

# MSL Telecom Automated Anomaly Detection

Ryan Mukai, Zaid Towfic, Monika Danos, Mazen Shihabi, David Bell  
Jet Propulsion Laboratory, California Institute of Technology  
4800 Oak Grove Drive  
Pasadena, CA 91109

**Abstract**—The Mars Science Laboratory (MSL) Telecom Operations Team at the Jet Propulsion Laboratory (JPL) has implemented a machine learning system in order to automate the anomaly detection process as a part of daily operations. Machine learning enables reliable detection of anomalies in Telecom-related telemetry and automated reporting of Telecom subsystem status, resulting in an 90% reduction in team workload and improved anomaly detection reliability.

At present, machine learning methods are used to detect:

1. Anomalous long-term trends in telemetry data
2. Anomalous time-domain evolution of telemetry values

Both types of anomalies pose their own unique challenges that are addressed in different ways. In the first case, long term trending of daily minima, maximum, and mean telemetry values in temperatures, currents, voltages, and radio frequency (RF) power levels is used in addition to hard threshold safety checks to look for changes in long-term equipment health and performance. Long-term trending methods allow for ordinary seasonal variations in these quantities caused by temperature changes over the course of the Martian year while allowing operators to determine whether current performance remains in line with historical values from previous years. Changes in long-term trends can provide important insights into the health and status of the rover’s on-board systems as well as valuable early warning if subtle degradation begins to take hold. But while trending of daily statistics is valuable, it does not detect anomalies in the short-term time evolution of data over the course of minutes or hours during a day, and this task is handled with short-term shape analysis. Principal components analysis (PCA) has been found to provide robust detection of short-term anomalies, and several examples of the use of PCA to detect actual anomalous events will be provided here. In using PCA, we use both the percentage of explained variance and also a log likelihood test on the PCA expansion coefficients to flag telemetry data for human review. Previous work in the field of spacecraft anomaly detection includes [1] for MSL and [2] for some other JPL missions.

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. MSL TELECOM SUBSYSTEM.....	2
3. DETECTION OF ANOMALOUS EH&A DATA .....	2
4. LONG-TERM TREND CHECKS .....	2
5. EH&A TIME DOMAIN CHECKS .....	4
6. SUMMARY .....	5
REFERENCES .....	5
BIOGRAPHY .....	6

## 1. INTRODUCTION

The MSL Engineering Operations (EO) team is responsible for analyzing data from spacecraft subsystems and assessing both health and performance. Subsystem assessments are an important input to the daily planning process in spacecraft operations. If all subsystems are healthy and performing properly, ordinary planning can proceed. Plans are changed appropriately if health and performance issues are detected with any of the subsystems.

The telecom subsystem handles all communications between MSL and Earth and plays a vital role in mission success. It enables commands to be sent to the spacecraft, scientific data to be sent back to Earth, and important engineering health and safety telemetry, including telemetry on itself, to be sent back to Earth as well. Anomalies in this subsystem can not only prevent downlink of scientific data but can also endanger the mission itself since mission success depends on both the ability to command the spacecraft and on the ability to receive science and engineering data from the spacecraft. Hence, timely and reliable detection of anomalous conditions in Telecom plays an important role in mission success.

The two primary types of data used for this purpose are engineering housekeeping and accountability (EH&A) data and event record (EVR) data. Anomalous EVRs and anomalous EH&A must be detected and reported in a timely fashion in order to detect changes in health and performance and in order to allow the mission operations team to respond appropriately.

Examples of EVR data would include events such as an Electra Lite Transceiver (ELT) being initialized in preparation for a communications relay session with a UHF relay orbiter or the start of a communications window or the receipt of data from either a UHF relay orbiter or the Earth. Examples of EH&A data would include temperature, voltage, current, power consumption, received signal-to-noise ratio (SNR), and output radio frequency (RF) power as a function of time.

On a given Martian day (Sol) Telecom must process approximately 3000 EVRs in addition to time domain data from 150 EH&A channels. A manual review of this much data is infeasible, and prior to the use of automation it was necessary to limit the number of channels reviewed. Not only did the daily review process require four to five hours per day, but anomalies were often missed. Indeed, some anomalies in the data were only discovered in retrospect after automated anomaly detection was introduced! Hence, automated anomaly detection results in greatly improved reliability, since all anomalies previously detected by human Telecom operators are consistently detected by the automated tools as well. In addition, many anomalies previously missed by human operators have been discovered by the automated tools. Moreover, the daily assessment process takes less than 10 minutes with the aid of these tools.

