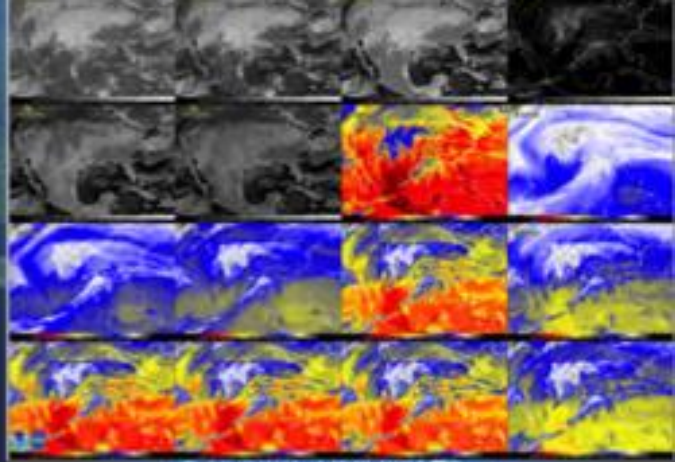


A Hybrid Machine Learning Approach to Anomaly Detections in Satellite Telemetry Data

Zhenping Li, ASRC Federal

Agenda

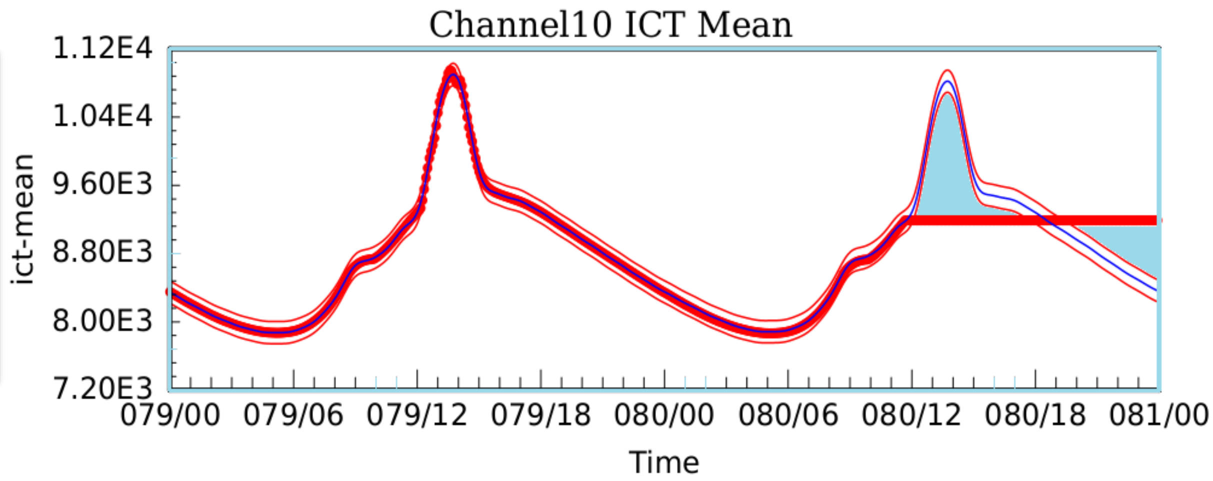
- Issues to be addressed in satellite telemetry data
- The hybrid approach
- Data training for LEO telemetry data
  - Challenges and requirements
  - Data training outputs for different data patterns
- How to characterize correlations in satellite telemetry
  - How to measure data pattern changes.
  - Hierarchical event vectors
- Hierarchical clustering for event classification and anomaly detections

The Machine Learning Approach to Space Missions

- An architecture model that defines the machine learning processes and interfaces.
  - Two key processes: data training and post training analysis
- A scalable and extensible enterprise architecture for a ML system that provides common infrastructure and services for ML algorithms that are treated as mission specific software components.
  - Implemented in Advanced Intelligent Monitoring System (AIMS)
- A ML algorithm library that provides efficient and accurate data training outputs for telemetry data and instrument calibration data.
  - Include ML models for telemetry data and clustering algorithms used for data quality evaluation and anomaly detection.
  - Software components with standard interfaces can be plugged into AIMS
- Application Portfolio
  - Satellite Instrument radiometric performance monitoring
  - Satellite health and safety telemetry monitoring
    - currently applied to LEO and GEO satellite

Challenges in Satellite Telemetry Datasets

Anomalies cause **unexpected** data pattern changes in datasets.



