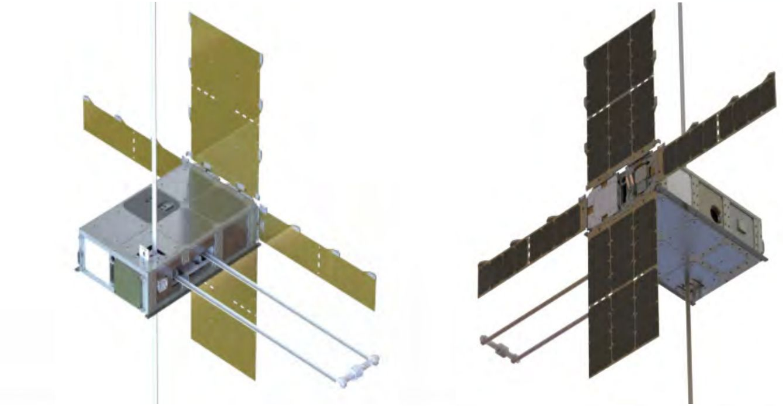
Methods for Data-centric Small Satellite Anomaly Detection and Fault Forecasting

Joseph Melville

UF | UNIVERSITY of FLORIDA

Objectives

1. Predict satellite faults before they occur
2. Relate faults to specific subsystems
3. Using real world small satellite bus telemetry



VPM Design Pictures: spacecraft rear (left), spacecraft front (right).
Photo credit: AFRL

Motivation

1. Inform subsystem fault isolation
2. Enable fault avoidance interventions
3. Decrease downtime for real world satellite systems

Prior work

- Treat anomalies as faults
- Expert systems, system thresholds
- Time-series forecasting, error thresholds
  - Recurrent Neural Networks (RNN)
  - Long Short Term Memory (LSTM)

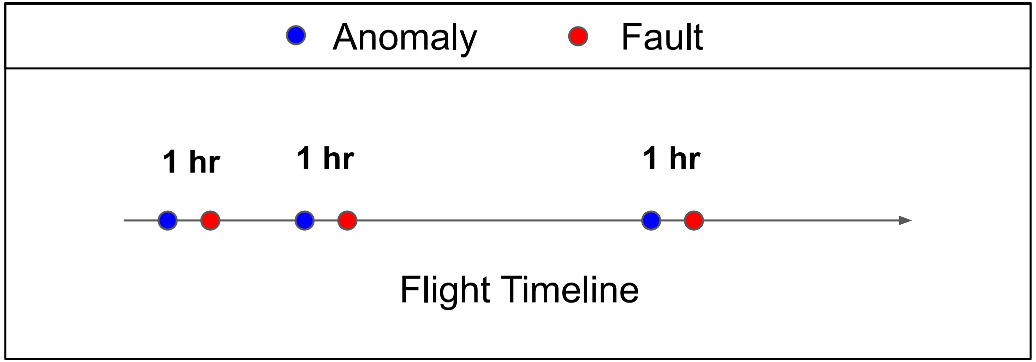
Limitations:

- Detect faults instead of projecting faults
- Use simulated data or lossless communication
- Do not clearly explain pre-processing methods

Current work

1. Use raw operational telemetry
2. Clearly outline data pre-processing methods
3. Define fault predictability metrics
4. Demonstrate fault forecasting using example anomaly/fault pairs

Fault Forecasting



Background

Definitions

- Fault - System malfunction
- Outliers - Unlikely occurrence (not necessarily a malfunction)
- **Anomaly** - Outliers from different data perspectives

Sequence → [15, **-64564**, 12, **14, 14, 14, 14, 14, 14, 14,** 13, 12, 16] → **Outlier**, **Anomalous**

