ATHMoS: Automated Telemetry Health Monitoring System at GSOC using Outlier Detection and Supervised Machine Learning

Corey O'Meara[*]   Leonard Schlag[†]   Luisa Faltenbacher[*]   Martin Wickler[‡]

Deutsches Zentrum für Luft- und Raumfahrt e. V., German Aerospace Center

Münchener Straße 20, 82234 Weßling, Germany

Knowing which telemetry parameters are behaving accordingly and those which are behaving out of the ordinary is vital information for continued mission success. For a large amount of different parameters, it is not possible to monitor all of them manually. One of the simplest methods of monitoring the behavior of telemetry is the Out Of Limit (OOL) check, which monitors whether a value exceeds its upper or lower limit. A fundamental problem occurs when a telemetry parameter is showing signs of abnormal behavior; yet, the values are not extreme enough for the OOL-check to detect the problem. By the time the OOL threshold is reached, it could be too late for the operators to react. To solve this problem, the Automated Telemetry Health Monitoring System (ATHMoS) is in development at the German Space Operation Center (GSOC). At the heart of the framework is a novel algorithm for statistical outlier detection which makes use of the so-called Intrinsic Dimensionality (ID) of a data set. Using an ID measure as the core data mining technique allows us to not only run ATHMoS on a parameter by parameter basis, but also monitor and flag anomalies for multi-parameter interactions. By aggregating past telemetry data and employing these techniques, ATHMoS employs a supervised machine learning approach to construct three databases: Historic Nominal data, Recent Nominal data and past Anomaly data. Once new telemetry is received, the algorithm makes a distinction between nominal behaviour and new potentially dangerous behaviour; the latter of which is then flagged to mission engineers. ATHMoS continually learns to distinguish between new nominal behavior and true anomaly events throughout the mission lifetime. To this end, we present an overview of the algorithms ATHMoS uses as well an example where we successfully detected both previously unknown, and known anomalies for an ongoing mission at GSOC.

I.   Introduction

In recent years, many operations teams, research groups, and space agencies have been looking to further automate the telemetry checking process as data retrieval, storage, and processing has become more ubiquitous. The human effort to manually inspect tens of thousands of parameters in detail is simply not an option. Additionally, a simple automated process such as the Out-Of-Limit (OOL) check, which only verifies if a certain parameter crosses soft(hard) upper or lower thresholds, is not robust enough to detect the various types of anomalies that may occur in the telemetry data. Many teams within the last 15 years have taken a different approach to detect more subtle changes in the data. Here, we focus on a select few to explain some fundamental concepts that we adapt in our approach.

---

[*]Mission Planning System Engineer, Mission Operations Department, German Space Operations Center
[†]MCDS (Mission Control and Data handling System Engineer) Mission Operations Engineer, Mission Operations Department, German Space Operations Center
[‡]Deputy Head of Mission Operations Department, German Space Operations Center

American Institute of Aeronautics and Astronautics

One particular system is the Inductive Monitoring System (IMS) developed at NASA-Ames[1] which uses clustering to extract a data model of the nominal telemetry data. For each timestamp, d telemetry parameter values are combined to create a d-dimensional vector, the collection of which are then ran through a clustering algorithm to identify nominal "regions" in the vector space where nominal data was usually observed. When new telemetry data arrives, it is then inserted into the vector space and its distance to the nearest nominal region is measured. This distance provides a measure of a points "outlierness" relative to the nominal regions. Since the IMS system was introduced, several evolutions of it have been created such as the Anomaly Monitoring Inductive Software System (AMISS) which is a collaboration between the Johnson Space Centre (NASA) and Ames to support the International Space Stations flight operations.[2]

Other existing anomaly detection systems implemented by NASA include the distance-based Orca,[3] Ames Research Centers support vector machine based Mariana,[4] and the Jet Propulsion Laboratories hybrid method Dynamical Invariant Anomaly Detector (DIAD).[5] For an overview of Orca and IMS and their application to NASA mission operations see Ref. 6, the references therein, and Ref. 7.

Another such technology was recently developed at the European Space Agency (ESA) by the Advanced Operations Concepts Office called the Novelty Detection System.[8] In this approach, telemetry data is first decomposed into sequential time intervals. For each time interval, a 4-dimensional feature vector is calculated and compared to a pre-labeled set of nominal data points. This comparison to the nominal data clusters is done via the density-based Local Outlier Probability (LoOP) algorithm[9] which assigns a probability to each new point that novel telemetry behaviour occurred during its respective time interval. As an application of the technology, the Columbus Control Centre (Col-CC) team at GSOC - DLR has implemented a framework called Sibyl[10] which automatically preprocesses the 4-dimensional feature vectors associated to each time interval for past data using the Density-Based Spacial Clustering of Applications with Noise (DBSCAN)[11] algorithm. Once the clusters of past data which correspond to nominal data are filtered out, they are then fed into the Novelty Detection system to compare to the new telemetry data.

ATHMoS is similar to the above systems in the sense that it computes large (dimension greater than 4) feature vectors from time intervals in the telemetry data, uses DBSCAN to filter nominal training data from past telemetry data but then creates three Classification Databases; Historic Nominal, Recent Nominal and Anomalies to compare the new telemetry data to. The comparison is done using a novel algorithm developed at GSOC named Outlier Probability Via Intrinsic Dimension (OPVID). OPVID has a central dependency on a statistical quantity known as the Intrinsic Dimension to obtain a probability that the time interval associated to a distinct feature vector contains a true anomaly. Furthermore, it allows us to analyse multiple telemetry parameters simultaneously, similar to the IMS and the AMISS systems, without significant loss of accuracy. To this end, we first must describe the notion of Intrinsic Dimensionality of a data set in order to define the OPVID algorithm.

II.   Outlier Probability Via Intrinsic Dimension

There are many different methods to determine outliers of a data set[12–14] and many ways to apply these techniques to identify anomalies in different types of data (see for example Refs. 15-17 and the references therein). As outlined in Section I, some of these anomaly detection methods have specifically been tailored to the space industry to investigate time series telemetry data. In this section, we show that a generalisation of an outlier detection method introduced in Ref. 18 can be used for such a purpose.

II.A.   Intrinsic Dimension of a Data Set

A central concept of data mining and machine learning is that of comparing data. That is, given some data set, we want to analyse its properties and use this analysis to create a predictive model of how it will change over time, or simply compare it to other existing data sets to see their similarities. The effectiveness of these types of predictive and comparison algorithms depend on the nature of the descriptive information gathered from the data sets. The amount of these features that are distilled from the data can vary greatly and the cardinality is usually referred to as the dimensionality of the data. As one can imagine, with increasing dimension, many of the most common outlier detection methods fail to be as effective as they are in lower dimensional spaces since data is distributed in a sparse manner. This observation is well known as the "curse of dimensionality" and ways to minimise its effects is currently a highly active field of research (see for example Refs. 19-21).

American Institute of Aeronautics and Astronautics

One possible solution to the problem is to make use of the concept of the Intrinsic Dimensionality (ID). Roughly speaking, the ID of a data set is the number of features (i.e. degrees of freedom) that are needed to accurately describe the data. There are several approaches to estimating the ID of a data set such as subspace projection methods[22–24] or fractal/geometric methods,[25–28] however here we are concerned with a recent ID model called the Continuous Intrinsic Dimensionality.

Suppose we have data points contained in a vector space which is equipped with a distance measure (e.g. the Euclidean norm). The distances from each point to some fixed query point q will all be positive numbers and thus a variable which takes on any such distance can be represented as a (continuous) random variable $\mathbf{X}$. Since $\mathbf{X}$ is a (continuous) random variable, we have an associated probability density function $f_{\mathbf{X}}(x)$ which describes the probability that the variable will be a certain value and additionally, the cumulative density function $F_{\mathbf{X}}(x) := P(\mathbf{X} \leq q)$ which describes the probability that $\mathbf{X}$ will take a value less than or equal to q. Using these functions we can define an abstract definition of the continuous ID[29,30] as

$$\mathrm{ID}_{\mathbf{X}} := \lim_{x \to 0^+} \frac{x \cdot f_{\mathbf{X}}(x)}{F_{\mathbf{X}}(x)} \ . \tag{1}$$

Remarkably, if we have clustered data in a d-dimensional subspace of an n-dimensional vector space then the continuous ID *is equal to the dimension of the cluster*, d. For this fact as well as others pertaining to details about the estimation and properties of both ID and continuous ID, we refer the reader to the above references as well as Refs. 31-34. For the purposes of this paper, it suffices to say that various methods of estimating the (local/continuous) ID of a data set exist as detailed in Ref. 35 and usually depend on a k-nearest neighbour parameter to compare local distances of points to one another. We call such a parameter the *context set k* and denote it k_c.

