# Implementation of data mining techniques for attitude and orbit control systems

**Xavier Rodrigues Henriques**
xavierhenriques@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

**November 2016**

**Abstract** – Given the complexity of operations and the indirect accessibility to data, data mining for space applications is yet at an early stage of development. This study, conducted in an entrepreneurial context, aimed to discuss the interest of data mining techniques for the attitude and orbit control subsystem (AOCS). In this context, two main applications were defined: feature reduction techniques for the uncertainties on a Monte Carlo spacecraft (S/C) model and anomaly detection techniques on historical telemetry (TM) data. Methods to identify which uncertainties are determinant for the pointing accuracy of a S/C can be used for increasing effectiveness of Monte Carlo (MC) simulation campaigns. We propose the application of two estimators for mutual information (MI) as measure of relevance between features and thus for feature selection. For our physical model in particular, uncertainties related to the thrust force half-cone angle (HCA) and the positioning of the center of mass (COM) are the most relevant for the generated torque. While the importance of the thrust direction was expected, the difference between the COM and the electrical thruster mechanism (ETHM) positioning was not anticipated. Detecting anomalies or novel behaviors in TM is a fundamental step for predictive maintenance. We adopt an out-of-limits (OOL) methodology to detect anomalous windows in TM, based on a three-step process: first, data is validated; second, segmented and summarized using descriptors; and finally, the anomalous events are detected. Results indicate our approach for automatic monitoring is capable of triggering alarms when trends are present in data, before multiple warning thresholds are reached. While applied to simple test cases, results issued from data mining techniques are thus promising for further applications and problems with increased complexity. It should be noted that the use of these methods of data mining was innovative in this field, with no knowledge of similar applications.
**Keywords** – attitude and orbit control subsystem; feature reduction; mutual information; anomaly detection; out-of-limits.

## 1 Introduction

During all phases of the mission, starting at S/C separation from the launch vehicle and throughout its operational lifetime, AOCS grants data acquisition from sensors and generates commands on actuators to ensure both attitude and orbital control. In addition, it must allow operations of the S/C from the ground and comprise a mechanism to acquire a safe attitude in case of emergency.

### 1.1 Big data

Big data analytics aim at transforming vast quantities of raw data into useful insights and critical information. *Big data* is the adequate term to describe datasets that are large or complex in way traditional data processing methods cannot be applied. *Machine learning* is a computational field comprising algorithms which learn from data and hence are able to enact intelligent action from it. *Data mining* can be defined as a process of exploratory data analysis in large datasets, extracting novel information and transforming it into understandable knowledge [Lantz, 2015].

### 1.2 Data mining for AOCS applications

This study was developed in the context of an internship done at Thales Alenia Space (TAS) Cannes. As its counterparts, Thales is investing in big data technologies in order to process the vast quantities of data generated by its own systems. The company interest was to take advantage of data mining techniques for benefit of the AOCS department of TAS, particularly for the Spacebus (SB) platform (P/F), at a time exploiting the works already in progress in the space operations field and focusing on research and development of new techniques, which were not yet applied in the domain. SB is a telecommunications P/F for geostationary Earth orbit (GEO) developed since 1980. SB Neo family was unveiled in 2014. Different thrust configurations are possible including an all-electric one [Bielecki and Lahlou, 2014]. For what concerns AOCS applications, vast quantities of raw data are collected mainly from two considerably distinct sources; for those, the main needs were identified by the company, for the purposed internship:

A. **Uncertainties reduction for the electrical station keeping mode:** MC simulation campaigns are performed in order to obtain security certifications for a given S/C. Our study will focus on MC simulations for the electrical station keeping (ESK) phase of SB Neo. We expect to find hidden relations between the different variables which are scattered in the simulations and possibly identify which uncertainties are determinant for the torque induced by the thrusters, and therefore the pointing accuracy of the S/C.

Never before data mining techniques were applied in the field and thus no *apriori* knowledge exists on whether some uncertainties can play a major role in this model or not;

B. **Novelty and anomaly detection for telemetry data:** TM signals are continuously verified on ground in order to check for abnormal behaviors on the S/C. Having available historical TM data for a few SB3000 and 4000 S/C and using data mining techniques, we expect to detect anomalies earlier than a ground operator would have done and hence give a larger time margin for the decisional chain. Anomaly detection is performed at these days using conventional techniques surveilled by ground operators. For what concerns the AOCS department, no data mining techniques are used for this surveillance. However, the anomalies are registered and processed *a posteriori* by the customer service and therefore are to be used as comparison for the results.

The remaining part of this document continues as follows: section 2 proceeds with a bibliographic study on data mining techniques which constitute the basis for the methodologies which will be applied in the two AOCS problems presented before; then, in sections 3 and 4 the two AOCS challenges are presented in detail and the suitable data mining techniques are chosen and applied; a elaborated explanation of the applied methodologies is provided; for both, results are therefore presented and discussed; finally, in section 5, we draw general conclusions about this study and point some possible paths for further development in the field of the data mining applications in the AOCS context.

## 2 Bibliographic review

A bibliography review was conducted in order to determine which techniques would be suitable for the proposed problems. In these section, we explore in detail different feature reduction techniques – to be applied on our ESK MC model – and methods for temporal series data mining, focusing on novelty detection – to be used on anomaly detection for TM data.

### 2.1 Feature reduction

Exploring feature reduction processes, it is possible to infer about the relative importance of features present in a dataset. This idea will be exploited in this work, as the objective itself is to rank the features of a given dataset – in order to reason about its properties – and not to apply a further data mining method to data. Having knowledge about which variables are more relevant, savings can be made in the next round of data collection, since the model has fewer features and redundant or irrelevant features were previously removed.

