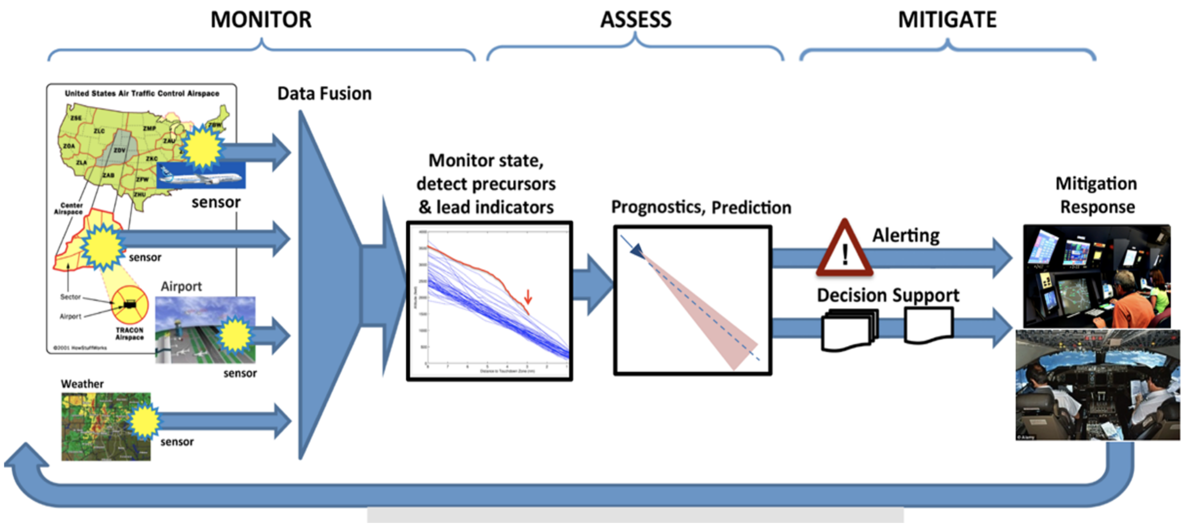
Machine Learning for Aviation Safety at NASA

Nikunj C. Oza, Ph.D.
Leader, Data Sciences Group
NASA Ames Research Center
nikunj.c.oza@nasa.gov

