with helpful information for decision support and enables more efficient and effective execution of their duties. The examples covered here and other similar experiments have shown the ability of these data driven anomaly detection techniques to characterize nominal interactions between multiple system parameters. This ability allows them to detect subtle anomalous parameter interactions that may not be apparent in more traditional single parameter data monitoring tools that are frequently used in space operations settings.

The utility and effectiveness of data driven system health monitoring methods have been demonstrated in three disciplines in NASA mission control, but the applications are not limited to just these examples. There are many areas with rich archived data repositories where these and similar techniques can be applied. Mission controllers from several additional ISS disciplines, including power management, communications, and life support, have expressed interest in developing similar system monitoring capability. Now that the software has been integrated with Mission Control data systems, expansion of the capability is primarily a matter of identifying relevant parameters to monitor and performing the archived data analysis. Following the successful demonstration of data driven tools, as described here, work has begun to develop real-time monitoring capability for ISS thermal control systems.

Eventually we plan to develop tools that allow mission control personnel to build and maintain their own data driven monitoring applications. Controllers will be able to specify which parameters to monitor, what time periods to include in the nominal training data, and any computations that should be performed on the raw telemetry data. The tool set will retrieve the desired archived training data, remove spurious data points using outlier detection, and build a new monitoring knowledge base and an appropriate monitoring application configuration to run on their control console.

A useful enhancement to the current monitoring software would be the ability to automatically detect operating mode changes in the monitored system and switch to a targeted monitoring knowledge base developed specifically for that mode. For instance, the ISS is flown in different orientations and configurations during different mission phases. The behavior of the CMGs can differ in the various configurations. Rather than building one large knowledge base per CMG that covers all cases, as in the current deployment, a separate knowledge base could be built from archived data collected during each ISS configuration, then consulted for real-time monitoring when the ISS is in that configuration. This would provide more accurate and efficient monitoring capability.

Another application of data driven monitoring to explore is the use of supervised learning methods to help identify fault signatures. If examples of fault behavior are available in the archived data, supervised learning algorithms, such as decision tree or support vector machine based techniques, may be able to analyze the data and distinguish between different types of fault behavior and normal operation. If the monitored system exhibits unusual behavior, fault characterizations from the supervised learning algorithm could help controllers identify the cause of the anomaly. These techniques could allow automated fault identification in cases that are too complex to be encoded using simpler schemes.

Acknowledgments

The author would like to thank the following collaborators who directly supported and contributed to the development and deployment of the mission control software described in this paper: From NASA Ames Research Center – Rick Alena, Charles Lee (EASI), Dawn McIntosh, Mark Schwabacher, Lilly Spirkovska, and William Taylor. From NASA Johnson Space Center – Christopher Craw (USA), Tatiana Dobrinskaya (USA), Ann Esbeck (USA), Jeff Fitch (USA), Tammy Gafka, Andrew Hillin (USA), Joshua Johnson, Keith Martin (TSI), Charles McCann, and James McCarroll (USA). Funding for this work was provided by the NASA Exploration Technology Development Program (ETDP) Integrated Systems Health Management project and the NASA Exploration Systems Mission Directorate (ESMD) Constellation Mission Operations project.

References

[1]Bay, S. D., and Schwabacher, M., "Mining Distance-Based Outliers in Near Linear Time with Randomization and a Simple Pruning Rule," *Proceedings of The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.

[2]Iverson, D. L., "Inductive System Health Monitoring," *Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI04)*, CSREA, Las Vegas, Nevada, June 2004.

[3]Columbia Accident Investigation Board Report, Volume 1, NASA, August 2003.

[4]International Space Station Thermal Control System Training Manual, NASA Document ISS TCS TM 21109, January 2004.

[5]Schwabacher, M., Oza, N., and Matthews, B., "Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring," *Proceedings of the AIAA Infotech@Aerospace Conference*, AIAA, Reston, VA, 2007.

[6]Martin, R. A., Schwabacher, M., Oza, N., and Srivastava, A., "Comparison of Unsupervised Anomaly Detection Methods for Systems Health Management Using Space Shuttle Main Engine Data," *Proceedings of the JANNAF Propulsion Meeting*, JANNAF, Denver, CO, 2007.