International Journal of Robotics and Mechatronics 2018, Vol. 5, No. 1

Data-Driven Models for Satellite State-of-Health Monitoring and Evaluation

Ahmad M. Al-Zaidy[†], Wessam. M. Hussein[†], Mahmoud M.A. Sayed[‡], Ibrahim El-Sherif[†]

[†]Military Technical College, Cairo, Egypt

[‡]Canadian International College, Cairo, Egypt

Abstract— Satellite state-of-health conditions and keeping up systems safe are the fundamental worry of space operations. Since managing the correlation structure between variables give greater information than individual variables or causes relationships, Data driven methodology presented as an answer of managing enormous information. In this article a survey of some recent related work is done. Then a mathematical foundation for latent variable modelling, support vector machines (SVMs) and artificial neural networks (ANN) is illustrated. Then a four data driven models are introduced for monitoring, diagnosis and prediction of the thermal control of satellite power supply system. Two kinds of PLS-Batch models were implemented for the first time with a satellite telemetry. After that a utilization of multivariate Shewart, DModX in addition to CuSum charts for system monitoring and early fault detection. Another two machine learning models were built and coded in Python. These two models are (SVMs) model and (ANN) model. Both of the last two models are integrated with PCA for feature extraction and dimensionality reduction for visualization. The outcomes confirmed the effectiveness of utilizing such models in monitoring, faults detection, and diagnosis. Also system health evaluation could be established within the operation through these models. In addition to prediction of system behavior and early faults detection.

Keywords—

DOI: https://doi.org/10.21535/ijrm.v5i1.982

I. Introduction

SPACE mission operations manages enormous amount of telemetry data identified with satellite systems states. They always seek to reduce time and cost in analyzing data. Taking predictive or corrective action relies on upon how much valuable information could be extracted from telemetry. Telemetry data are of three classes: parameters, activities and alarms. Parameters are static objects change its status or value with time. Parameters introduce sensory data as well as modes of operations flag.

Attempting to extrapolate systems behavior some time, recently, amid and subsequent to taking any action is a major concern to space operations. This obliged models to be checked while taking any action. These models ought to be able to foresee the system status while running up with a particular operation. There are two methodologies of building such models. The first is to assemble a model of the system itself relying upon the design and configuration perspectives. The second is to build a model of genuine telemetry identified with system behavior amid genuine operations. This methodology could present a superior representation for genuine state particularly when managing all data immediately. Analyzing all data together may discover extreme and moderate outliers that would not appear when working with single variables.

Data mining with the data driven approach presented as an intense telemetry demonstrating and analyzing methodology. Data driven analysis with the two branches of machine learning and multivariate analysis. Data driven analysis like any statistical approach didn't deal with physical cause and effect between phenomena, but it deals with the correlation between variables as most techniques make data preprocessing to remove units and equalize the range of values. Each of both methodologies has multiple techniques to enhance analyzing, information extraction and decision making while working with all data together. This will diminish time consumed and enhance investigating capacity, consistency and predictability of the telemetry model. Both approaches have the unsupervised and supervised learning methodology relying upon the existence of desired output. Unsupervised learning is to extract information or features from unlabeled data as in principal component analysis PCA and partial least squares projections to latent structure PLS utilized for further classification or analysis. On the other hand, supervised learning has a labeling for each class of data, each class related to a certain region or mode of operation as PLS-DA –discriminant analysis- and the machine learning techniques support vector machine SVMs and artificial neural networks ANN. supervised models used to ensure the data will be coordinated to the targeted behavior of the system.

In this paper four telemetry modelling techniques is applied to the telemetry of the thermal control of NHBM. The first is PLS-Bach modelling performed in SIMCA-P using 10 good batches of 14999 observations which mostly have normal data to model the evolution of "good batches". This is done by fitting

Corresponding author: Ahmad M. Al-Zaidy
(e-mail: ahmadalzaidy46@gmail.com)

This paper was submitted on Oct 23, 2017; and accepted on Oct 23, 2017.

a PLS model with Y virtual batch time to 60 variables as Xs. This observation level model is used to monitor the evolution of the new batches (11 to16) of 10000 observations each of which have faults through scores and distance to model plots. The second model is built by fitting the batches to PCA making a model used to classify the upcoming batches to normal or faulty. In both previous models contribution plots are used for interpretation of anomalies. The batch process in the previous models is first time applied to satellite telemetry as the author's best knowledge.

Shewart and distance to model charts used for monitoring the telemetry observations, then CuSum charts is used to show how early we can detect the anomalies and preserve the system.

The rest two models are built and coded in Python. The models are built through two machine learning techniques SVMs and ANN. The two techniques are combined with PCA for dimensionality reduction and features extraction. PCA is chosen after a comparison with PLS-DA for which make better data representation in our case. The two techniques are built for multiclass classification with 3-classes of normal, deviated and up-normal. The models are built by 250 observations as training set containing 3-classes and same for testing.

II. Related Works

Researchers always attempts to find more accurate analysis techniques that can detect faults, make diagnosis, and prediction using multivariate analysis and machine learning approaches.

Tianshe Yang et al. [1] proposed several data mining methods for satellite telemetry data. Threshold, multi-threshold, and dynamic adaptive detection methods were used. According to the author the disadvantage of threshold detection is the disability to detect parameters' abnormal changes within the threshold; therefore; it's not suitable for early fault detection.

Lin Su et al. [2] the technique of PCA was applied to detect faults. Also, PCA fault isolation approach is improved to satellite attitude control system. According to the author due to faults complexity and disturbances existence, some knowledge should be combined with contribution plots to explain the static results in faults isolation.

Bo Lee et al. [3] apply PCA method to micro-satellite power supply system. According to the author the analysis was only based on single sensor failure. Therefore, the case of fault detection and reconstruction for multiple sensor failure should be studied.

Macro Vieira et al. [4] study the hypothesis of using clustering algorithms to perform telemetry analysis. Two clustering algorithms (K-means and Expectation Maximization) are applied. Regarding to the author the clustering algorithms are sensitive to data changes related to satellite aging. This way, permanent learning is required, which could lead to the learning of anomalies as normal behavior.

Yu Gao et Al. [5] and Xia et al. [6] introduced an integration of PCA and kernel PCA with SVMs respectively. They first build a binary SVMs binary model for fault detection, and the diagnosis if done through a multi class SVMs based on previous faults identification. However, any new type of faults or merged of two identified faults could not be easily diagnosed.

Fang et al. [7] studied the power supply system health evaluation based on fuzzy analytic hierarchy process (AHP), and the component health based on the statistical analysis, Bayesian network and Support Vector Machines. They use SVMs for 2-class classification for faults detection and behavior check. Regarding to the authors they didn't make health evaluation for subcomponents which is very important in faults diagnosis and accurately health assessment.

III. Latent Variable Modelling

A. Principal Component Analysis (PCA)

There are a lot of difficulties while dealing with a system of large number of sensors which form high dimensional matrix. Different variables ranges and units in addition to Visualizing sensory data for further analysis and interpretation are the most difficult issues inherent multivariate statistics.

