

Machine Learning Solution for Space Missions

Zhenping Li
ASRC Federal

@2019 ASRC Federal



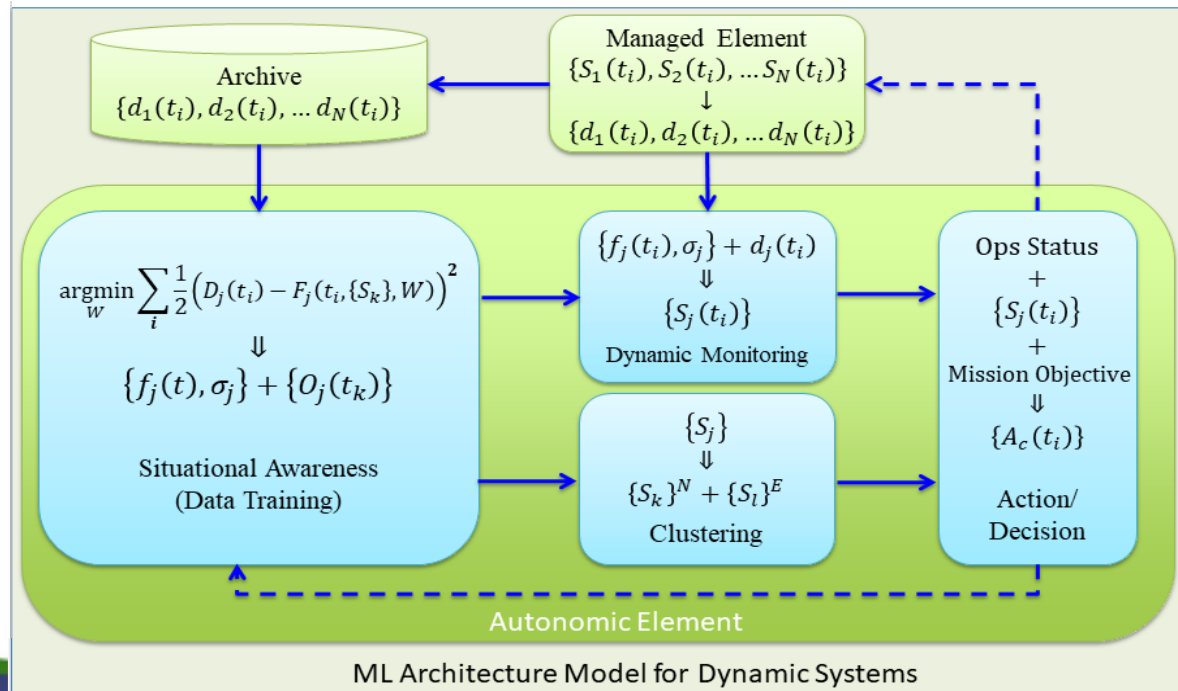
- Machine Learning for Space Missions Overview
- ML Architecture Model
- Data Training for satellite datasets
- Anomaly detection and Characterization
- Post Training Analysis
- Enterprise Architecture for ML solutions in space mission

- ML Architecture Model
 - How a ML interacts with space and ground assets
 - What are machine learning processes and techniques involved.
 - What are the functions and operational concept for a machine learning systems
- Algorithms and Formulism
 - How to apply the existing machine learning techniques to the problems in dynamical systems.
 - Data representations, and data training algorithms.
 - How to quantitatively measure data quality, system performance and operational status.
 - To obtain the actionable information
- ML Platform
 - How to address the challenges of ML in an operational environment.
 - How to develop a flexible, scalable, and extensible machine learning platform
- Portfolios
 - Applying the machine learning framework to a specific domain.
 - Satellite with different orbit characteristics: GEO, LEO.
 - Systems with large number of sensors.