↓

Number frequency perspective → [1, 1, 1, **7,** 1, 1, 1] → **Outlier**

Dataset structure

Packet Type: ADCS Sensors

Items

Individual **packets**

ID	Time	x	y	z	x_vel	y_vel	…
1							
2							
3							

…

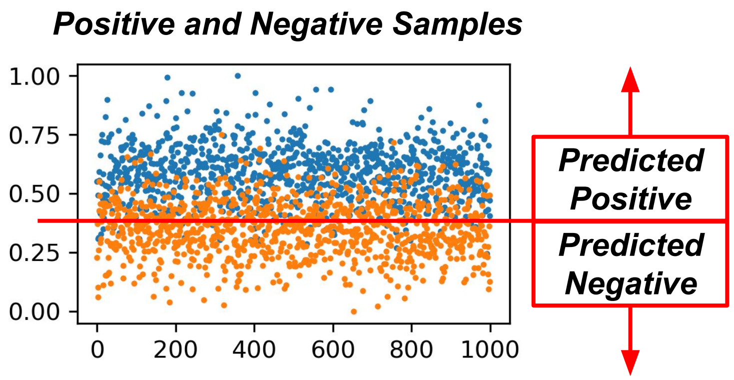
Binary Classification

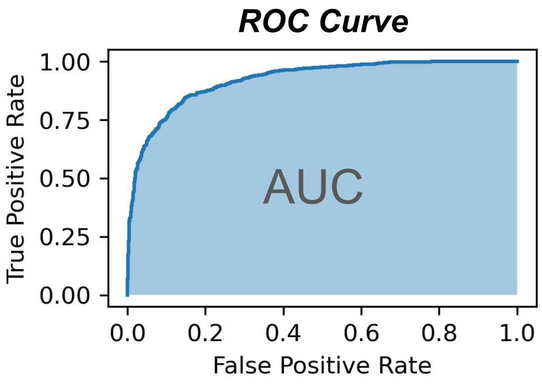
- Precision
  - P = TP/(TP+FP)
  - False positive rate = 1 - Precision
- Recall
  - R = TP/(TP+FN)
  - True positive rate
- F1 score
  - F1 = 2*P*R/(P+R)

		Actual	
		True	False
Predicted	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Receiver Operating Characteristic (ROC) Curve

- True positive rate
  - Recall
- False positive rate
  - 1 - Precision
- ROC curve
  - Draw many thresholds
  - Each threshold is a single point in the curve
- Area under the ROC curve (AUC)
  - Measure of separability

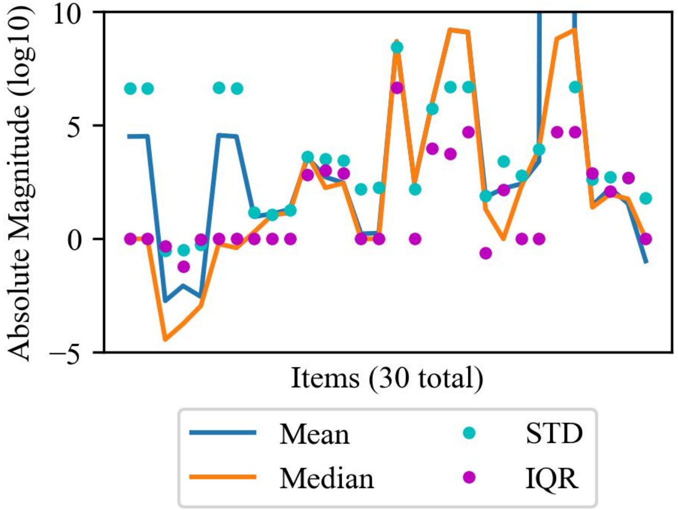




VPM Dataset

- Disparate data (text, timestamp, binary…)
- Extreme outliers
- Non-uniform and inconsistent timelines

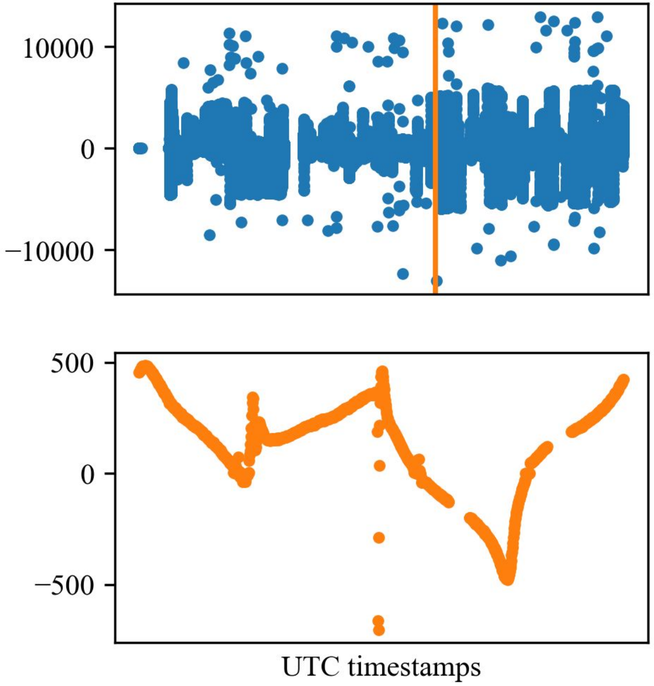
Statistics from ADCS Overview packet type