## 2. MSL TELECOM SUBSYSTEM

A detailed description of the MSL Telecom Subsystem is provided in [3]. This subsystem consists of the X-Band subsystem for direct communication with Earth and the UHF subsystem for communications with Mars relay orbiters, including the Mars Reconnaissance Orbiter (MRO), Odyssey (ODY), Maven (MVN), Mars Express (MEX), and the Trace Gas Orbiter (TGO). Most commanding is done by X-band uplink directly from Earth via the Deep Space Network (DSN), but nearly all data downlinks are via UHF relay orbiters.

The X-Band subsystem at a high level contains:

- Small Deep Space Transponder (SDST): The X-Band communications radio capable of full duplex communications with the DSN on Earth.
- Solid State Power Amplifier (SSPA): The X-Band power amplifier that amplifies the SDST's output signal for transmission to Earth.
- Waveguide Transfer Switch (WTS): Although two WTS units are in the X-Band subsystem, in the surface phase of the mission only one unit is used. This unit selects between the two X-Band antennas described in the next two items.
- High Gain Antenna (HGA): This is a fully steerable antenna used for higher communication rates at X-Band.
- Rover Low Gain Antenna (RLGA): This is a nearly omnidirectional antenna used for low rate X-Band communications, including "safe mode" emergency communications in anomaly situations.

The UHF subsystem at a high level contains:

- Two ELT radios: These are two identical UHF communications radios. ELT-A is the prime radio, and ELT-B is the backup.
- A UHF Coaxial Transfer Switch (CTS): This switch is used to control which radio's input and output will connect to the UHF antenna
- Rover UHF Antenna (RUHF): This is the only UHF antenna available in the surface phase of the mission.

Component health, especially the health of the UHF radios, the SDST, and the SSPA, must be continuously monitored. Additionally, subsystem performance must be continuously trended in order to detect unexpected changes in performance that may indicate a change in hardware health. Each Sol is approximately 24 hours, 39 minutes, and 25 seconds long. Each session in which the rover communicates either with Earth via X-Band or with a UHF relay orbiter is called a communications window. Health telemetry is trended either on a per-Sol basis or on a per window basis.

It is important to note that some UHF relay windows, especially for MRO or ODY, tend to occur within a certain band of local mean solar time (LMST) ranges. For example, all MRO morning (in LMST) windows will occur within a certain LMST time band each Sol, and all MRO afternoon (in LMST) windows also occur within a certain LMST time band each Sol due to the sun synchronous nature of the MRO orbit. This statement is not true of MVN and TGO, however.

Most UHF data are trended on a per window basis because the ELT-A (primary UHF) radio is only powered on and collecting health and performance data during the course of a window. Moreover, for MRO and for ODY, these data are often separated for AM and PM passes, and this is especially important for temperature and RF power telemetry.

On the other hand, most X-Band data tend to be trended on a Sol-by-Sol basis since there is normally only one X-Band window per Sol and since the X-Band radio is powered on and listening whenever the rover is powered on, except in cases where the rover is engaged in UHF communications, in which case the rover is powered on but the X-Band radio is powered off.

## 3. DETECTION OF ANOMALOUS EH&A DATA

EH&A data provides vital insight into the health and performance of the MSL spacecraft. There are three types of anomaly detection applicable to EH&A data:

1. *Hard limits*: For example, a given temperature may have an acceptable range from -20 degrees Celsius to +50 degrees Celsius. A violation of a hard limit is always detected and reported by the automated system. Hard limit detection is an established technique used for many decades in mission operations.
2. *Long-term trend limits*: Even if a quantity, such as a temperature, is well within the allowable hard limits, that quantity may be out-of-trend. For example, the daily minimum, mean, and maximum of a given temperature channel will often exhibit seasonable trends. A maximum temperature of +35 degrees Celsius may be entirely normal during the Martian summer but may signify a hardware health issue during the Martian winter, leading to the need to detect data that are out-of-trend.
3. *Time domain data shape*: Even if the minimum, maximum, and mean quantities taken over a day or taken over a communications session are well within both safety limits and long-term trend bounds, anomalous evolution of the data as a function of time may signify a health or performance problem, necessitating detection of anomalous time domain shapes.