Pre-treatment of data through scaling to unit variance and mean centering in order to unify the length of variables and their numerical values. Also enhances the interpretability of the generated model.

The aim of PCA is to find a principal component of direction that minimize the residual distance of each point to this line. This minimization is done through maximizing the variance of scores. The second component shall represent the second direction of higher variance. The second component should be perpendicular to the first component, where a score value t_i of each observation is the distance from the origin, along the direction vector p_i as in Figure 1.

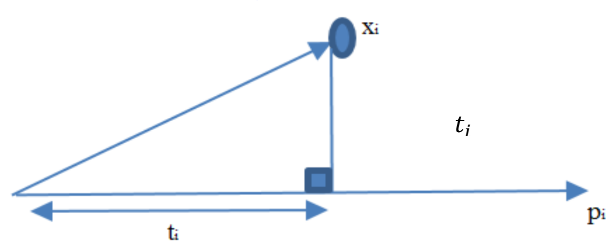

Figure 1

For entire matrix of data X, the scores matrix can be calculated as:

$$T = XP \quad \textbf{(1)}$$

$$(N * A) = (N * K)(K * A) \quad \textbf{(2)}$$

To compute PCA of matrix X there are different algorithms found in [9], [11] are used such as:

- Eigenvalue decomposition.
- Singular value decomposition.
- Nonlinear iterative partial least squares (NIPALS) which used through the upcoming work.

Cross validation found in [9], [10], [12] also is used to compute how many components that are needed for good estimate through the values of R^2 and Q^2, where R^2 estimates goodness of fit, and Q^2 estimates goodness of prediction.

B. Projections to Latent Structure (PLS)

Projection to Latent Structures (PLS) deals with data matrix X and respond matrix Y.

In PCA, the objective was to best explain X data. To obtain this scores and loadings are calculated. However, PLS calculate two sets of scores and loadings for X data and Y data.

As in PCA, data will be scaled to unit variance and mean centered, then NIPALS algorithm is applied to the two matrices X and Y to obtain the loadings and scores matrices such that

$$X = 1 * x^{\mathrm{T}} + T * PT + E \tag{3}$$

$$Y = 1 * y^{\mathrm{T}} + U * C^{\mathrm{T}} + F \tag{4}$$

So,

$$Y = 1 * y^{\mathrm{T}} + T * C^{\mathrm{T}} + G \tag{5}$$

Due to inner relation $U = T + H$,

$$E = X - T * P^{\mathrm{T}} \tag{6}$$

$$F = Y - T * C^{\mathrm{T}} \tag{7}$$

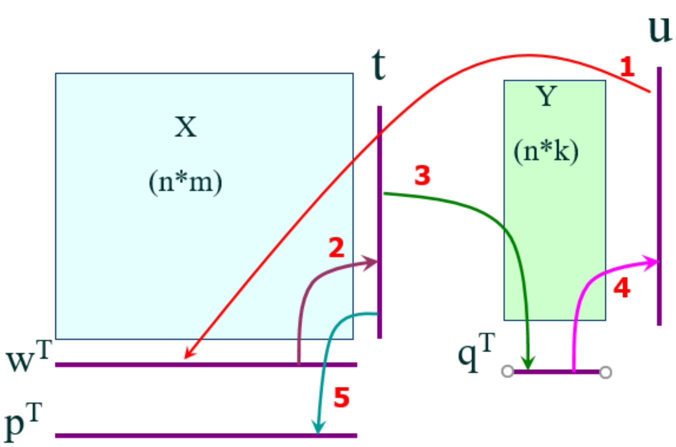

Figure 2 NIPALS algorithm procedure

The algorithm starts by selecting a column from Y_a as initial estimate for u_a. The X_a and Y_a matrices are just the preprocessed version of the raw data when a = 1.

Arrow 1:
$$w_a = \frac{u^T x}{u^T u} \tag{8}$$

Arrow 2:
$$t_a = \frac{x_a w_a}{w_a^T w_a} \tag{9}$$

Arrow 3:
$$c_a = \frac{Y_a^T t_a}{t_a^T t_a} \tag{10}$$

Arrow 4:
$$u_a = \frac{Y_a c_a}{c_a^T c_a} \tag{11}$$

Arrow 5:
$$p_a = \frac{t_a^T X}{t_a^T t_a} \tag{12}$$

This is one round of the NIPALS algorithm. The iterations are done through these 4 arrow steps until the u_a vector does not change much. Then, we store these 4 vectors: w_a, t_a, c_a, and u_a, which jointly define the a_{th} component.

C. Multivariate batch process analysis

Batch process data is always a three-way data structure as seen from Figure 3. The present approach uses all available data in the analysis by unfolding the three-way matrix and using PCA or PLS for the modeling of the process. Batch modeling is mainly based on two levels, the observation level and the batch level. The aim of the observation level is to build a model based on normal batches and use it for monitoring and fault detection, as outlined in this application. Figure 4 illustrates the unfolded matrix configuration of the available data which is to be used for modeling at the observation level. The main interest here lies in evaluating the individual observations and to specify the batch development by understanding their progress. On the other hand, the batch level focuses mainly on using all available data (initial conditions, variables, response variables like quality characteristics) in modeling and using this model for classifying new batches. In addition, batch level allows the understanding of the nature of the relation between the output response variable with available data (initial condition) during the progress of the batch process.

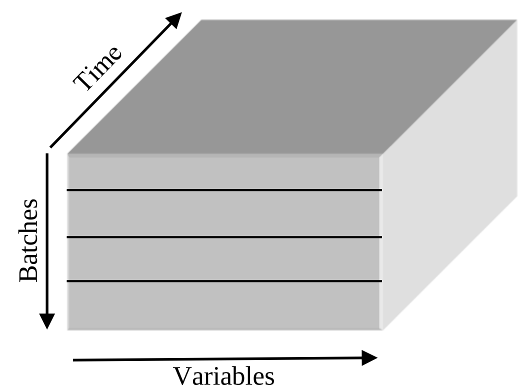

Figure 3 Data table of the batch process data

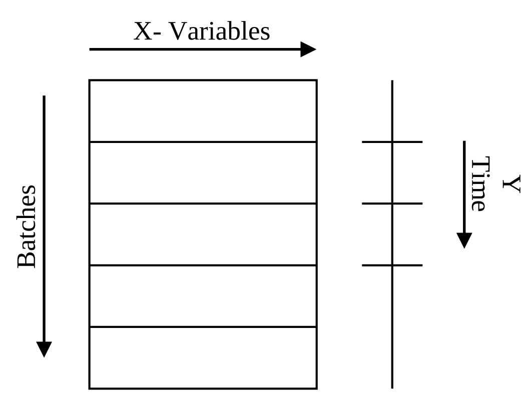

Figure 4 Unfolded data table preserving the direction of variables

In this application, the analysis aims at building a monitoring system for early fault detection by using observation level modelling. By unfolding the matrix data, one gets a two-way matrix with training batches times samples results in observations. A dummy y-variable is added to include local batch time in the model (automatically generated in SIMCA-P).