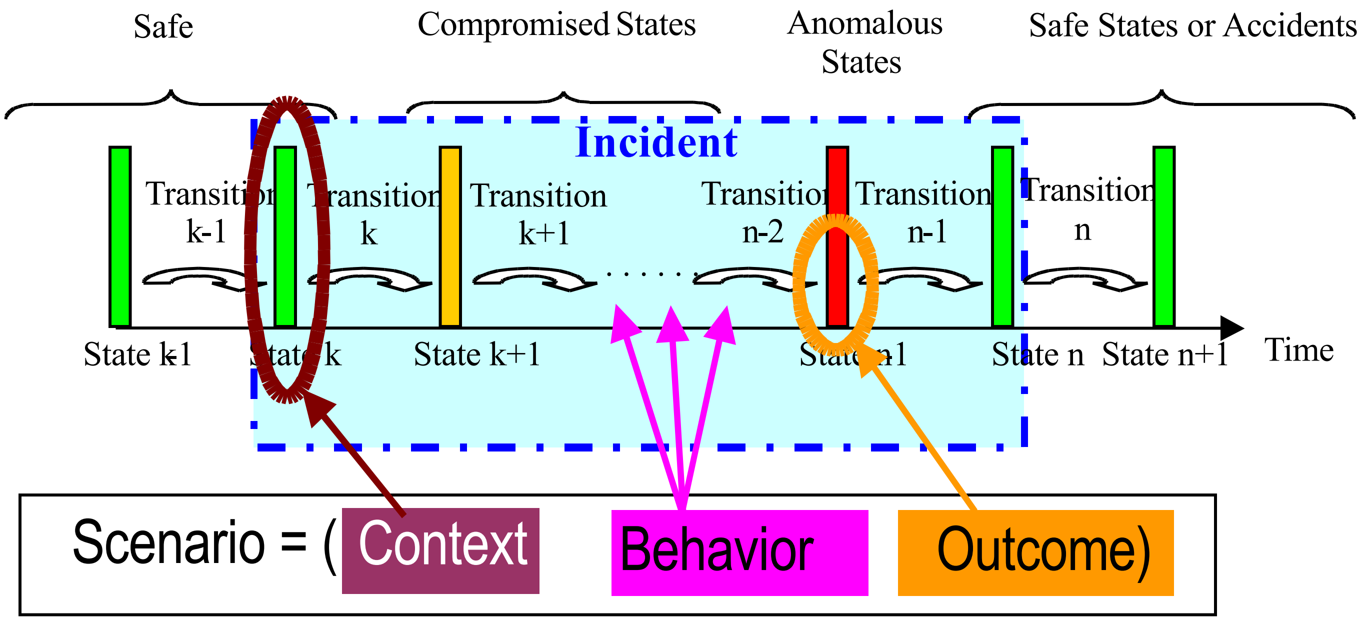
NASA System-Wide Safety Project

Technical Challenge 1: Integrated Risk Assessment for the Terminal Area

- Develop tools to collect and analyze relevant data, identify incidents
- Integrate tools into dashboards for risk assessment



The Anatomy of an Aviation Safety Incident



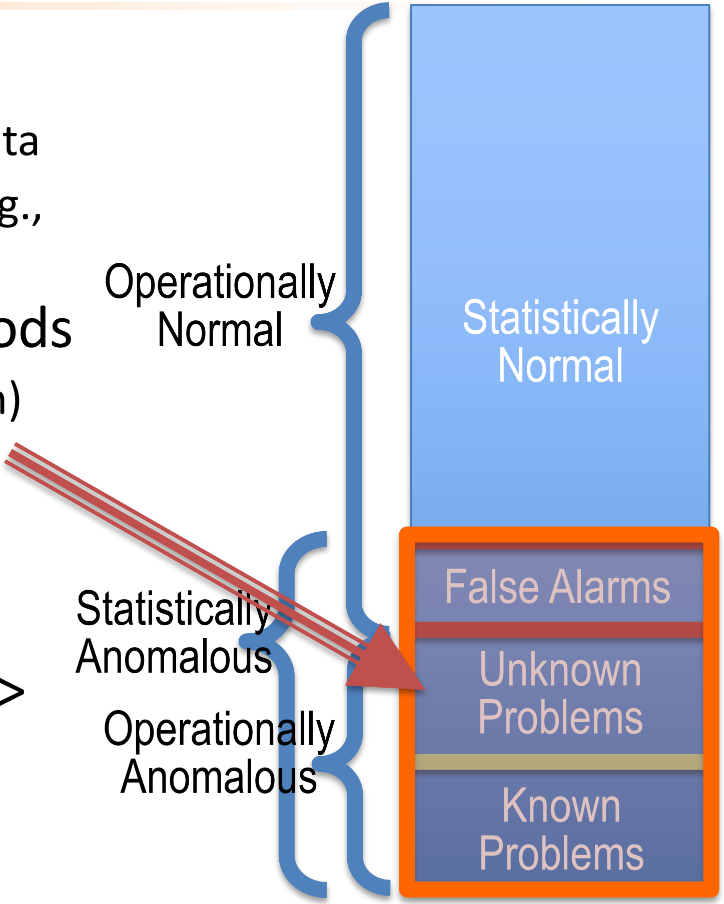
From Irving Statler, Aviation Safety Monitoring and Modeling Project

Range of Problems

- Types of data
  - Flight Operations Quality Assurance (FOQA) / FDM
  - Trajectory/Radar Track Data
  - Safety Reports
  - Weather
  - Others
- Types of problems
  - Anomaly Detection
  - Precursor Identification
  - Text: Classification, topic identification