- Separate normal operations from anomalies
  - Both normal operations, such as orbit maneuver, and anomalies lead to data pattern changes in telemetry data
  - Anomaly detections without separations expected from unexpected data pattern changes lead to false positives
- Correlations among datasets must be considered in analyzing telemetry data
  - Interactions among subsystems in a satellite lead to strong correlations in telemetry datasets.
  - Data pattern changes due to normal satellite operations or anomalies generally involves multiple telemetry datasets in multiple subsystems.

The Hybrid Approach to Satellite Health and Safety Telemetry

- Detect data pattern changes in telemetry datasets through a supervised learning
- Separate anomalies from normal operations with an unsupervised learning
  - Define a quantitative metric to measure data pattern changes in a data sets
  - Develop a mathematical representation to characterize correlations among datasets
  - Develop a clustering algorithm to capture signatures of normal operations and detect anomalies

The event clustering not only detects anomalies but also capture signatures of normal operation events.

Supervised Learning

Data Training

$$\underset{W}{argmin} \sum_i \frac{1}{2}\left(d_j(t_i) - f_j(t_i, W)\right)^2$$

$\Downarrow$

Data Model

$$\{f_j(t), \sigma_j\}$$

$\Downarrow$

Data Pattern Change Detection

$$|f_j(t_i) - d_j(t_i)| < N\sigma_j$$

$\Downarrow$

Data pattern change metric

$$\{\psi_j\}$$

$\Rightarrow$

Unsupervised Learning

Hierarchical Event Vectors

$$\boldsymbol{e}(\psi_S) = \left\{\frac{\psi_{S_1}}{\psi}, \frac{\psi_{S_2}}{\psi}, \dots \frac{\psi_{S_n}}{\psi}\right\}$$

$\Downarrow$

Hierarchical Clustering

$$\boldsymbol{e}^{\boldsymbol{i}}(\psi_S) \cdot \boldsymbol{e}^{\boldsymbol{j}}(\psi_S) = \frac{1}{\psi^i \psi^j} \sum_S \psi_S^i \psi_S^j \geq \alpha^{th}$$

$\Downarrow$

Event Cluster & Anomalies

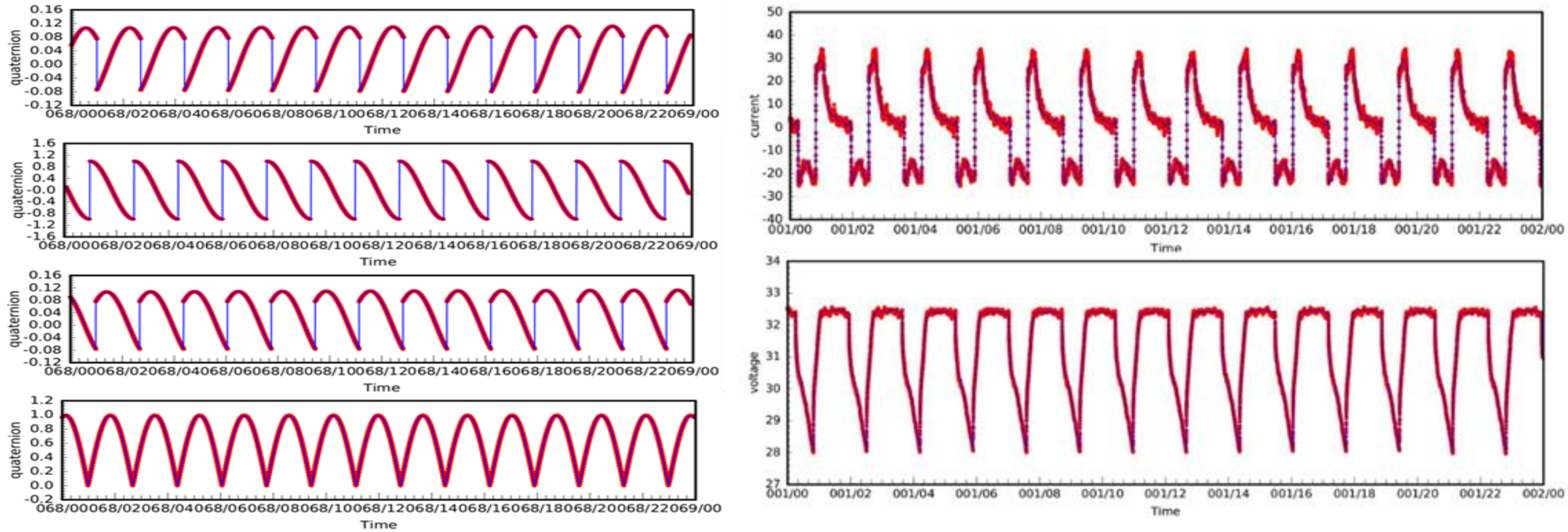
$$\{\boldsymbol{e}^{\boldsymbol{i}}(\psi_S)\}^N + \{\boldsymbol{e}^{\boldsymbol{i}}(\psi_S)\}^A$$