VPM Dataset

- Disparate data (text, timestamp, binary…)
- Extreme outliers
- Non-uniform and inconsistent timelines

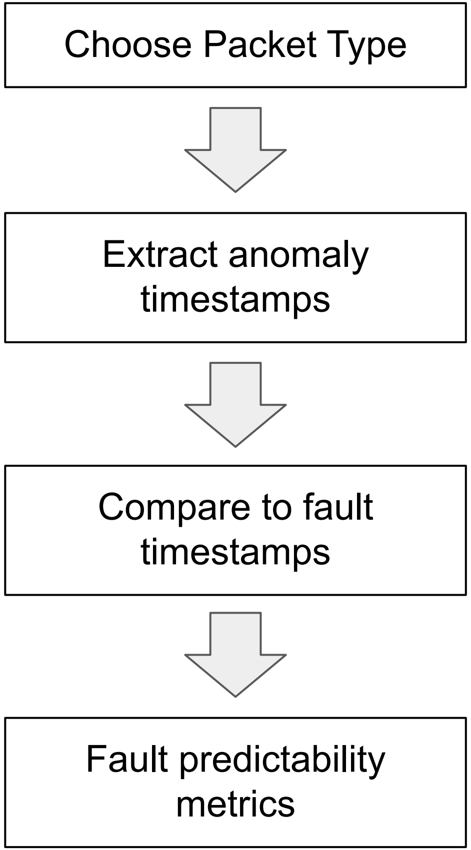
Single item from ADCS Overview packet type



Methods

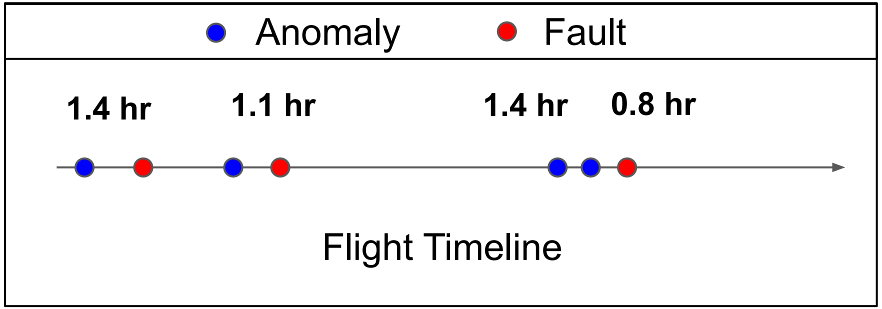
Extract timestamps

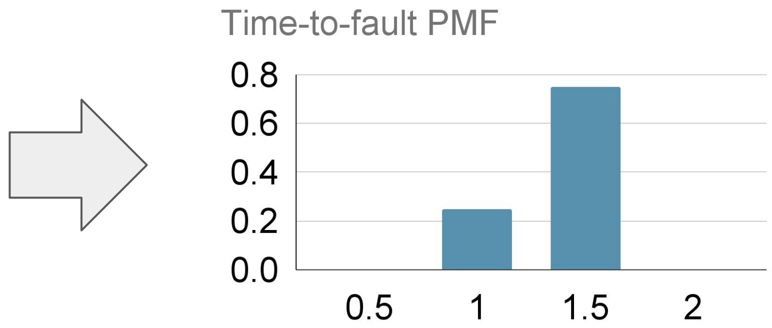
- Faults
  - Reboot
  - Demote to sunsafe
  - Momentum too high
- Anomalies (applied to each packet type)
  - Blip
  - Isolation Forest (IF)
  - Principal Component Analysis (PCA) → IF
  - PCA → Local Outlier Factor (LOF) → IF