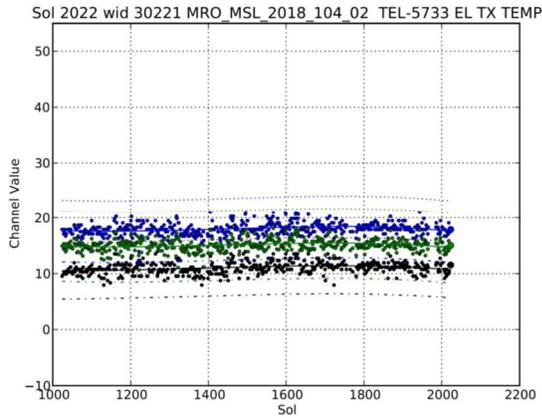
While hard limits are straightforward, long-term trend bounds and time domain data shape checks represent items that must be learned from actual data over time. This paper focuses on long-term trend bounds and time domain data shapes.

## 4. LONG-TERM TREND CHECKS

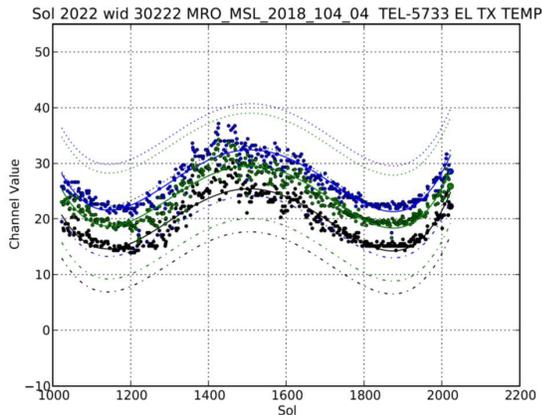
On either a daily basis or on a per-communications-window basis, MSL telecom automatically gathers data on the minimum, mean, and maximum values of telecom-related EH&A channels. In Figures 1 and 2, we are trending a temperature within the ELT-A UHF radio for morning UHF relay sessions and for afternoon MRO UHF relay sessions, respectively. In each Figure, we trend the per relay session minimum, mean, and maximum temperature in degrees Celsius.

We trend the minimum (black), mean (green), and maximum (blue) temperatures on a window by window basis. We observe a large difference between the thermal behavior of morning and afternoon UHF windows with afternoon windows showing a very strong dependence on Martian season. However, all temperatures lie well within hardware safety limits. Colored dots are actual data points while broken lines signify a four-sigma bound.

Although morning UHF sessions in Figure 1 exhibit very steady behavior in minimum, mean, and maximum temperatures, there is a very clear seasonal variation in Figure 2. Indeed, it is safe to say that a maximum temperature of +33 degrees Celsius would be well within normal behavior for



**Figure 1.** UHF Radio Morning Temperatures (deg C). Minimum (black), Mean (green), and Maximum (blue) temperatures are plotted.

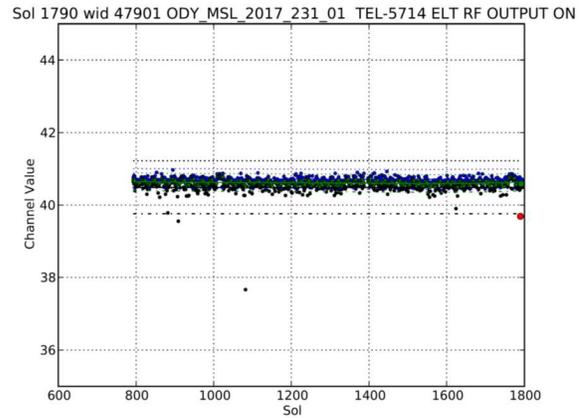


**Figure 2.** UHF Radio Afternoon Temperatures (deg C). Minimum (black), Mean (green), and Maximum (blue) temperatures are plotted.

this radio during the Martian summer for an afternoon UHF session. However, even in Figure 2 this would be clearly anomalous during a Martian winter afternoon UHF session. It would also be anomalous during the morning MRO UHF relay sessions, regardless of season, as illustrated by Figure 1.