Data Training for LEO Satellite Telemetry Data

ASRC FEDERAL

- Challenges
  - More diverse data types
  - High complexity in data patterns
  - Relationships exists among datasets
- Requirements for Data training in an operational environment
  - Efficient:
    - Data training for a dataset should be completed in seconds or minutes instead of hours or days.
  - Accurate:
    - Essential in detecting data pattern changes in datasets.
  - Robust:
    - The training data may contain outliers that distort data training outcome
- The flexibility in selecting different data models for different patterns and noise level is critical
- Data training are performed for the telemetry data from NOAA Suomi National Polar-orbiting Partnership (NPP) satellite telemetry data

Data Training Outputs for the data with orbital patterns



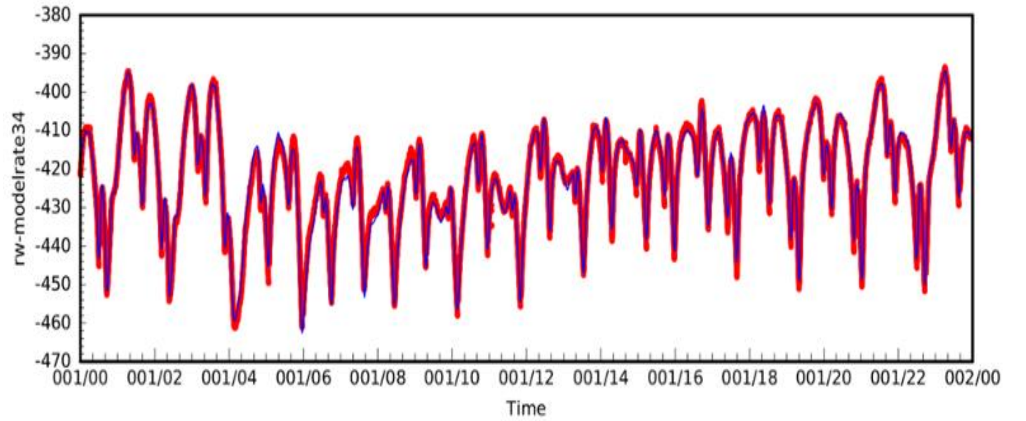
The ML model (blue lines) and data (red dots) for the current and voltage in the power system (left) the quaternions (bottom)

- The Data patterns follows the orbit characteristics.
  - Approximately periodic and follows the orbit period.
  - Include
    - the current and voltage in power systems
    - The temperature profile.
    - The quaternions
- The neural networks and Fourier expansion models are implemented

The outputs for the datasets with different patterns

The data patterns are periodical with longer period.

