Trusted AI Frameworks for Change and Anomaly Detection in Observed ISR

A SmartSat CRC Project at the University of Adelaide

Matthew Roughan <matthew.roughan@adelaide.edu.au>

July 24, 2025

THE UNIVERSITY *of* ADELAIDE

SMARTSAT Cooperative Research Centre

Introduction

People

The University of Adelaide

- Matthew Roughan
- Tat-Jun Chen
- **David Shorten**
- Wathsala Karunarathne
- Melissa Humphries
- John Maclean

The University of New South Wales

- Yang Yang

BAE Systems

- Will Heyne

Defence Science Technology Group

- Kruger White
- Han Gaetjens

Australian Space Agency

- Marianne Hawthorne

SmartSat CRC

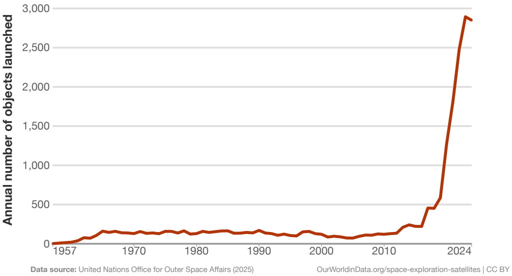
- Andrew Barton
- Peter Kerr

High-Level Overview

- SmartSat Cooperative Research Centre project
- Focussed on detecting satellite manoeuvres and anomalies using Two-Line-Element data (TLEs)
- Investigated different techniques:
  - Baseline
  - Particle Filters
  - Transformer Neural Networks
- Developed a chat interface
- Presented as a web app

Motivation

- Growing need for Space Situational Awareness (SSA)
  - Knowledge of the orbits of satellites allows for mission planning, resource management and strategic countermeasures.
- Our subproblem: detect
  - manoeuvres (even station keeping) for large constellations
  - unexpected (anomalous) behaviour
  - failures, collisions, ...
  - patterns of behaviour



Number of space objects regularly tracked by Space Surveillance Networks $\simeq 42,000$

Estimated number of break-ups, explosions, collisions, or anomalous events > 650

Need for SSA of increasing importance given the growing number of space objects and the growing technological importance of satellites

Challenges

- Scale
  - 10s of thousands of objects
  - time-scale of years
- Data
  - We need ground truth to train and test
- Communicating results
  - Are dashboards enough?
- **Trust**
  - AI systems have a bad rep for promising everything, but failing in delivery

Trusted AI

Factors that lead to trust

- Explanability
  - Interpretability
  - Transparency
- Benchmarking
  - Accuracy (known if not perfect)
  - Robustness (generalizability, out of distribution, new behaviour)
- User experience
  - Reliability
  - Safety
  - Balance between FA and FD

Anomaly detection usually falls down on benchmarking (lack of data) and user experience, but even explanability is getting worse these days with a proliferation of black-boxes.

Two-Line Element Data

TLEs

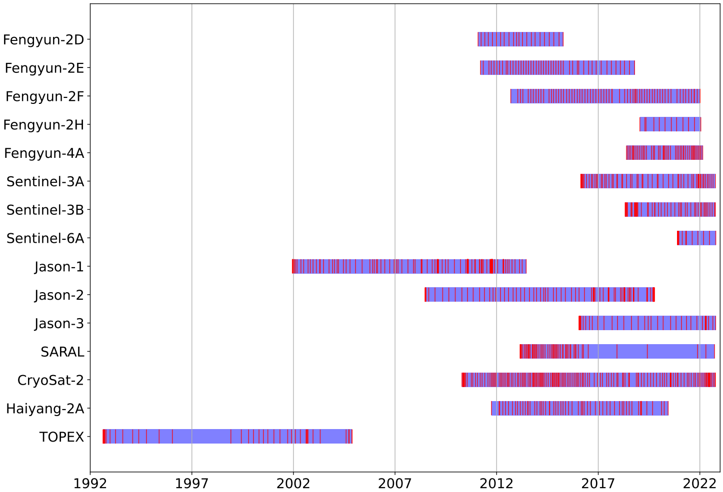
- Provided by the USA's 18th Space Defence Squadron
- Contains mean Keplerian orbital elements
- Updated roughly daily
- Comprehensive
- Low sample rate (roughly daily)