Similarly, a peak temperature of +17 degrees Celsius would be entirely normal for a morning relay session with MRO anytime during the Martian year as shown in Figure 1. But as shown in Figure 2, this would clearly be a significant outlier for the summer during an afternoon relay session. Indeed, +17 degrees Celsius would be an anomalous minimum temperature during an afternoon session in the Martian summer.

The radio temperature figures illustrate an important point regarding anomaly detection: it depends heavily on historical trends. A temperature that is entirely normal in one context could be abnormal in another.



**Figure 3.** Minimum value of RF Power Output in dBm Long Term Trend

In the case of ELT-A radio temperatures, separate data series have been defined for morning and afternoon passes due to large differences in morning and afternoon thermal behaviors, especially as a function of Martian season.

Once sufficient data have been gathered, the minimum, mean, and maximum temperatures are trended as a function of the Martian day (Sol). Each of the three data series has a fifth-order trending polynomial fitted to it, and the standard deviation of the error (actual data minus polynomial fit) is computed. This error is not always Gaussian, so pursuing a Gaussian-fit statistic test as a bound does not always yield an accurate false alarm probability. In practice, however, a Gaussian fit with a three-sigma or four-sigma bounds have often worked well.

If a given data point lies within the defined three- or four-sigma bound of the trend line, and if it also lies within hardware safety limits, it is classified as “normal.” A point that only lies out-of-trend but not outside of hardware safety limits is flagged for review by a human operator with a notice flag in the automated report. A point that violates a hard safety limit, regardless of whether a trending bound has been violated, is flagged as a red warning for human review in the automated report.

An example of anomaly detection involving UHF radio RF power output is shown in Figure 3. This is a plot of minimum, mean, and maximum RF power output for each UHF relay window over 1000 Sols. We observe a point in red highlighting the anomaly on Sol 1795, in which the minimum RF power output fell to slightly under 40 dBm. Although this is within the acceptable safety limits for a running RF amplifier for the Electra Lite radio, it is also out-of-trend, which is why it was flagged. Please note that the system flags out-of-trend points for the present day’s report. We see in the same figure that another point slightly below 38 dBm is *not* flagged because it was not part of the present day’s report, and that point had been identified as a harmless outlier nearly two years earlier. In general, the software searches for anomalies occurring on reporting day and does not place old anomalies in the report since reporting is handled on a daily basis.

## 5. EH&A TIME DOMAIN CHECKS

Although the foregoing long-term trend checks are good at catching abnormal minimum, maximum, or mean data, such checks can easily miss anomalous time domain evolution of data. If we plot an EH&A quantity as a function of time, then an abnormal shape, even one with entirely normal minimum, maximum, and mean value, must also be flagged for review.

There are many possible approaches to detected such shape anomalies. In this paper, we describe a principal component analysis (PCA) approach.

The key observation to the approach is that when the data shape as a function of time is not “white noise,” we expect that the autocovariance matrix of the time-domain signal will be non-diagonal. This is due to neighboring time-domain samples being relatively highly correlated over small time differences.

For this reason, we first compute the autocovariance matrix of the data to be tested. This is then followed by the extraction of eigenvalues and eigenvectors of the autocovariance matrix to yield an orthonormal basis over which each time-domain vector can be expanded. The key insight, gained by analyzing normal data, is that for a normal data shape, most (i.e. 95% or more) of the “energy” of the vector expansion will be contained in only the first few coefficients of expansion. This implies that projecting onto these “dominant” eigenvectors should preserve the energy of the test signal. Previously collected and analyzed data are used to determine the number  $n$  of expansion coefficients to use over the  $N$ -dimensional basis (i.e., how many eigenvector filters to consider “dominant”).

The preprocessing steps are listed in Algorithm 1.