#### 2.1.1 Feature selection

Feature selection algorithms intend to select a subset of features from an initial large set, without changing the nature of collected data. These methods fall into three conceptual frameworks – the filter, wrapper and embedded methods – and can rely on two main notions: feature ranking and subset selection.

**Filter methods:**filter methods do not try to optimize the performance for any particular data mining algorithm as they do not have any information about the model [Kantardzic, 2011]. These methods rely therefore on a ranking function – which measures the relevance of each one of the features when relating to another – and a search method – in order to select a subset of features, maximizing a function according to an evaluation function score, based on this ranking.

Different filter methods can be applied using a variety of search heuristics, such as correlation coefficients or MI, which we present and compare in the following sections.

***Algorithms based on correlation:***correlation coefficients would be the first choice for estimating relations between two or more variables.

The major drawback while using Pearson's correlation-based ranking algorithms is its nature by itself, since it only reflects the linear relations between variables – though dependencies between features will be nonlinear in most cases [Kantardzic, 2011].

***Algorithms based on mutual information:*** a number of relevance indexes were introduced to compute the goodness of features for predicting decisions – its relevance in a dataset – in order to sidestep the shortcomings of Pearson's correlation coefficient, as MI, dependency or consistency [Hu et al., 2011].

Regarding feature selection, MI has proven to be an effective measure. MI is derived directly from entropy concept and measures the mutual dependence between two variables. Its value quantifies the knowledge of a variable obtained through another one. In an intuitive way, the MI between two variables measures the shared information among them and how much the information about one of them reduces the uncertainty about the other.

The **mutual information** between two variables $X$ and $Y$, with $p(x_i, y_j)$ the joint probability function and $p(x_i)$, $p(y_j)$ the probability function for $X$ and $Y$, respectively, is given by equation (1):

$$MI(X,Y) = -\sum_{i,j} p(x_i, y_j) \log \left( \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right). \quad (1)$$

An important characteristic of MI is its intrinsic relationship with Shannon's entropy definition [Shannon, 1948]: the MI between two variables is directly related to the entropy of each one individually and the joint entropy between by equation (2):

$$MI(X,Y) = H(X) + H(Y) - H(X,Y). \quad (2)$$

where $H(X)$ and $H(Y)$ are the entropies of $X$ and $Y$, respectively and $H(X,Y)$ their joint entropy.

For discrete data, most methodologies proceed by estimating the probabilities by relative frequencies.

For continuous data, the simplest way is to keep the methodology above: binning thus discretizing data and then go on as for discrete variables. The naive or empirical estimator for the entropy of a continuous variable is the entropy of the empirical histogram distribution, which is itself a maximum-likelihood estimate of the discretized frequency distribution. Miller [1955] and Chao and Shen [2003] propose corrections to the naive estimator, which can be used to estimate entropy and therefore MI.

**Kraskov et al. $k$-NN MI estimators:** avoiding the discretization problems, Kraskov et al. [2004] present two classes of estimators for MI,based on entropy estimates from $k$-nearest neighbor distances. The parameter $k$ defines the size of the neighborhood to take in account for the computation of the MI and has to be chosen. They propose to use typically integers values between $k = 2$ and 4.

**Hu et al. neighborhood estimators:** also using a non-binning approach, Hu et al. [2011] generalize Shannon's information entropy to a neighborhood information entropy and therefore propose the concept of neighborhood mutual information, introducing a parameter $\delta$ to control the granularity for data classification problems: the coarser the granularity, the larger the decision boundary region would be. According to observations made by Hu et al. [2008], $\delta$ should take values in the range $[0.1, 0.2]$.

**Minimal redundancy maximum relevance:** having the MI estimated for all variables in a dataset, different approaches can be followed to rank and select features, which would use either MI directly or derived measures. The first and basic approach is to select the top-ranked features. Ding et al. [2005] point out this simple ranking approach could lead to selecting only variables correlated among them, raising the redundancy issue. Therefore, they propose a new approach, requiring features to be at the same time maximally dissimilar between each other. This approach is called the minimal redundancy maximum relevance (mRMR). This mRMR score is computed for each feature; thus a significance ranking is obtained. The issue of selecting the number of features for the minimal set, is kept and has to be overcome with some choosing algorithm.

## 2.2 Temporal data mining

Temporal data mining can be considered as two equally important parts: a preliminary phase where data is processed; and the actual mining phase. After preprocessing, learning tasks for non-temporal data can be applied for temporal series. One of the most frequent tasks is either related with clustering the data and detecting frequently appearing patterns or, in opposition, anomalies. Novelty and anomaly detection concepts will be exploited in this work, as the objective itself is to detect behavioral changes in temporal series.

### 2.2.1 Preliminary processing of time series

Representation, summarization and segmentation techniques are preponderant steps before actually mining temporal data and will be discussed in the following sections.

**Segmentation:** segmentation can either refer to construct a model from $K^*$ piecewise segments, where $K^* \ll N$, closely approximating the original time series or simply partitioning the series into $K^*$ internally homogeneous sections [Ratanamahatana et al., 2010].

**Representation:** a proper representation can decrease the size of data in a way it becomes treatable. Multiple solutions exist to approximate time series for different applications and domains. Among data adaptive techniques, there are the singular values decomposition (SVD) and the piecewise polynomial approximations. Discrete Fourier transform (DFT) and discrete wavelet transform (DWT) are non-data adaptive techniques [Ratanamahatana et al., 2010]. The relevance of representation is crucial as storage, transmission and further computations with raw data is prohibitive in most cases.