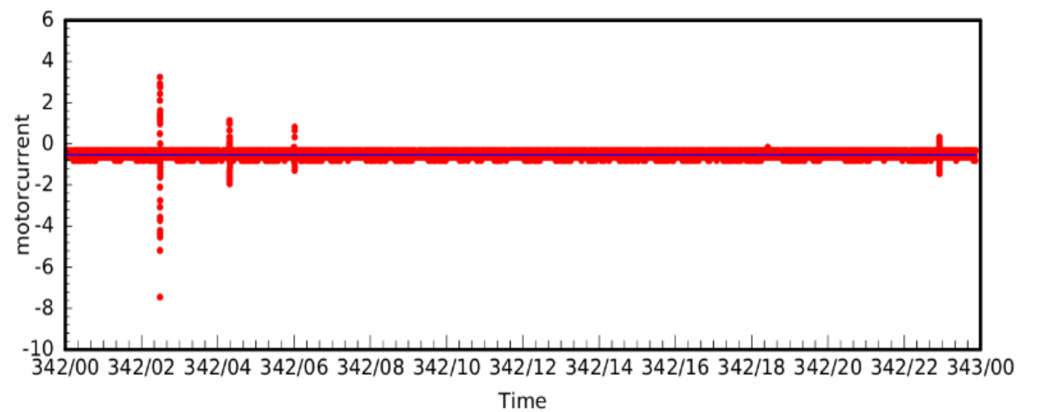
- Momentum or speed profiles in reaction wheels.
- Fourier Expansion models are used for data training.



Reaction Wheel Momentum profile

The datasets are constants but noisy

- Motor currents in reaction wheels is an example.
- Data Training is the same as collecting statistics.
- Vertical lines are treated as outliers



Reaction Wheel Motor Current

How to measure data pattern changes

- Data pattern change metric:

$$\psi_j = \frac{1}{T}\sum_i \frac{1}{2f_s}\left(\left|O_j^N(d(t_i))\right| + \left|O_j^N(d(t_{i-1}))\right|\right)\delta\left(t_i - t_{i-1} = {}^1/_{f_s}\right)$$

- O_j^N is the **normalized outlier:**

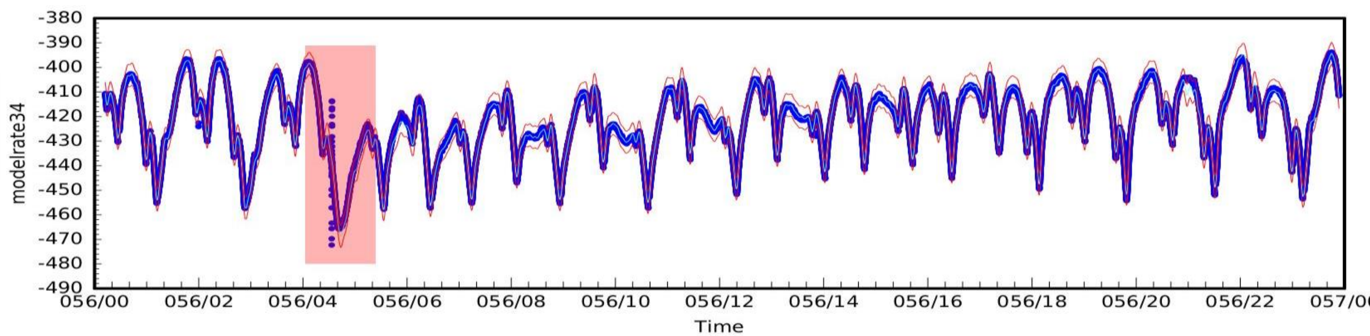
$$O^N(d(t_i)) = \frac{\delta(|\Delta(t_i)| > N\sigma)}{N\sigma}\{\Delta(t_i) - sign(\Delta(t_i))\cdot N\sigma\}$$

  - where

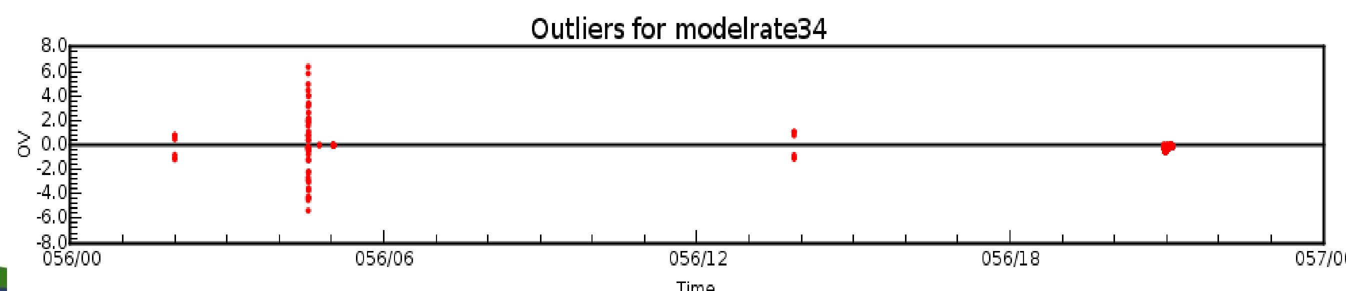
$$\Delta(t_i) = f(t_i,\{S_k\}) - d(t_i)$$

  - Normalized outlier is a dimensionless quantity so that different telemetry data points can compare with each other.

