# Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding

Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell and Tom Soderstrom

NASA Jet Propulsion Laboratory

California Institute of Technology

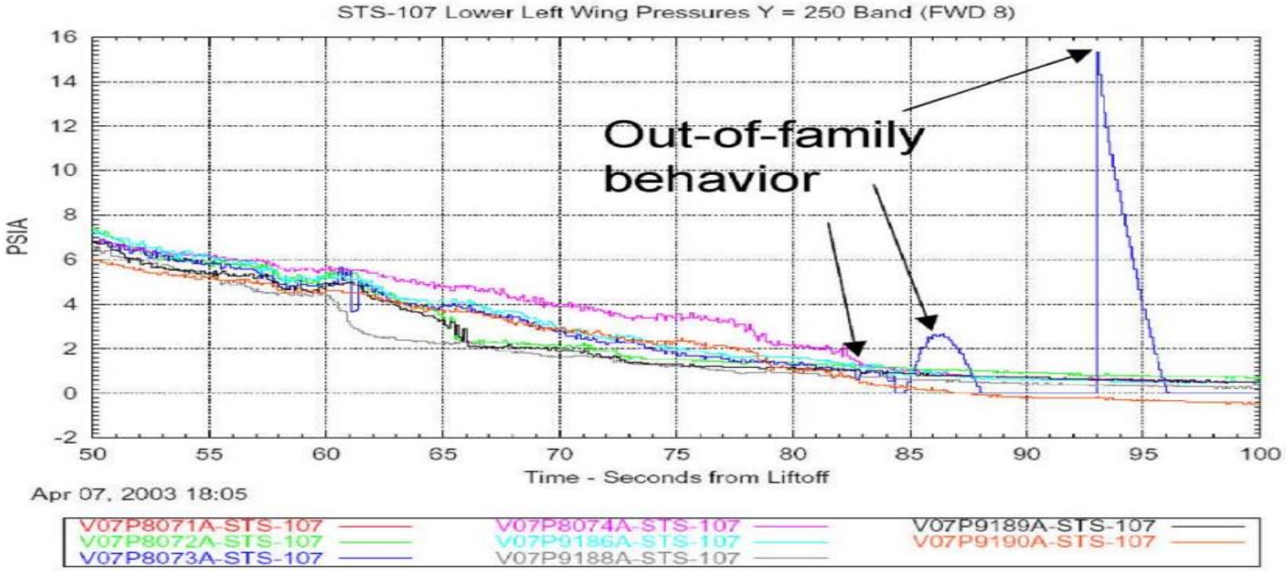
Presenter: Avisha Bhiryani

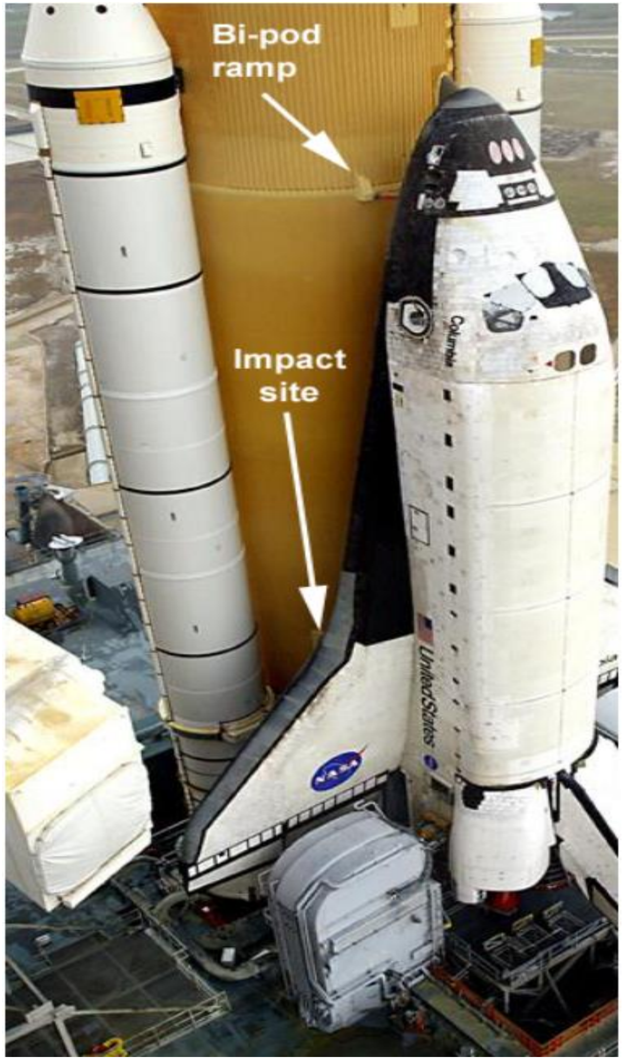
# Table of Contents:

- Motivation
- Data at Hand
- Anomalies in Time Series Data
- Anomaly Detection in Aerospace
- Long Short Term Memory
- 3-step Procedure for Anomaly Detection using LSTMs
- Experiment Setup and Result
- References

# Columbia STS-107 Tragedy

- Picture on the right is the very reason for the loss of Columbia STS-107 space-shuttle along with the 7 member crew on board.
- Purpose of the Modular Auxiliary Data System(MADS) on board the Columbia Orbiter was to support annunciation of vehicle system failures and out-of-tolerance system conditions.