D. Projections to Latent Structure Discriminant Analysis (PLS-DA)

The purpose of PLS-DA is separating classes of observations through extension a PLS with encoded Y matrix with zeros and ones according to each class. Such that for class I the column I in Y matrix shall be one for observations in the class and zeros for the others.

E. Support Vector Machines

Support vector machines was developed by Vapnik in 1990s to be considered one of the powerful techniques in data mining to be used for classification and regression.

The aim of support vector machines is to separate two classes with maximum margin by identifying a hyperplane that can make this separation.

The hyperplane is identified with a weight vector w perpendicular to the hyperplane and a constant b such that:

$$w^T x + b = 0 \tag{13}$$

So, for an unknown sample x_i

$$w^T x + b \geq 1, \text{ for class } 1 \tag{14}$$

$$w^T x + b \leq -1, \text{ for class } 2 \tag{15}$$

Finding the optimal hyperplane of maximum margin of separation is subjected to maximizing $\frac{2}{||w||}$ which is equivalent to minimizing $\frac{1}{2}||w||^2$ instead of $\frac{1}{2}||w||$ for mathematical convenience. By using Lagrange multipliers with α_i as a Lagrange multiplier. For y_i such that $y_i = +1$ for class 1 and $y_i = -1$ for class 2,

$$L = \frac{1}{2}||w||^2 - \sum \alpha_i [y_i(w^T x + b) - 1] \tag{16}$$

$$L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j (x_i^T x_j) \tag{17}$$

Since the maximization depends on x_i, x_j only, so for unknown sample *u*:

$$\sum \alpha_i y_i (\mathrm{x}_i^T u) + b \geq 1, \text{ for class } 1 \tag{18}$$

$$\sum \alpha_i y_i (\mathrm{x}_i^T u) + b \leq -1, \text{ for class } 2 \tag{19}$$

In case of non-linear separable data transformation to a higher dimension to find a hyperplane that could make the separation. This transformation is done using kernel function based on the inner product of data [18][19][20]. So,

$$L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i^T x_j) \tag{20}$$

And the decision function will be

$$\sum \alpha_i y_i \mathrm{k}(x_i^T u) + b \geq 1, \text{ for class } 1 \tag{21}$$

$$\sum \alpha_i y_i \mathrm{k}(x_i^T u) + b \leq -1, \text{ for class } 2 \tag{22}$$

In case of multiclass classification one-versus-rest OVR technique is used. By considering each class *i* with the others as class *j* OVR is applied. However, a one-versus-one OVO technique by considering each class pairs seems to be capable of reducing the unclassifiable region than OVR [22][23], not even fuzzy SVMs [24].

F. Artificial Neural Networks

Since neural networks as an aspect in biology with axon and dendritic tree transformed with mathematics into a model of neurons, neural networks consider a powerful technique in classification.

For a neural network with multi inputs x_s, multi outputs y_s and (1) hidden layers: weights w_s and biases b_s shall initialized with according to the number of inputs with z_i and a_i as the input and output of hidden layer, where a_i is the activation function applied to z_i. Activation function could be tanh or sigmoid or ReLUs. The output y_i is calculated through a softmax function which used to convert values to action probabilities.

$$z_1 = x.\mathrm{w}_1 + \mathrm{b}_1 \tag{23}$$

$$a_1 = f(z_1) \tag{24}$$

$$z_2 = a_1.\mathrm{w}_2 + \mathrm{b}_2 \tag{25}$$

$$Y = a_2 = \mathrm{softmax}(z_2) \tag{26}$$

Softmax function:

$$S(z_i) = \frac{e^{z_i}}{\sum_{j=0}^{k} e^{z_k}} \tag{27}$$

The procedure above called forward propagation. Then the error is calculated between actual and desired outputs. Cross entropy method is commonly used with softmax function for error calculating. The cross entropy loss is calculated with N number of examples and C number of classes:

$$S(z_i) = \frac{e^{z_i}}{\sum_{j=0}^{k} e^{z_k}} \tag{28}$$

Gradient decent is applied to minimize error with stochastic or batch gradient decent [25][26] gradients will be calculated through back propagation algorithm using chain rule.

$$\delta_3 = Y - y \tag{29}$$

$$\delta_2 = f'(a_1) * w_2^T \tag{30}$$

$$\frac{\partial L}{\partial w_2} = a_1^T.\delta_3 \tag{31}$$

$$\frac{\partial L}{\partial b_2} = \delta_3 \tag{32}$$

$$\frac{\partial L}{\partial w_1} = x^T.\delta_2 \tag{33}$$

$$\frac{\partial L}{\partial b_1} = \delta_2 \tag{34}$$

Then the gradients is updated fist by adding regularization term λ which is often a small number to weights, this helps to avoid over fitting by making weights small. Then weights and biases are updated with a specific learning rate ε.

$$\frac{\partial L}{\partial w_2} = \left(\frac{\partial L}{\partial w_2}\right)_{old} + \lambda * w_2 \quad (35)$$

$$\frac{\partial L}{\partial w_1} = \left(\frac{\partial L}{\partial w_1}\right)_{old} + \lambda * w_1 \quad (36)$$

So the new parameters are

$$w_1 = (\mathrm{w}_1)_{old} - \varepsilon * \frac{\partial L}{\partial w_1} \quad (37)$$

$$b_1 = (\mathrm{b}_1)_{old} - \varepsilon * \frac{\partial L}{\partial b_1} \quad (38)$$

$$w_2 = (\mathrm{w}_2)_{old} - \varepsilon * \frac{\partial L}{\partial w_2} \quad (39)$$

$$b_2 = (\mathrm{b}_2)_{old} - \varepsilon * \frac{\partial L}{\partial b_2} \quad (40)$$

IV. SPACE SYSTEM CONFIGURATION

The space system as in Figure 5 consists of space segment and ground segment. The space segment may have one or more spacecraft.

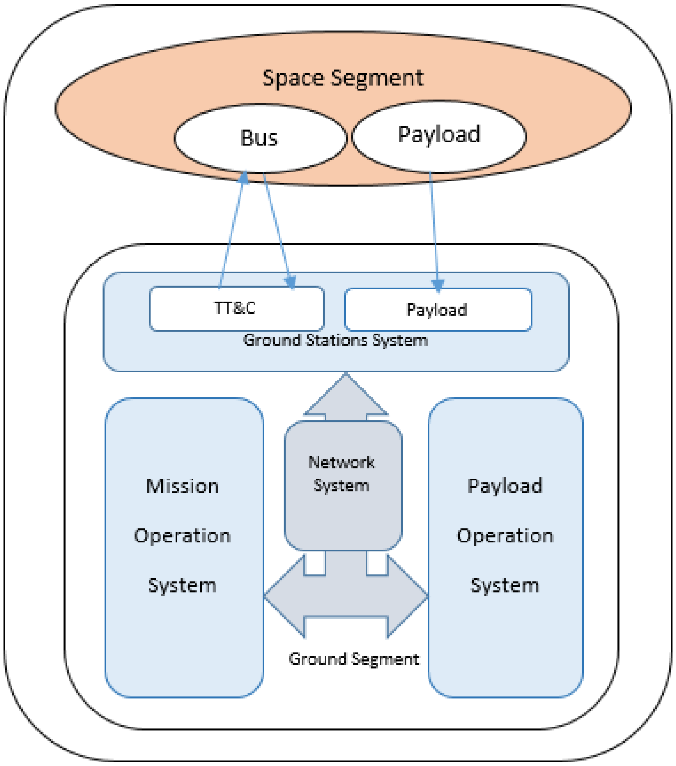

Figure 5 Space system

A. Ground Segment

The ground segment consists of the following:

Mission operation system, payload operation system, ground stations, and ground network system.