**Summarization:** summarization can be described as creating an approximation which retains the essential features of data [Ratanamahatana et al., 2010]. Given the possible size of time series, a previous summarization step can be not only useful but rather obligatory before proceeding with mining techniques. Common statistical summaries of data are the mean, the standard deviation, among others. Tourneret [2016] suggests an extensive list of summary variables including statistical summaries, extreme values and parameters derived from the frequency domain.

### 2.2.2 Novelty detection

Kandhari et al. [2009] attempt to review broadly the anomaly detection problem, regarding different application domains and research areas. Pimentel et al. [2014] address the novelty detection task, providing a structured review of the most recent studies also for different application domains.

The most straightforward approach is to learn a description of the normality, using normal behavior samples, and then comparing unseen patterns with this normality. The normal behavior can be defined as a region; observations which do not fall into this region are considered anomalies. In the following, we focus on statistical-based algorithms for anomaly detection in time series.

**Statistical-based algorithms:** anomaly detection algorithms based on probabilistic approaches are based on estimating in some way the generative probability density function of the training data. Then, a threshold can be defined to limit the boundaries of normality. The OOL concept relies on defining upper and lower thresholds for samples: whenever a sample goes out of these *apriori* defined limits, it is considered anomalous.

The simplest methods are based on statistical hypothesis tests. As first approach, assuming normality

of data, the Grubbs' test and its variants compute the deviation from the mean of the samples to each sample; any sample which is far than a given threshold $c$ is considered an outlier or anomaly [Grubbs, 1969]. Outliers are thus defined to be any samples outside the range:

$$\tilde{R} = [\bar{x} - c\hat{\sigma_X},\ \bar{x} + c\hat{\sigma_X}]\,. \tag{3}$$

Winters et al. [2014] present two simple approaches for anomaly detection for predictive maintenance based on OOL and the Grubbs' test.

Instead of being based on the most typical summaries like the mean and the standard deviation, other methods flag samples based on other variability measures. Tukey [1977] proposed a method based on the interquartile range, defining an outlier to be any observation outside the range:

$$\tilde{R} = [\mathrm{Q}_1 - e\,\mathrm{IQR},\ \mathrm{Q}_3 + e\,\mathrm{IQR}]\,, \tag{4}$$

where $e$ is any non-negative constant. Tukey [1977] defined the boundaries at $e = 1.5$ for *outliers* and $e = 3$ for *far out* data points.

Martínez-Heras et al. [2012] refered that one shortcoming is that some OOL alarms are completely expected, previously known, and therefore they should not be counted as anomalies.

## 3 Uncertainties reduction for the electrical station keeping mode

MC simulations rely on randomness to solve problems which can be deterministic or not, by repeated random sampling in order to obtain numerical results. The basic procedure is to sample from a probability distribution for each input variable to produce hundreds or thousands of different outcomes. These results are then analyzed to infer about the multiple scenarios which can occur. Particularly, this study is placed on the case of MC simulations for modeling behaviors with significant uncertainties in inputs, as we are particularly focused on the ESK phase of the SB Neo P/F. We aim at using feature reduction techniques to identify which uncertain parameters are determinant for this model, particularly, which ones will be more *related* with the estimated torque and thus the pointing accuracy.

### 3.1 Electrical station keeping for the Spacebus Neo

During ESK phase, the S/C is in Earth pointing attitude and all appendices are deployed. The attitude control loop is performed on the well know star tracker (STR)-reaction wheel (RW) closed loop: the inertial attitude is obtained using STRs and its attitude is kept with RWs. Four electric propulsion systems (XPSs) are used, all connected to a power processing unit (PPU) and an inert gas supply.

### 3.2 Electrical station keeping model

For station keeping (SK), one of the four available Hall effect thrusters (HETs) is sequentially fired. The generated torque depends on three physical quantities: the thrust applied, the thrust application point and the S/C COM, symbolically corresponding with equation (5):

$$\mathbf{M} = \mathbf{GP} \times \mathbf{F}, \tag{5}$$

where $\mathbf{M}$ is the generated torque, $\mathbf{GP}$ a vector from the COM toward the force application point and $\mathbf{F}$ the thrust force. Knowing the S/C configuration, it is then possible to compute the ETHM positioning angles in order to have thrust pointing towards the on-board known COM.

For each firing, a unique thruster is used. Thrust norm and direction will undergo evolution over time after firing and as consequence the generated torque will fluctuate. The application point of the force and the COM are taken constant for each firing. The thrust vector orientation and norm are picked randomly at each new burn. For about [X]minutes, a transient behavior is present for both orientation and norm. Steady state behavior follows a particular periodic pattern and it is here approximated, for simplicity, by a sinusoidal signal, with random frequency. Using suitable rotation matrices between the different frames involved and using the information about the ETHM angles for the thruster being used at each firing, the computation of the thrust direction is done assuming to target the COM. The $\mathbf{GP}$ vector is thus obtained. Having $\mathbf{GP}$ and full knowledge about $\mathbf{F}$ the torque is computed using equation (5).

### 3.3 Uncertainties model

The model uncertainties can be reduced to the electrical thrust uncertainties (force, orientation and application point) and to the COM uncertainties.

Considering the COM, three sources of uncertainties are identified:

- the dry measurement accuracy, $\Delta G^D$;
- the filling ratio of propellant reservoirs, $\Delta G^F$;
- the xenon volume distribution over its tanks, $\Delta G^X$.

Thrust application point uncertainties are directly related with the ETHM motor angles uncertainties:

- the thermoelastic geometry uncertainties, $\Delta\Theta^T$;
- the initial position, which is not known with precision, $\Delta\Theta^0$;
- the backlash effect, $\Delta\Theta^B$.