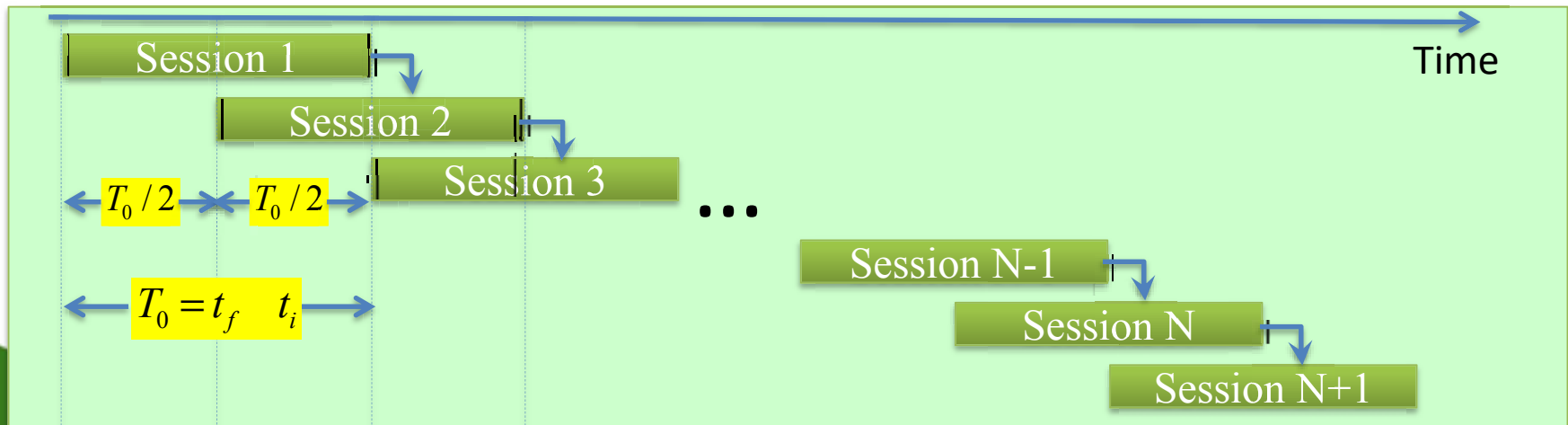
ML Architecture Model

- Situation Awareness is the **perception** of the elements in the environment within a volume of time and space, the **comprehension** of their meaning, and the **projection** of their status in the near future
- The architecture model defines an information loop between a machine learning system and its managed element.
 - The managed element provides the real-time data to a machine learning system.
 - Machine Learning system determine the state of dynamical system and taking appropriate actions.

ML creates the situational awareness for both space and ground assets that provides anomaly detection, data monitoring, sensor quality assessment.



- The data training in operational environments is performed periodically, in sessions, to ensure the time dependent trend captures both short term data patterns and long term changes
 - The trending period must be long enough to capture the data patterns
 - Two neighboring trending sessions overlap to ensure the continuity and the stability of the data training outcomes
 - The output of the trending session N is used as the input of the trending session N+1 to improve the trending efficiency
- The post training analysis is performed after data training.
- The data models used for data monitoring and filtering is updated after each training session



- The Gaussian probability distribution for a continuous dataset $\{d_j(t_i)\}$ can also be characterized by

- The time dependent function:

$$S_j = f_j(t - t_0, \{S_k\})$$

- And the noise level σ (Standard Deviation):

$$\sigma_j = \sqrt{\frac{1}{N} \sum_i \left(f_j(t_i - t_0, \{S_k\}) - d_j(t_i) \right)^2}$$