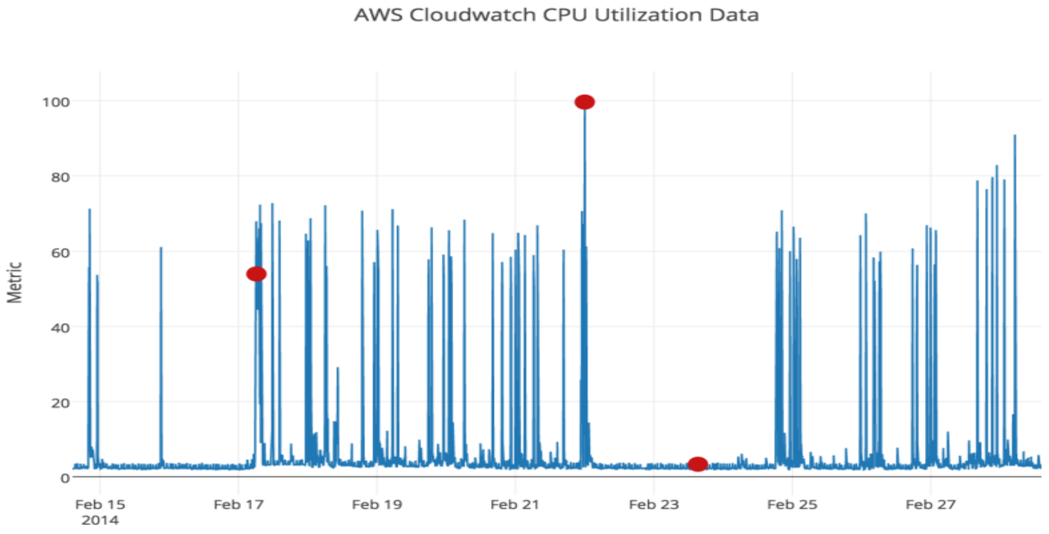
[Left]:https://history.nasa.gov/columbia/Troxell/Columbia%20Web%20Site/CAIB/CAIB%20Website/CAIB%20Report/Volume%204/part03.pdf
[Right]https://www.cbsnews.com/network/news/space/home/memorial/107.html

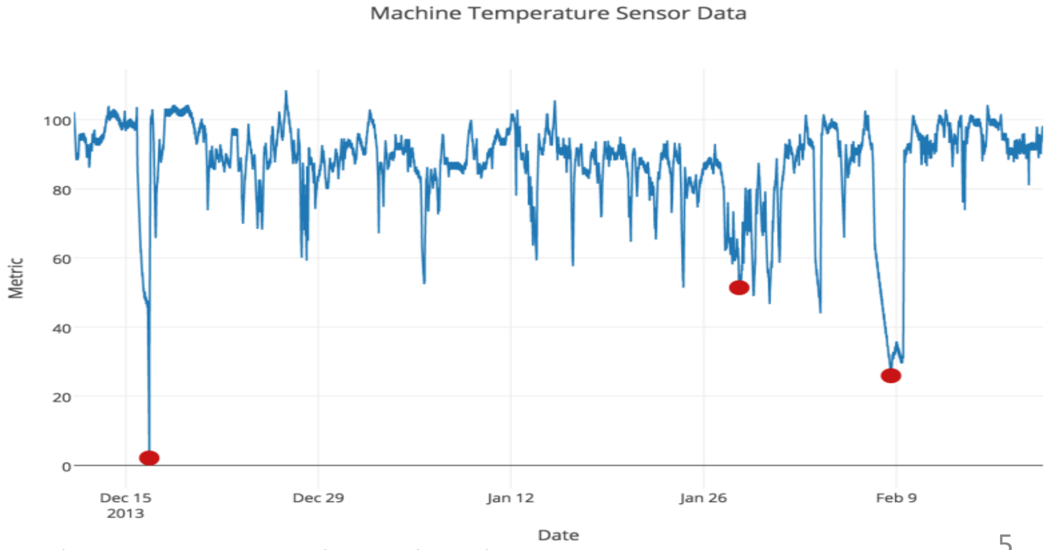
# Data at Hand

- Thousands of telemetry channels detailing aspects such as: Temperature, Radiation, Power, Instrumentation and computational activities.
- Data at hand is multivariate time series data.
- Challenges:
  - Lack of labelled anomalies
  - Noisy and high dimensional data
  - Context dependent

# Introduction to Anomalies in Time Series

- Anomalies in time series data are the patterns that do not conform to the past behavior of the data.
- There are 3 categories of anomalies
  - Point: Single values that fall within low density region of values.
  - Collective: Sequence of anomalies rather than a single one.
  - Contextual: They do not fall in regions of low density values but are anomalous w.r.t the local values