For clarity we provide one particular approximation to make the application of the Eq. (1) more concrete. Let q be the query point whose ID score we wish to determine. Set $k_c = 10$ such that we want to approximate the points ID score relative to its 10-nearest-neighbours and let $x_1, x_2, \ldots, x_{10}$ denote the (increasing) ordered distances from the point q to its 10-nearest-neighbours. The Maximum-Likelihood-Estimator of the ID score of the point q is given by

$$\mathrm{ID}(q) = -\left(\frac{1}{10} \cdot \sum_{i=1}^{9} \ln \left(\frac{x_i}{x_{10}} \right) \right)^{-1} \ . \tag{2}$$

II.B. Application to Outlier Detection

The insight that the (continuous) ID score of a point q is equal to the dimension of the cluster it belongs to and increases if the point lies outside a cluster was a central motivation behind the Intrinsic Dimensionality Outlier Score (IDOS).[18] In the associated reference, the authors first approximated the infamous Local Outlier Factor (LOF) outlier score[36] in terms of ID and then modified this approximated LOF formula to only be dependent on the ID score of each data point. This yielded the IDOS score

$$\mathrm{IDOS}(q) := \frac{\mathrm{ID}(q)}{k_r} \sum_{p \in N_{k_r}(q)} \frac{1}{\mathrm{ID}(p)} \ , \tag{3}$$

where $k_r \geq 1$ such that $N_{k_r}(q)$ is the set of k_r-nearest neighbours of the query point q and $\mathrm{ID}(p)$ is the ID Score (see for example the estimator given by Eq. (2)) of the point p. Similar to the LOF score, we say that a point q is an outlier if $\mathrm{IDOS}(q) >> 1$ and an inlier if $\mathrm{IDOS}(q) \approx 1$. Furthermore, the authors provided a detailed accuracy comparison and efficiency tests against current state-of-the-art outlier detection methods LOF, SOD,[37] and FastABOD[38] where they concluded that IDOS was superior in terms of both accuracy and efficiency.

For our purposes, the IDOS algorithm is insufficient due to the fact that we want to run such an outlier detection algorithm for *tens of thousands* of different data sets and for each data set the notion of an outlier score "much much greater than 1" varies. That is, for one particular telemetry data set, an outlier score for a query point q may correspond to an outlier if its score is 5, whereas for another telemetry data set, a value of 15 may only indicate that the point is truly an outlier. These differences imply we must have a normalized
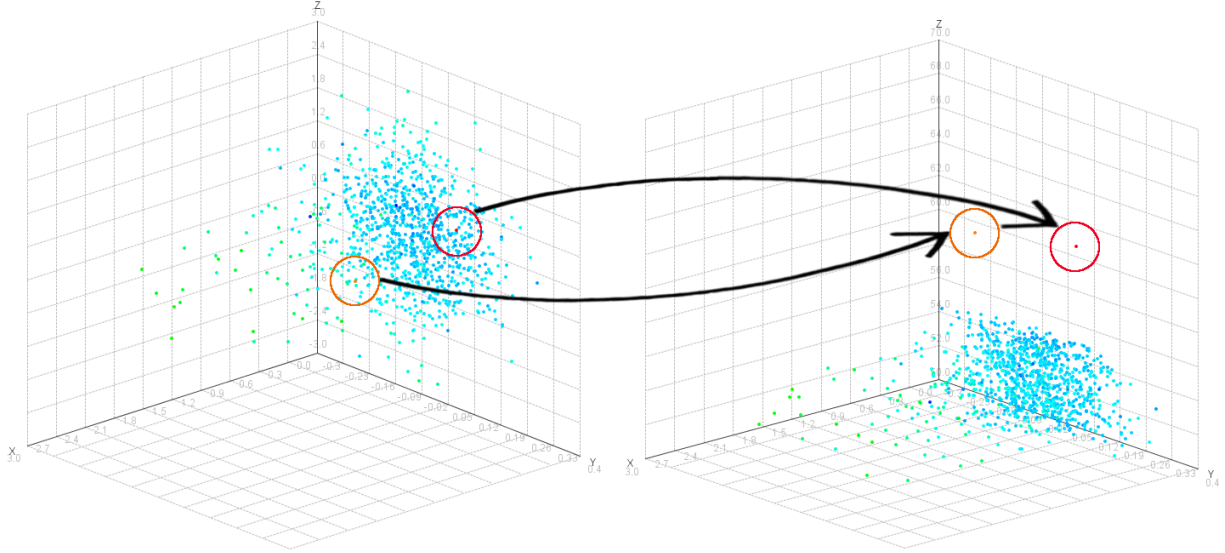
American Institute of Aeronautics and Astronautics

Figure 1. A 7-dimensional data set projected into two different 3-dimensional spaces with a color gradient representing the ID scores of each point. In the left we see that in one possible 3-dimension projection it appears that there are no outliers, whereas in the right image we see that projecting to a different basis vector set reveals two outliers. This illustrates the difficulty of interpreting which points are outliers in higher dimensional vector spaces and how the use of ID of a point can help circumvent such difficulties.

outlier score such that we will always know whether a point has a high likelihood of being an outlier for any given data set.

In light of these considerations we propose a new outlier score, the Outlier Probability Via Intrinsic Dimension (OPVID) score, which serves exactly this purpose. The general outline is as follows and makes use of a normalization technique used in Ref. 9 to obtain the probability measure.

After determining the ID score of every point in the data set using for example one of the methods described in Ref. 29 or Ref. 35, we determine the difference between a points own ID score and the average of its local k_r-nearest neighbours ID scores, i.e.

$$\text{PIDOS}(q) := \text{ID}(q) - \frac{1}{k_r} \cdot \sum_{p \in N_{k_r}(q)} \text{ID}(p) , \tag{4}$$

which we call the *Probabilistic Intrinsic Dimension Outlier Score* (PIDOS) and will be close to zero for points which belong to clusters. As a side remark, we note that this this score is *not* a probability score yet since it clearly can be negative. Now, using the set of all PIDOS scores we can define the Outlier Probability Via Intrinsic Dimension score of a query point q as

$$\text{OPVID}(q) := \max \left\{ 0, \text{erf} \left(\frac{\text{PIDOS}(q)}{\lambda \cdot \sqrt{2} \cdot \text{PIDOS}_{\text{STD}}(q)} \right) \right\} , \tag{5}$$

where λ is a scaling factor, $\text{PIDOS}_{\text{STD}}(q)$ is the standard deviation of the PIDOS scores of the query points local k_r-nearest neighbours, and *erf* is the Gaussian error function. We note that the application of the Gaussian error function assumes that the PIDOS scores are normally distributed around zero, which, for points which have an ID score close to the average of its neighbours is a just assumption. Therefore, the OPVID score for a point which belongs to a local cluster and hence has each of its neighbours sharing a close ID score will be close to zero whereas a point which strays outside of a local cluster will have a score closer to one.

For an insight into some preliminary benchmarking results that will be published elsewhere, we briefly describe an example of one of our tests. Given a starting data set where each point is pre-labelled as either

American Institute of Aeronautics and Astronautics

"inlier" or "outlier", we run the outlier detection algorithm against it and then obtain a Receiving Operating Characteristic (ROC) curve which plots the true positive verses false positive rates of detecting outliers. By computing the area under the curve (AUC) we then obtain a measure of accuracy of the algorithm. Here we use the Amsterdam Library of Object Images (ALOI) data set[39] which is a collection of 110,250 images of 1000 objects where each feature vector that corresponds to a single image is 64-dimensional. Using a subset of the ALOI data set which consists of 10,000 data points with 100 known outlier points we compared OPVID with four different k_c values (which are used to calculate the ID score of each point) against LoOP and LOF for a range of k_r-nearest neighbour values. Figure 2 provides a preliminary look into the results where it is seen that OPVID outperforms the two other methods in correct outlier identification.
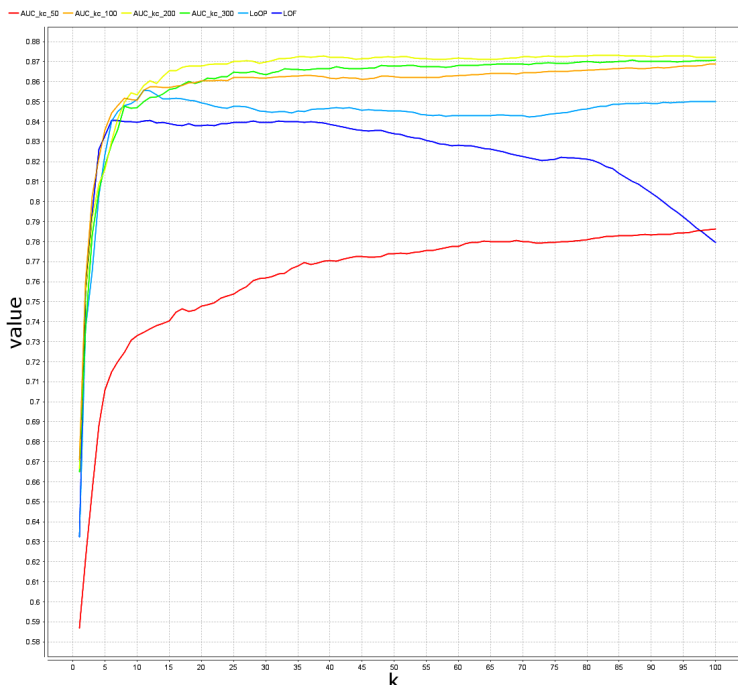

Figure 2. Accuracy of OPVID for various k_c values (green, yellow, orange and red lines) that are used to estimate the ID score of each point against LoOP (light blue line) and LOF (dark blue line) in a ROC-AUC plot for a sub-sampled ALOI data set. The higher the curve, the higher accuracy in correctly identifying the known outliers in the data set.