But TLEs aren't enough for training and testing – we need "ground" truth

Benchmark Dataset [4]

Name	SATCAT Number	TLE Epochs	Observed Manoeuvres Num	First	Last
Fengyun-2D	29 640	1178	22	2011-02-01	2015-04-10
Fengyun-2E	33 463	2367	48	2011-03-17	2018-10-15
Fengyun-2F	38 049	2977	68	2012-09-11	2022-01-05
Fengyun-2H	43 491	1043	12	2019-01-18	2022-01-18
Fengyun-4A	41 882	1296	49	2018-05-22	2022-02-21
Sentinel-3A	41 335	2226	64	2016-02-23	2022-10-07
Sentinel-3B	43 437	1416	56	2018-05-01	2022-10-07
Sentinel-6A	46 984	566	18	2020-11-24	2022-10-14
Jason-1	26 997	4000	119	2001-12-12	2013-06-14
Jason-2	33 105	3946	111	2008-06-24	2019-10-05
Jason-3	41 240	2315	43	2016-01-20	2022-10-11
SARAL	39 086	3165	62	2013-02-28	2022-09-22
CryoSat-2	36 508	4243	168	2010-04-16	2022-10-06
Haiyang-2A	37 781	3025	58	2011-09-29	2020-06-10
TOPEX	22 076	4156	43	1992-08-18	2004-11-18

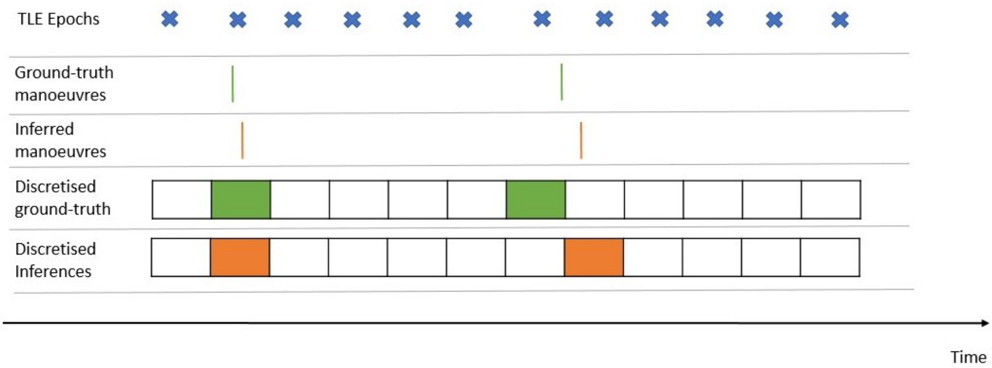
github.com/dpshorten/TLE_observation_benchmark_dataset



github.com/dpshorten/TLE_observation_benchmark_dataset

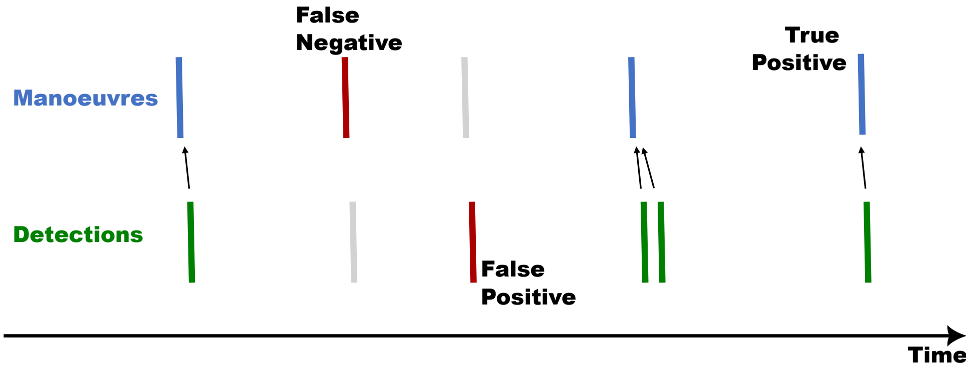
Manoeuvre Detection Techniques

Evaluation (simple)