---

### Algorithm 1 Anomaly Detection Preprocessing

---

**Require:** test sample  $\mathbf{x} \in \mathbb{R}^N$

- 1: Compute the mean  $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$  by averaging over adjacent time samples.
  - 2: Compute the autocovariance matrix  $\boldsymbol{\Lambda} = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]$  by averaging over adjacent time samples.
  - 3: Obtain the  $N$  eigenvalues  $\{e_i\}$  and corresponding eigenvectors.  $\{\mathbf{v}_i\}$  and sort in order of descending eigenvalues. Since  $\boldsymbol{\Lambda}$  is an autocovariance matrix, its eigenvalues are real and non-negative, making this sorting operation possible. Furthermore, the eigenvectors are orthogonal, forming a useful  $N$  dimensional basis: the PCA basis.
  - 4: Select  $n$  such that an explained variance of at least 95% is achieved over the input training set of vectors.
  - 5: Compute the projection vector  $\mathbf{y} = \mathbf{V}^T \mathbf{x}$ , where  $\mathbf{V} \in \mathbb{R}^{N \times n}$  denotes the top  $n$  eigenvectors of the autocovariance matrix  $\boldsymbol{\Lambda}$ .
- 

The number of principal components  $n$  is selected to achieve at least 95% explained variance across the training set of input vectors, and this number varies for each channel, with 3 to 10 principal components being typical for most channels.

Now,  $\mathbf{y}$  denotes the vector of PCA expansion coefficients. Even if the explained variance threshold is passed successfully, it is necessary to determine whether or not  $\mathbf{y}$  itself is out-of-family. In practice, for the channels used by the MSL Telecom subsystem for health and performance analysis, it is found that  $\mathbf{y}$  does not form multiple widely spaced clusters in the 3- to 10-dimensional vector space of expansion coefficients but usually forms a single cluster. So, in our approach:

---

### Algorithm 2 Anomaly Detection Approach

---

- 1: We compute  $\boldsymbol{\mu}_y = \mathbb{E}[\mathbf{y}]$ , the expected value.
- 2: We compute  $\boldsymbol{\Lambda}_y$ , the autocovariance matrix of  $\mathbf{y}$ .
- 3: Assuming a vector Gaussian probability density function, the log-likelihood is (to within an additive constant):

$$L = -(\mathbf{y} - \boldsymbol{\mu}_y)^T \boldsymbol{\Lambda}_y^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \quad (1)$$

- 4: Declare anomalous if

$$L < \gamma \quad (2)$$

or

$$\frac{\|\mathbf{y}\|_2^2}{\|\mathbf{x}\|_2^2} < \delta \quad (3)$$


---

Observe that  $\boldsymbol{\Lambda}_y$  is a different, and much smaller, autocovariance matrix than the one computed over the raw data vectors prior to PCA basis projection in Alg. 1.

Log-likelihoods in (1) are computed over the training set of coefficient vectors. Out of thousands of log-likelihoods, far fewer than one-percent are anomalous, and a percentile threshold under 1% is used to determine a log-likelihood threshold. In addition, the threshold in (3) ensures that the explained variance is sufficiently high (i.e.,  $\delta = 90\%$ ).

An example of anomaly detection via data shapes is shown in Figures 4 through 6, which plot successful frame UHF transmissions over a link between MSL and MRO. In a typical link, the frame counter begins at zero and climbs upward during the relay pass as frames are transmitted. Once the link finishes as the orbiter sets on the Martian horizon, the frame counter stops incrementing and remains at its final value until the radio is powered off. Since the link operates at maximum efficiency when the orbiter is high in the sky, the frame counter typically climbs steeply during this period. The resulting “S-curve” is shown in Figure 4.

In anomalous cases due to software resets in the radio, one of two behaviors may appear. Sometimes the frame counter will immediately drop to zero as shown in Figure 5. In other cases, the frame counter may actually climb to a very high but completely false value before finally dropping to zero, as shown in Figure 6. Both of the anomaly cases differ visually from the typical case shown in Figure 4.

Applying algorithm 1 with three eigenvectors, the data shape in Figure 4 can be described with an explained variance exceeding 98%, which is typical for nearly all cases involving this telemetry channel. However, for Figure 5, the explained variance was 13%. It was only 2% for Figure 6.