III.   Automated Telemetry Health Monitoring System

In this section we explain the basic concepts and notions behind ATHMoS to provide an overview of the entire system. We first describe our approach to computing a non-trivial feature vector by analysing the raw time series data as well as its smoothed form which is computed via Fourier transform. We then discuss our approach to obtaining three so-called Classification Databases; the Historic Nominal, Recent Nominal, and Anomaly Databases. Once we have these three databases for each telemetry parameter we wish to monitor, we make use of the OPVID, LoOP and DBSCAN algorithms to compare new test data to them in order to determine behaviour of the new data. Finally, we finish with an illustrative example where we were able to detect with high precision a known anomaly from the TET-1 satellite that wasn't detected by the OOL-check.

III.A.   Feature Vectors and Summary Statistics

As mentioned in Section II.A, a central aspect of data mining is to identify statistical features for a given data set that best describe the data. A common approach is to make use of summary statistics. That is, for a set of observations, one tries to describe the data in terms of a measure of the shape of its distribution, statistical dispersion, central tendency and bounds. Recalling that the ID score of a data point provides a measure of the subspace dimension of the cluster it may belong to, it gives us a nice opportunity to create high dimensional feature vectors made up of several different summary statistics.

American Institute of Aeronautics and Astronautics

The idea is as follows: for a fixed telemetry parameter e.g. from a LEO mission, we can split the time series data into 90-minute time intervals (approximate time for one orbit) and compute many summary statistics after carefully interpolating the data to a fixed data sample rate. Such summary statistics include: maximum, minimum, average, median, first quartile, third quartile, midrange, median absolute deviation, standard deviation...etc. From here and henceforth we refer to these statistics as *basic summary statistics*. Thus, for each 90-minute time interval of a fixed telemetry parameter we obtain a corresponding feature vector of the form

$$V_n = (\text{Max, Min, Avg, Med, Q1, Q3, MidR, } \dots), \tag{6}$$

which can then be considered as a point in a vector space and thus allows us to determine areas where these points cluster together, i.e. clusters of "orbit" data points. Clearly, when a point lies far enough away from clusters, then the corresponding orbit can be flagged as behaving non-nominally relative to the data available in the clusters. After a thorough analysis of which summary statistics to choose as elements of our feature vectors, we determined that noise in the data was having a large effect on these statistics and caused many false positives to be reported. In order to overcome this limitation for non-discrete-valued telemetry data (i.e. parameters which do not have step-function behaviour) we first apply a smoothing algorithm to isolate the noise from the data.

A common technique for smoothing time series data is to apply a discrete Fourier transform to the data, mapping it into a frequency space which provides an insight into the individual frequencies which the original signal is composed of (See Fig. 3). Once in the frequency space, we set the amplitudes of high frequencies to zero, *cutting* them off, and then apply the inverse Fourier transformation to map it back to its original domain. The smoothed signal we obtain this way does not contain any of the high frequency noise. The difficulty lies in finding an optimal cutoff value that is low enough to remove possible noise contained within high frequencies and high enough to include the inherent frequency of the underlying raw time series data if there is any. In case no dominant underlying frequency can be detected, a sufficiently high default cutoff value is used to prevent the resulting smoothed signal to be void of any information contained in the raw time series data. Details of how to obtain the optimum frequency cutoff value is outside the scope of this paper. However, a general outline can be seen in Figs. 4, 5 and their captions.
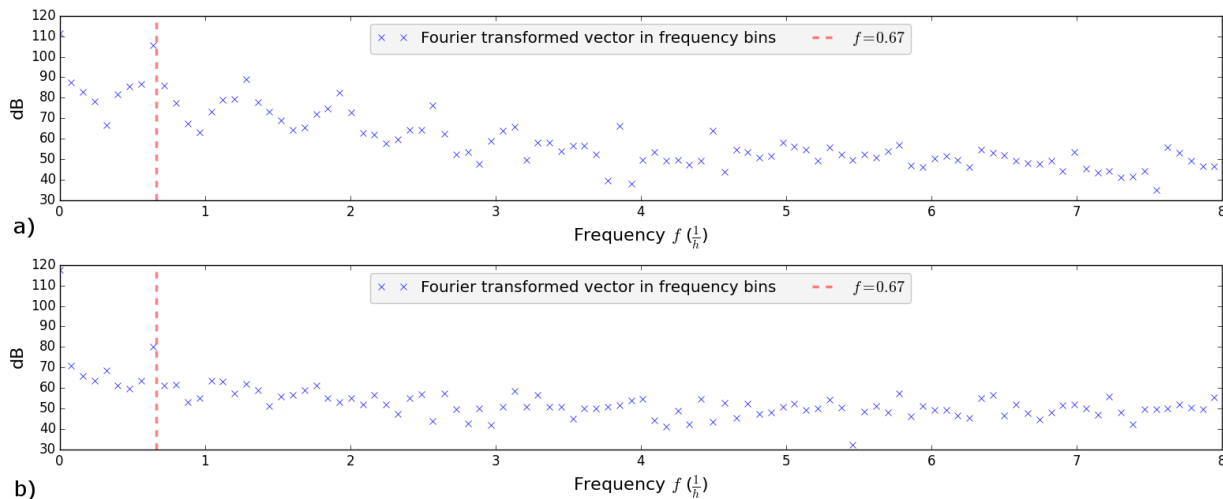

Figure 3. The first part of the vector representing the Fourier transformation (using a real fft) of the raw telemetry data from Fig. 5, where Figs. 3a) and 3b) correspond to the raw data in Figs. 5a) and 5b), respectively. For these two sets of Fourier transformed telemetry data, a peak at a frequency around 0.67 $\frac{1}{h}$, which corresponds to one cycle per orbit ($\sim$90-minutes per orbit), is visible. This periodicity becomes obvious when looking at the corresponding raw telemetry data. However, this peak might not always be sufficient to determine the inherent frequency of a telemetry parameter and a generalized approach, roughly depicted and described in Fig. 4, is used in the ATHMoS Algorithm.

Let $\mathbf{y} := (y_1, y_2, \dots, y_n)$ be the vector of all data points for a fixed telemetry parameter within a given 90-minute time interval (representing the time period of one LEO orbit) and let $\mathbf{y}' := (y'_1, y'_2, \dots, y'_n)$ be the new, smoothed vector after cutting off the higher order frequencies and transforming it back to the original domain. Defining a relative root-mean-squared (*rrms*) error as

American Institute of Aeronautics and Astronautics

$$\text{rrms}(\mathbf{y}') := \sqrt{\frac{1}{n} \cdot \sum_{i=0}^{n} \left(\frac{y_i' - y_i}{y_{\max} - y_{\min}} \right)^2}, \tag{7}$$

allows us to measure the effects of removing each individual higher order frequency. By plotting the rrms-error against the cutoff value, we can obtain a visualization of the inherent frequency in the original data and those which we want to consider as noise (see Fig. 4). Once we obtain the optimal cutoff value and smoothen the data, we arrive at the situation depicted in Fig. 5.

a)