Note that detection doesn't point to the exact time of the event

Evaluation (improved)

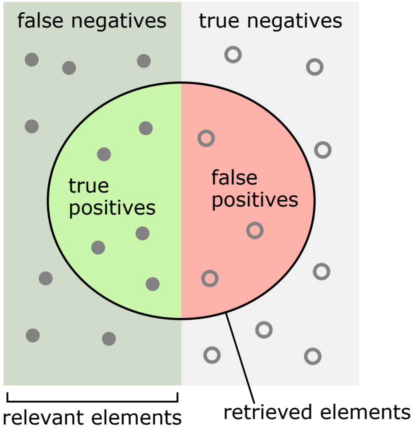


See Zhao [5] and note relative rarity of events

Threshold choice

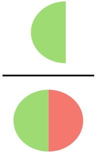
- Every technique returns an anomaly statistic (s_k) for each inspected epoch k.
  - Higher values are more indicative of an anomaly/manoeuvre
- Need to choose a threshold above which we predict a manoeuvre
- Try all possible thresholds
  - Each will have an associated precision and recall
  - Choose the tradeoff
  - Plot curve for testing
  - Note the idea of a "true negative" is a little fuzzy here because there isn't a 1:1 mapping from "non-event" to "non-detection", so conventional ROC curves don't work, but we can do a precision-recall curve
- We can preserve the values to prioritise

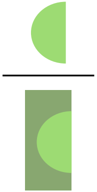
Precision and Recall



How many retrieved items are relevant?

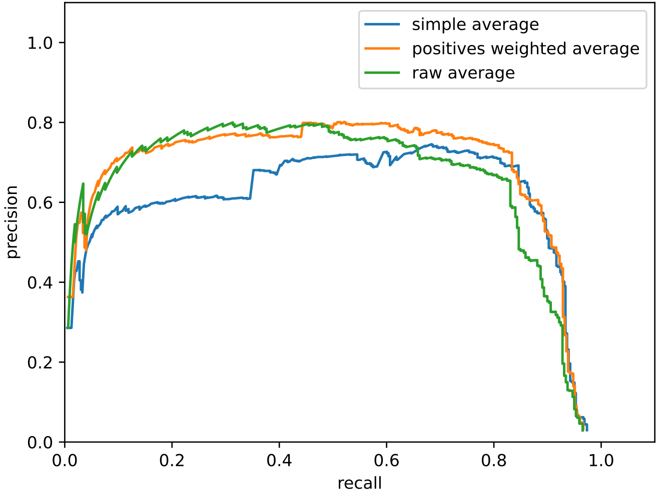
How many relevant items are retrieved?