AWS Cloudwatch CPU Utilization Data

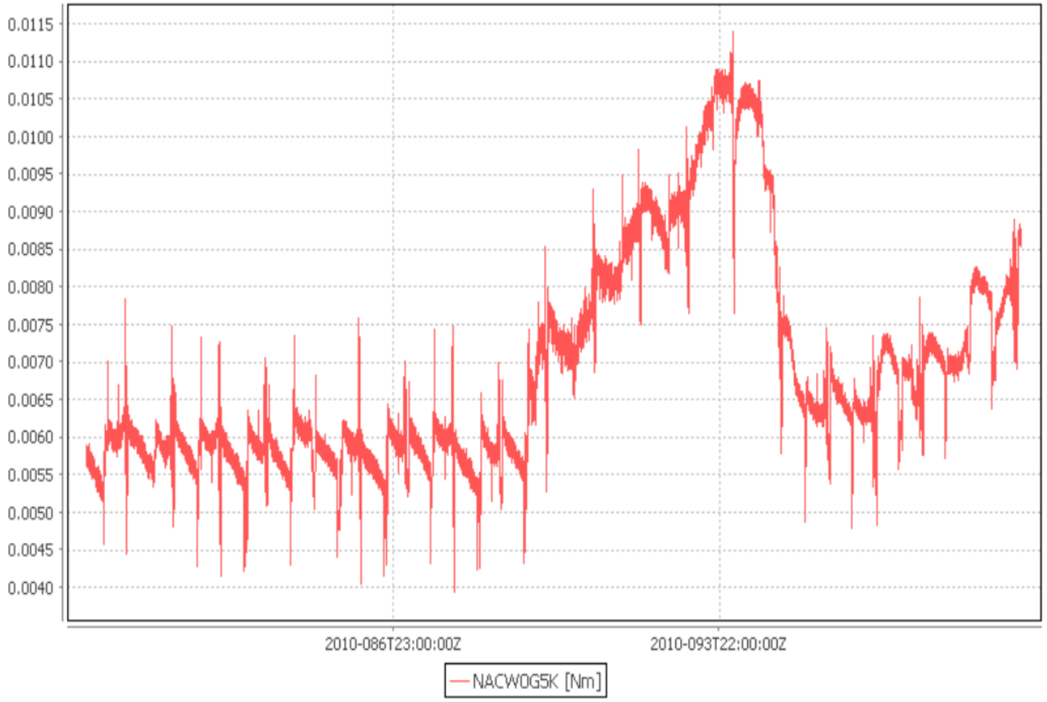

Machine Temperature Sensor Data

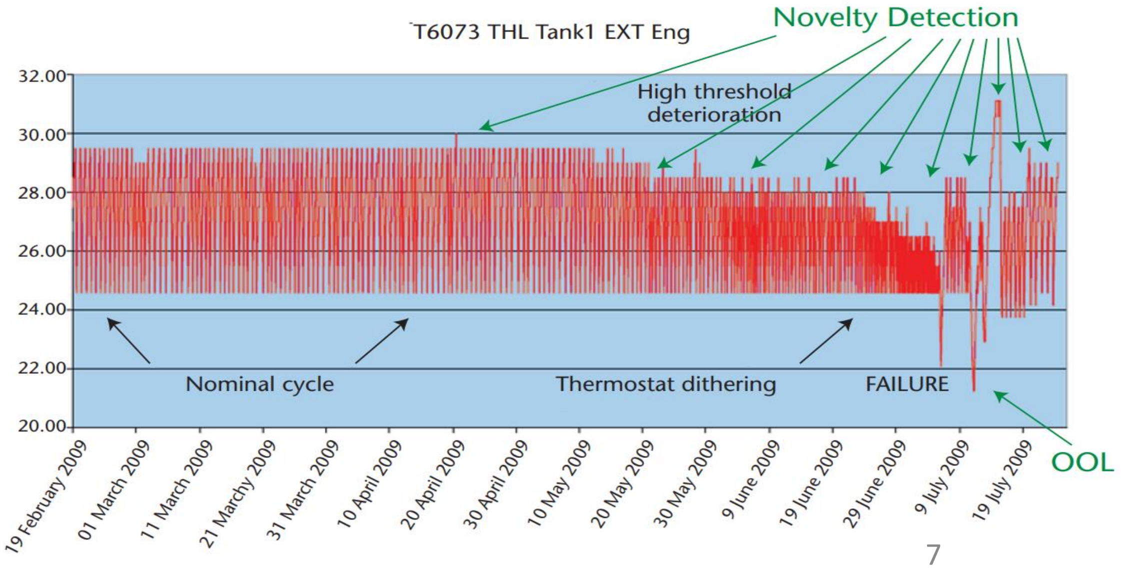
Source: A. Lavin and S. Ahmad, Evaluating Real time Anomaly Detection Algorithms – the Numenta Anomaly Benchmark

# Anomaly detection in Aerospace

- Different Approaches to detect anomalies in spacecraft telemetry data are as follows, but not limited to:
  - OOL(Out of limits)
  - Clustering
  - Nearest Neighbor
  - Expert systems
  - Deep Learning

- Out of Limits(OOL):
  - Alarm is triggered when a measurement goes beyond the set lower or upper threshold.
  - Lot of manual intervention needed:
    - Engineers will inspect the parameter that is out of limits and determine whether it is an anomaly or not and decide which action to take.
    - OOL bounds are not defined for every parameter.
  - Most popular approach:
    - Low computational expense, broad and straight-forward applicability, and ease of understanding.