b)
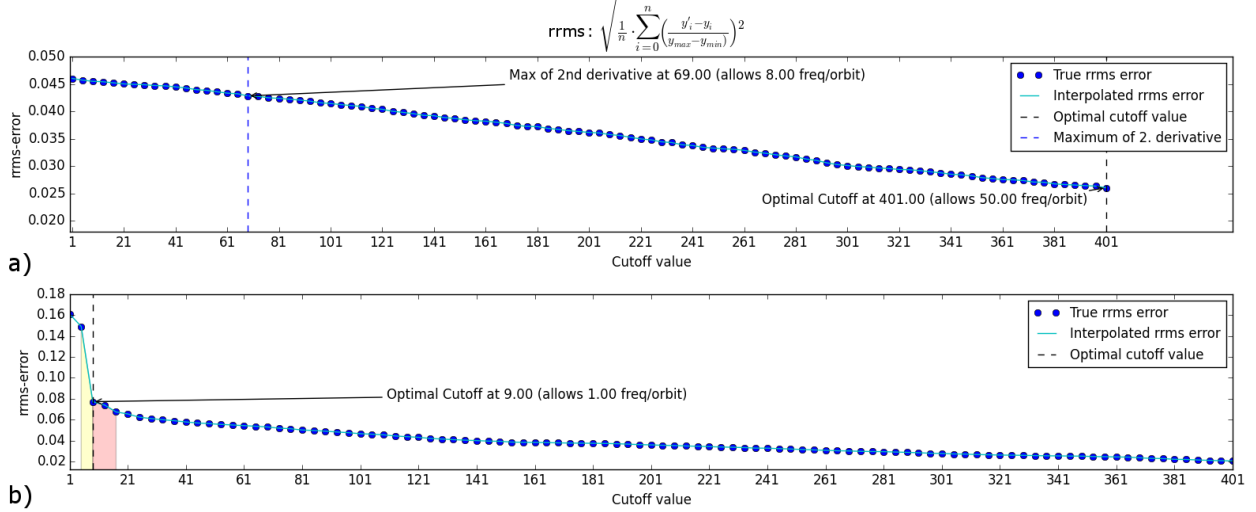
Figure 4. The rrms-error obtained at varying frequency cutoff values for a sample size of eight orbits. A higher cutoff value means less frequencies are cut off in the frequency space before applying the inverse fourier transformation to acquire the new signal. The more frequencies are used to compute the new signal via the inverse Fourier transformation, the lower the rrms-error (see Eq. (7)) will be since the new signal more closely resembles the original data. We assume that a significant drop in the rrms-error signifies the inherent frequency or multiples of the inherent frequency. In Fig. 4a), we see no distinct drop in the rrms-error. Here, the maximum of the discrete 2nd derivative, which can be used to locate the largest drop in the rrms-error, does not represent a significantly large drop and thus, our algorithm computing the optimal cutoff defaults to a pre-configured default value of 401. In Fig. 4b), a significant sharp drop in the rrms-error at a cutoff value of nine (which corresponds to one frequency per orbit) is clearly visible and automatically detected by our algorithm. This frequency represents the periodic nature of the signal within its time interval. Furthermore, comparing Fig. 4b) with its raw telemetry data visualized in Fig. 5b) allows us to visually verify the optimal cutoff value of nine. The plot of this raw data shows that it includes eight 90-minute orbits (~12-hours), meaning we want a resolution of at least eight frequencies in our smoothed vector. Cutting off all frequencies above and including nine leaves us with a resolution of eight frequencies which is exactly what we want our optimal cutoff value to be.

In addition to removing the noise using the previously outlined method using Fourier transformations and thereby gaining additional features stabilizing our feature vector, the features derived from the smoothed signal indirectly contain information about its inherent frequency and allow changes in its frequency to be detected. Also, by smoothing the original raw telemetry data, a natural step is to investigate the noise that was filtered out. By taking the difference of the original data for a given time interval $\mathbf{y}$ and smoothed data $\mathbf{y}'$ we obtain a noise vector

$$\mathbf{y}_N = (y_1 - y_1', y_2 - y_2', y_3 - y_3', \dots, y_n - y_n'), \tag{8}$$

which enables us to detect major changes in the noise of a telemetry parameter.

Finally, for a fixed time interval, we now compute the basic summary statistics over the raw data, smoothed data and noise data given by $\mathbf{y}$, $\mathbf{y}'$ and $\mathbf{y}_N$ respectively to obtain one large feature vector composed of these summary statistics. Once we have computed these feature vectors for each time interval for a given telemetry parameter, we can begin to compare new feature vectors to them to determine the likelihood that the new point represents an anomaly in the described time interval.

III.B. The ATHMoS Algorithm

There are three main types of machine learning approaches to anomaly detection; supervised, unsupervised and semi-supervised. In supervised anomaly detection, we have two sets of labelled data; nominal and
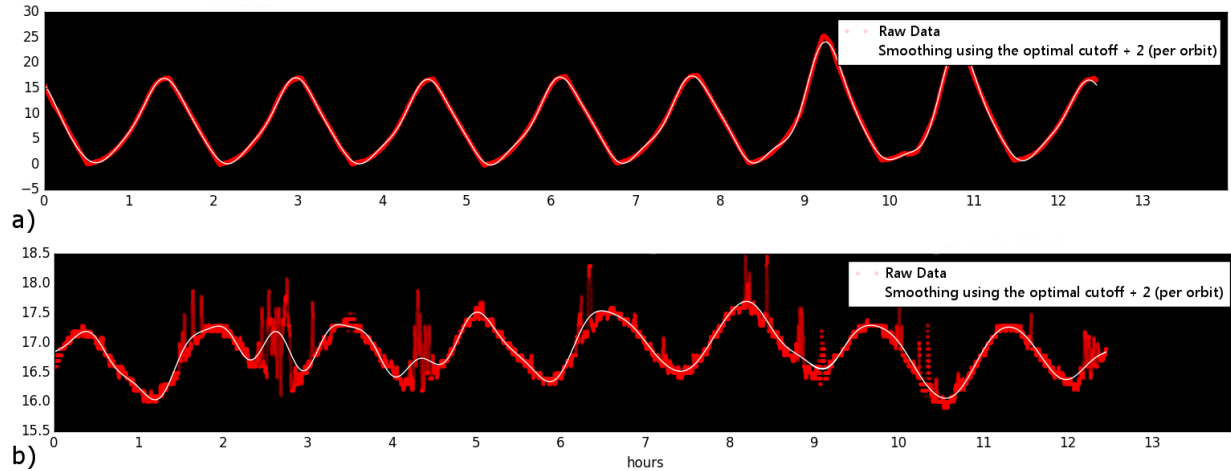
American Institute of Aeronautics and Astronautics

Figure 5. The original raw data (red) plotted against the smoothed data (white) for two different telemetry parameters over a ~12-hour time interval. The smoothing was acquired by cutting off all frequencies above and including 25 in the frequency space and applying the inverse Fourier transformation to the resulting frequency vector. We arrived at the cutoff value of 25 by automatically computing the optimal cutoff value which was nine for both of the plotted parameters and adding an additional two frequencies per orbit (16 in total) to the optimal cutoff value. Allowing two extra frequencies per orbit in the frequency space ensures that the resulting smoothed data still includes some of the details of the raw data and, for parameters which do not have a large amount of noise and behave similar to a sinus wave, closely follows the raw data as is the case for Fig. 5a).

anomaly sets. When new data is obtained, it is compared to both of the sets and an algorithm decides which data set each new data point belongs to. In semi-supervised anomaly detection, we are only provided with a labelled nominal data set and then compare the new test data to this set. In unsupervised anomaly detection we have no labelling of any initial data at all and simply compare new test data to it. The downside to such an approach is that there may have been several anomalies that occurred in the past and therefore formed a cluster in the high dimensional feature space. When we compare the new test data to it, it appears to be part of said cluster and therefore is not identified as a statistical outlier. As referenced in Section I, a notable example in the spacecraft operations community which makes use of the semi-supervised approach is the Novelty Detection System developed at ESA.[8] Here we take the supervised approach.

III.B.1. Classification Databases

For each new telemetry parameter which will be monitored by ATHMoS, an initial configuration run must be completed where we compute the Classification Databases; the Historic Nominal, the Recent Nominal, and the Anomaly Databases. That is, we take unlabelled feature vectors from past data and separate them into these three categories. This step is accomplished by putting all the data together and determining the clusters that are present in the data by using the Density-Based Spacial Clustering of Applications with Noise (DBSCAN) algorithm - one of the most widely used clustering algorithms.[11] It is very useful since unlike other clustering algorithms, it does not require the user to input a set amount of clusters that must be found within the data set. Through an extensive analysis and corroboration with operations engineers, we concluded that any clusters that contain less than 5% of the total data set can be separated out and stored in the *Anomaly Database*. The remaining data points are kept and stored in the *Historic Nominal* and the *Recent Nominal* Databases. As the name suggests, the Recent Nominal Database contains points which are more recent relative to the soon to be tested new test data. An example of the automatic cluster detection output from DBSCAN is given in Fig. 6.