Time-to-fault PMF

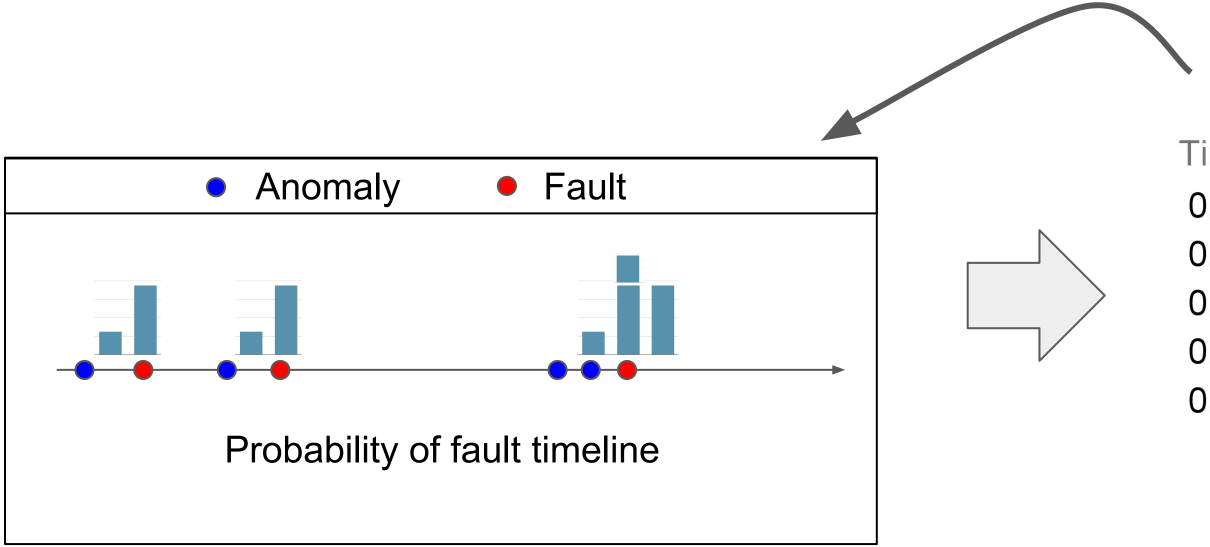
1. Measure hours from each anomaly to next fault
2. Find probability mass function (PMF)
   - Histogram (bins defined by desired sampling rate)
   - Normalize sum of bins to one

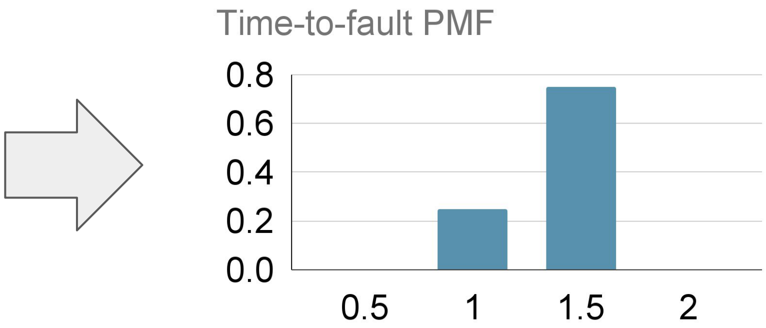




Probability of fault timeline

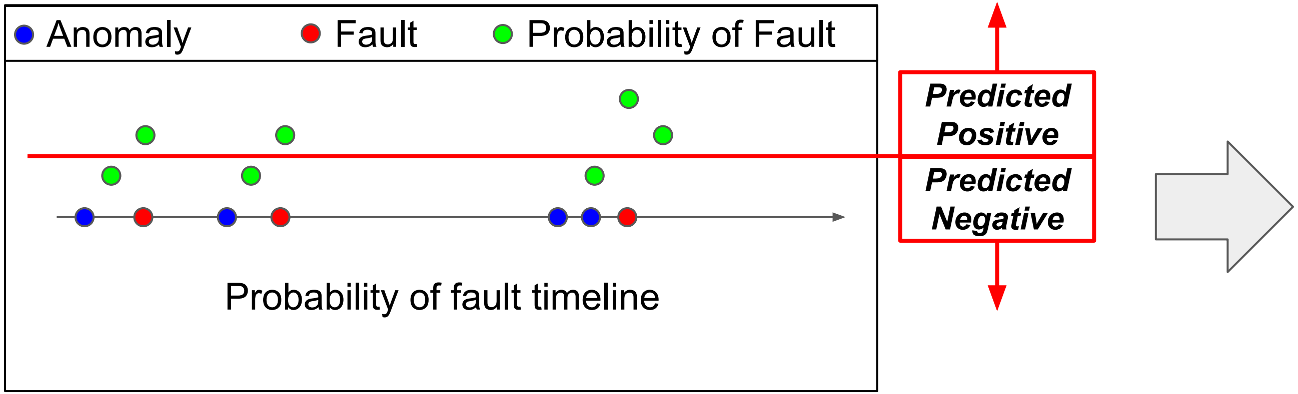
1. Define uniform timeline (match histogram bin size)
2. Overlay time-to-fault PMF at each outlier