Precision = 

Recall = 

0.0 is bad, 1.0 is good

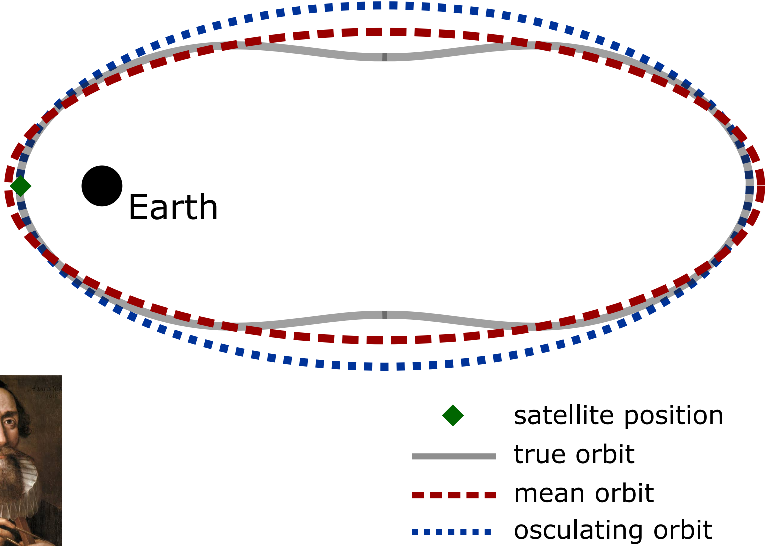
Example precision-recall curve



Baseline Method

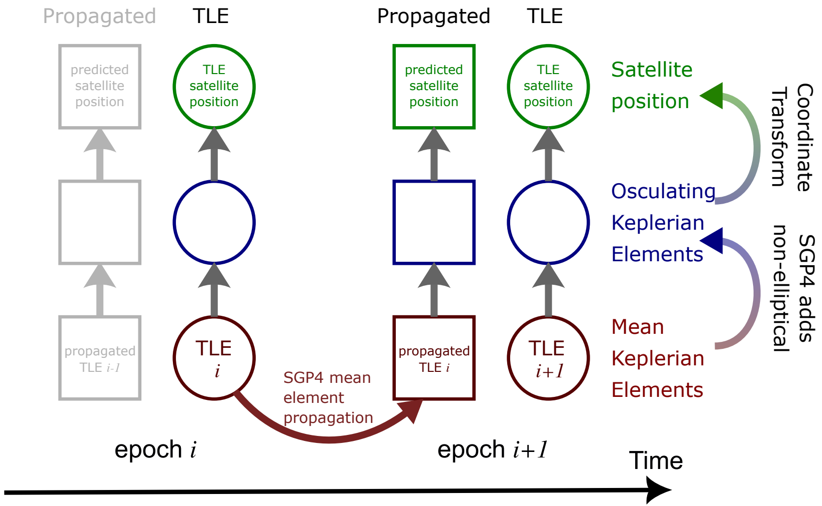
- Strawman from the literature
  - improved (see next slides)
- Propagate to the next epoch using SGP4
- Compare with the observed TLE at the next epoch
  - remember many orbits in 1 day
  - TLEs contain (unknown, non-Gaussian) noise
  - care needed between mean and osculating orbit

Care Needed with TLEs



Johannes Kepler (1571–1630)

Care Needed



Propagated
TLE
Propagated
TLE
predicted satellite position
TLE satellite position
predicted satellite position
TLE satellite position
Satellite position
Coordinate Transform
Osculating Keplerian Elements
SGP4 adds non-elliptical
propagated TLE i-1
TLE i
SGP4 mean element propagation
propagated TLE i
TLE $i+1$
Mean Keplerian Elements
epoch i
epoch $i+1$
Time

Particle Filters [3]

- Assume that the TLEs are noisy observations of true mean elements
- Filtering aims to find these true mean elements
    - along with an uncertainty
    - can be non-linear, non-Gaussian
- From our belief in one epoch, can find distribution of expected orbit in next epoch
    - track a swarm of particles
- If observation is sufficiently unlikely, according to this distribution, designate an anomaly

Particle Filters

Track a swarm of particles each following:

$$\boldsymbol{x}_k = f_{\mathsf{SGP4}}\left(\boldsymbol{x}_{k-1}, t_{k-1}, t_k\right) + \boldsymbol{\xi} \qquad (1)$$

where $\boldsymbol{\xi} \sim \mathcal{N}(0, \boldsymbol{Q})$ and t_k is the epoch time associated with the k^{th} TLE. $\boldsymbol{x}_k$ is the vector of underlying mean orbital elements and we observe