The data training output and actual data



The outlier plot based on normalized metric $O^N(d(t_i))$ only highlights time and strength of data pattern changes

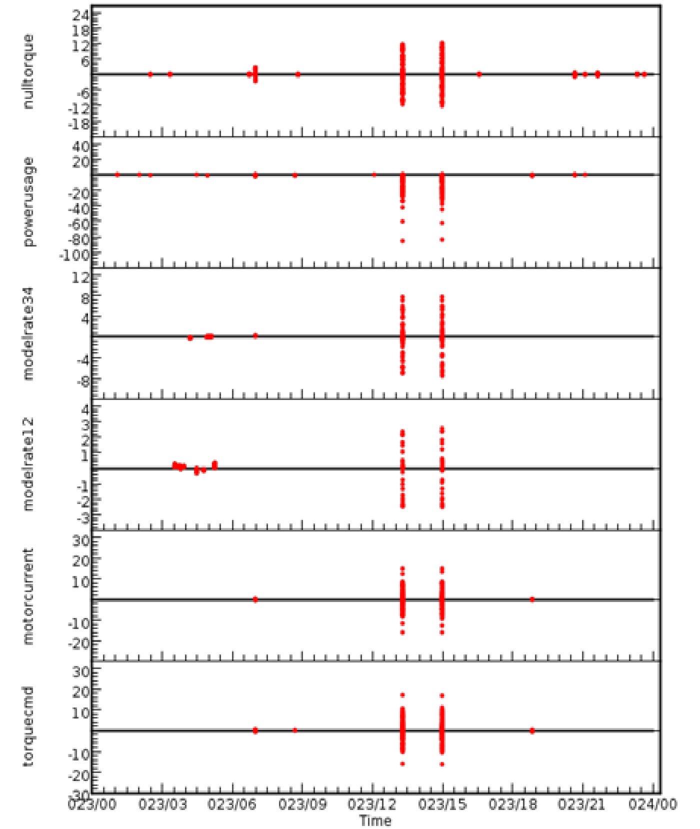


Event Vectors in Mathematical and Graphical Form

- Both normal operations and anomalies can be regarded as events
- Events are characterized by event vectors in an outlier space

$$\boldsymbol{e}(t_i, t_f) = \left\{\frac{\psi_1}{\psi}, \frac{\psi_2}{\psi}, \dots \frac{\psi_n}{\psi}\right\}$$

  - t_i, t_f: start and time of an event
  - ψ_i : the data pattern change metric value for telemetry dataset.
  - $\psi = \sqrt{\sum_i \psi_i^2}$
  - $\boldsymbol{e}(t_i, t_f)$ is an event profile that defines the composition and relative strength of data pattern change metrics.
  - Normal operations and anomalies have different event profile.
- The unit vector $\boldsymbol{e}(\psi_S)$ provides a mathematical representation in machine learning for event classifications and anomaly detections



Two events with outliers aligned at the exact same time for multiple datasets in the reaction-wheel subsystem. The outlier plot provides insights into the correlations among datasets in multiple subsystems.

Clustering for Event Classifications and Anomaly Detections

- Clustering Criteria: Two events belong to the same cluster if

$$\boldsymbol{e^i}(\psi_S) \cdot \boldsymbol{e^j}(\psi_S) = \frac{1}{\psi^i \psi^j} \sum_S \psi_S^i \psi_S^j \geq \alpha^{th}$$

  - The value for $\boldsymbol{e^i}(\psi_S) \cdot \boldsymbol{e^j}(\psi_S)$ has the range from 0 to 1.
  - The value of α^{th} is between 0.95 to 0.98.
- The events as part of normal operations are generally repeatable and happen regularly
  - Each type of events form their own clusters
- The events corresponding to anomalies are generally not repeatable and happens rarely with their own characteristics.
  - Part of noise in event clustering
  - Anomaly detections: determine if there are events belonging to noises.
- Clustering algorithm selections:
  - Can separate noise events (don't belong to any cluster) from normal clusters.
    - DBScan Clustering
    - Mean Shift Clustering