For the electrical thrusters, two main categories of uncertainties have to be taken into account: the thrust force, $F$, uncertainties and the thrust direction uncertainties, which can be described in terms of a HCA, $\theta$, and an orientation angle, $\varphi$. For each, three main uncertainty sources are considered:

- reproductibility, the recurrence observed from one thruster to another, for a given S/C, $\Delta\bar{F}$, $\Delta\bar{\theta}$ and $\Delta\bar{\varphi}$;
- repeatability, the repeatable extent from burn to burn, $\Delta\tilde{F}$, $\Delta\tilde{\theta}$ and $\Delta\tilde{\varphi}$;
- stability during time, $\Delta F$, $\Delta\theta$ and $\Delta\varphi$.

The total number of simulations $N$ is defined. The number of simulations for each S/C $N_{S/C}$ is also fixed – as the reproductible values will change between S/C and between thrusters but not between firings.

### 3.4 Feature reduction using mutual information

We took the presented uncertainties as input features and used the torque norm and three Cartesian components, in the S/C frame, as target features. We started by analyzing the results for the torque norm. The goal was to infer which feature or which subset of features would be more significant for the torque norm behavior.

After justifying the shortcomings of using correlation-based algorithms – since no linear relationships were detected using scatter plot matrices (SPMs) and correlation maps – and conducting a variability study – in order to infer which features induce more variability in the target feature, we followed by applying and evaluating multiple MI estimator.

We remark that all results that were obtained present low values for the MI estimations (all MI based rankings were normalized with respect to the MI between the target feature and itself, which would rank 1).

We concluded that estimators based on bins, which discretize continuous variables, are not adequate for this problem and therefore chose to proceed with two others, which follow different approaches: the first $k$-NN estimator from Kraskov et al. [2004] and the NMI estimator from Hu et al. [2011]. These estimators depend on the selection of parameters $k$ and $\delta$ and a generic algorithm was applied.

Regarding the obtained results, we note that Kraskov et al. estimator is sensible to the modeling which was retained for the different reproductible uncertainties, whereas the Hu et al. is not. The issue of selecting the number of features for the minimal set is kept.

In order to validate the different methods used, we introduce an additional control variable in the initial dataset. This control variable, noted $U$, corresponds to a random uniform distribution between 0 and 1.

Using this control variable $U$ would be a good way to select the subset – considering all features which rank higher than $U$ – however we observe that in most cases this number is still large.

MI rankings, figure 1, allow inferring about the relevance of features regarding the target feature, not taking in account possible redundancy between these features. Regarding the torque norm, we observe that the reproductible uncertainty of the thrust HCA is clearly the most relevant feature. Since rankings for the different estimators do not perfectly match, we can select a common subset of features. Selecting the 10 top-ranked features and considering only the ones which are present for both estimators we have the subset from table 1.

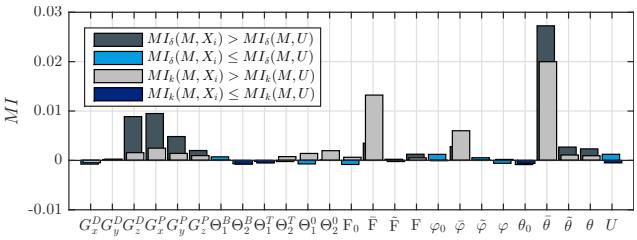



**Figure 1:** MI ranking between the torque norm and each input feature. Comparison between Kraskov et al. estimator with $k = 19$ and Hu et al. estimator with $\delta = 0.17$ (noted $MI_k$ and $MI_\delta$, respectively). MI values are normalized using the target feature entropy. Bars are colored in either light gray or dark gray if they rank higher than the control variable $U$ and light blue or navy blue otherwise.

The high relevance of the thrust HCA was expected. Indeed, a major challenge is how to estimate both reproductible and repeatable uncertainties for the thrust HCA, provided that are no measures obtained for the thrust direction itself. In opposition, the significant influence of the position of the COM was not anticipated; particularly, the large relative difference between almost all COM related uncertainties and those associated with the ETHM position was quite surprising. We remark the specific effect of the filling ratio tanks as well as the xenon distribution: the three Cartesian coordinates figure in the selected set. We note though that the estimations used for the range of this uncertainties were given with little or no prior study, contrarily to the ones related to the dry measurement error which were already further studied.

Furthermore, we remark that results concerning the MI between the uncertainties and the torque norm highly correspond with the results issued from a variability analysis done preliminary: features which induce greater variability in the torque norm are those which rank higher considering MI.

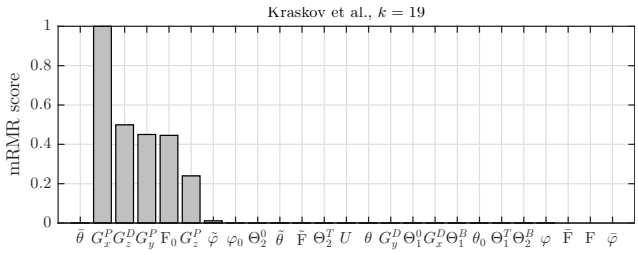
**Table 1:** Selected features regarding relevance.

| Uncertainties | Short description |
|---|---|
| HET firing | Reproductible HCA |
| COM position | Tanks filling ratio and xenon distribution<br>Dry accuracy measurement error<br>Tanks filling ratio and xenon distribution |
| HET firing | Repeatable HCA |
| COM position | Tanks filling ratio and xenon distribution |

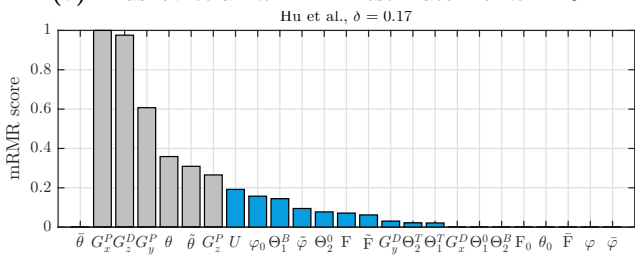
### 3.5 Feature reduction using mRMR

In addition to MI, we obtained and tested the results using the mRMR criterion, which derives directly from it. This first ranking approach can lead to select only variables related among themselves that are redundant, since no redundancy measure was taken in account. Therefore, we followed by restraining features to be at the same time maximally dissimilar between each other and maximally relevant. This corresponds to the mRMR approach, searching to reduce possible redundancy in the selected set of features.