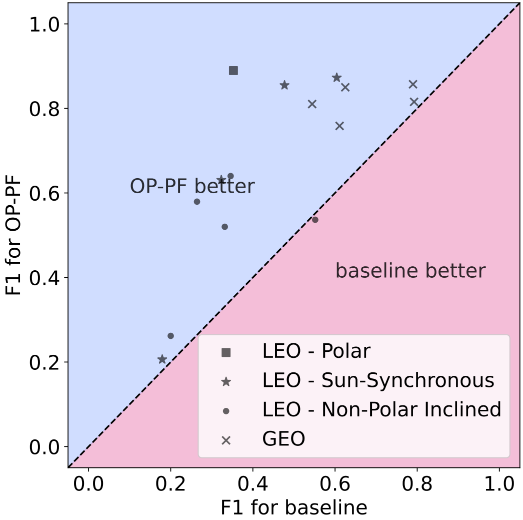
$$\boldsymbol{y}_k = \boldsymbol{x}_k + \boldsymbol{\eta_k} \qquad (2)$$

where $\boldsymbol{\eta_k} \sim \mathcal{N}(0, \boldsymbol{R})$

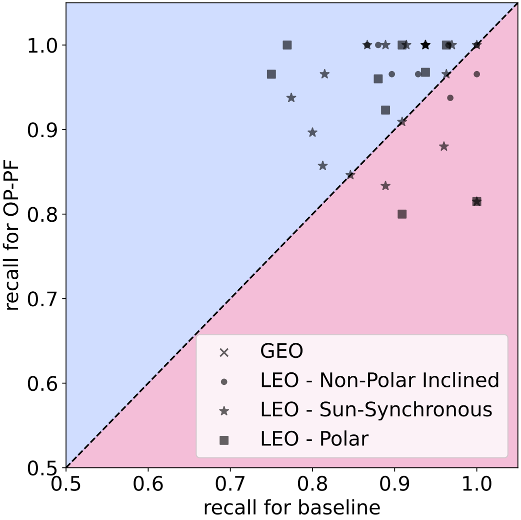
Uncertainty

- SGP4 has substantial errors in the argument of perigee and mean anomaly.
  - Although their sum is more accurate
- Much focus was therefore applied to finding appropriate uncertainty covariance matrices.
  - *e.g.* degenerate parameters

Some Results (benchmark)



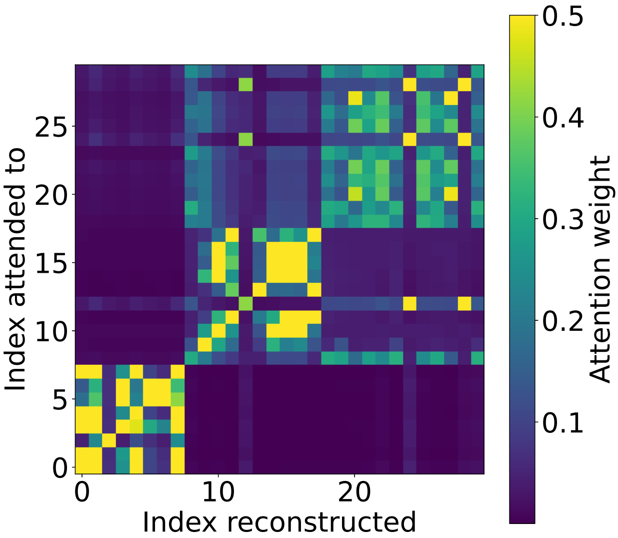
Some Results (simulated)