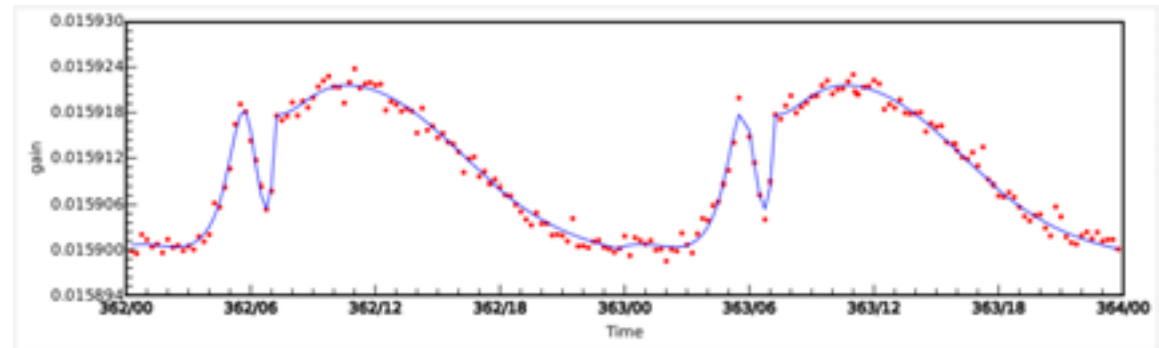
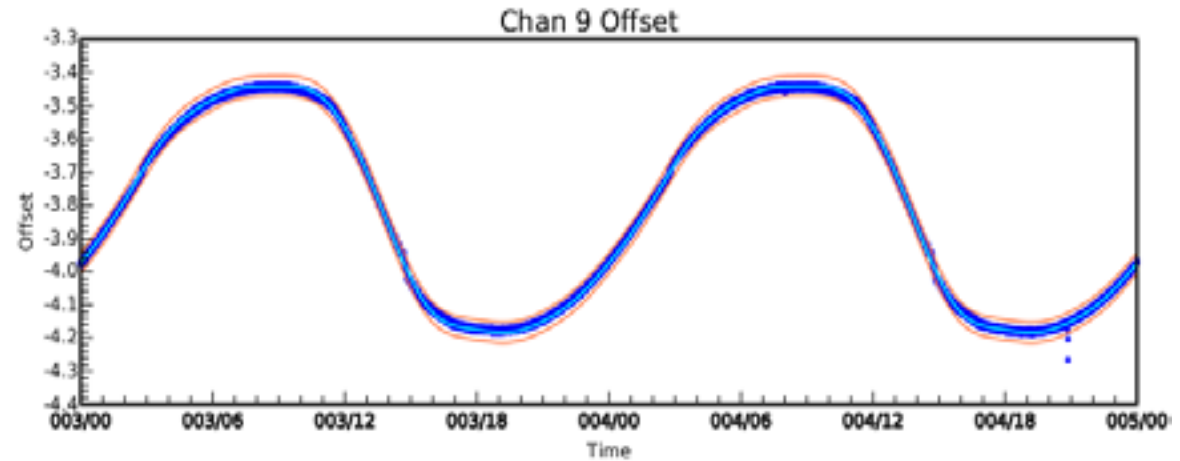
- t_0 is a reference time for a specific data pattern
- The combination $\{f_j(t - t_0, \{S_k\}), \sigma_j\}$ is defined as the **time dependent trend**
 - The noise level σ_j defines the quality of a dataset.
- Define a data model $F(t_i - t_0, \{S_k\}, \{w_j\})$ for $f(t_i, \{S_k\})$ in time-dependent trending, where $\{w_j\}$ is the parameter set to be determined.
- Develop a data training algorithm to minimize the error function

$$e = \frac{1}{2} \sum_i \left(D(t_i) - F(t_i - t_0, \{S_k\}, \{w_j\}) \right)^2$$

with respect to the parameter set $\{w_j\}$, which is the least-square fitting.