Since we applied two different MI estimators, two mRMR rankings were also obtained (figures 2a and 2b) Kraskov et al. estimations are biased for the reproductible uncertainties for the thrust norm and orientation angle relevant; however, these are not highly ranked using mRMR score which means they are redundant with the thrust HCA factor. Results using Hu et al. scores show that all features selected in table 1 are relevant and non-redundant among them as they are the top ranked in terms of mRMR score (only the stability uncertainty of the HCA of thrust is added). Other than this, we remark that these are the only features which rank better than the control variable and thus the subset of table 2 could be potentially representative.




**(a)** Kraskov et al. $k$-NN MI estimator for $k = 19$.




**(b)** Hu et al. NMI estimator for $\delta = 0.17$. Features are normalized between 0 and 1.

**Figure 2:** mRMR scores for 5000 samples. MI computed between the torque norm and each input feature. The first-selected feature corresponds with the maximal MI and therefore there is no computed mRMR score. The remainder mRMR scores are normalized with respect to the second-selected feature. Bars colored in light gray rank higher than the control variable $U$ (otherwise colored in light blue).

**Table 2:** Selected features regarding both relevance and redundancy.

| Uncertainties | Short description |
|---|---|
| HET firing | Reproductible HCA |
| COM position | Tanks filling ratio and xenon distribution<br>Dry accuracy measurement error<br>Tanks filling ratio and xenon distribution |
| HET firing | Stability HCA<br>Repeatable HCA |
| COM position | Tanks filling ratio and xenon distribution |

## 4 Novelty and anomaly detection for telemetry data

Predictive maintenance techniques allow conveniently schedule corrective maintenance maneuvers and prevent unexpected equipment failures. Novelty or anomaly detection techniques are closely related with predictive maintenance, in the field of space applications. We distinguish anomalies from novel observations considering the former can be considered a newly observer *normal* behavior and therefore should be included in the normality models. The basic procedure to be followed is to acquire the observations and apply a dimensionality reduction process, based on temporal windows. Having then the time series represented as a smaller set, different methodologies to detect novel observations can be applied. Particularly, this study on anomaly and novelty detection in TM time series is aimed at detecting behavioral changes which can lead or not to posterior failures, namely the transition to a safe mode (SM) via a failure detection isolation & recovery (FDIR) procedure, for the SB family of S/C.

### 4.1 Available telemetry data

Since data is not collected for exclusive purpose of mining, we are constrained to a set of parameters related with the RWs, including, for each one of the four wheels, for 4 different S/C: the speed, motor current, temperature and estimated friction torque. TM data is available in different time periods, for at least 6 consecutive years.

### 4.2 Anomaly detection using a conventional approach

Surveillance of the TM signals is done by the AOCS customer service team. For RWs, the motor current and the estimated friction torque are monitored since these are the parameters which are mostly determinant for the triggering of the FDIR. There are two levels of alarm: the warning threshold, detected by ground operators, which is an informative level; the second level is the one at which the S/C would trigger the FDIR procedure by itself; thus we call it FDIR threshold. Since all S/C we are dealing with are in GEO, parameters are evaluated in a daily basis, using the mean as summary variable; thus. Corrective actions are decided when the warning threshold is attained or a clear tendency of future crossing of this is detected.

### 4.3 Temporal data classification

Having the signals analyzed using the DFT and the autocorrelation function, we want to infer about their properties in an automatized way, obtaining information on whether the temporal series is periodic or not, and if it is, having an estimation of its period. The identification of the type of temporal series is important in the measure of choosing a reduced set of features to summarize it. The categorization follows a defined decision tree and signals can be classified as:

- zero;
- constant;
- binary;
- classes;
- square aperiodic wave (SAW);
- noise-like signal (NLS);
- sinusoidal periodic wave (SPW).

If the signal does not fall in any labeled condition in the decision tree, it stays unidentified.

Having identified the characteristics of the different available TM signals and grouped them into broad classes of time series, we can thus proceed with the temporal segmentation and then the application of summarization techniques, choosing an adequate and small set of descriptors, based upon these categories.

### 4.4 Telemetry segmentation

In a segmentation step, we simply partition the series into internally homogeneous sections, for a piecewise linear approximation. Following the approach of Martínez-Heras et al. [2012], we will divide the TM time series in equal-period windows. Having in mind that we are dealing with GEO spacecraft, we choose to split the time series in windows of 1 day each.

### 4.5 Telemetry summarization

Since the number of features is important for the performance of mining algorithms, inferring about which features are more prominent for each TM parameter in the dataset can be of great interest for posterior anomaly detection. Thus, we are interested in choosing the right summaries or descriptors for each class of TM signal identified.