In practice, the explained variance threshold method of equation (3) has been highly successful in detection of real-life anomalies, as in the preceding example. False positives are rare, but they can be dispositioned quickly by a human analyst when they are flagged by the system. The authors find it interesting that the log-likelihood in equation (1) has not, as of this writing, caught any actual anomalies. Nevertheless, the authors also believe that an anomaly detection system that does not analyze  $y$  beyond simple explained variance would be incomplete, and there remains a possibility that the log-

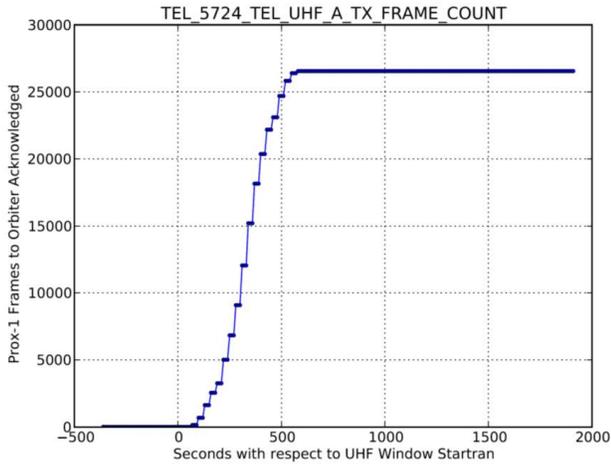


Figure 4. UHF Relay Packet Counter

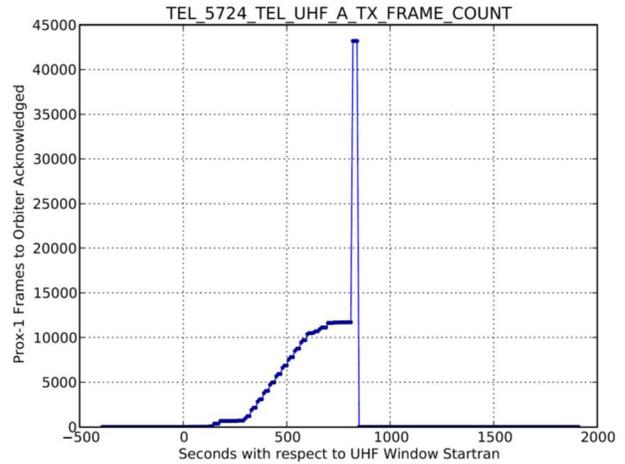


Figure 6. UHF Relay Packet Counter

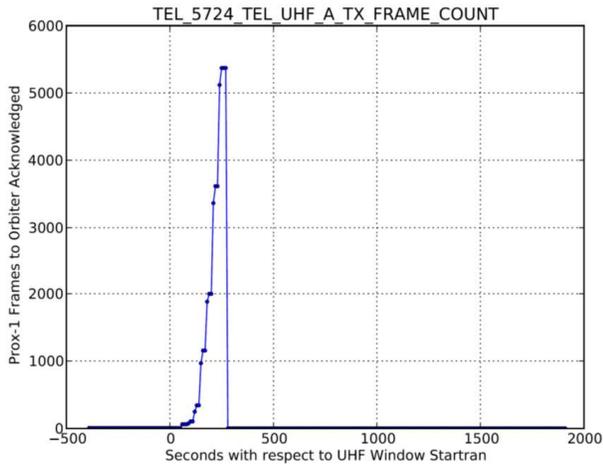


Figure 5. UHF Relay Packet Counter

likelihood test will catch unforeseen anomalies in the future although, again, all time-domain anomalies thus far were discovered through explained variance using equation (3).

## 6. SUMMARY

Automated application of long-term trending analysis and PCA-based anomaly detection enables rapid, reliable detection of anomalous data from the MSL Telecom subsystem, significantly reducing the time for a human analyst to detect an anomaly from hours to under 10 minutes. Moreover, since human analysts were unable to process 150 channels within a four- to five-hour period, the new system enables a far more thorough review of the data than would be possible without automatic anomaly detection.

There are multiple possible future directions, including, but not limited to, the following:

- Related channels, including closely related temperature channels, could be combined to generate combined channel vectors  $\mathbf{x}$  to be assessed using PCA. In this way, deviations in typical channel-to-channel behavior can be detected more reliably since the autocovariance matrix used to generate the PCA expansion would now include channel-to-channel covariances as well.
- Use of neural-network based autoencoder models [4] may provide an even more powerful method of checking for time-domain anomalies than the present system does. This could be applied to raw data vectors  $\mathbf{x}$  or to PCA coefficient vectors  $\mathbf{y}$ .
- Dictionary learning is another approach that has yet to be tried on this problem [5].
- For time-domain data anomalies, some exploratory work using the LSTM neural network approach of Hundman et al. [2] is in progress.