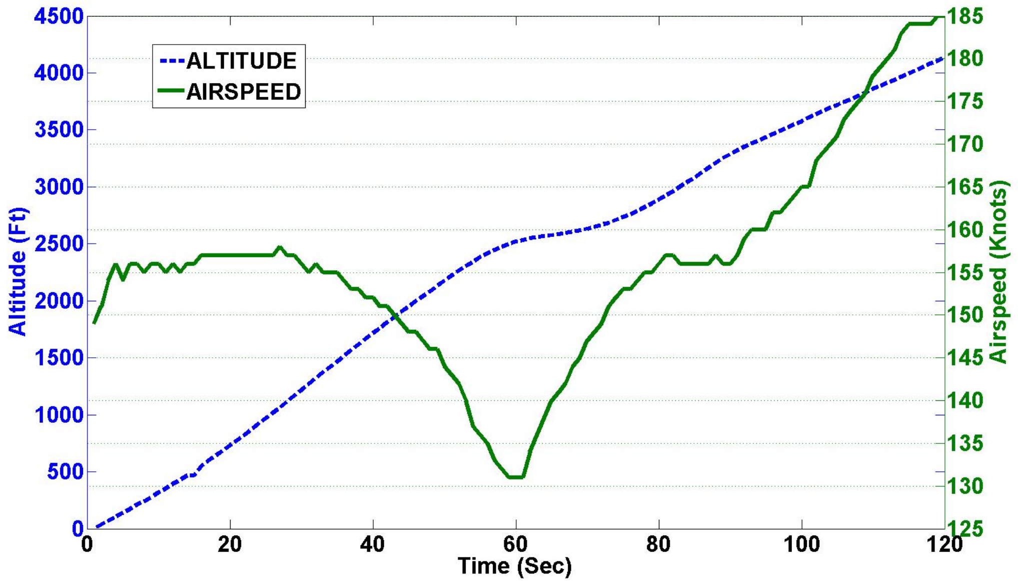
Data-Driven Methods

- DISCOVER anomalies by
  - learning statistical properties of the data
  - finding which data points do not fit (e.g., far away, low probability)
- Complementary to existing methods
  - Lower false negative (missed detection) rate
  - Higher false positive rate (identified points/flights unusual, but not always operationally significant)
- Data-driven methods -> insights -> modification of exceedance detection



Drop in airspeed

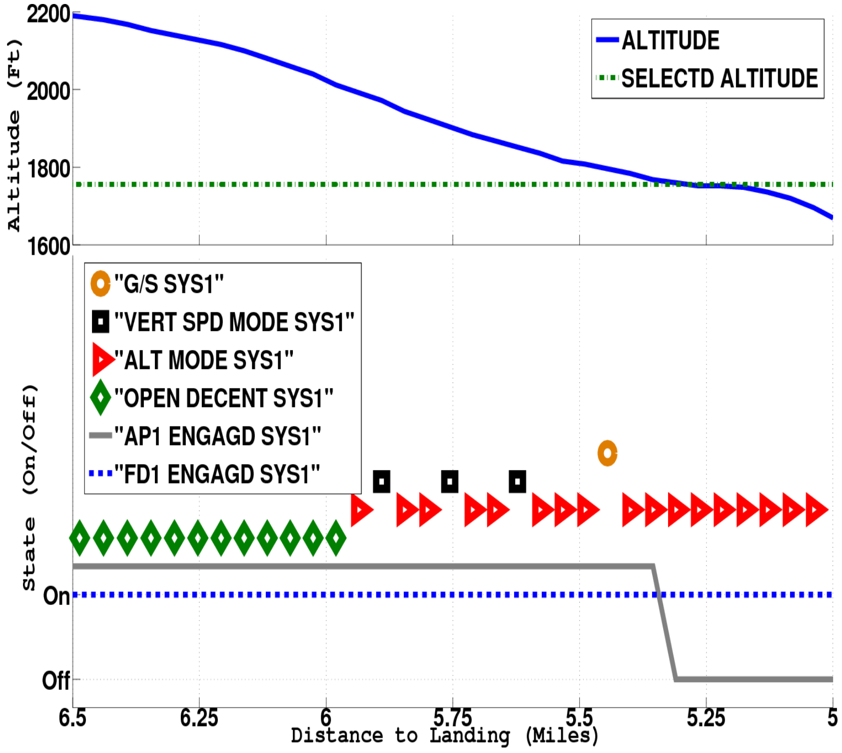
30 seconds after takeoff, drop in airspeed to 12 knots above estimated stall speed. Drop in airspeed continued for 30 seconds.