Manufactured data is used to introduce and evaluate a methodology to choose the more pertinent descriptors for the identified signals. We define a parameter which, intuitively, the closer it is to one, the larger is the average difference between one abnormal and one normal window, using a given descriptor. Contrarily, if the descriptor does not differ a lot (in average) between normal and abnormal windows, this parameter value tends to zero. We start with the initial descriptors set presented by Tourneret [2016] and we try to select a subset which should summarize the information contained in each time window for a given time series.

**Table 3:** Telemetry parameters descriptors.

| Periodicity | Class | Descriptors |
|---|---|---|
| **Aperiodic** | Zero | $\bar{x}$, $\max X$, $\min X$ |
| | Constant | $\bar{x}$, $\max X$, $\min X$ |
| | Binary | $\max X$, $\min X$ |
| | Classes | $\bar{x}$, $\max X$, $\min X$ |
| | SAW | $\bar{x}$, $s_X$, $E_X$, $N_{\bar{x}}$ |
| | NLS | $\bar{x}$, $s_X$, $E_X$, $N_{\bar{x}}$ |
| **Periodic** | SPW | $\bar{x}$, $E_X$, $\tilde{E}_X$, $N_{\bar{x}}$ |
| | NLS | $\bar{x}$, $E_X$, $\tilde{E}_X$, $N_{\bar{x}}$ |

To each category of TM signals identified, we associate no more than 4 descriptors from the initial set (table 3). Regarding the results obtained, we expect the selected subset of descriptors to be enough to characterize the parameters of one given category in a way anomalies can be detected using different methodologies. Having the series partitioned into internally homogeneous sections, a summarization step follows which intends to create an approximation which retains the essential features of data [Ratanamahatana et al., 2010], using the selected descriptors.

### 4.6 Anomaly detection using an automatic out-of-limits approach

The first and simplest concept for detecting anomalies in data is the OOL methodology. This technique is based on defining upper and lower thresholds for samples. Then, whenever a sample goes out of this *apriori* limits, the sample is considered anomalous. Intuitively, complexity can vary in an important way, since different methods to define the thresholds can be applied. Our first approach is based on the simplest methods, assuming normality of data and computing the standard deviation from the mean of the dataset to each sample. Then, we will propose a method based on the interquartile range, as Tukey [1977]; Solberg and Lahti [2005]. For this approach, the thresholds are computed based on the first and third quartiles; the IQRis used as measure of scale.

#### 4.6.1 Triage and regrouping anomaly detections

Given the huge quantities of data and the computational resources available, detections are done in fixed windows of one year; in addition, we have the advantage of having one TM parameter re-classified at each year, allowing for a better detection. For both methods, common criteria are used to sort the detected outliers. Firstly, the outliers are detected using one of the two approaches, for a given set of parameters; the detections are therefore sorted by date in a list; then, the list is screened and split when two consecutive detections differ more than 1 day. The grouped detections are then sorted using multiple exclusion criteria.

#### 4.6.2 Gaussian-based OOL technique

We start by applying a Gaussian-based OOL method to our segmented and summarized data. This method assumes normality of data and then computes the z-score of each sample that is the number of standard deviations the sample differs from the mean:

$$z_i = \frac{|\bar{x} - x_i|}{s_X} \tag{6}$$

We decide to perform the detection considering thresholds of 3, 4 and $5s_X$ apart from the mean. Each window will be considered an outlier if its z-score is greater than these values.

Figure 3 corresponds to the estimations of the mean and the $5s_X$ boundaries on the wheel speed TM parameter for the SAT3 S/C (classified as SPW and thus described by 4 summarization variables), where this method is applied.

After applying the triage process and grouping the anomalies into events or anomalies, we are able to reduce greatly the number of detections. However, we remark that the number of alarms is still important. Moreover, we expect to find an important part of this outliers which were completely anticipated, previously known, and therefore they should have not been

counted as outliers (as pointed previously by Martínez-Heras et al. [2012] as a shortcoming of this method). Since we possess listings from historical events detected using traditional methods, we can compare those with the ones obtained and infer about the success of applying this methodology.

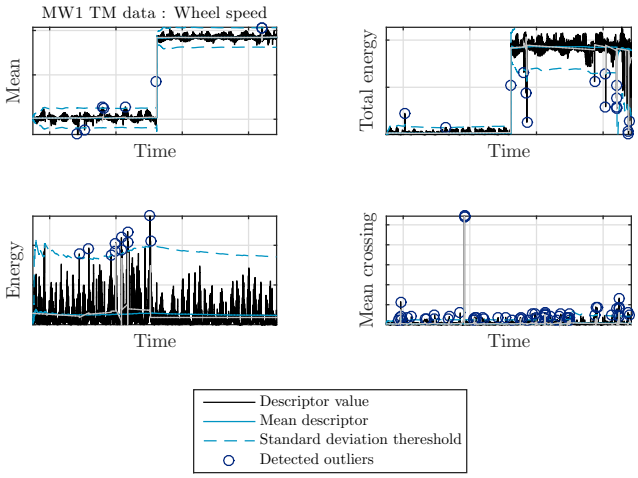



**Figure 3:** Gaussian-based OOL anomaly detection for the momentum wheel (MW)1 speed from SAT3. The mean, the total energy, the energy in the signal band and the mean crossing counting are used as descriptors. Outlier boundaries are computed using the mean and the the $5s_X$ boundary for each summary feature.

On table 4, we present the summary of the concordant events found by this approach for each S/C, within a day or a week interval. We note that, using the Gaussian-based approach, we are able to detect a part of the relevant events – related with the RW behavior. However, results are not consistent among S/C. We observe that relaxing the *detection window*, that is the difference between the listed dates and the ones at which we are able to find anomalies, to a week gives mixed results. As last remark, we note that we are able to detect primarily brutal behavior changes (the replacement of a fixed momentum wheel (FMW), the reaching of a FDIR threshold) and not its causes – in general increasing trends in the motor current, the estimated friction torque or the temperature – with few exceptions.