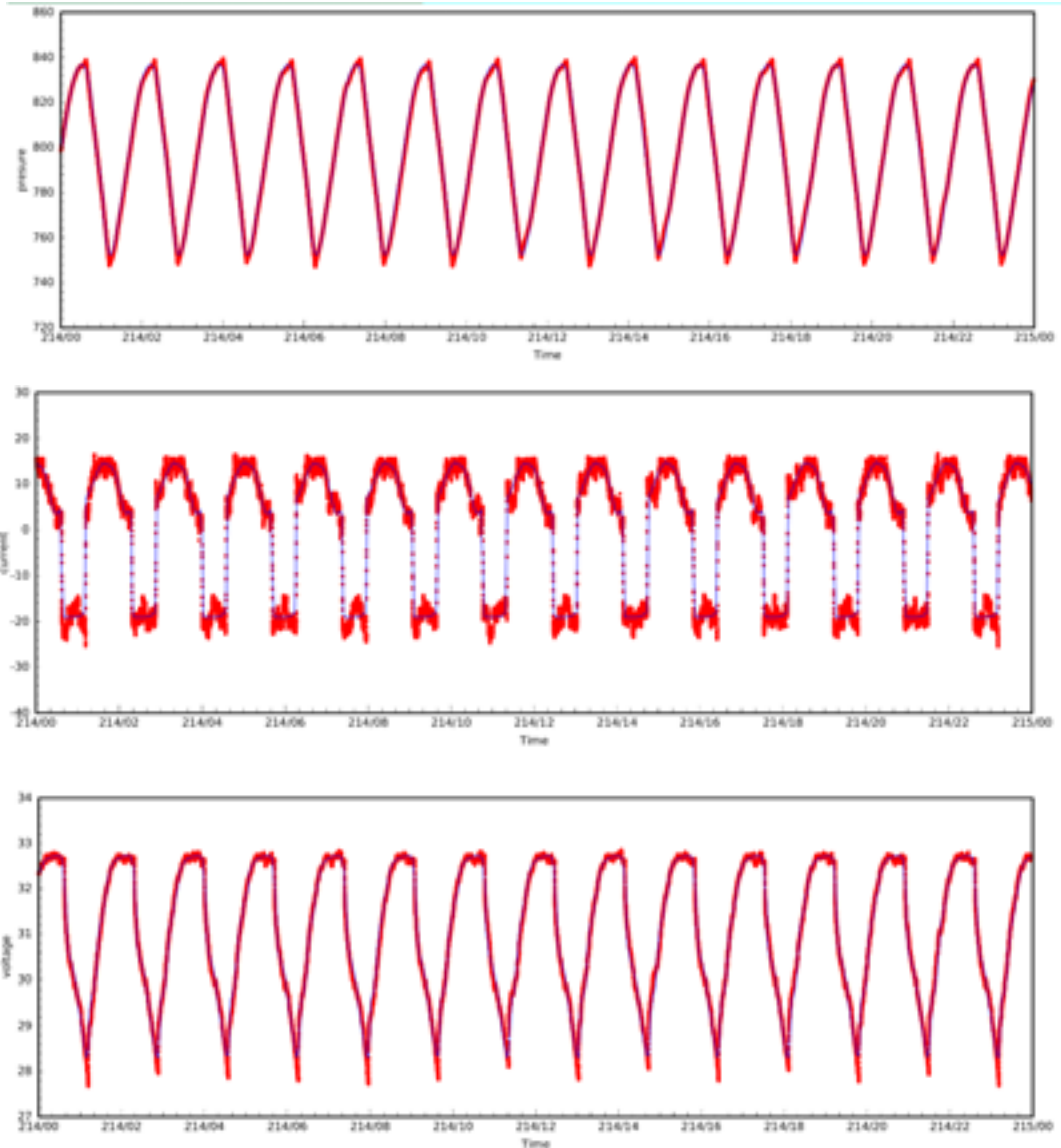
- Determined by the Gaussian Probability Distributions.

- Improving the data training efficiency while maintaining the accuracy in the training outcome is critical
 - The linear type data model is preferred.
- The Fourier Expansion Model (linear) and Neural Network (nonlinear) are implemented
 - Fourier expansion model is good for data patterns with dominant low frequency components.
 - Neural net is good for any pattern, but less efficient in data training.



Fourier Model (top), and neural network model (simple neural network with two hidden layers) (bottom). The data comes from GOES instrument data

- Neural networks with the same network structure are used as the data model for all three mnemonics
 - Two hidden layer network with 4 nodes in the first hidden layer and 2 nodes in the second hidden layer
- The mnemonics are the voltage, current, and pressure of the battery
- Two line element is used to calculate the reference time and the pattern period.



- A data point with the value $d(t_i)$ satisfies

$$|f(t_i, \{S_k\}) - d(t_i)| > N\sigma$$

is defined as outlier $O(d(t_i))$ above its noise level.