III.B.2. Comparing New Data to the Classification Databases

When new telemetry data is available, we compute the feature vectors for each predetermined time interval length as detailed in Section III.A, i.e. 16 time intervals corresponding to 90-minutes each for a 24-hour time period (for a LEO mission). Once we have obtained this set of new data points, the so-called *Test Data*, we compare each individual Test Data point against the Classification Databases. For comparison to the Historic
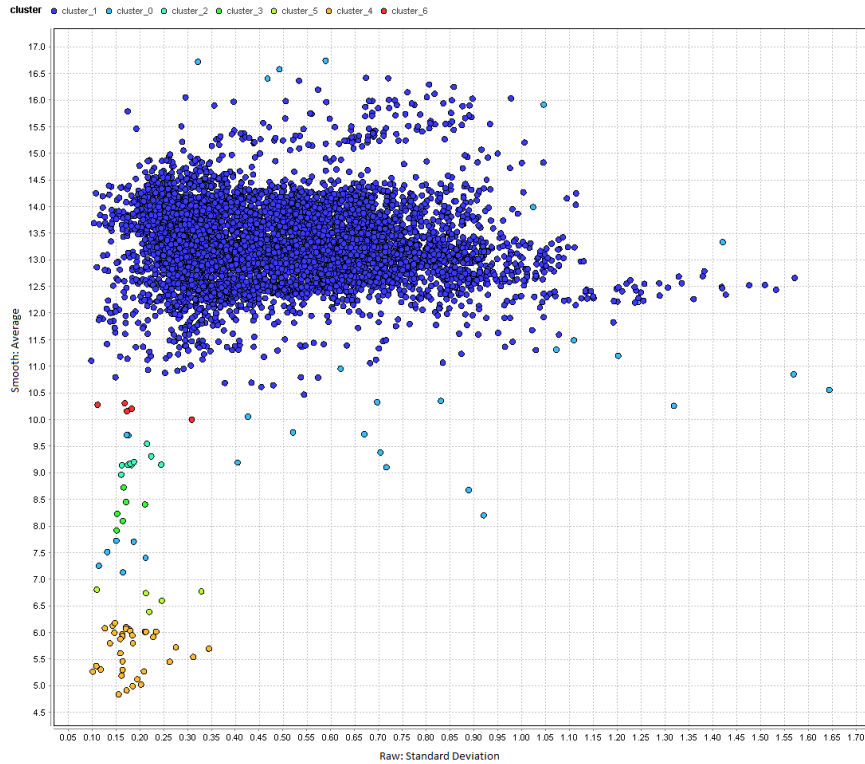
Figure 6. Each point corresponds to a 90-minute time interval for a fixed parameter and is a 15-dimensional vector. In total, this is one years worth of orbits. This figure depicts a two dimensional projection of this high dimensional data to the standard deviation of the raw data verses the average of the smoothed data and the clusters (seven) that DBSCAN detected. We assume that any cluster that contains less than 5% of the total number of points represents outliers. Thus, these are filtered out and only "cluster 1" (dark blue) is used as a starting nominal data set which is then time-wise separated into Historic and Recent Nominal Databases.

and Recent Nominal Databases, we use the OPVID algorithm described in Section II.B since (for ongoing missions) we have a large amount of data contained in them which lends itself to provide an accurate ID score. This differs from the comparison to the Anomaly Database, where the amount of labelled data is not high enough to obtain an accurate predication with OPVID so there we make use of the LoOP algorithm and DBSCAN algorithms to provide us with a labelling of each point as either Nominal or Anomaly *with respect to the data set of known anomalies.* At the end of this procedure, we have eight possible different labels of a test data point, each with varying degrees of certainty of whether it corresponds to a true anomaly event. We then add points which are considered nominal to the Recent Nominal Database and those which are anomalies to the Anomaly Database. Separately from this addition, we define High Priority Anomalies (HPA) as those which have been added to the Anomaly Database and are genuine new anomalous behaviour, and Low Priority Anomalies (LPA) which are points that follow a trend and may converge to a known anomaly. These two classes of anomalies are then reported to the subsystem engineers so they can investigate the HPAs and know to monitor the LPAs in the future. If the subsystem engineers determine that a behaviour should be labelled differently, their label change is then fed back into the system and dispersed to the correct Classification Database which creates a higher degree of accuracy in the overall performance and estimation capabilities of the system. Furthermore, as the system continues to run, the Recent Nominal Database gets filled with more points and eventually data which lies too far in the past must be transferred to the Historic Nominal Database. This creates a recursion that allows us to continually compare new Test Data to the Recent Nominal Database and keep track of recent trends.

III.B.3.   *Overview of System Architecture and Implementation*

The implementation of such a data pipeline provides several architectural challenges. Handling multiple missions where each satellite has thousands of parameters requires a system architecture which is distributed

American Institute of Aeronautics and Astronautics

across a cluster of servers such that individual process components can be duplicated and scaled for performance. The prototype currently being developed is making use of state-of-the-art big data technologies such as Apache Spark[40,41] for data processing and Apache Cassandra[42,43] for storage. For maximum modularity of components to facilitate individual deployment and specific multimission configurations, we are implementing the system using a microservice architecture[44,45] in the Scala programming language. An overview is provided in Fig. 7.
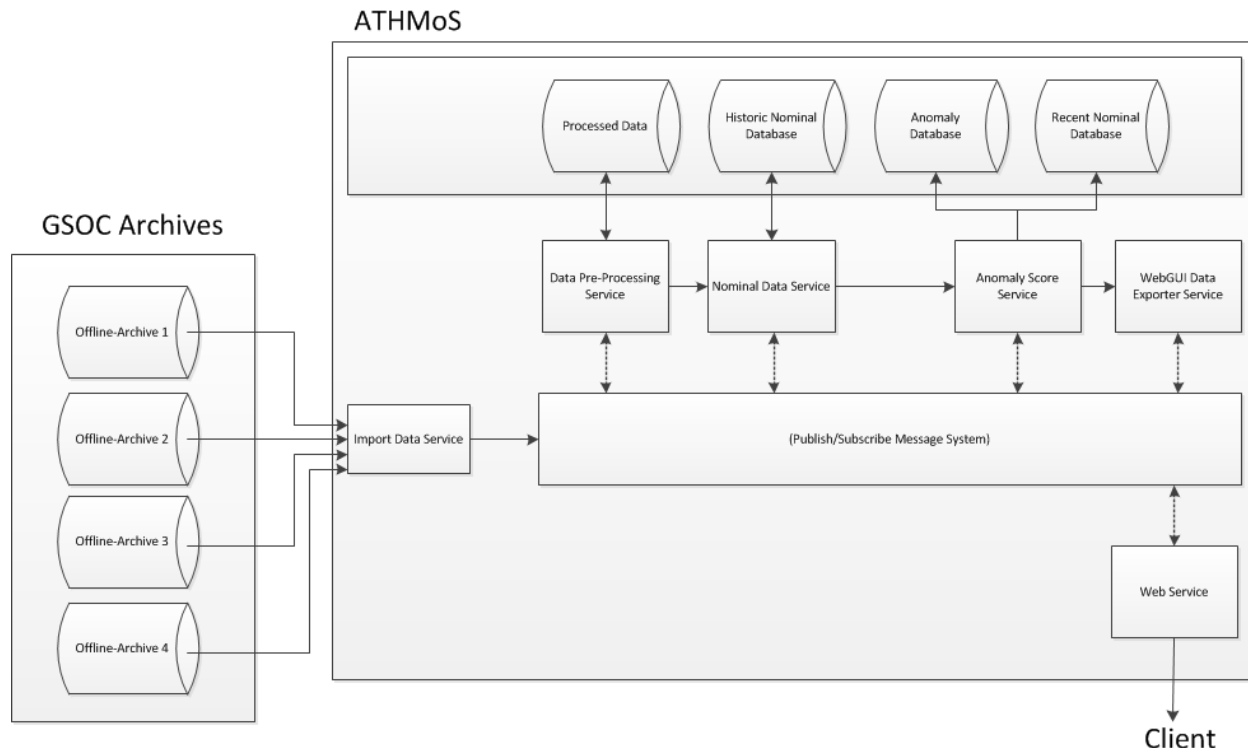

Figure 7. A high level view of the system architecture of ATHMoS currently being developed at GSOC. Here we can see the various types of possible data input as well as the individual classification databases and sub-applications which compute the various parts of the overall algorithm. Operation engineers then access the system to review the latest detection results where they can for example plot the various statistics computed for the feature vectors or check any new clustering behaviour any of the Classification Databases.

III.B.4. *Correlated Multi-Parameter Applications*

So far, this overview has focused on computing the feature vectors for a given time interval for a *single* telemetry parameter. Since we use the OPVID algorithm which bases its outlier score on the ID of a data point, this opens up the possibility for *multi-parameter* anomaly detection. Recall that the ID of a data point which lives in an n-dimensional vector space is approximately the number d if that point belongs to a d-dimensional cluster. Two extreme situations may occur when increasing the dimension of the initial vector to include the feature vector elements of another (possibly many) telemetry parameter(s). First, if the additional dimensions which are added are correlated to the original ones, then the distance between points will in increase, however the points may form new clusters in new subspace projections which now lie between the original dimensions and the newly added ones. In this case, the ID score of such points will increase to reflect the fact that the cluster size that the query point belongs to has increased. In the second scenario, one can add additional dimensions which arise from summary statistics from an uncorrelated telemetry parameter. Adding these additional dimensions creates a more sparse distribution of data points thereby separating the existing clusters in certain subspaces - this leads to a larger increase in ID score than the first case due to the sparseness of data. Although ID scores of the points are increased in both cases, the overall OPVID scores which are computed by comparing the Test Data to the Historic and Recent Nominal Databases behave as one might expect: the score provides a probability that the selected grouping of telemetry parameters behaved differently *as a whole*, relative to the nominal databases.