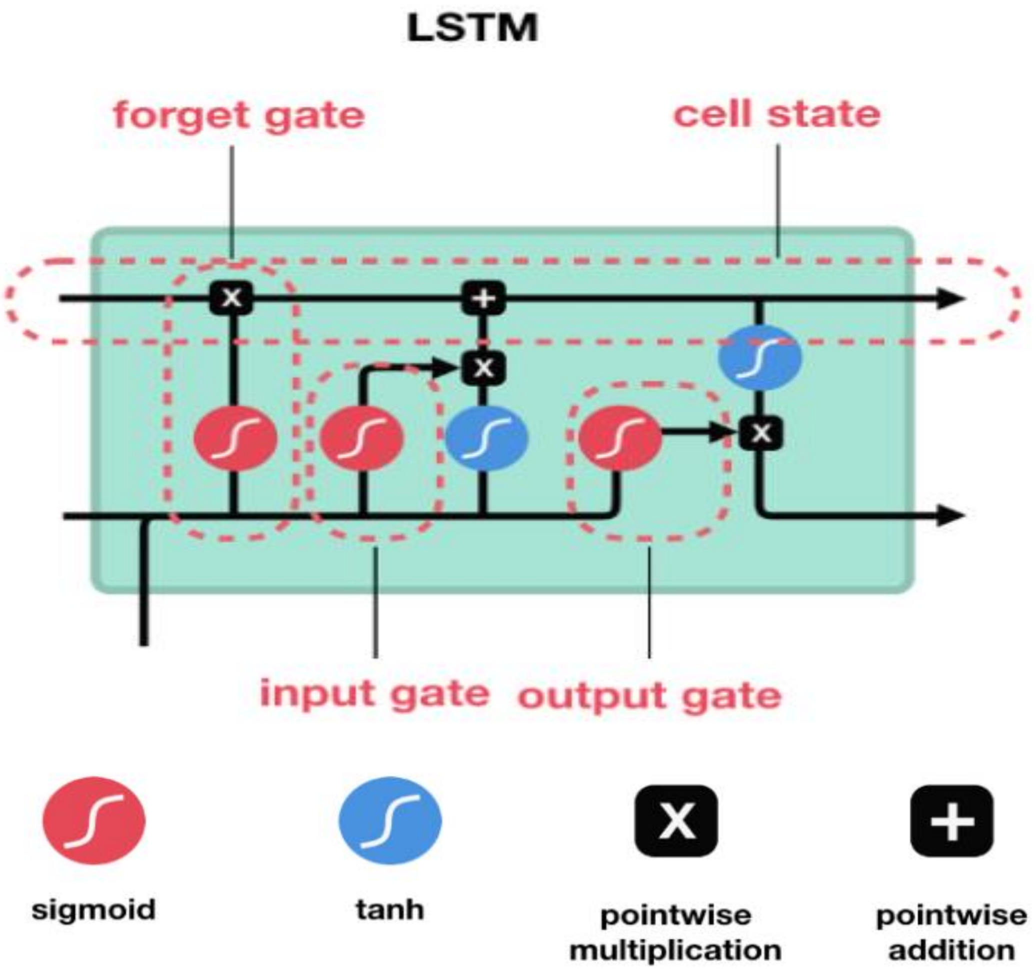
Source:Jose MartÃŋnez-Heras and Alessandro Donati. 2014. Enhanced Telemetry Monitoring with Novelty Detection.

- Nearest Neighbor Approach:
  - Orca(currently being used on International Space Station) uses a nearest-neighbour approach to search for anomalies by calculating the distance of each data point from neighbouring points.
  - The program outputs a score for each point representing the average distance to the nearest k-neighbours. Points that have a larger average distance to their nearest neighbours than most other points in are considered outlier.
  - Issues:
    - Amount of past data to consider needs to be decided.
    - High computational complexity and memory consumption
    - Can only find global outliers
- Clustering based Approach:
  - The Inductive Monitoring System(IMS)(currently being used on ISS) tool uses clustering to analyze archived spacecraft data and characterize nominal interactions between selected parameters. This model, is compared with real time or archived system data to detect off nominal behaviour.
  - IMS returns the distance that vector falls from the nearest nominal operating region.

# Deep Learning Approach: Long Short Term Memory

- Core Concept:
  - Cell State: Memory of the network
  - Gates: Information get’s added or removed to the cell state through gates.
- Forget Gate:
  - It decides what is relevant to keep from prior steps.
- Input Gate:
  - It  decides what information is relevant to add from the current step.
- Output Gate:
  - The output gate determines what the next hidden state should be
- Advantages of using LSTM:
  - Overcomes the Vanishing Gradient problem, so can learn long term correlations in a sequence.
  - LSTM networks obviate the need for a pre-specified time window.

# Long Short Term Memory:



Source: https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

# Core Components:

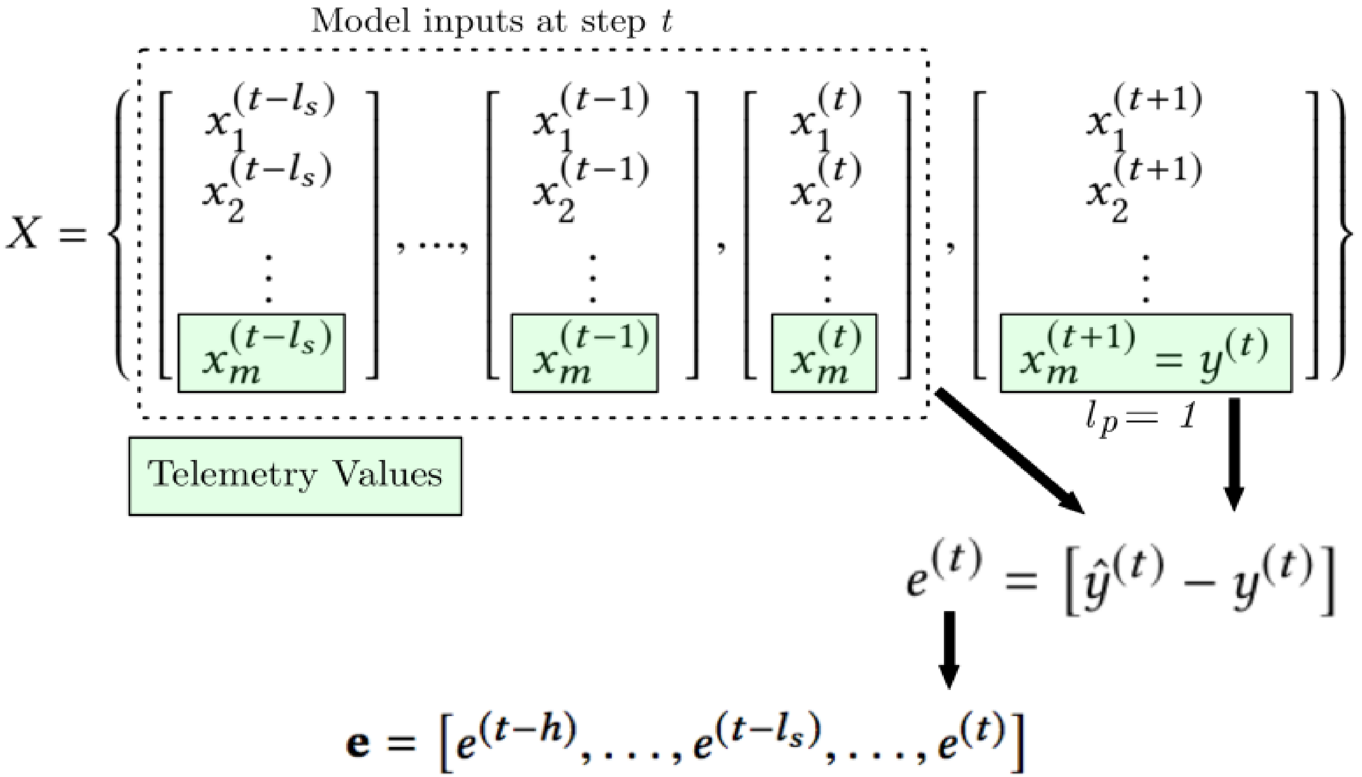
- LSTMs are used to predict telemetry data by learning from normal command and telemetry data