- An anomaly means persistent **data pattern change** that leads to a cluster of consecutive outliers
 - The data model based on the existing pattern is no-longer valid
- The cluster metric for the dataset j based on the outlier cluster,

$$\chi_j^O = \sum_i \left(\frac{\delta_i^W}{T} \right) + \frac{N^E}{N^W} \sum_i \left(\frac{\delta_i^E}{T} \right)$$

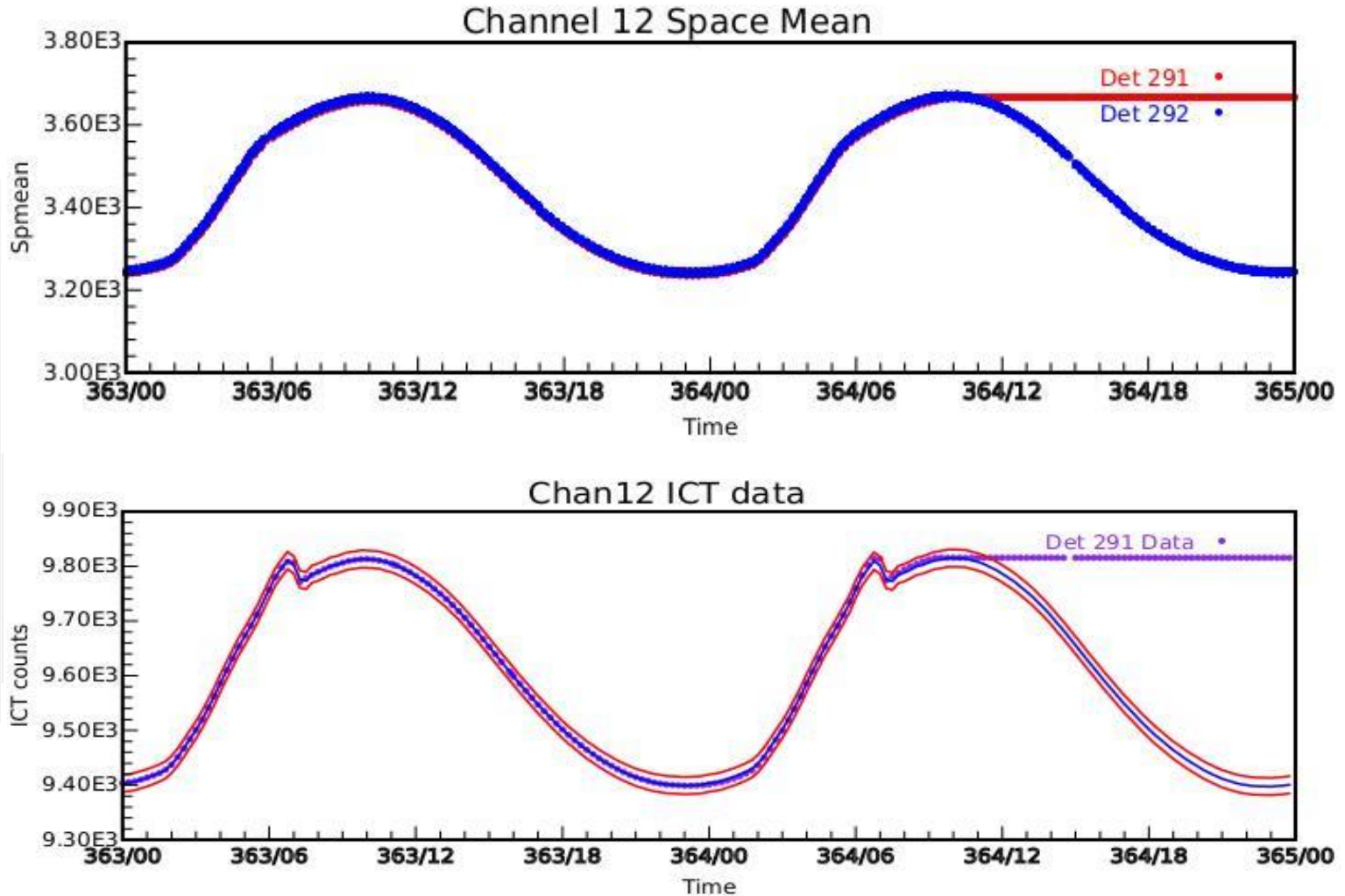
- δ_i^W and δ_i^E are the period for the warning and error outlier clusters $C_j^E(\delta_i^W)$ and $C_j^E(\delta_i^E)$ respectively, and

$$\delta_i^W = \delta_i^E = 0$$

for a single outlier

- N^W and N^E are the warning and error threshold parameters
- T represents the time scale of persistent outliers determined by the sampling frequency of datasets
- χ_j^O is dimensionless