Fault forecasting metrics

- Standard binary classification metrics now work
  - Binary classification thresholds
  - Precision, recall, F1 score
  - ROC, AUC
- Times to true positive from most recent anomaly (TTP)

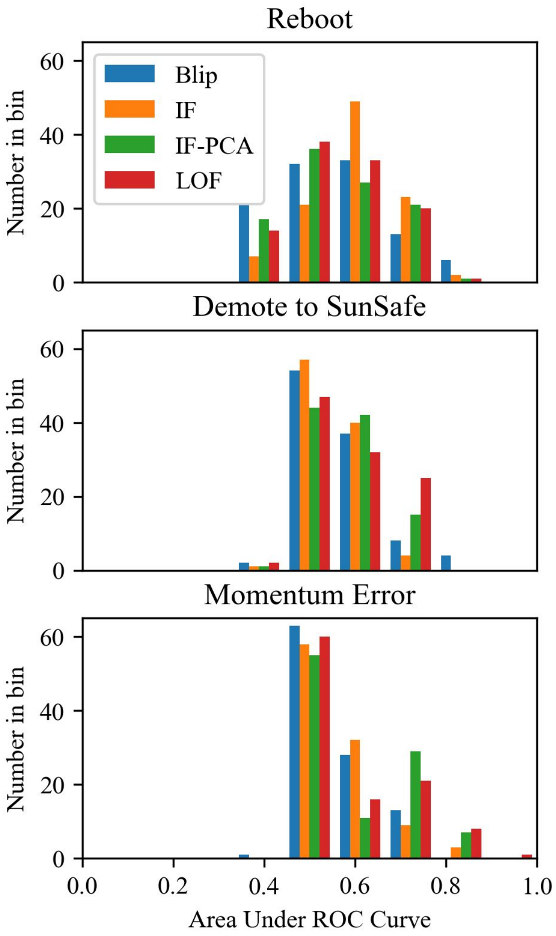


Precision = 0.75
Recall = 1
F1 score = 0.86
TTP (avg) = 1.18 hr
TTP (std) = 0.29 hr

Results

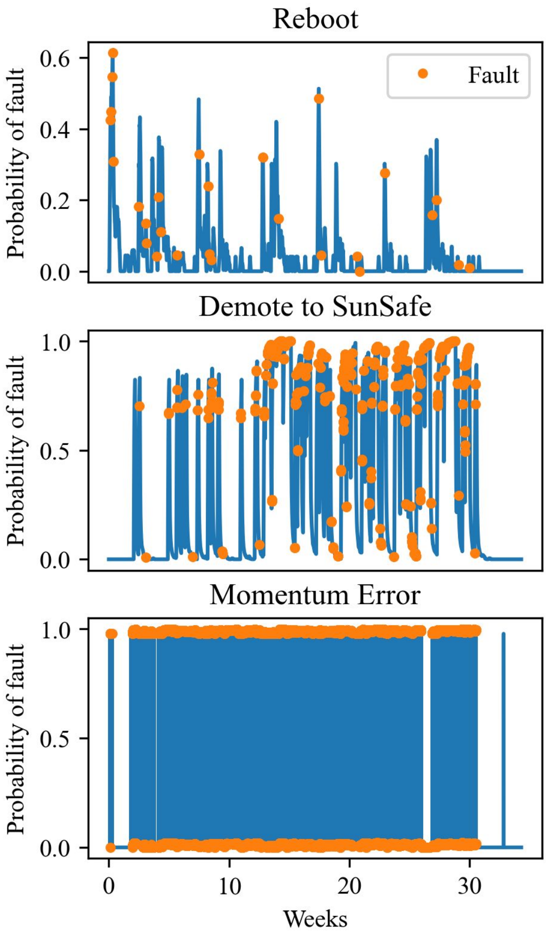
AUC histograms

- Separability of fault and non-fault timesteps
- For each combination of:
  - Packet type
  - Anomaly detection method, and
  - Fault type



Compare with precision and recall

- Reboot
  - AUC = 0.840
  - Precision = 0.007
  - Recall = 0.037
- Demote to SunSafe
  - AUC = 0.787
  - Precision = 0.004
  - Recall = 0.202
- Momentum too high
  - AUC = 0.927
  - Precision = 0.829
  - Recall = 0.929