**Table 4:** Concordance between the observed and detected anomalies.

| S/C | Total | Gaussian-based | | IQR-based | |
|---|---|---|---|---|---|
| | | Day | Week | Day | Week |
| **SAT1** | 2 | 1 | 1 | 1 | 1 |
| **SAT3** | 3 | 2 | 5 | 2 | 4 |
| **SAT2** | 7 | 3 | 4 | 1 | 1 |
| **SAT4** | 5 | 0 | 1 | 0 | 0 |

Concordance between the observed events and the anomalies detected using both the Gaussian-based and the IQR-based OOL methods. Boundaries of $5s_X$ and 3 IQR respectively.

#### 4.6.3 IQR-based OOL technique

As second approach, we apply an IQR-based OOL method to our available data. In a similar way to the computation of the z-score of each sample, here we compute the distance to the boundaries, defined by $\mathrm{Q}_1 - e\,\mathrm{IQR}$ and $\mathrm{Q}_3 + e\,\mathrm{IQR}$ as:

$$w_i = \begin{cases} \frac{|\mathrm{Q}_1 - x_i|}{\mathrm{IQR}} & \text{if } x_i < \mathrm{Q}_1 \\ \frac{|\mathrm{Q}_3 - x_i|}{\mathrm{IQR}} & \text{if } x_i > \mathrm{Q}_3 \end{cases} \tag{7}$$

The thresholds are defined for each fixed $e = 1.5$, 3, 4.5 and 6. When comparing with the previous results, we remark that the number of detections is even larger, which leads to conclude that even tighter boundaries should have been used (more than 6 IQR for the limits). Given the observations, we remark that this method fails to find the most preeminent events, performing worse than the Gaussian-based technique. This can be justified by the huge number of detections when compared with the precedent methodology.

### 4.7 Predictive analysis

Our ultimate goal using data mining techniques for anomaly detection is to create tools for predictive maintenance which would be able to detect early signs of deterioration before the conventional methods. As test case, we used the detection of an increasing trend in the friction torque on FMW3 for the SAT2, noted by the customer service of AOCS department. We remark that this anomalous event is being detected using the Gaussian-based but not the IQR-based OOL method. Thus, we consider the Gaussian-based OOL method, with the $5s_X$ boundaries based on the mean and the standard deviation. The particular detection of this increasing trend in the friction torque is followed by another one which corresponds to another event detected using conventional techniques.

For predictive maintenance, we shall focus on the first detection window, in order to infer about when we could trigger an alarm, using the results of this technique. Ideally, this alarm would be triggered before the warning threshold; otherwise, the anomaly would be detected first by ground operators and only then by the algorithm. However, a single alarm is in general not considered sufficient by them for taking corrective actions.

We re-run our algorithm using the Gaussian-based OOL method using all data available until the day before the event. If our method is capable of finding this particular anomaly using this data, we are capable of providing an early alarm than the one provided by the warning threshold. Using this procedure, we simulate the possible application of this method, receiving TM data and performing the computation at each day.

We observe that the warning threshold is indeed reached before our method is capable of triggering an alarm. However, we have seen that one single crossing of the thresholds is not considered a valid argument for any FDIR procedure, since signals can frequently

pass this value and thus warnings are generally based on daily means. For this situation, the daily mean is never superior to the warning threshold.

The AOCS costumer service has no records of an increasing trend on the estimated torque increasing before the 9th July, even if the warning thresholds was reached several times. Since we need the whole window to apply the summarization and then the detection steps, the first alarm could be triggered on the 4th July. Additionally, a second alarm would be triggered at the 8th July. Therefore, our method is capable of detecting this increasing trend and trigger an alarm before this date.

## 5 Conclusion

Given the interest nowadays in big data challenges, this study – conducted in an entrepreneurial internship context – arises from the idea of taking advantage of data mining techniques in benefit of the AOCS, particularly for the SB S/C of Thales Alenia Space, focusing on research and development of new approaches. While some data mining applications can be found in the space industry, no data mining techniques are yet applied for the AOCS field in particular. Two main applications were defined: the first, using feature reduction techniques for the uncertainties on the ESK phase model; the second, applying novelty and anomaly detection techniques on historical TM data.

Methods to identify which uncertainties are determinant for the pointing accuracy of a S/C can be used to increase effectiveness of MC simulation campaigns. MC simulation campaigns are used to infer properties about a S/C behavior, with significant uncertainties on the inputs. Particularly, our study focused on the ESK phase of a SB Neo.

We used a simplified MC model and feature reduction techniques to find hidden relations between the different uncertainties which are present in the model, and thus spread-out in the simulations, and the computed torque. We began by introducing our model for the ESK phase with uncertainties on the inputs and we explored different possibilities for feature reduction techniques. Measures of relevance between features play an important role on feature selection and MI has been proved to be a satisfactory measure for relevance. Nevertheless, there are limitations in computing continuous numerical features due to problems in estimating the probability density functions.

In this work, we studied the application of two estimators for MI. In contrast to conventional estimators, based on binnings, no discretization is required to compute relevance when using the proposed algorithms. The first is based on entropy estimations from k-nearest neighbor distances and the second generalizes Shannon's information entropy to neighborhood information entropy. While the first implies choosing the size of the neighborhood to consider, the second introduces a parameter which allows controlling the granularity in analyzing data. We provided an algorithm to efficiently choose these parameters in order to maximize concordance between results. Then, we compared the algorithms and showed that neighborhood MI and $k$-nearest neighbor estimators provide nearly the same results. We combined both MI estimations with redundancy measures in order to proceed to feature selection, using the maximal relevance minimum redundancy criteria. Finally, the algorithms were used on the model to infer about which are the key uncertainties among the initial set.