The Unexpected Pattern Change Example



The data pattern change here can not be detected through the static limit monitoring process

The pattern change can not be detected with the static red/yellow limits

- Data quality metrics measures quality (or data pattern change) of datasets for a training session.
 - Quantitative and actionable.
- Three metrics are defined:
 - Outlier clusters:

$$\chi_j^O = \sum_i \left(\frac{\delta_i^W}{T} \right) + \frac{N^E}{N^W} \sum_i \left(\frac{\delta_i^E}{T} \right)$$

- $0 \leq \chi_j^O < \infty$

- The temporal change measure the change in $\{\sigma_j\}$ in the consecutive training sessions

$$\chi_j^T = \frac{\sigma_j^M}{\sigma_j^{M-1}}$$

- $0 \leq \chi_j^T < \infty$

- Significant increase in the metric may be caused by the data pattern changes.

- The spatial change measures the relative data quality for a dataset group:

$$\chi_j^S = \frac{\sigma_j}{\frac{1}{N} \sum_{k \in \{J\}} \sigma_K}$$

- A group has a set of datasets with the same patterns and scales

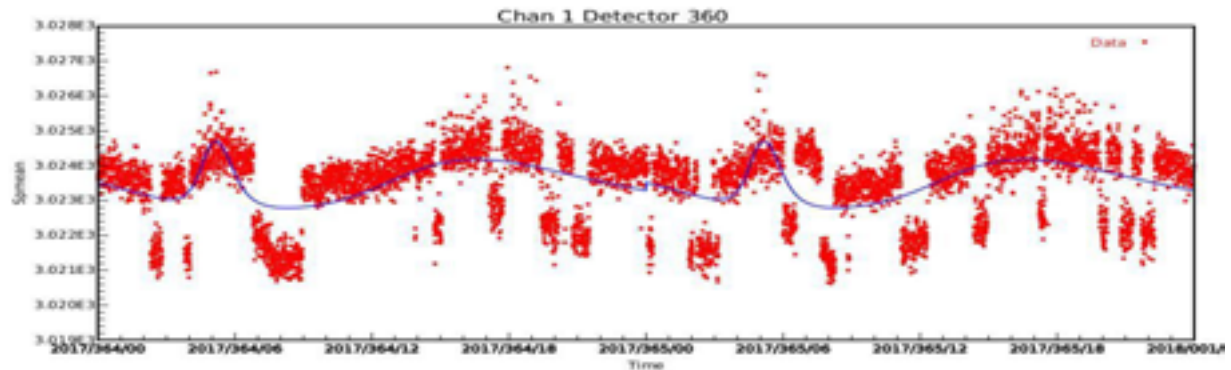
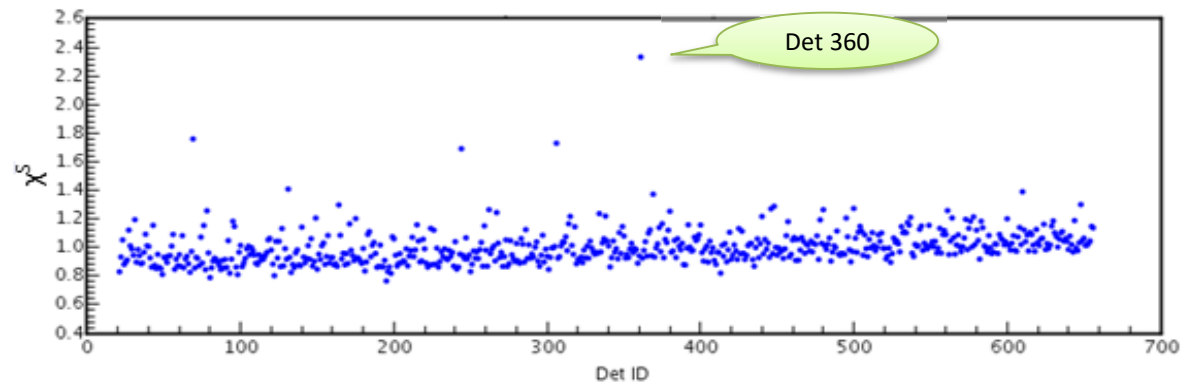
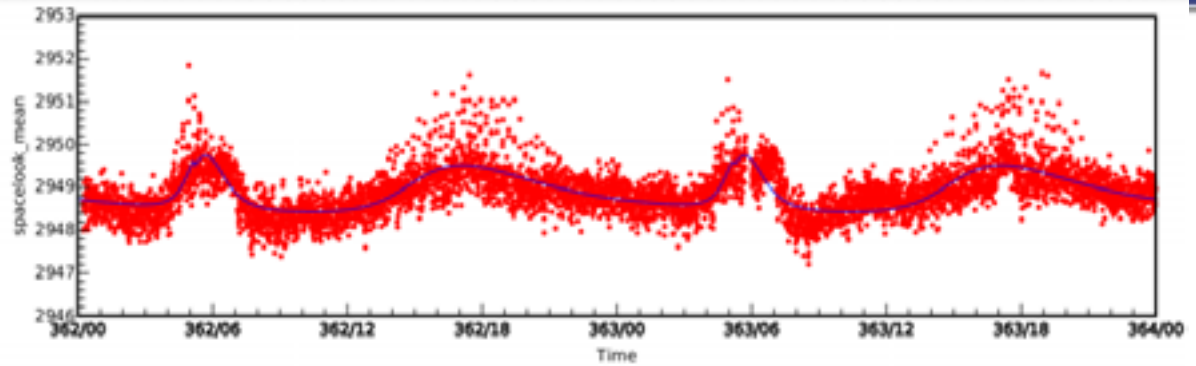
- The mnemonics for detectors within the same spectral channels belongs to the same dataset group.