- An unsupervised thresholding method is then used to automatically assess hundreds of diverse streams of telemetry data and determine whether resulting prediction errors represent spacecraft anomalies.

- In order to mitigate false positive errors, certain strategies are employed.

# Telemetry Value Prediction using LSTM:

- A single model is created for each telemetry channel and each model is used to predict values for that channel.
- This is because:
  - LSTMs struggle to accurately predict high dimensional data.
  - Modeling each channel independently also allows traceability of the issue down to the channel level.
  - If the system were to be trained to detect anomalies at the subsystem level without this traceability, operations engineers would still need to review a multitude of channels and alarms across the entire subsystem to find the source of the issue
  - Early stopping can be used to limit training to models and channels that show decreases in validation error.



Model inputs at step $t$

$$X = \left\{ \begin{bmatrix} x_1^{(t-l_s)} \\ x_2^{(t-l_s)} \\ \vdots \\ \boxed{x_m^{(t-l_s)}} \end{bmatrix}, \ldots, \begin{bmatrix} x_1^{(t-1)} \\ x_2^{(t-1)} \\ \vdots \\ \boxed{x_m^{(t-1)}} \end{bmatrix}, \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \\ \vdots \\ \boxed{x_m^{(t)}} \end{bmatrix}, \begin{bmatrix} x_1^{(t+1)} \\ x_2^{(t+1)} \\ \vdots \\ \boxed{x_m^{(t+1)} = y^{(t)}} \end{bmatrix} \right\}$$

$l_p = 1$

Telemetry Values

$$e^{(t)} = \left[\hat{y}^{(t)} - y^{(t)}\right]$$

$$\mathbf{e} = \left[e^{(t-h)}, \ldots, e^{(t-l_s)}, \ldots, e^{(t)}\right]$$

# Dynamic Error Thresholding

- Only abnormally high or low smoothed prediction errors should be considered as spacecraft anomalies.
- Set of errors are then smoothed to dampen spikes in errors that frequently occur with LSTM-based predictions.
- To determine the threshold:
  - Values evaluated for ϵ are determined using z, where z is an ordered set of positive values representing the number of standard deviations above the mean of smoothed error vector.
  - A threshold is found that, if all values above are removed, would cause the greatest percent decrease in the mean and standard deviation of the smoothed errors
  - Then the highest smoothed error in each sequence of anomalous errors is given a normalized score based on its distance from the chosen threshold.

$$\mathbf{e}_s = [e_s^{(t-h)}, \ldots, e_s^{(t-l_s)}, \ldots, e_s^{(t-1)}, e_s^{(t)}]$$

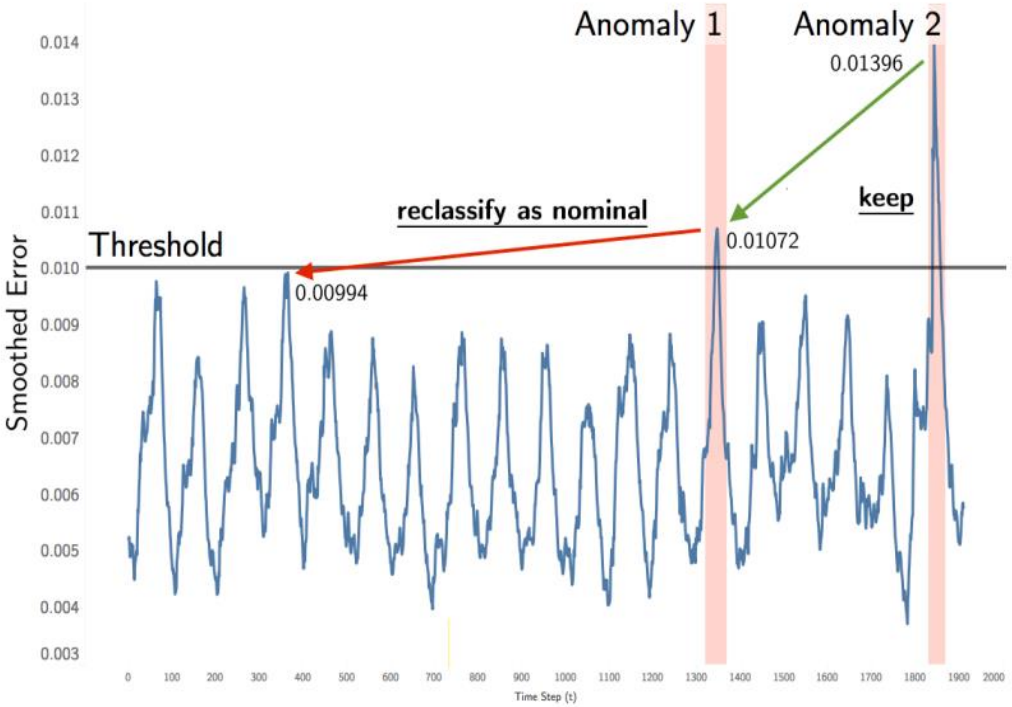
$$\boldsymbol{\epsilon} = \mu(\mathbf{e}_s) + \mathbf{z}\sigma(\mathbf{e}_s)$$