While the work presented here has already produced large improvements in anomaly detection reliability along with substantial workload reductions on human personnel, it is very much a starting point, and new approaches will be explored and introduced over time.

## ACKNOWLEDGEMENT

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

## REFERENCES

- [1] M. M. Fernandez, Y. Yue, and R. Weber, "Telemetry anomaly detection system using machine learning to streamline mission operations," in *2017 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, Sep. 2017, pp. 70–75.
- [2] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Söderström, "Detecting spacecraft anomalies

using lstms and nonparametric dynamic thresholding,” *CoRR*, vol. abs/1802.04431, 2018. [Online]. Available: <http://arxiv.org/abs/1802.04431>

- [3] A. Makovsky, P. Ilott, and J. Taylor, “Mars science laboratory telecommunications design,” 2009. [Online]. Available: [https://descanso.jpl.nasa.gov/DPSummary/Descanso14\\_MSL\\_Telecom.pdf](https://descanso.jpl.nasa.gov/DPSummary/Descanso14_MSL_Telecom.pdf)
- [4] S. Ger and D. Klabjan, “Autoencoders and generative adversarial networks for anomaly detection for sequences,” *CoRR*, vol. abs/1901.02514, 2019. [Online]. Available: <http://arxiv.org/abs/1901.02514>
- [5] H. Ren, W. Liu, S. Olsen, S. Escalera, and T. Moeslund, “Unsupervised behavior-specific dictionary learning for abnormal event detection,” 09 2015.

## BIOGRAPHY



**Ryan Mukai** earned a B.S. and an M.S. in electrical engineering in 1997 and a Ph.D. in electrical engineering in 2003, all from the University of California at Los Angeles. He has been working as a telecommunications engineer at the Jet Propulsion Laboratory since 1999. Dr. Mukai is presently supporting the Mars Science Laboratory project. He previously supported the Mars Exploration

Rovers, Dawn, Cassini, and Juno.



**Zaid Towfic** holds a B.S. in Electrical Engineering, Computer Science and Mathematics from the University of Iowa. He received his Electrical Engineering M.S. in 2009 and Ph.D. in 2014, both from UCLA, where he focused on signal processing, machine learning, and stochastic optimization. After receiving his Ph. D., Zaid joined the MIT Lincoln Laboratory where he

worked on distributed beam forming and geolocation, interference excision via subspace methods, simultaneous communication, and electronic warfare. Zaid joined the Jet Propulsion Laboratory in January of 2017 and has been focused on machine learning and signal processing efforts.



**Monika Danos** obtained her BSEE from Cornell University (Ithaca, New York) in 1995 and her MSEE with a concentration in communications from the University of Southern California (Los Angeles) in 1998. In 1999 she joined the Jet Propulsion Laboratory where she has worked as a telecommunications systems engineer on various projects including the Mars Exploration Rovers, Juno, and

Mars Science Laboratory.



**Mazen Shihabi** holds B.S. and M.S. from University of Southern California, and PhD from University of California at Irvine, all in Electrical Engineering with emphasis on digital communications. He joined the Jet Propulsion Laboratory (JPL) in 1990 as a member of the Communications Research Section where he worked on modem design, bandwidth efficient modulations, and arraying techniques applicable to flight and ground communication systems. From 1996 to 2008, he worked in the commercial sector on telecom technologies related to aerospace, cellular service, and remote patient monitoring. Dr. Shihabi rejoined the JPL, Flight Communications Section in 2008 where he is currently the Group Supervisor of the Communications Architecture and Operations and focusing on new communications technologies and services.



**David Bell** is the section manager for the Flight Telecommunications Section at JPL which designs, builds, and operates telecommunications subsystems on JPL spacecraft. The section also designs and builds unique antennas, high voltage power supplies and power amplifiers for radar science payloads. Beyond support to existing flight missions the section is involved in technology development efforts for antennas, radios and transponders, RF amplifiers, power supplies for flight and automated ground support equipment for test. Prior to his role as section manager, Mr. Bell was group supervisor the Flight Telecom Systems Engineering and Operations Group and was the system design lead for the first software/firmware defined deep space flight radio, *Electra*, that has been deployed on multiple NASA and ESA flight missions to Mars.