- $0 \leq \chi_j^S < \infty$.

- Spatial change is generally used for sensor quality evaluation

An example of data quality metrics

- Spacelook datasets use the neural network (4 nodes in the 1st hidden layer, 2 nodes in the 2nd).
- The data point spatial change χ_j^S value 2.4 corresponds to detector 360.
- The plot below represents the spacelook data points for detector 360 during the trending period. The metric provides a measure of the detector's data quality.
 - The pattern is referred as the burst (or popcorn) noise.
 - May causes the occasional striping.



- Big Data Challenge
 - Large Number of Datasets (Mnemonics) and the ML
 - The number of datasets or mnemonics is at order of $\sim 10^4$ or more for a mission.
 - Large Data Volume
 - The data volume for telemetry and instrument meta data is at the order of gigabytes.
 - The data training must be completed in a very short period
 - Diverse Data Types and patterns
 - Consists of both static (no noise) and dynamic (noisy) datasets
 - Datasets used for data training could be defective
 - The data used in data training may contain outliers that distort the training outcome
- A ML System should have
 - Flexibility in selecting different data models.
 - Extensibility to meet the mission specific requirements.
 - Scalability to handle large number of datasets

- Based on the ML Architecture Model with component, function, service and infrastructure layers
- Separate the data model specific logic and mission specific logic from the common infrastructure and services.
 - The common infrastructure and services are provided by ML platform
 - The data model specific logic and mission specific logic are implemented as components
- Standard Interfaces for components
- Benefits
 - Reduce the development cost and risk
 - Allow developer to focus on the components to meet the mission specific requirements



Advanced Intelligent Monitoring System (AIMS) provides an implementation of the ML enterprise architecture.

- The ML Architecture Model has well defined common functionalities, operation concepts
- An anomaly in a dataset causes the unexpected data pattern change, and can be detected in real time monitoring and post training analysis
- The ML enterprise architecture is an software implementation of the ML architecture model
 - Could be an industry level standard for developing the machine learning solutions in space missions
- The ML system, Advanced Intelligent Monitoring System (AIMS), has been developed.
 - Deployed in GOES-R for monitoring the instrument radiometric performance
- More works are in progress
 - Algorithm development for monitoring the satellite health and safety
 - Architecture development and interface definition
 - Developing ML portfolios for various missions