1) The mission operations system (MOS) has the following components:
   a) The mission control system (MCS), which includes:
      - ➢ The spacecraft database which includes the definitions of telemetry and telecommand with respect to the raw data.
      - ➢ The telemetry system.
      - ➢ The commanding system prepare and verify parameters necessary to uplink.
      - ➢ The sessions archiving database.
      - ➢ Mission planning system (MPS).
      - ➢ The on-board software maintenance.
      - ➢ Spacecraft online monitoring.
      - ➢ Spacecraft performance and evaluation.
      - ➢ Spacecraft shift flight director (SFD).
   b) The simulation.
   c) The flight dynamics system.
2) The payload operations system.
3) The ground station system.
4) The ground network system.

B. Space Segment

Space segment have only one satellite in operation. The satellite consists of bus and payload. The spacecraft bus function is to maintain the operation of the payload without anomalies.

Spacecraft bus consists of on-board computers, Guidance control and navigation system, thermal control system, propulsion system, power supply system and communication system.

Spacecraft thermal control system (TCS) purpose is to maintain equipment temperature in specified ranges. This will guarantee optimum performances when operating, avoid damage when non-operating.

The power supply system (PSS) is designed to provide constant current power to other bus-systems and payload. PSS consists of solar panels, power control and regulation unit, NHBs each battery consists of 18 cells, and battery commutation unit used as an interface for ground testing.

NHBs purpose is to provide power during eclipse to other systems. During lit-part of the orbit if the power of solar panels are not sufficient to perform functions. Each battery has its thermal control system which consists of heat pipes, radiators, temperature sensors, heaters and thermal cooling modules.

The concern will be for 2-NHBs. Each with the telemetry of 4-temperature sensors, 4-pressure sensors, 3-battery voltage sensors, 18-cells voltage sensors, and battery capacity which is calculated on-board.

Satellite state when receiving telemetry: The observations are taken before, during and after the two batteries tests. The purpose of the test is to check the performance of the battery and equalize the charge level of 18-cells of each battery. The test procedure as follows:

1) Charging algorithm starts to charge the battery by current of (15 ± 1) A.
2) Charging cuts-off if the pressure and temperature is more than the allowable values.
3) Test algorithm starts by discharging by fixed current of about (40 ± 1,5) A until operation of minimum voltage sensor triggered (down to 17 V) or until exceeding the temperature limit.
4) After automatic completion of discharge test algorithm is off, and there is a break for NHBM temperature decreasing.
5) After the break is finished a command for discharge resistance activation is generated onboard automatically. Complete discharging of battery is implemented by low current.
6) After completion of NHBM discharging on DR, charging algorithm is activated, forced charging of battery by current of (15,0±1,0) A is activated up to the pressure corresponding to the upper charge level of battery.
7) Charging is finished on achieving battery upper charge level, automatic onboard generation of a command for turning on compensation charging (compensation mode) and battery charging by low current of about (2 ± 0,2) A.
8) Compensation charging is turned off if the temperature exceeded the limits.
9) The test algorithm is activated then battery is forcibly discharged at fixed discharge current of about 40 A until operation of minimum voltage sensor triggered (down to 17 V) or until exceeding the temperature limit.
10) After automatic completion of discharging the battery testing is finished. The capacity is calculated and the battery returns to its normal state of operation.

A sudden temperature rise happened in battery-1 during the test leads to voltage drop in cell 13. The analysis of this off-nominal situation will be analyzed in the upcoming sections.

V. FIT THE OBSERVATION LEVEL MODEL

The purpose is building a model based on good batches. This model will be used to check the upcoming batches. But before start in building the model, the whole data is fitted to a PCA model. Data is used as X matrix. The X matrix is scaled and centered to unit variance. By using cross validation, the first two components are used to fit the model by R2X (cum): 0.714, Q2 (cum): 0.692. The model shows a good percentage of explanation of the X-variables R2Xcum, together with a good percentage in the prediction ability as shown by Q2cum. Shown below in Figure 4 the line plot of the first component (t1) of all observations.

The Figure 6 shows four areas of interest:

1) First 5000 observations which have observation number 348 as largest outlier below -3σ.
2) Observations from 5000 to 20000 as the normal operation.
3) Observations from 20000 to 30000 show a lot of disturbances.
4) Observations after the off nominal region from 31000 to the end.

The contribution plot of observation 348 shown below in Figure 7 shows drop in voltages of cells of the two batteries with low capacity and in temperature but not as the voltages.

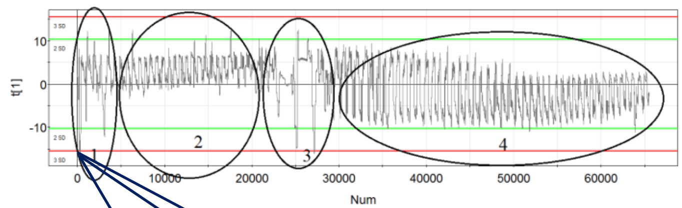

Figure 6 Line plot of t1

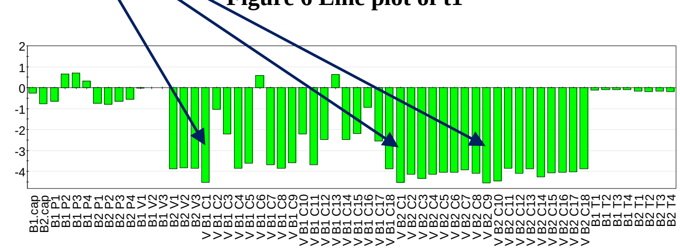

Figure 7 Contribution plot of observation 348

B1C1, B2C2, and B2C9 show less voltage of 1.2V, but battery 1 shows less drop than battery 2 on the other cells which shown also by less drop in voltage sensors of whole battery1 than battery2.

The drop in voltage is still within the specifications limits and especially in this case where a lot of payload activities during eclipse.

The next step is to build a PLS model of 10 good batches. The batches are taken from the normal region from observation 5000 to 14998. The batches represent the X matrix while the generated relative batch time represented as Y matrix.

The batch control chart of normal batches as a model in Figure 8 shows how t1 (left) and t2 (right) varies with time. A good new batch should evolve in the same way and its trace should be inside the control limits.