Event clustering are hierarchical

- A dataset can be defined by its hierarchical path

  Subsystem/mnemonic/index

- An event vector can be defined at subsystem level or mnemonic level by aggregation of event vectors at lower levels

$$\psi_S = \sqrt{\sum_M \psi_{S,M}^2}$$

and

$$\psi_{S,M} = \sqrt{\sum_I \psi_{S,M,I}^2}$$

- Event vectors at higher levels reduces the space dimensions while keeping sufficient information for event classification and anomaly detections
- Hierarchical clustering are performed from subsystem level to mnemonic level
  - Top-down clustering

$$\boldsymbol{e}(t_i, t_f) = \left\{ \begin{array}{l} \dfrac{\psi_{S_1}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_1,M_1}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_N,M_1,I_1}}{\psi} \\ \cdots \\ \dfrac{\psi_{S_N,M_1,I_l}}{\psi} \end{array} \right. \\ \cdots \\ \dfrac{\psi_{S_1,M_k}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_1,M_k,I_1}}{\psi} \\ \cdots \\ \dfrac{\psi_{S_1,M_k,I_l}}{\psi} \end{array} \right. \end{array} \right. \\ \cdots \\ \dfrac{\psi_{S_N}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_N,M_1}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_N,M_1,I_1}}{\psi} \\ \cdots \\ \dfrac{\psi_{S_N,M_1,I_l}}{\psi} \end{array} \right. \\ \cdots \\ \dfrac{\psi_{S_N,M_k}}{\psi} \left\{ \begin{array}{l} \dfrac{\psi_{S_N,M_k,I_1}}{\psi} \\ \cdots \\ \dfrac{\psi_{S_N,M_k,I_l}}{\psi} \end{array} \right. \end{array} \right. \end{array} \right.$$

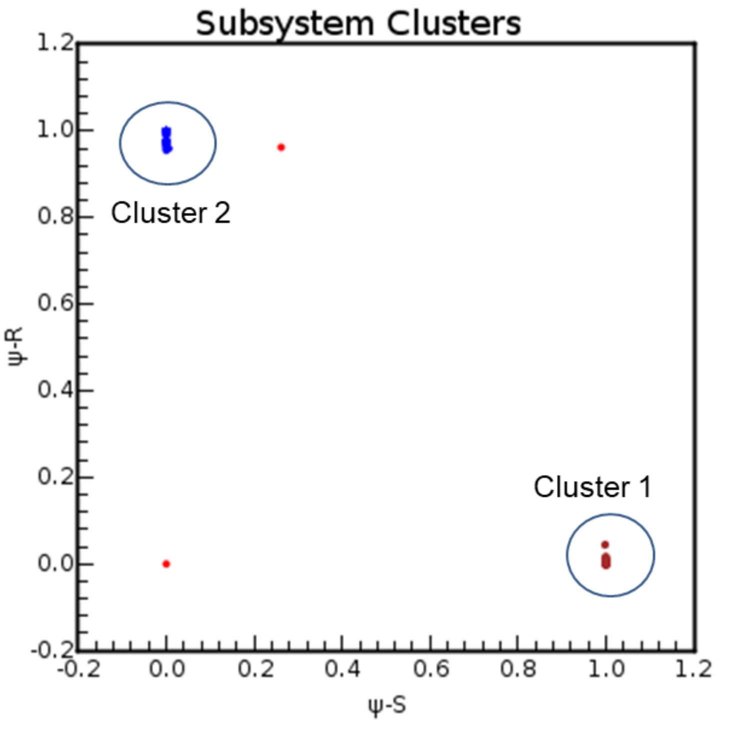
Clustering for Event Classifications and Anomaly Detections

Hierarchical Clustering Algorithm

Perform clustering for all events at subsystem level
Output noise events
For each cluster at subsystem level
Perform clustering for events mnemonic level
Output noise events