$$\epsilon = argmax(\boldsymbol{\epsilon}) = \frac{\Delta\mu(\mathbf{e}_s)/\mu(\mathbf{e}_s)) + (\Delta\sigma(\mathbf{e}_s)/\sigma(\mathbf{e}_s)}{|\mathbf{e}_a| + |\mathbf{E}_{seq}|^2}$$

$$s^{(i)} = \frac{max(\mathbf{e}_{seq}^{(i)}) - argmax(\boldsymbol{\epsilon})}{\mu(\mathbf{e}_s) + \sigma(\mathbf{e}_s)}$$

# Mitigating False Positives

- The precision of prediction-based anomaly detection approaches heavily depends on the amount of historical data (h) used to set thresholds and make judgments about current prediction errors.
- At large scales it becomes expensive to process historical data in real-time scenarios and a lack of history can lead to false positives that are only deemed anomalous because of the narrow context in which they are evaluated
- To mitigate false positives and limit memory and compute cost, a pruning procedure has been introduced



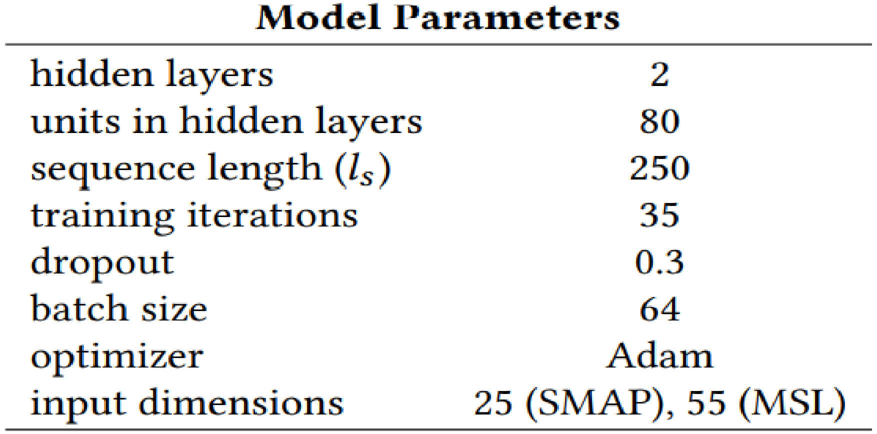
$$\mathbf{e}_{max} = [0.01396, 0.01072, 0.00994]$$

$$p = 0.1.$$

$$d^{(1)} = 0.23 > p \qquad d^{(2)} = .07 < p$$

# Experiment Setup

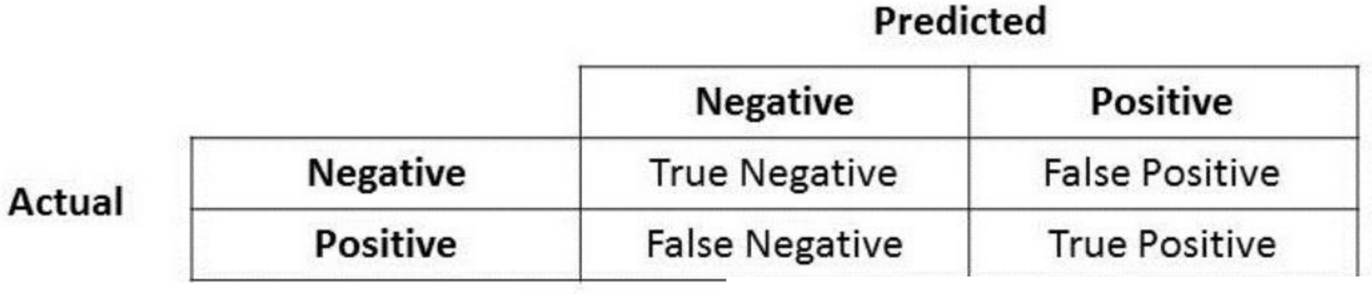
- Spacecraft data used: SMAP, MSL
- Incident Surprise, Anomaly Reports(ISA) were used to obtain labeled anomalies
- A 5-day span around anomalies is selected to get deeper insight into precision and also reasonable computational cost.
- Sequence Length and the amount of prior values that are used to evaluate a batch of values are selected through experiment result.



| Model Parameters | |
|---|---|
| hidden layers | 2 |
| units in hidden layers | 80 |
| sequence length ($l_s$) | 250 |
| training iterations | 35 |
| dropout | 0.3 |
| batch size | 64 |
| optimizer | Adam |
| input dimensions | 25 (SMAP), 55 (MSL) |

# Metrics

- True Positive: Only one true positive is recorded even if portions of multiple predicted sequences fall within a labeled sequence.
- False Positive: Only one true positive is recorded even if portions of multiple predicted sequences fall within a labeled sequence.



|  |  | **Predicted** |  |
|---|---|---|---|
|  |  | **Negative** | **Positive** |
| **Actual** | **Negative** | True Negative | False Positive |
|  | **Positive** | False Negative | True Positive |