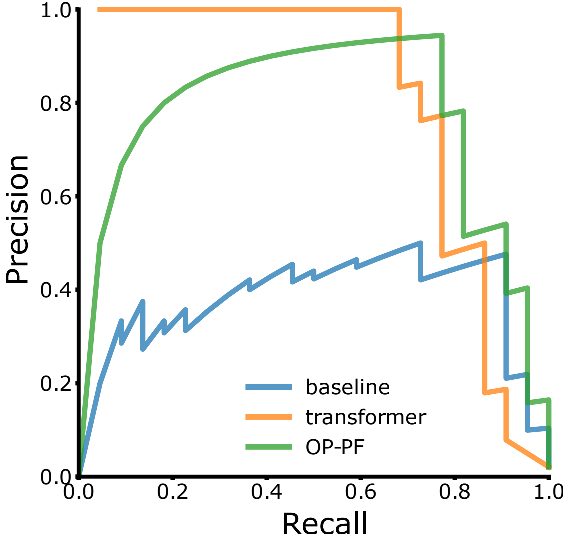
Transformers

- Train a neural network to perform reconstruction of windows of TLEs
- Lose information through initial dimensionality reduction
- Must use neighbouring epochs to reconstruct the TLE at a given epoch
- Through the attention mechanism we can track which TLEs are used
- Expect an interruption at manoeuvres

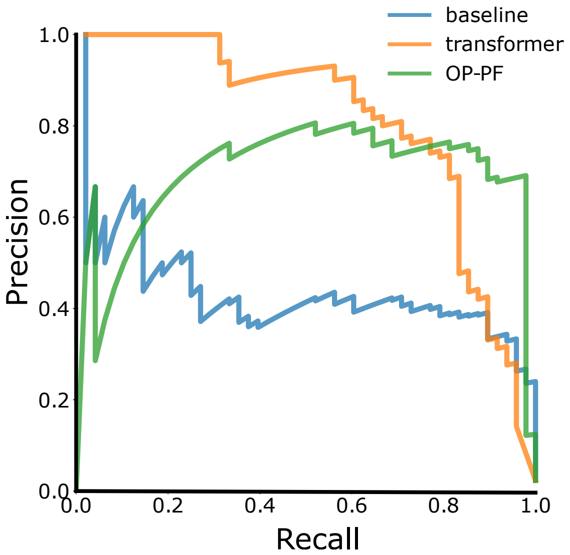
Transformer



Transformer Results: Fengyun-2D



Transformer Results: Fengyun-2E



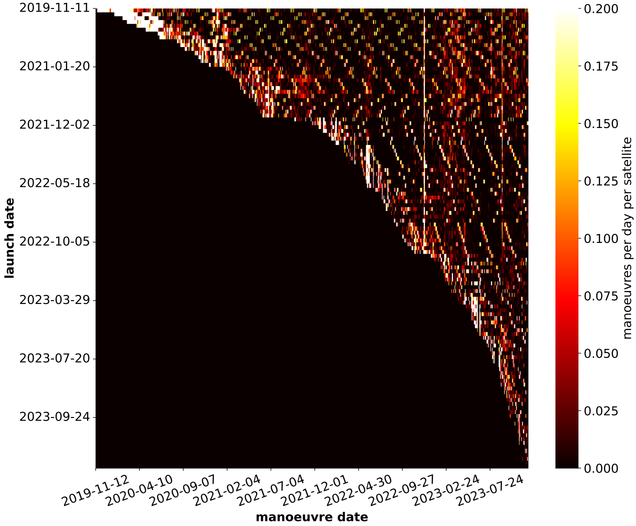
Manoeuvres of Starlink Satellites

Starlink Constellation [2]

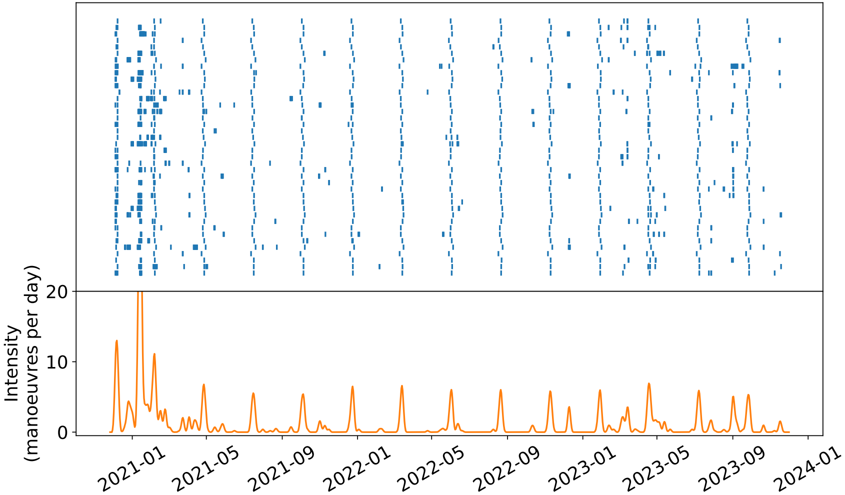
So let's use the detector to look for patterns