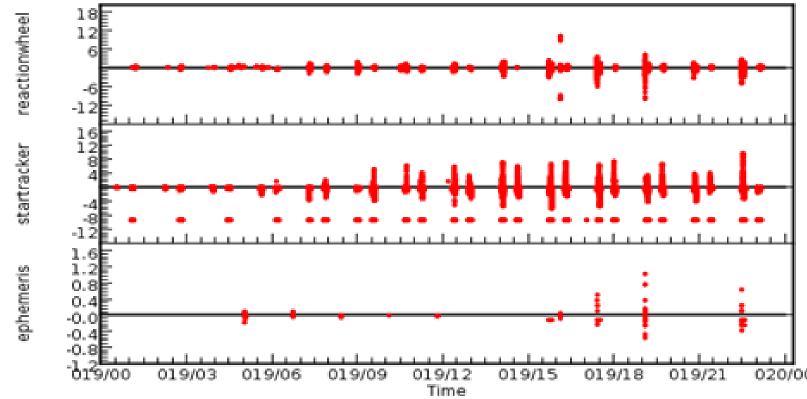
- Top-down hierarchical clustering
- DBScan clustering is implemented at each hierarchical level

- Event clustering are performed for all events accumulated in previous and current training sessions
  - The more events in clustering, the better clustering outcomes
- Outputs from NPP Telemetry Data:
  - The cluster dominated by data pattern change in star tracker subsystem.
  - The cluster dominated by reaction wheel subsystems.
  - Noise events (red dots) are potential anomalies

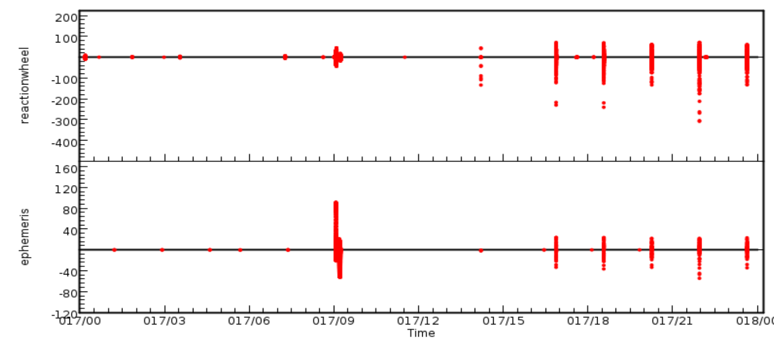


Event Clusters in the Outlier Plots at the subsystem level

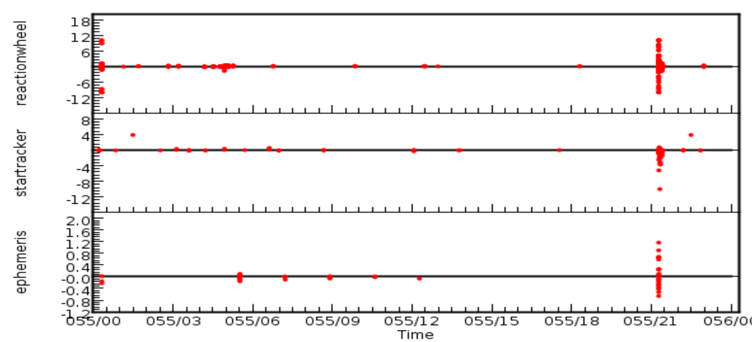
The cluster dominated by data pattern changes in the Star Tracker subsystems, which happens regularly. The reaction wheel and ephemeris subsystems are impacted.



The event at around 09 Z has the signature of orbit maneuver. The other events dominated by data pattern changes in the reaction wheel subsystem, which belong to another cluster.



The event at around 21 Z does not belong to any cluster, which is a potential anomaly.



Summary

- Detecting data pattern changes in telemetry datasets are critical first step for anomaly detections.
  - It is achieved with supervised learning of normal data patterns
- Correlations among telemetry datasets must be taken into account in anomaly detections to separate normal events from anomalies
  - Event vectors provides a mathematical representation to characterize correlations
  - Outlier plots provides a graphical view of correlations among different telemetry datasets
  - The pattern of correlations determines the nature of an event
- The separation of normal events from anomalies is achieved by hierarchical event clustering.
  - Also capture signatures of normal operation events
- The algorithms is implemented in Advanced Intelligent Monitoring System (AIMS)
  - A flexible and extensible ML platform for time series data