American Institute of Aeronautics and Astronautics

Currently, we envisage ATHMoS to primarily work in a parameter by parameter basis except for certain subsystems that engineers are particularly interested in and thus must provide the system with lists of telemetry parameters they also wish to combine. By checking each parameter individually as well as combinations of certain select parameters, we will be able to more accurately discern which anomalies occur as stand-alone events (i.e. are associated to a single telemetry parameter), and those which are the result of the interplay between related systems. One of our active areas of research is investigating this process, and namely, how to automate the process of combining telemetry parameters based on correlation coefficients and changes in ID scores for various parameter combinations. See Fig. 8 for more details.
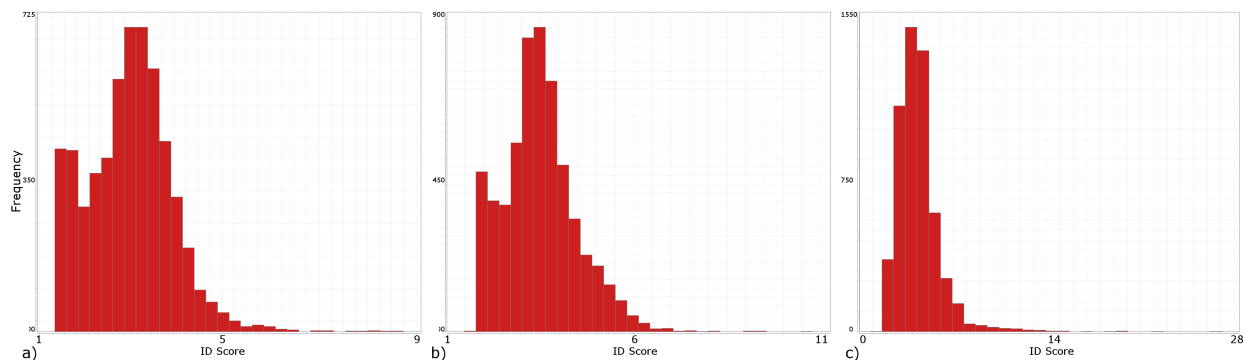

Figure 8. Frequency of ID scores for data points from Historic Nominal Databases for varying combinations of combined feature vectors. The histogram in Fig. 8a) depicts the ID score distribution for a fixed single telemetry parameter - we take this to be the default ID score distribution. Figure 8b) shows the effect when we augment the original feature vector with the corresponding nominal elements from another feature vector that describes different *correlated* telemetry parameter. There we see that that the average score and distribution stays almost the same, albeit slightly increased. In Fig. 8c), we augmented the original feature vector with feature vector elements from another *uncorrelated* telemetry parameter (again from a respective Historic Nominal Database which has the same time intervals). This time the distribution as well as the maximum and minimum values of the ID scores are greatly affected. This is due to the fact that the additional dimensions added to the data cause a higher degree of sparsity within the new vector subspaces and hence we lose some partial cluster dimension information that we had in the original data set.

III.B.5. Example: TET-1 Reaction Wheel Anomaly

On February 22nd 2015, the TET-1 satellite experienced an increase in temperature in one of its reaction wheel sensors. This increase was not enough to trigger the OOL-check to send an automatic notification to the operations team. The temperature spike can be seen in Fig. 9. One of the first proof of concepts of ATHMoS was to check if this anomaly could have been detected as well as identifying any possible precursor behaviour which hay have predicted this temperature spike.

As a starting point, we decided to take all the data for the telemetry parameter from the year 2014 and construct the Historic Nominal, Recent Nominal and Anomaly Databases in order to compare one week intervals of Test Data from January 1st to March 7th 2015. Running the DBSCAN algorithm on the 5843 data points for the year (after splitting into 90-minute time intervals) showed that there were six clusters of behaviour, with one central cluster containing approximately 98% of the total data set. Filtering out the five small data clusters results in the two data sets shown in Fig. 10. We then separated out the nominal data from December 2014 and stored it as the Recent Nominal Database.

Once we have obtained these three initial Classification Databases, we now compare new Test Data for each week from January 1st to March 6th 2015 to these databases using a combination of the OPVID, LoOP and DBSCAN algorithms (depending on the database). Figure 11b) depicts all test data points that have an OPVID score of over 90% when compared to only the Historic Nominal Database. The choice of a cutoff value of 90% for a point to be a potential anomaly was the result of several benchmarking tests completed with cooperation of GSOC subsystem engineers. We can clearly see that every orbit within the anomaly period from February 22nd to March 6th is an anomaly, however some of the points have higher anomaly scores than others. This can be seen as a direct consequence of the high temperature fluctuations that occur after the initial temperature spike on February 22nd (see the telemetry highlighted in red in Fig. 12). Most importantly, OPVID correctly identified a *new* anomaly of the same type (a temperature spike) which occurred January 23rd 2015, *31 days before* the known anomaly. Interestingly, at the time of this anomaly, the operations team successfully identified several other anomalies which all occurred concurrently onboard.

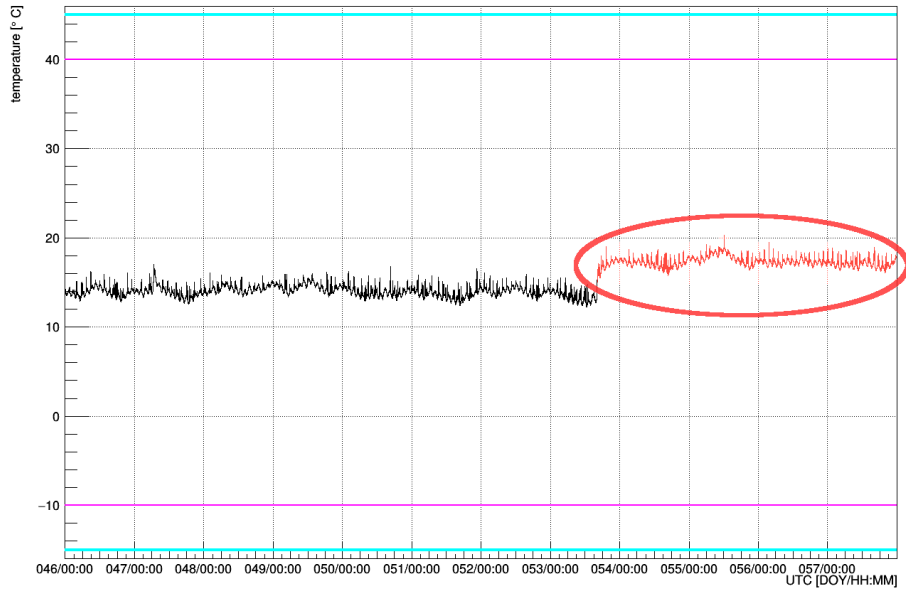Reaction Wheel Temperature (DOY 15046-15057)

Figure 9. An 11-day plot of the temperature of a reaction wheel for the TET-1 satellite. The temperature spike anomaly which can be seen on DOY 53 in 2015 (red) was not close to the hard (cyan) or soft (pink) upper and lower limits that are used by the OOL-check.

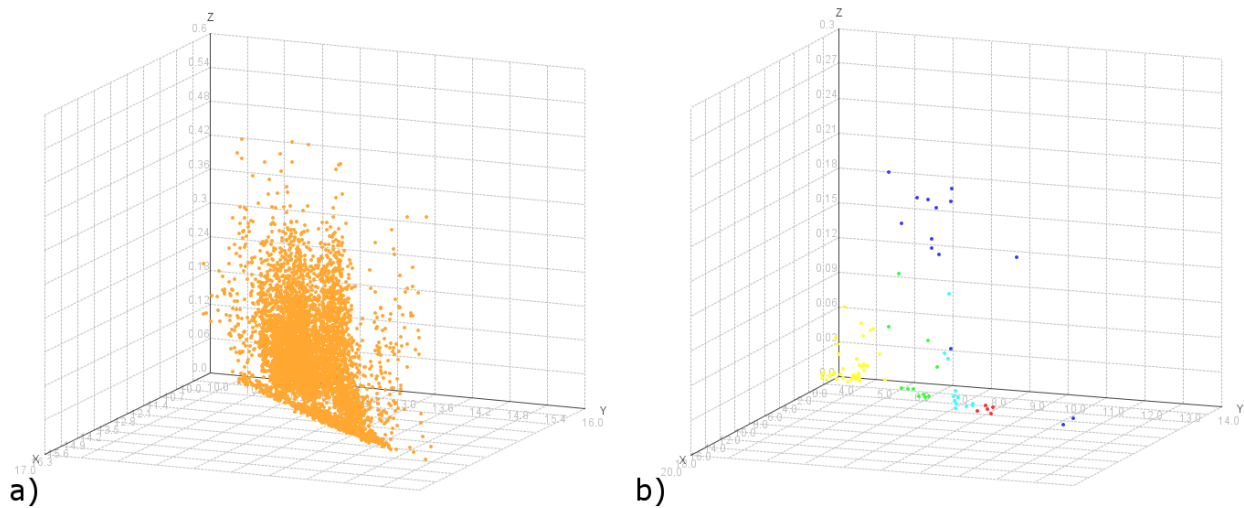a)   b)

Figure 10. Two 3-dimensional projections of the filtered clusters after running DBSCAN on the 2014 data. Figure 10a) shows the cluster of points that was filtered out from the unlabelled 2014 data which was higher than 5% of the total number of 5843 data points. This corresponds to the nominal data and hence is split time-wise into the Historic Nominal and Recent Nominal Databases. Figure 10b) shows the remaining clusters which make up the initial Anomaly Database.