- Around 5,000 satellites in StarLink (at the time of writing)
- 7 Orbital shells
    - Within shell: same altitude and inclination
- Each shell can be split into launch groups

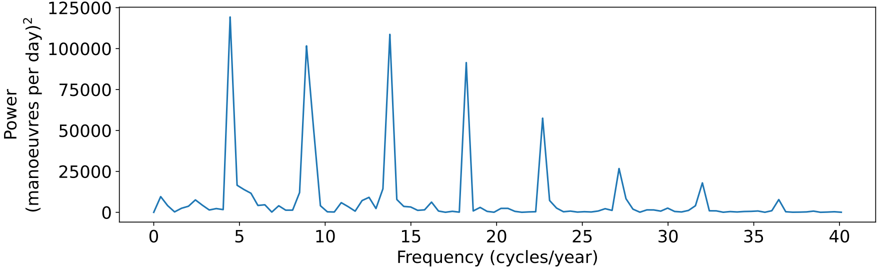
Overview of Manoeuvres



Launch Group 15 of Shell 1

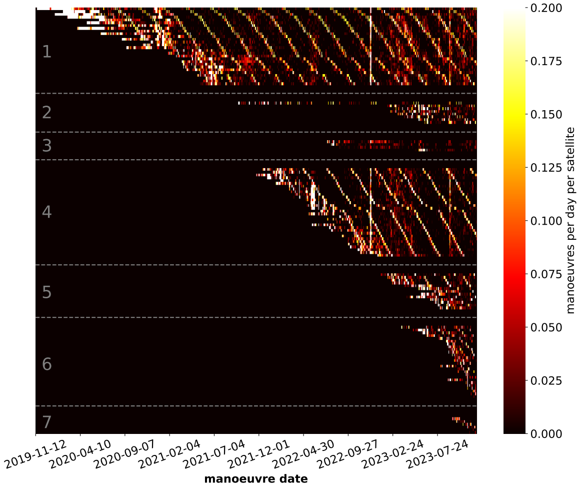


Launch Group 15 of Shell 1 (FFT)

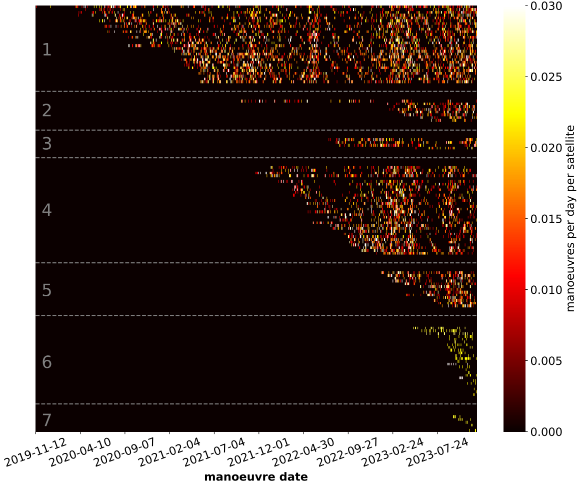


Fundamental frequency at around 4.5 cycles per year, representing an interval of around 80 days between station-keeping manoeuvres.