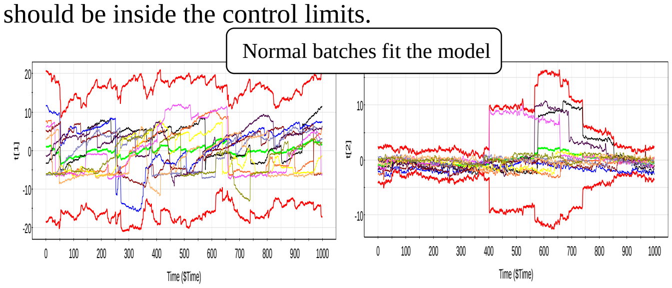

Figure 8 (1:10) Batches control chart of t1 (left) and t2 (right)

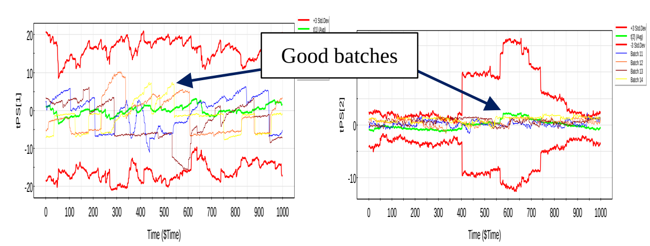

Figure 9 (11:14) Batches control chart of t1 (left) and t2 (right)

All reference batches are within the control limits. The constructed control charts are then used to monitor the upcoming batches. Below in Figure 9 the score control chart of t1 (left) and t2 (right) showing the batches 11,12,13,14 and they are all within limits.

The constructed control charts are then used to monitor the remaining batches from 15 and 16 are shown below in Figure 10 the score control chart of t1 (left) and t2 (right) showing that they are out of limits in many regions.

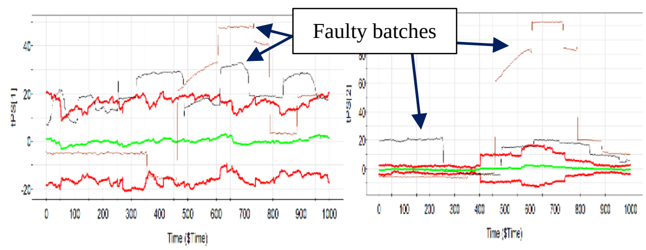

Figure 10 (15,16) Batches control chart of t1 (left) and t2 (right)

Batch 15 as shown in Figure 11 has extreme going out of limits. The batch develops correctly and shows a deviation starting from 460-time point in the score plot of t1 and t2 thus alarming the existing of a process shift as shown below in Figure 12.

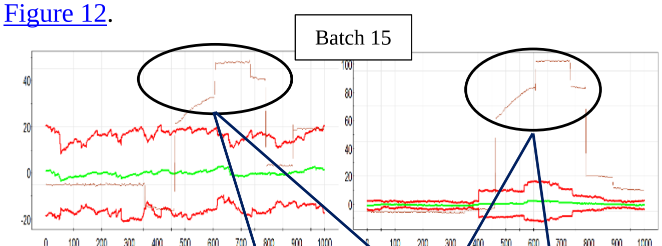

Figure 11 Predicted score batch plot of t1 (left) and t2 (right) for batch 15.

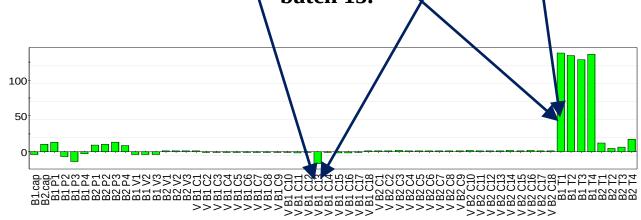

Figure 12 Contribution plot of batch 15 according to scores.

The contribution plot in Figure 12 shows high temperature of battery 1 at the four sensors of (59.7, 58.7, 58.5, 58.7) C which leaded to voltage drop of cell 13 of battery 2 to zero voltage, Which will be continued with batch 16 as shown below in Figure 13 for batch 16 in score and DModX control charts.

The contribution plot of batch 16 in Figure 14 shows the drop in cell 13 of battery 1 which drops to zero voltage. The battery will be applied to change some algorithms like charge phases and minimum voltage stopping trigger to overcome the consequences of missing this cell.

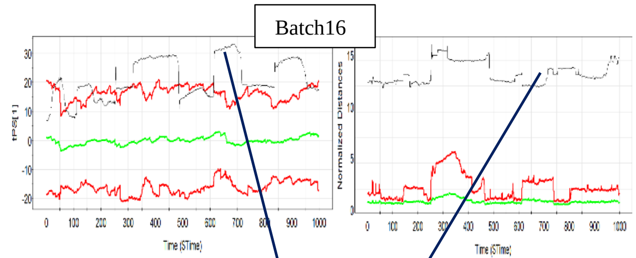

Figure 13 Predicted score batch plot of t1 (left) and DModX (right) for batch 16

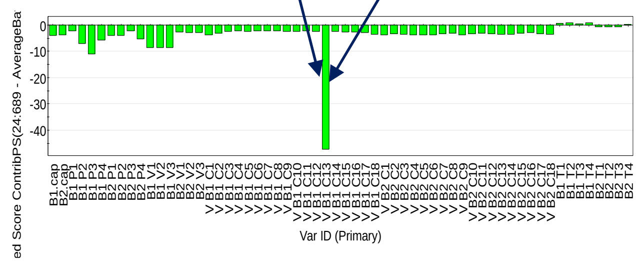

Figure 14 Contribution plot of batch 16 according to scores

VI. Whole Batch Model

The batches will be fitted to PCA model with two components of cross validation R2X (cum):0.549 and Q2 (cum):0.164. The scores scatter plot of t1 against t2 shown below in Figure 15 of first 10 batches span the space with no outliers.

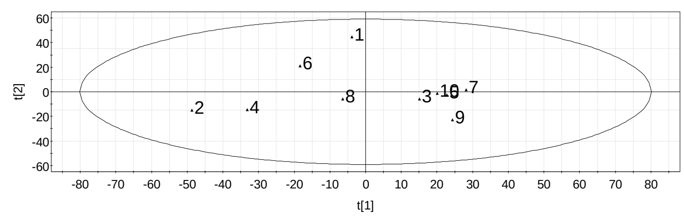

Figure 15 Scores of t1 vs. t2 of 10 good batches

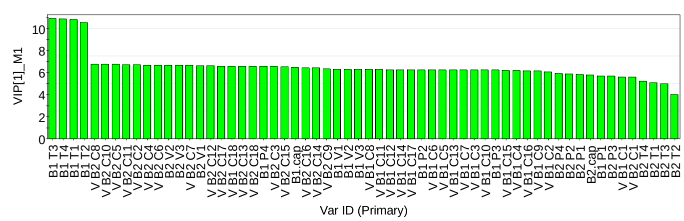

Figure 16 Batch variable importance

The batch variable importance related to t1 shown below in Figure 16. This plot, by combining the importance of the scores in the batch level model, with the weights w* derived from observations level model, displays the overall importance of the measured variables in the whole batch model.

As in figure all variables are important and are highly correlated. Also shown that the four temperature sensors of battery 1 with maximum importance.