American Institute of Aeronautics and Astronautics

Next we compare our method to a setup reminiscent of the Novelty Detection System[8] which uses a 4-dimensional feature vector and the LoOP algorithm to obtain the probability scores in visible in Figure 11a). There we see that there are many more points which are above the 90% (or even 95%) threshold and occur the day before the OPVID algorithm had its first data point with a score above 90%. After comparing to our training data it was concluded that the real anomaly occurs on the 23rd of January as highlighted in blue in Fig. 12 since the slight increase in temperature that occurred on the 21st of January (highlighted in green in the same figure) was in fact nominal (these and other false positives are circled in Fig. 11).
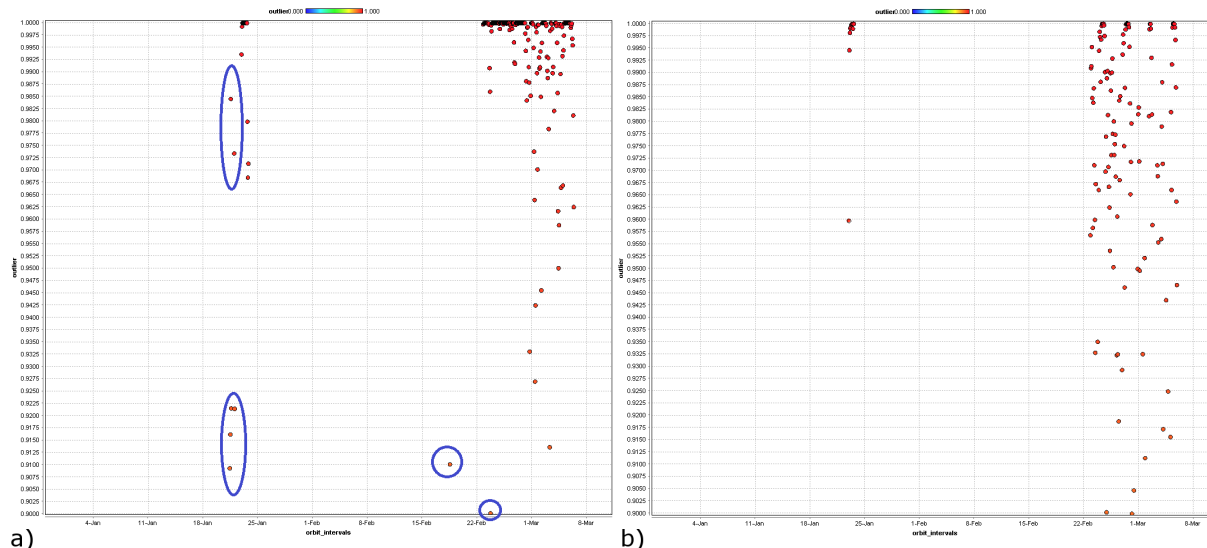

Figure 11. A comparison between LoOP and OPVID scores for two different feature vectors. Figure 11a) shows the LoOP scores of the test data that are over 90% when the feature vector is 4-dimensional and consists of the minimum, maximum, average and standard deviation, whereas the Fig. 11b) shows the OPVID scores over 90% where each point is computed from the generalised feature vectors described in Section III.A. After consulting with operations engineers, the ATHMoS method in Fig. 11b) has significantly less false positives while at the same providing a more accurate representation of the spread of anomalous behaviour during the temperature spike phase that was persistent for 13 days. Moreover, the anomaly that occurred on January 23rd is visible in both approaches however the LoOP approach contains several false positives (circled in blue).

IV.   Conclusion

In this paper we introduced GSOCs approach to automated telemetry monitoring as the Automated Telemetry Health Monitoring System (ATHMoS). ATHMoS is novel in the space community for two main reasons.

First, it uses supervised learning methods to attempt to solve the anomaly detection problem by comparing new telemetry data with past nominal and anomaly data, where additionally, the past nominal data is divided into recent nominal and historic nominal data. Using a combination of comparison techniques between test data and these Classification Databases, ATHMoS assigns a label to each new data point, say for example a High Priority Anomaly (HPA) or Low Priority Anomaly(LPA). Based on these labels, subsets of the test data are grouped together and then stored in the various Classification Databases they belong to. Then, the operations team is notified of certain classes of points along with the time intervals they correspond to. Based on the engineers feedback, the points may be relabelled and fed back into the system for higher accuracy in future computations.

And second, in order to determine if there are anomalies in the new test data, we used concepts behind the Local Outlier Probability (LoOP) and Intrinsic Dimension Outlier Score (IDOS) outlier detection algorithms to create the Outlier Probability Via Intrinsic Dimension (OPVID) algorithm. This algorithm combines aspects of these two state-of-the-art outlier detection techniques by assigning an outlier *probability* that is computed via the local (continuous) Intrinsic Dimension (ID) of each point. Our investigations have revealed that for non-discreet time series telemetry, OPVID is more accurate than the LoOP algorithm, especially for high dimensional data. Using OPVID also allows us to combine feature vectors of multiple correlated

American Institute of Aeronautics and Astronautics

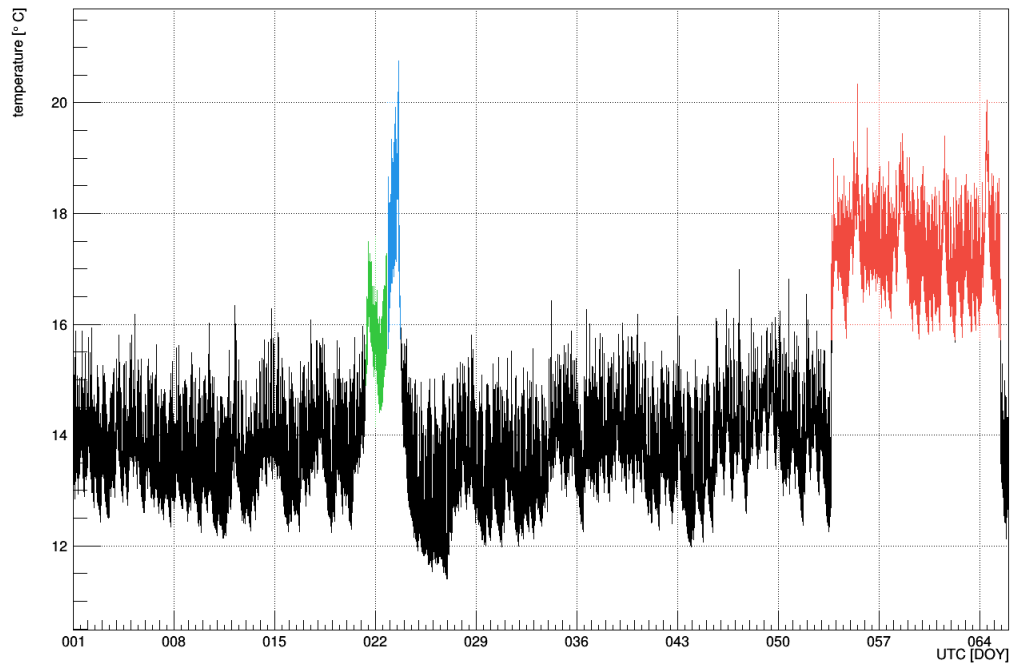## Reaction Wheel Temperature (DOY 15001-15065)

Figure 12. In the left of the figure (blue colored) we see the *newly* detected anomaly event that occurred on January 23rd 2015 which ATHMoS identified. A separate anomaly occurred onboard that affected several subsystems. However, it was not immediately identified that this particular parameter was affected. The small temperature spike (green colored) just to the left of the anomaly was identified as being an outlier by using the LoOP algorithm as described in the caption of Fig. 11, however, relative to our training data it is actually a false positive. The right of the figure shows the known anomaly event (red colored) which occurred from February 22nd to March 6th of 2015. We see that there is an inherent frequency in the telemetry once the temperature spike has occurred, and these fluctuations are precisely seen in the OPVID outlier scores in Fig. 11. For these reasons, this example illustrates the accuracy benefits of using the OPVID algorithm.

American Institute of Aeronautics and Astronautics

parameters to help detect anomalies which might only be visible when considering the multiple telemetry parameters simultaneously.

Lastly, we provided a quick high level introduction to the design and architecture of ATHMoS. Namely, the system is designed to be fault-tolerant, scalable and highly available in terms of its distributed computing and microservice architecture which makes use of the data processing framework Apache Spark and clustered database Apache Cassandra.

Glossary

ATHMoS Automated Telemetry Health Monitoring System

GSOC German Space Operation Center

LEO Low Earth Orbit

OOL Out-Of-Limit (telemetry limit check)

DBSCAN Density-Based Spacial Clustering of Applications with Noise (clustering algorithm)

LOF Local Outlier Factor (outlier detection algorithm)

LoOP Local Outlier Probability (outlier detection algorithm)

ID Intrinsic Dimension (a type of statistical quantity assigned to a data point and data set)

IDOS Intrinsic Dimension Outlier Score (outlier detection algorithm)

OPVID Outlier Probability Via Intrinsic Dimension (outlier detection algorithm)

HPA High Priority Anomaly (a label for new telemetry data that is assigned by ATHMoS)

LPA Low Priority Anomaly (a label for new telemetry data that is assigned by ATHMoS)

TET-1 The "Technologieerprobungsträger" is a small German satellite for technical experiments in space, the abbreviation means Technology Experiment Carrier

Acknowledgements

References

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

[2]"ODVEC Monitoring System Certified and Deployed in JSC Mission Control Center," NASA Press Release, December 2012.