Momentum too high top results

Anomaly Method	Packet Type	AUC	Precision	Recall	TTP (hrs)	
					AVG	STD
LOF	ADCS Sensors	0.927	0.829	0.929	0.008	0.063
LOF	ADCS Overview	0.888	1.000	0.776	0.000	0.000
IF-PCA	ADCS Verify	0.849	0.899	0.736	0.006	0.029
LOF	I2C Power	0.818	0.676	0.959	0.178	0.699

20 second to 10 minute faut forecasting

Too large

Sampling rate vs predictability metrics

- Lower sampling rate (larger timeline step sizes)
- Higher predictability (AUC , F1 score…)
- Lower resolution
  - Fault is 30-60 minutes instead of 39-40 minutes

Conclusion

Contributions

1. Novel method to calculate **future fault probabilities** from **raw telemetry**
2. Evaluation **metrics for anomaly fault predictability**
3. Relate packet types to fault types
4. Can be implemented onboard to inform fault avoidance

Limitations

1. Fault labels are needed
2. No evidence for transferability between systems
3. Anomalies not specifically derived to forecast faults

Future work

- Cross validation
- Explore cross correlations instead of time-to-fault PMF
- Investigate transferability between systems

- Opportunities:
  - Active learning for fault labeling
  - RNN to find better informed fault forecasting features
  - MDP for fault avoidance

Code availability

- Select code for resampling, PMF estimation, and predictability metrics
- Demonstration using synthetic data

Acknowledgments

- The Small Satellites Portfolio, Space Vehicles Directorate, Air Force Research Lab provided facilities, VPM small satellite data, and expertise in support of this work
- This material is funded by the U.S. Department of Defence through a Science, Mathematics, and Research for Transformation (SMART) scholarship

Questions?

Appendix

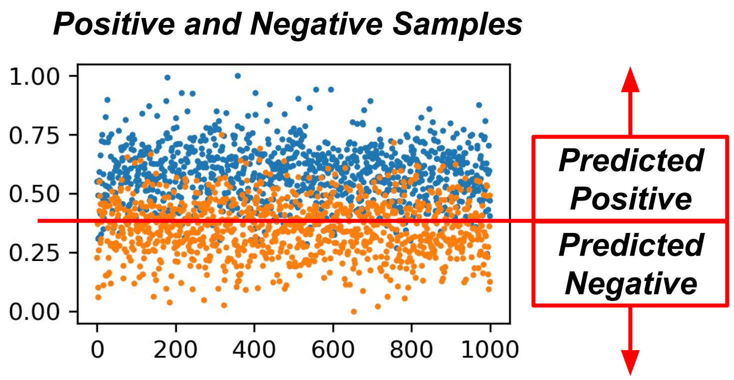
Binary Classification

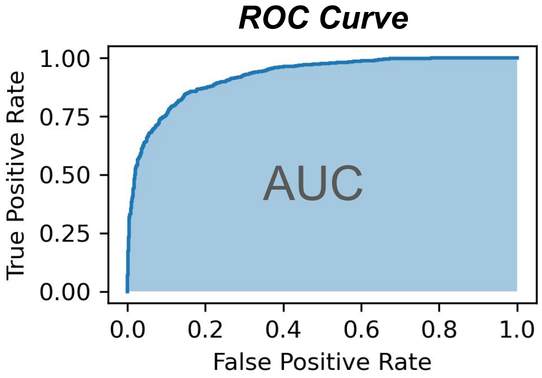
- Precision
  - P = TP/(TP+FP)
  - False positive rate = 1 - Precision
  - Fraction of positive predictions that are correct
- Recall
  - R = TP/(TP+FN)
  - Detection rate
  - Fraction of actual positives identified
- F1 score
  - F1 = 2*P*R/(P+R)
  - Harmonic mean between precision and recall

		Actual	
		True	**False**
Predicted	**True**	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Receiver Operating Characteristic (ROC) Curve