The batch level model will be used to monitor the upcoming batches as shown below in Figure 17. It is clearly shown in Figure 17 that batches 11:14 are inside the Hotelling T^2 ellipse, on the other hand batches 15 and 16 shown as strong outliers.

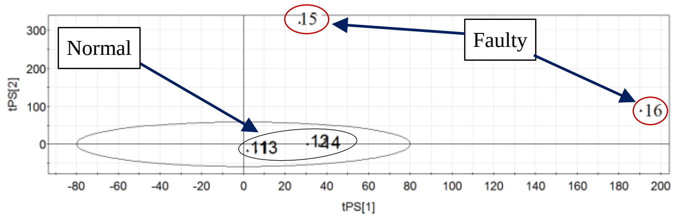

Figure 17 Score plot of next 11:16 batches fitted to the model

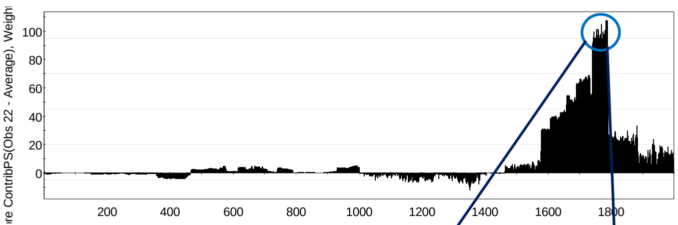

Figure 18 Score contribution plot of batch 15

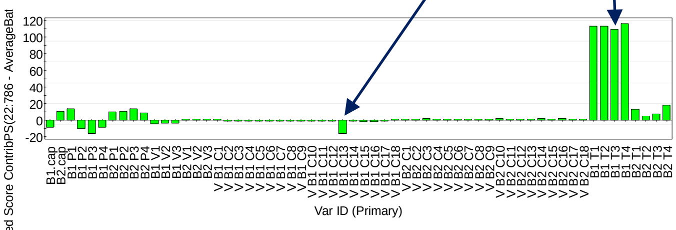

Figure 19 Resolved score contribution plot of batch 15

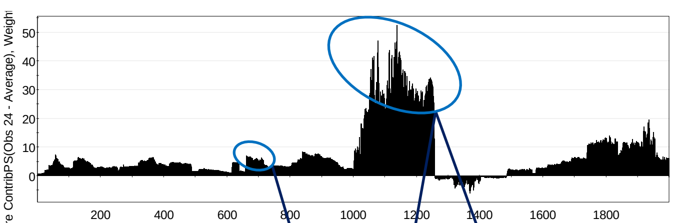

Figure 20 Score contribution plot of batch 16

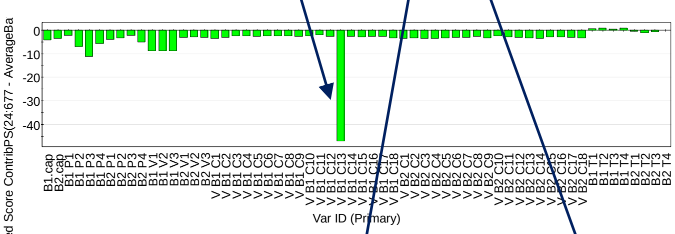

Figure 21 Resolved score contribution plot of batch 16

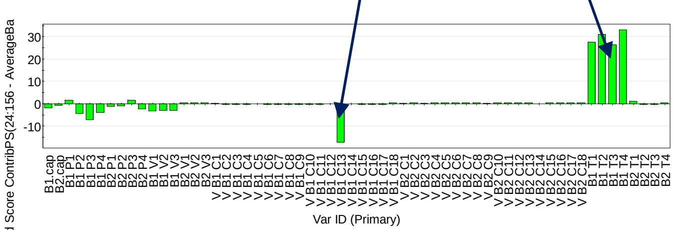

Figure 22 Resolved score contribution plot of batch 16

The contribution plot of batch 15 shown below in Figure 18 and is resolved with respect to original variables in Figure 19.

Also the contribution plot shows the high temperature of battery 1 which leads to cell 13 voltage drop. The contribution plot of batch 16 shown below in Figure 20 and is resolved with respect to original variables in Figure 21 and Figure 22.

Cell 13 voltage drop still exist with some temperature rise but not as extreme (20 C maximum) as shown above in Figure 21 and Figure 22.

VII. Fit the Data to Control Charts

Monitoring charts are a graphical tool, aiming to rapidly detect a problem by visual analysis. A Shewhart chart monitors that a process variable remains on target and within given upper and lower limits. It is a monitoring chart for location. It answers the question whether the variable's location is stable over time. The scores and Hotelling's T^2 can be plotted in multivariate Shewhart charts. Shewhart chart of scores shows how the process evolves over time in the respective PCA – or PLS model dimension.

$$T^2 = \sum_{a=1}^{A} \frac{t_{ia}^2}{s_{ta}^2} \tag{41}$$

content. s_{ta}^2 is variance of t_a according to the class model

$$s_{ta}^2 = T_i^2 * \frac{N(N-A)}{A(N^2-1)} \tag{42}$$

F is distributed with A and N-A is degree of freedom

N is number of observations in the model

A is number of components in the model

Hence if

$$T_i^2 > \frac{A(N^2-1)}{N(N-A) * \left(F_{\text{critical}(P=0.05)}\right)} \tag{43}$$

then observation is outside 95% confidence region.

The main advantage of the T^2 chart is the ability to simplify calculations down to simple single-number criteria. It produces in-control run length and maintains the original data means, variances, and correlation.

Below in Figure 23 Shewhart chart of Hotelling's T^2 (above) against DModX (below) shows the drift of process between 20000 and 30000 observations which make some changes in the process behavior appears after that.

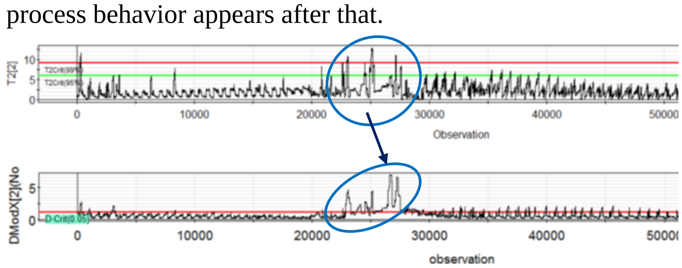

Figure 23 Shewhart chart (above) against DModX (Below)

The Shewhart plot of t1 summarizes the systematic variation in the data. Tiled underneath is a plot of DModX, which enables an overview of unsystematic variation. The Shewhart chart is not too sensitive to predict shifts in the mean.

Instead of plotting the individual observations Xs, their accumulative deviation from target value is plotted. In multivariate CuSum the model scores are plotted.

$$S_{\mathrm{h}(i)} = \max\left(0, S_{\mathrm{h}(i-1)} + X_i - \mathrm{Target} - 0.5 * \mathrm{STD}\right) \quad \textbf{(44)}$$

$$S_{\mathrm{l}(i)} = \min\left(0, S_{\mathrm{h}(i-1)} + X_i - \mathrm{Target} - 0.5 * \mathrm{STD}\right) \quad \textbf{(45)}$$

where S_{h} is the high cumulative sum, S_{l} is the low cumulative sum, and STD is standard deviation.