[3]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*, August 2003, pp. 29-38.

[4]https://ti.arc.nasa.gov/opensource/projects/mariana/

[5]Park, H., Mackey, M., James, M., Zak, M., Kynard, M., Sebghati, J., and Greene, W., "Analysis of Space Shuttle Main Engine Data Using Beacon-Based Exception Analysis for Multi-Missions," *IEEE Aerospace Conference Proceedings*, Vol. 6, pp. 2835-2844, 2002.

[6]Iverson, D. L., "System Health Monitoring for Space Mission Operations," *IEEE Aerospace Conference*, March 2008, pp. 1-8.

[7]Spirkovska, L., Iverson, D. L., Hall, D. R., Taylor, W. M., Patterson-Hine, A., Brown, B. L., Ferrell, B. A., and Waterman, R.D., "Anomaly Detection for Next-Generation Space Launch Ground Operations," *Proceedings of the AIAA SpaceOps 2010 Conference*, April 2010.

[8]Martínez-Heras, J. A., Donati, A., Kirsch, M. G., and Schmidt, F., "New Telemetry Monitoring Paradigm with Novelty Detection," *Proceedings of the AIAA SpaceOps 2012 Conference*, 2012, pp. 11-15.

[9]Kriegel, H. P., Kröger, P., Schubert, E., and Zimek, A., "LoOP: Local Outlier Probabilities," *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, November 2009, pp. 1649-1652.

[10]Verzola, I., Donati, A., Martinez-Heras, J. A., Schubert, M., and Somodi, L., "Project Sibyl: A Novelty Detection System for Human Spaceflight Operations", *Proceedings of the AIAA SpaceOps 2016 Conference*, May 2016.

[11]Ester, M., Kriegel, H. P., Sander, J., and Xu, X., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Kdd*, Volume 96, No. 34, pp. 226-231, August 1996.

[12]Schubert, E., Wojdanowski, R., Zimek, A., and Kriegel, H. P., "On Evaluation of Outlier Rankings and Outlier Scores," *Proceedings of the 2012 SIAM International Conference on Data Mining*, April 2012, pp. 1047-1058.

[13]Zimek, A., Schubert, E., and Kriegel, H. P., "A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data," *Statistical Analysis and Data Mining*, Vol. 5, No. 5, pp. 363-387, 2012.

[14]Achtert, E., Kriegel, H. P., Reichert, L., Schubert, E., Wojdanowski, R., and Zimek, A., "Visual Evaluation of Outlier Detection Models," *Database Systems for Advanced Applications*, pp. 396-399, Springer Berlin-Heidelberg, April 2010.

[15]Chandola, V., Banerjee, A., and Kumar, V., "Anomaly Detection: A Survey," *ACM Computing Surveys (CSUR)*, Vol. 41, No. 3, p. 1-58, July 2009.

[16]Gogoi, P., Borah, B., and Bhattacharyya, D. K., "Anomaly Detection Analysis of Intrusion Data Using Supervised & Unsupervised Approach," *Journal of Convergence Information Technology*, Vol. 5, No. 1, pp. 95-110, 2010.

[17]Görnitz, N., Kloft, M. M., Rieck, K., and Brefeld, U., "Toward Supervised Anomaly Detection," *Journal of Artificial Intelligence Research*, 2013.

[18]von Brünken, J., Houle, M. E., and Zimek, A., "Intrinsic Dimensional Outlier Detection in High-Dimensional Data," *National Institute of Informatics Technical Report 2015-003E*, Tokyo, Japan, March 2015.

[19]Pestov, V., "On the Geometry of Similarity Search: Dimensionality Curse and Concentration of Measure," *Information Processing Letters*, Vol. 73, No. 1, pp. 47-51, 2000.

[20]Houle, M. E., Kriegel, H. P., Kröger, P., Schubert, E., and Zimek, A., "Can Shared-Neighbor Distances Defeat the Curse of Dimensionality?," *Scientific and Statistical Database Management*, pp. 482-500, Springer Berlin-Heidelberg, June 2010.

[21]Karger, D. R., and Ruhl, M., "Finding Nearest Neighbors in Growth-Restricted Metrics," *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, May 2002, pp. 741-750.

[22]Bouveyron, C., Celeux, G., and Girard, S., "Intrinsic Dimension Estimation by Maximum Likelihood in Isotropic Probabilistic PCA," *Pattern Recognition Letters*, Vol. 32, No. 14, pp. 1706-1713, 2011.

[23]Roweis, S. T., and Saul, L. K., "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, Vol. 290, Issue 5500, pp. 2323-2326, 2000.

[24]Karhunen, J., and Joutsensalo, J., "Representation and Separation of Signals Using Nonlinear PCA Type Learning," *Neural networks*, Vol. 7, No. 1, pp. 113-127, 1994.

[25]Gupta, A., Krauthgamer, R., and Lee, J. R., "Bounded Geometries, Fractals, and Low-Distortion Embeddings," *44th Annual IEEE Symposium on Foundations of Computer Science*, October 2003, pp. 534-543.

[26]Levina, E., and Bickel, P. J., "Maximum Likelihood Estimation of Intrinsic Dimension," *Advances in Neural Information Processing Systems*, pp. 777-784, 2004.

[27]Grassberger, P., and Procaccia, I., "Measuring the Strangeness of Strange Attractors," *The Theory of Chaotic Attractors*, pp. 170-189, Springer New York, 2004.

[28]Camastra, F., and Vinciarelli, A., "Estimating the Intrinsic Dimension of Data with a Fractal-Based Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 10, pp. 1404-1407, 2002.

[29]Amsaleg, L., Chelly, O., Furon, T., Girard, S., Houle, M. E., and Nett, M., "Estimating Continuous Intrinsic Dimensionality," *National Institute of Informatics Technical Report*, Tokyo, Japan, March 2014.

[30]Houle, M. E., "Dimensionality, Discriminability, Density and Distance Distributions," *IEEE 13th International Conference on Data Mining Workshops*, December 2013, pp. 468-473.

[31]Kégl, B., "Intrinsic Dimension Estimation Using Packing Numbers," *Advances in Neural Information Processing Systems*, pp. 681-688, 2002.

[32]Fukunaga, K., and Olsen, D. R., "An Algorithm for Finding Intrinsic Dimensionality of Data," *IEEE Transactions on Computers*, Vol. 100, No. 2, pp. 176-183, 1971.

[33]Houle, M. E., Kashima, H., and Nett, M., "Generalized Expansion Dimension," *IEEE 12th International Conference on Data Mining Workshops*, December 2012, pp. 587-594.

[34]Hein, M., and Audibert, J. Y., "Intrinsic Dimensionality Estimation of Submanifolds in R-d," *Proceedings of the 22nd International Conference on Machine Learning*, August 2005, pp. 289-296.

[35]Amsaleg, L., Chelly, O., Furon, T., Girard, S., Houle, M. E., Kawarabayashi, K.I., and Nett, M., "Estimating Local Intrinsic Dimensionality," *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2015, pp. 29-38.

[36]Breunig, M. M., Kriegel, H. P., Ng, R. T., and Sander, J., "LOF: Identifying Density-Based Local Outliers," *ACM Sigmod Record*, Vol. 29, No. 2, pp. 93-104, May 2012.

[37]Kriegel, H. P., Kröger, P., Schubert, E., and Zimek, A., "Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data," *Advances in Knowledge Discovery and Data Mining*, pp. 831-838. Springer Berlin-Heidelberg, 2009.

[38]Kriegel, H. P., and Zimek, A., "Angle-Based Outlier Detection in High-Dimensional Data," *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2008, pp. 444-452.

[39]Geusebroek, J. M., Burghouts, G. J., and Smeulders, A. W., "The Amsterdam Library of Object Images," *International Journal of Computer Vision*, Vol. 61, No. 1, pp. 103-112, 2005.

[40]http://spark.apache.org

[41]Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., and Stoica, I., "Spark: Cluster Computing with Working Sets," *HotCloud*, Vol. 10, pp. 10-10, 2010.

[42]http://cassandra.apache.org

[43]Lakshman, A., and Malik, P., "Cassandra: A Decentralized Structured Storage System," *ACM SIGOPS Operating Systems Review*, Vol. 44, No. 2, pp. 35-40, 2010.

[44]Newman, S., "Building Microservices, " O'Reilly Media, Inc., 2015.

[45]Fernández Villamor, J. I., Iglesias Fernandez, C. A., and Garijo Ayestaran, M., "Microservices: Lightweight Service Descriptions for REST Architectural Style," 2010.