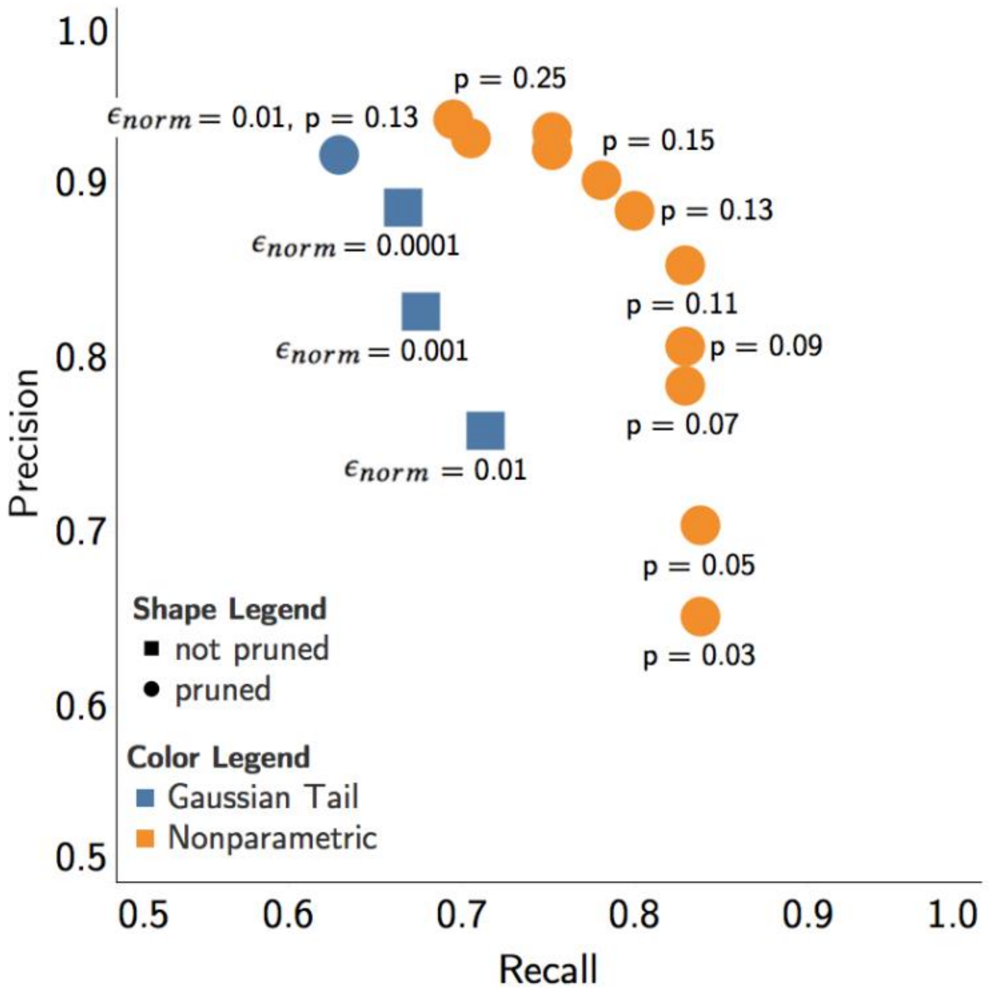
$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

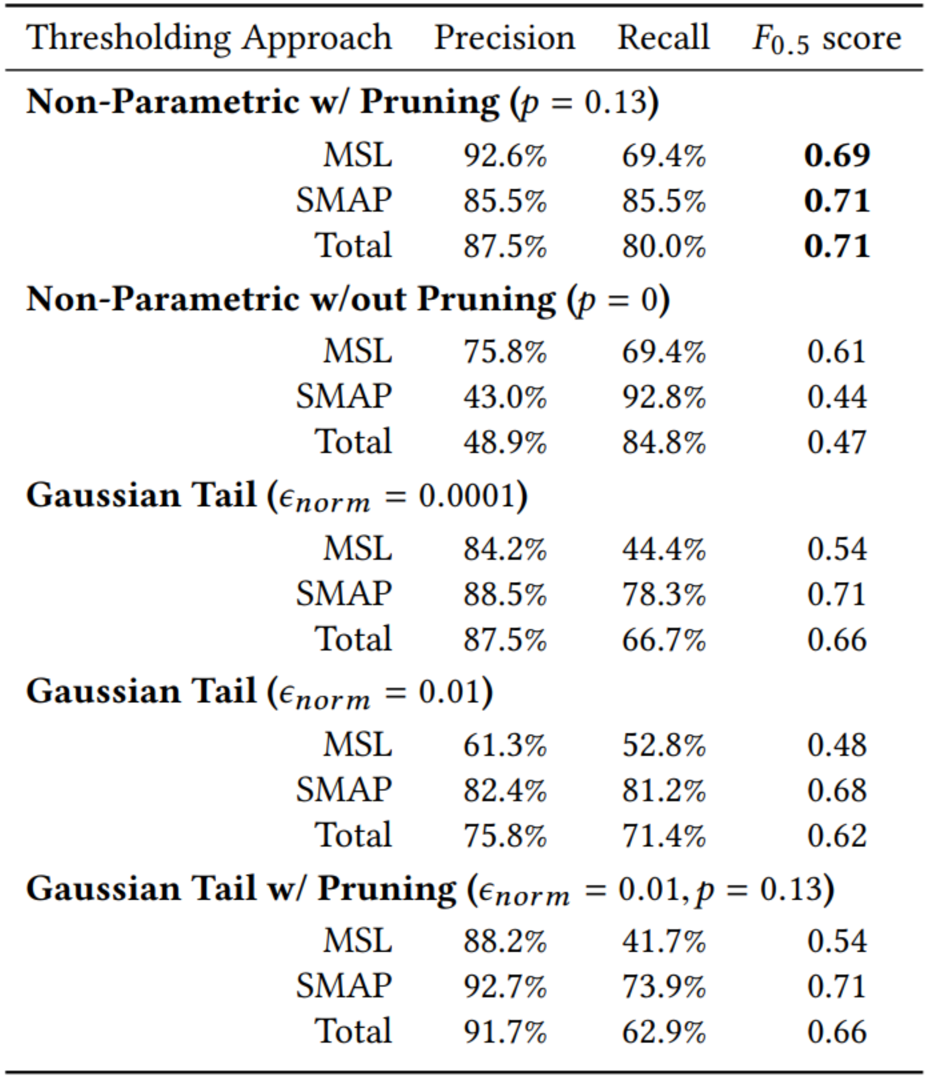
$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Source: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