The cumulative sum chart, or CuSum chart, allows more rapid detection of these shifts away from a target value because the CuSum chart is extremely sensitive to small changes due to accumulation action of any shift in both directions high or low shifts. The two CuSum charts of first component t1 and Hotelling's T^2 in Figure 24 and Figure 25 give early warning of problems especially t1.

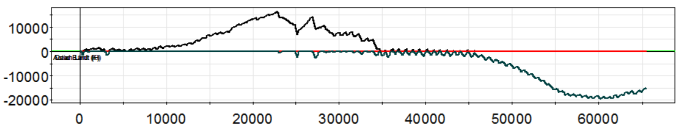

Figure 24 CuSum of t1

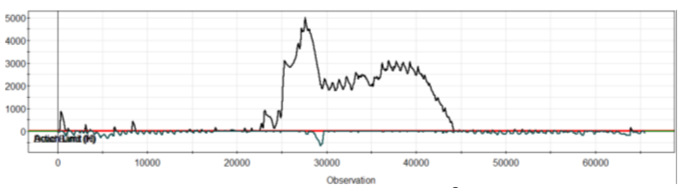

Figure 25 CuSum of T^2

The CuSum chart can easily predict the high shift in process from target, as shown in Figures 24 from observation 9000 and quit earlier the process started to go strongly away from target. This information was not easily retrievable from the previous Shewhart charts. As the shift starts and increases it indicates the starting of an off-nominal situation.

As a recommendation, all testing activities of PSS should be stopped at time of continuous shifting starts; cooling procedures should be achieved like activating the thermal cooling module, which should enhance the operation of contour heat pipes. The contour heat pipes can remove all heat increase through the radiation panels; this should save the battery cells from the extensive temperature increase and maintain all cells in operation.

VIII. Building Support Vector Machines Model

Support vector machines as a classifier is used to build a model with three regions: normal, deviation and faulty. The multiclass support vector classifier with the technique of one vs one is applied to the three classes of telemetry. The model shall be used in further analysis and system behavior checking.

Four one-versus-one SVMs models were built and coded in Python with 250 observations as training set with three classes 100 normal, 50 deviating and 100 faulty. And another set of 249 observations as training set with three classes 100 normal, 49 deviating and 100 faulty for testing.

Dimensionality reduction is applied to data for visualization and features extraction. Visualization is very important not only for enhancement decision making, but also for facilitating and speeding up analysis process. The importance of visualization. The dimensionality reduction is applied through supervised technique PLS-DA and unsupervised technique PCA to the whole data set. The training set is shown in Figure 26, left for PCA and right for PLS-DA.

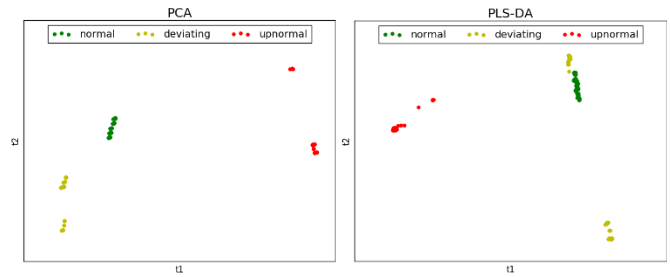

Figure 26 Scores scatter plot or PCA (left) and PLS-DA (right)

Figure 26 shows that PCA make better representation of the three classes of data especially deviating class, as PLS-DA divide the up-normal class into two separate regions around normal class.

The previous step should be checked with every new set to ensure the technique with better representation of data as the new set shall influence the latent model structure.

PCA is applied to the data then, the first two scores used with SVMs. Four models as in Figure 27 built with SVM regularization C = 1, and with linear SVC, linear kernel, rbf kernel with γ = 0.7, and polynomial kernel with degree 3.

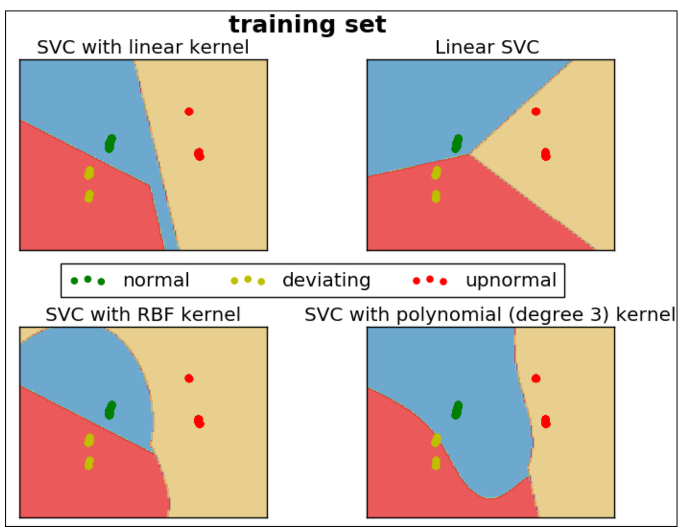

Figure 27 Four SVMs models with the training set

Figure 27 above shows SVMs models classify the data with three regions with its decision boundaries. The testing set is fitted to the regions within the decision boundaries as in Figure 28.

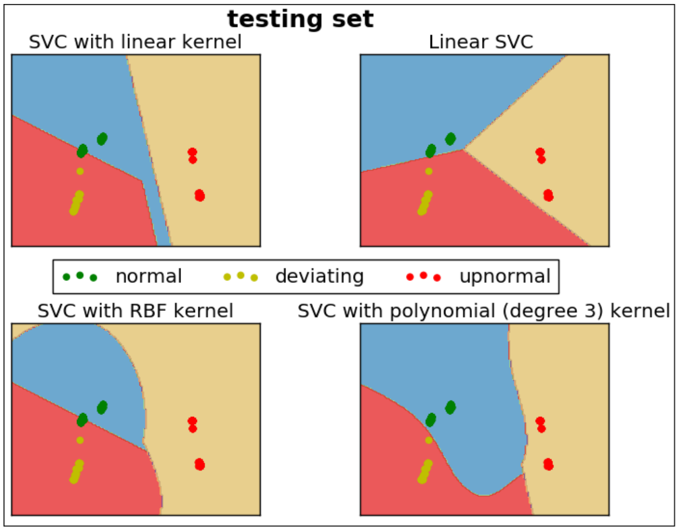

Figure 28 Testing set is fitted to the models

The SVMs models will be used for further offline analysis. The upcoming telemetry observations will be combined and fitted to the model to check the normality of the data as the observations should lie in the normal region. The direction of observations towards the deviation or faulty regions shall give an indication of system health and could be used for anomalies predictions with the system specific operation.

IX. Building Artificial Neural Network Model

Another model is built with data with artificial neural network with 3-layer network. The input layer of 2-nodes defining the data after dimensionality reduction taking the first two scores of PCA as in previous section. The hidden layer of 5-nodes. The output layer of 3-nodes defining the number of classes. A forward propagation with tanh as an activation function is done then the activation function for the output layer is SoftMax to convert values to action probabilities. The cross-entropy loss is used with SoftMax to calculate the error. After that batch gradient decent is performed with chain rule through the back propagation with optimization of parameters using Quasi-Newton (BFGS) [27] for minimizing the error. Learning is performed with regularization term of 0.01 and 0.01 learning rate.