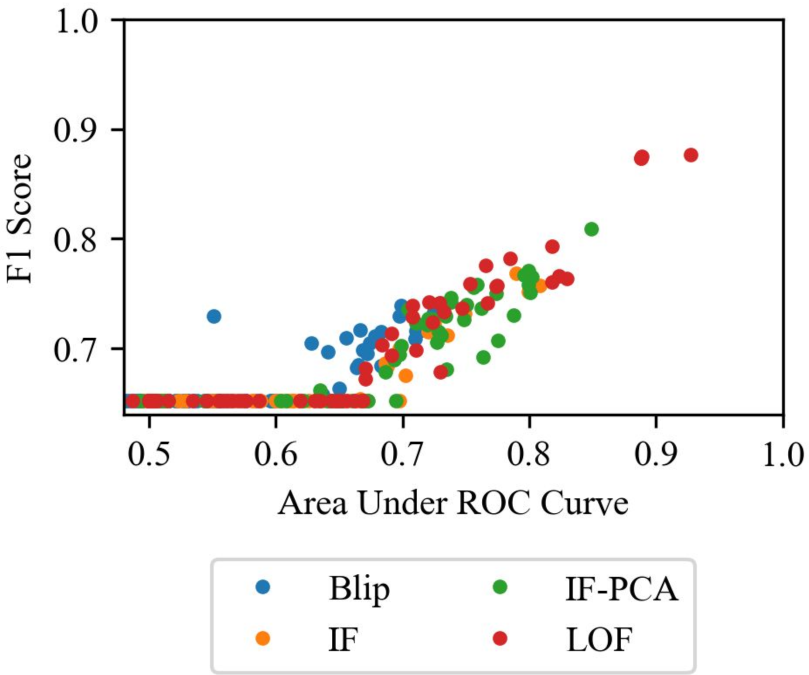
- True positive rate
  - TPR = Recall = Detection rate = TP/(TP+FN)
  - Fraction of actual positives identified
- False positive rate
  - FPR = FP/(FP+TN)
  - Fraction of actual negatives predicted positive
- ROC curve
  - Draw many thresholds
  - Each threshold is a single point in the curve
- Area under the ROC curve (AUC)
  - Measure of separability





AUC and F1 relationship

Momentum too high



Motivation for automated FDIR

Small Satellites:

- **Vital capabilities** →Observation, communications, tracking, science, security
- **Vulnerable to faults** →Collisions, internal errors, radiation, adversaries
- **Isolated** →Few operators, communication delay, limited contact

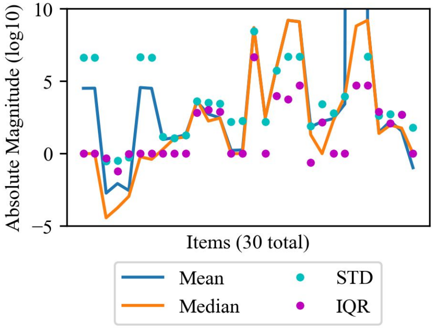
Automated FDIR enables:

- Quick, appropriate responses **despite vulnerabilities and isolation**
- Fault forecasting and avoidance to **preserves vital capabilities**

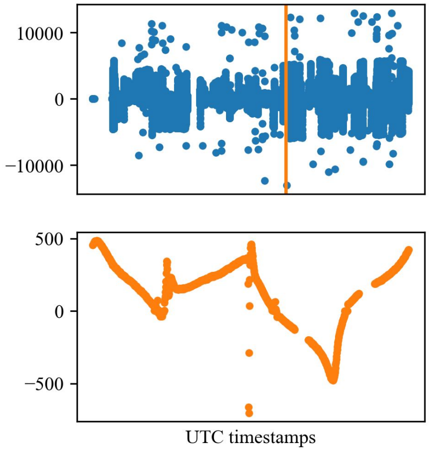
Challenges of real world telemetry

- Large amounts of disparate data
  - Text, binary, continuous, formatted timestamps
  - 123 packet types, 15 items and 250338 packets each (avg)
  - 3.9 GB
- Extreme outliers
- Sampling inconsistencies
  - Non-uniform sample rates within packet types
  - Inconsistent timelines between packet types
  - Gaps from dropped packets and reboots

Statistics from ADCS Overview packet type



Single item from ADCS Overview packet type



Outlier and fault timestamps

- Outlier detection methods
  - Blip
  - Isolation forest (IF)
  - IF with principal component analysis (PCA)
  - Local Outlier Factor (LOF) with PCA and IF
- Fault categories
  - Reboot
  - Demote to sunsafe
  - Momentum error
- Assess anomaly fault predictability
  - 4 outliers detection methods
  - 123 packet types
  - 3 fault categories
  - 1476 combinations to asses