The results show that uncertainties related to the thrust force HCA and the positioning of the COM are the most relevant for the model. While the importance of the thrust direction was expected, the difference between the COM and the ETHM positioning was not anticipated.

Detecting anomalies or novel behaviors in TM data is a fundamental step for predictive maintenance. Predictive maintenance allows scheduling in advance corrective maneuvers and preventing unexpected equipment failures. Advantages include increased equipment lifetime, fewer failures and thus cost savings. TM signals are continuously verified on the ground in order to check for abnormal behaviors on the S/C. These temporal series can be processed using knowledge-discovery techniques in order to detect early signs of degradation, thus preventing in advance the S/C to trigger a FDIR procedure. In this context, novelty or anomaly detection techniques are closely related with predictive maintenance.

Particularly, our study focused on the detection of anomalies on TM data. We adopted an OOL methodology to detect and classify anomalous windows in TM based on a three-step process: first, validation of temporal data; second, the segmentation and summarization of the time series, using descriptors; and finally, the anomalous event detection. We started by performing a temporal and spectral analysis which provides data for the application of the anomalies detection method. We provided an approach to split the time series into windows and classify the temporal series. This classification was later used to choose the adequate summary descriptors for temporal series. Having the time series represented as an ensemble of windowed descriptors, we analyzed the data using an OOL approach. This method consists of defining upper and lower thresholds for the descriptors and, when a descriptor goes above the upper limit or below the lower one, an alarm is generated. Using the TM data, we apply different thresholds based firstly on the mean and the standard deviation, and secondly, on the interquartile range, the first and third quartiles. Using observations, we chose the adequate thresholds in order to maximize the concordance between the observed results and the ones previously obtained. We compared the results of both algorithms and showed that the results are consistent, the first approach performing better than the one based on the interquartile range. Finally, the Gaussian-based method is used on a specific

case to infer about whether it would be possible or not to use these techniques for triggering alarms for ground operators.

Results indicate that this approach for automatic TM monitoring is capable of triggering alarms when trends are present in data, before multiple warning thresholds, provided by conventional methods for anomaly detection, are reached. However, the number of irrelevant detections is important and the number of detections where preventive actions could be taken is too small for taking conclusion on whether this method would allow to efficiently prevent failures.

Following the study on feature reduction techniques, it would be interesting to infer the results of using the maximal dependency criteria. Regarding more sophisticated data mining techniques, the direct estimation of the distribution and then the computation of the MI could also be a path to follow. Regarding this model in particular, there is interest by the company to proceed this study and extend the model to compute directly the pointing error, instead of the torque.

Given the provided methodology for anomaly detection in time series, more sophisticated data mining approaches can be applied in future work, using the preprocessing methods developed in this study. As suggestion, we point the local outlier probabilities (LoOP), a local density based outlier detection method, and the one-class support vector machine (SVM), based on normality regions. Both are being nowadays tested for space operations applications in a research context and being considered by the company as a path to further studies.

## References


S. Bielecki and T. Lahlou. The Spacebus family is growing with Spacebus NEO, 2014.

A. Chao and T.-J. Shen. Nonparametric estimation of Sannon's index of diversity when there are unseen species in ample. *Environmental and Ecological Statistics*, 10:429–443, 2003.

C. Ding, H. Peng, and H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(2):185–205, 2005.

F. E. Grubbs. Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1–21, 1969.

Q. Hu, D. Yu, J. Liu, and C. Wu. Neighborhood rough set based heterogeneous feature subset selection. *Information Sciences*, 178(18):3577–3594, 2008.

Q. Hu, L. Zhang, D. Zhang, W. Pan, S. An, and W. Pedrycz. Measuring relevance between discrete and continuous features based on neighborhood mutual information. *Expert Systems with Applications*, 38(9):10737–10750, 2011.

R. Kandhari, V. Chandola, A. Banerjee, V. Kumar, and R. Kandhari. Anomaly detection: A Survey. *ACM Computing Surveys*, 41(3):1–6, 2009.

M. Kantardzic. *Data Mining: Concepts, Models, Methods, and Algorithms.* 2011.

A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69(6 2):1–16, 2004.

B. Lantz. *Machine Learning with R.* Packt Publishing, Birmingham, 2015.

J.-A. Martínez-Heras, A. Donati, M. G. F. Kirsch, and F. Schmidt. New Telemetry Monitoring Paradigm with Novelty Detection. *SpaceOps 2012 The 12th International Conference on Space Operations*, pages 1–9, 2012.

G. Miller. Note on the Bias of Information Estimates. *Information Theory in Psychology: Problems and Methods*, pages 95–100, 1955.

M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.

C. A. Ratanamahatana, J. Lin, D. Gunopulos, E. Keogh, M. Vlachos, and G. Das. Mining Time Series Data. *Data Mining and Knowledge Discovery Handbook*, pages 1049–1077, 2010.

C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

H. E. Solberg and A. Lahti. Detection of Outliers in Reference Distributions: Performance of Horn's Algorithm. *Clinical Chemistry*, 51(12):2326–2332, 2005.

J.-Y. Tourneret. Analyse TM pour Diagnostics. 2016.

J. W. Tukey. *Exploratory Data Analysis.* Addison-Wesley Publishing Company, 1977.

P. Winters, I. Adae, and R. Silipo. Anomaly Detection in Predictive Maintenance: Anomaly Detection with Time Series Analysis, 2014.