The output for classification is calculated through forward propagation and argmax function which returns the indices of probabilities which is the classes in this case as in Figure 29 shows the training set making the model (left) and testing set fitted within decision boundaries (right).

Figure 29 shows the classification regions using the network building a model for further classification of upcoming data as SVMs models in the previous section. Neural networks work as a black box; however, the nonlinearity of the network is very useful in classification in multiclass cases.

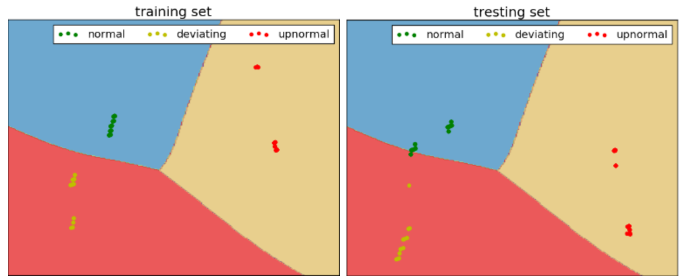

Figure 29 Training set model (left), and testing set (right)

Since the features extraction is done through PCA with ANN and SVMs in previous section, the diagnosis could be done through contribution plots as in section V and section VI.

X. Conclusion

Data mining with the data-driven approach could be used in anomalies identification and diagnosis. This is performed through building telemetry models of real system behavior.

A novel application of batch-process model and whole batch model was established between the sensory data and the virtual time of the batch process. The two models were used to monitor new batches and successfully perform anomaly detection. The anomaly diagnosis was done through contribution plots.

Shewart and DModX of scores and Hotelling's charts could be used in monitoring the upcoming observations in addition to detecting anomalies, but not in early detection.

The two CuSum charts give early warning of problems especially the first component. The CuSum chart is designed to detect shifts in the mean from the target value. So, with the help of CuSum charts, the temperature rise during battery-1 test could be avoided if this technique was used. This could save the battery from missing any cell.

Another two models were built in SVMs an ANN. These models could be used in further system analysis and health evaluation to ensure system behavior shall lie in the normal region, in addition to fault detection in case of off-nominal situations.

Telemetry models were built to the real system states during real operation, so it's powerful not only in anomaly detection, but also in identifying system behavior deviation from targeted behavior, and furthermore for prediction. Dealing with all data at once reduce time and enhance the analysis process through extracting information from the correlations between data.

REFERENCES

[1] Tianshe Yang, Bin Chen, Yu Gao, Junhua Feng, Hailong Zhang, Xiaole Wang “Data Mining-Based Fault Detection and Prediction Methods for In-Orbit Satellite” Measurement, Information and Control (ICMIC), 2013 International Conference.

[2] Lin Su; Chaoxuan Shang ; Yunhong Su ; Yihua Zhai “Fault detection and isolation based on multivariate statistical analyzing for the satellite attitude control system”Electronic Measurement & Instruments, 2009. ICEMI '09. 9th International Conference.

[3] Bo lee, Xinsheng Wang “Fault Detection and Reconstruction for Micro-satellite Power Subsystem Based on PCA”Systems and Control in Aeronautics and Astronautics (ISSCAA), 2010 3rd International Symposium.

[4] Marco Vieira,Denise Rotondi Azevedo, Ana Maria Ambrósio. “Applying Data Mining for Detecting Anomalies in Satellites”.2012 Ninth European Dependable Computing Conference.

[5] Yu Gao, Tianshe Yang, Nan Xing, Minqiang Xu “Fault Detection and Diagnosis for Spacecraft using Principal Component Analysis and Support Vector Machines”. 7th IEEE Conference on Industrial Electronics and Applications.

[6] Keqiang Xia, Baohua Wang, and Ganhua Li “An Integrated Fault Pattern Recognition Method of Satellite ControlSystem Using Kernel Principal Component Analysis and SupportVector Machine”.2014 IEEE Chinese Guidance, Navigation and Control Conference.

[7] Hongzheng Fang, Hui Shi1, Yi Xiong, Rui Li, Ping Wang “The Component-level and System-level SatellitePower System Health State Evaluation Method”2014 Prognostics and System Health Management Conference.

[8] Wessam M. Hussein “Machining process monitoring using multivariate latent variable methods”. Hamilton, Canada 2007.

[9] Kevin Dunn, “Process Improvement using data”. McMaster University. 18 August 2015.

[10] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, and S. Wold. “Multi- and Megavariate Data Analysis”. Part I: Basic Principals and Applications.

[11] J.E.Jackson, "A User's Guide to Principal Components," Wiley, New York, 1991.

[12] Svante Wold research group for chemometrics Umea university Sweden “Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models”Technometrics Vol. 20, No. 4, Part 1 (Nov., 1978), pp. 397-405.

[13] L. Sirnar and W.Hardle, "Applied Multivariate Statistical Analysis," Tech method and data technologies, Springer Verlag, Berlin and Louvainla-Neuve, 2003.

[14] "onymolls. SIMCA·P manunl. Umctrics A B.

[15] Willey J. Larson and Daryl G. Boden, “Cost-Effective Space Mission Operations”

[16] Willey J. Larson and James Wertz “Space Mission Analysis and Design”

[17] “Satellite Mission operations best Practices”. Assembled by Space Operations and Support Technical Committee and American Institute of Aeronautics and Astronautics.

[18] J. MacGeorge and P. Nomikos, “Multivariate SPC Monitoring Batch Process”, Technimetrics, Vol.37 No.1 (1995)41-57.

[19] Noordwijk. The Netherlands, European Cooperation for Space Standardization (ECSS) Secretariat, ESA-ESTEC, Requirements & Standards Division. ECSS-E-ST-70C. “Space Engineering. Ground Systems and Operations”.

[20] The Consultative Committee for Space Data Systems (CCSDS). Informational report concerning space data systems standards. CCSDS 520.0-G-3. “Mission Operations Services Concept”. Green book, December 2010.

[21] Xindong Wu, Vipin Kumar “The Top Ten Algorithms in Data Mining” CRC press.

[22] Lutz Hamel, University of Rhode Island, “Knowledge Discovery with Support Vector Machines”, Wiley.

[23] Shigeo Abe, “Support Vector Machines for Pattern Classification”, second edition, Springer.

[24] Christian Moewes, Rudolf Kruse, “On the Usefulness of Fuzzy SVMs and Extraction of Rules from SVMs”, 7th conference of the European Society for Fuzzy Logic and Technology.

[25] Ian T. Nabney, “NETLAB: Algorithms for Pattern Recognition” Springer.

[26] Sebastian Raschka “Python Machine Learning”, PACKT.

[27] LeCun, Yann A., Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. "Efficient backprop." In Neural networks: Tricks of the trade, pp. 9-48. Springer Berlin Heidelberg, 2012.