# Experiment Result:

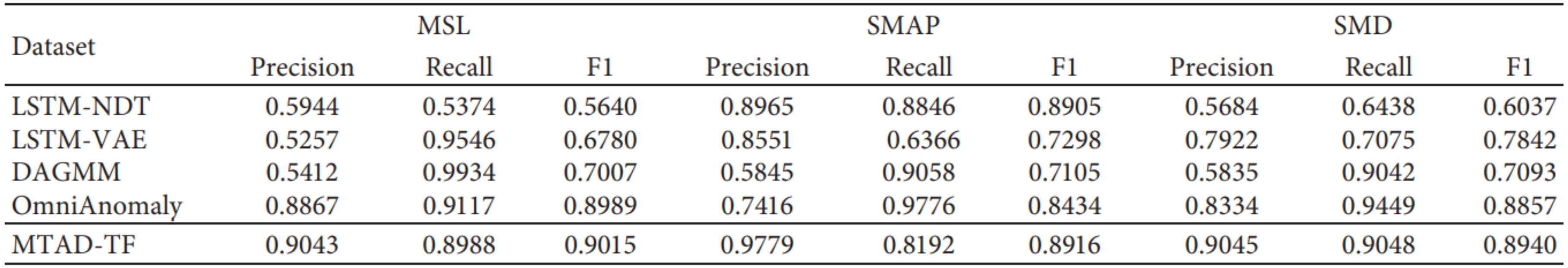




| Thresholding Approach | Precision | Recall | $F_{0.5}$ score |
|---|---|---|---|
| **Non-Parametric w/ Pruning** ($p$ = 0.13) | | | |
| MSL | 92.6% | 69.4% | **0.69** |
| SMAP | 85.5% | 85.5% | **0.71** |
| Total | 87.5% | 80.0% | **0.71** |
| **Non-Parametric w/out Pruning** ($p$ = 0) | | | |
| MSL | 75.8% | 69.4% | 0.61 |
| SMAP | 43.0% | 92.8% | 0.44 |
| Total | 48.9% | 84.8% | 0.47 |
| **Gaussian Tail** ($\epsilon_{norm}$ = 0.0001) | | | |
| MSL | 84.2% | 44.4% | 0.54 |
| SMAP | 88.5% | 78.3% | 0.71 |
| Total | 87.5% | 66.7% | 0.66 |
| **Gaussian Tail** ($\epsilon_{norm}$ = 0.01) | | | |
| MSL | 61.3% | 52.8% | 0.48 |
| SMAP | 82.4% | 81.2% | 0.68 |
| Total | 75.8% | 71.4% | 0.62 |
| **Gaussian Tail w/ Pruning** ($\epsilon_{norm}$ = 0.01, $p$ = 0.13) | | | |
| MSL | 88.2% | 41.7% | 0.54 |
| SMAP | 92.7% | 73.9% | 0.71 |
| Total | 91.7% | 62.9% | 0.66 |

# Criticism

- Does not generalize well.
- Along with the telemetry data, channel specific command information is also an input to the model that needs to be refined.



| Dataset | MSL | | | SMAP | | | SMD | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall | F1 |
| LSTM-NDT | 0.5944 | 0.5374 | 0.5640 | 0.8965 | 0.8846 | 0.8905 | 0.5684 | 0.6438 | 0.6037 |
| LSTM-VAE | 0.5257 | 0.9546 | 0.6780 | 0.8551 | 0.6366 | 0.7298 | 0.7922 | 0.7075 | 0.7842 |
| DAGMM | 0.5412 | 0.9934 | 0.7007 | 0.5845 | 0.9058 | 0.7105 | 0.5835 | 0.9042 | 0.7093 |
| OmniAnomaly | 0.8867 | 0.9117 | 0.8989 | 0.7416 | 0.9776 | 0.8434 | 0.8334 | 0.9449 | 0.8857 |
| MTAD-TF | 0.9043 | 0.8988 | 0.9015 | 0.9779 | 0.8192 | 0.8916 | 0.9045 | 0.9048 | 0.8940 |

# References:

- Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. 2018. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In KDD ’18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3219845
- A. Lavin and S. Ahmad, “Evaluating Real-time Anomaly Detection Algorithms – the Numenta Anomaly Benchmark,” in 14th International Conference on Machine Learning and Applications (IEEE ICMLA’15), 2015
- Columbia Accident Investigation Board Report, Volume 3, NASA, August 2003
- David Iverson. 2008. Data Mining Applications for Space Mission Operations System Health Monitoring. In SpaceOps 2008 Conference. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2008-3212
- Q. He , Y. J. Zheng , C.L. Zhang ,and H. Y. Wang, MTAD-TF: Multivariate Time Series Anomaly Detection Using the Combination of Temporal Pattern and Feature Pattern in Hindawi Complexity Volume 2020, Article ID 8846608, 9 pages https://doi.org/10.1155/2020/8846608
- Donald P. Tallo, John Durkin, and Edward J. Petrik. 1992. Intelligent fault isolation and diagnosis for communication satellite systems. Telematics and Informatics 9, 3-4 (jun 1992), 173–190. https://doi.org/10.1016/s0736-5853(05)80035-8
- Jose MartÃŋnez-Heras and Alessandro Donati. 2014. Enhanced Telemetry Monitoring with Novelty Detection. 35 (12 2014), 37–46