Re-ordered launch groups

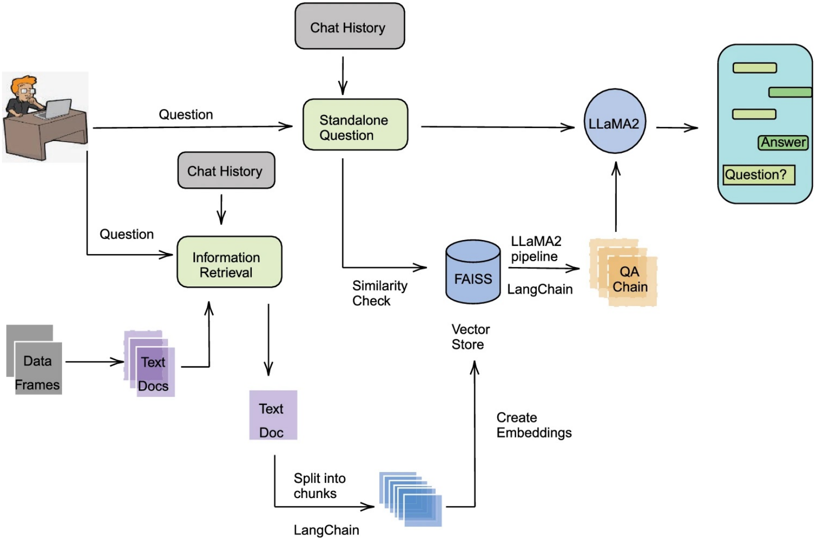


Re-ordered launch groups (station-keeping removed)



Web App

Architecture [1]



Demo

Star

Matthew Roughan <matthew.roughan@adelaide.edu.au>

Conclution

Concluding Remarks

- Good detection of events
  - preferance for particle filters
  - more explainable than transformers, but for some settings transformers are better
- Chat interface demonstrates a more general idea, but is still very much a prototype
- We are continuing this work
  - Funding news soon
  - Detection could be improved, but other topics are more urgent, *e.g.*, working with other sources of data and event classification
  - Threshold determination in practice

Thanks!!

References

W. Karunarathne, D.P. Shorten, and M. Roughan.
An intelligent conversational agent for querying satellite manoeuvre detections: a case study.
Discover Artificial Intelligence, 2025.

David Shorten, Wathsala Karunarathne, and Matthew Roughan.
How is Starlink manoeuvring? an analysis of patterns in the manoeuvres of Starlink satellites.
In *Proceedings of the 9th International Conference on Internet of Things, Big Data and Security - Volume 1: IoTBDS*, pages 174–184. INSTICC, SciTePress, 2024.

David P. Shorten, Melissa Humphries, John Maclean, Yang Yang, and Matthew Roughan.
Optimal proposal particle filters for detecting anomalies and manoeuvres from two line element data.
Acta Astronautica, 228:709–723, 2025.

David P Shorten, Yang Yang, John Maclean, and Matthew Roughan.
Wide-scale monitoring of satellite lifetimes: Pitfalls and a benchmark dataset.
Journal of Spacecraft and Rockets, pages 1–5, 2023.

Liang Zhao.
Event prediction in the big data era: A systematic survey.
ACM Computing Surveys (CSUR), 54(5):1–37, 2021.

Extra Slides

TLE

Originally designed for use with 80-column punched cards

```
1 27651U 03004A 07269.09107561 .00000015 00000-0 17636-4 0 4191
2 27651 039.9956 188.8112 0026975 282.9289 076.8483 14.81973121252789
```

Orbital Elements

They have weird names

- **Eccentricity:** how circular or elliptical (0.0 – 1.0)
- **Semi-major axis:** long axis of ellipse (half the distance between the apoapsis and periapsis) (
- **Inclination:** vertical tilt of the orbital plane with respect to the equator (0.0 – 90.0 degrees)
- **Longitude of the ascending node:** the direction the ellipse points.
- **Argument of periapsis:** rotation of ellipse around its plane relative to reference (perigee for Earth)
- **Mean motion:** average angular speed
- **Epoch:** time
- **Mean anomaly:** where we are on the orbit

TLEs use these, but there are alternative parameterisations.

F1 Score

F1 score is a way to combine prevision and recall into one number

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

With

$$\text{precision} = TP/(TP + FP)$$
$$\text{recall} = TP/(TP + FN)$$