Mode Confusion

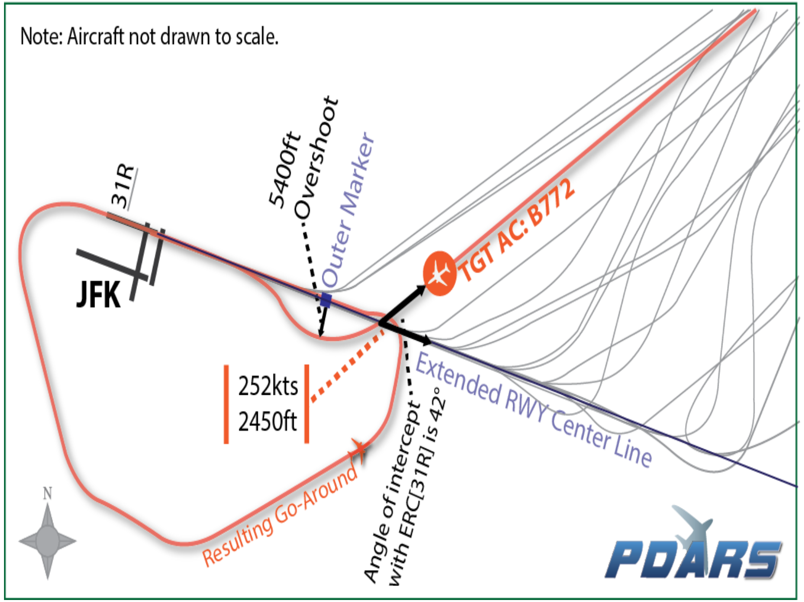
- Unusual switching between
  - Vertical Speed Mode
  - Altitude Hold Mode

- Results in recycling flight director to fix the conflict



High Speed Go-Around

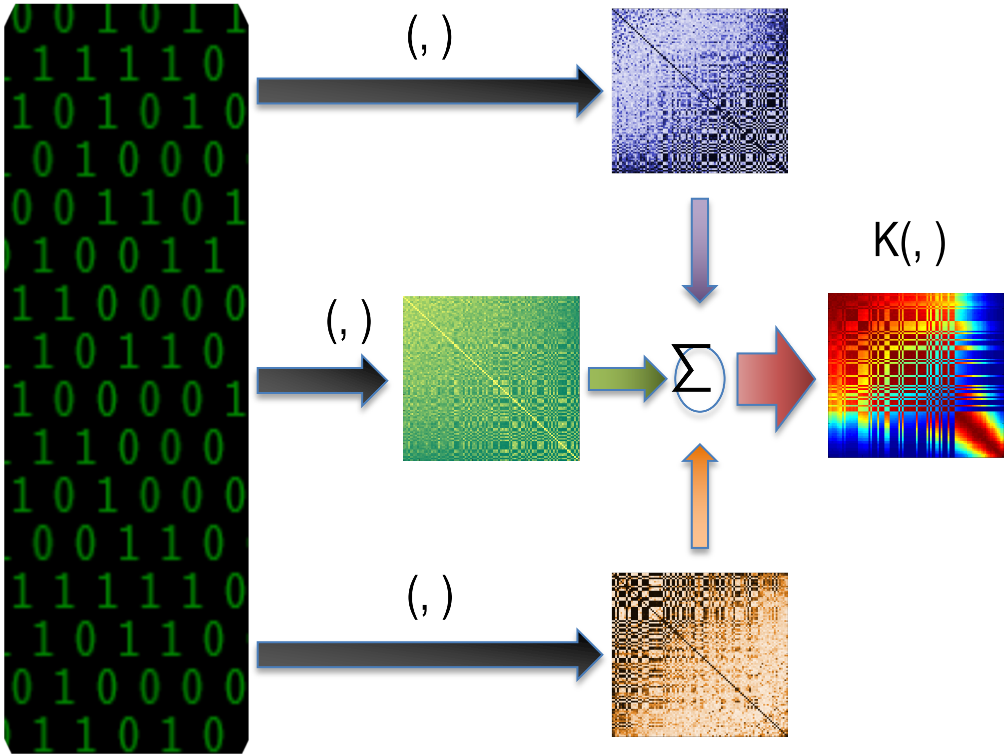
- Overshoots Extended Runway Centerline (ERC) by over 1 SM
- Over 250 Kts @2500 Ft.
- Angle of intercept > 40°
- Overshoots 2nd approach



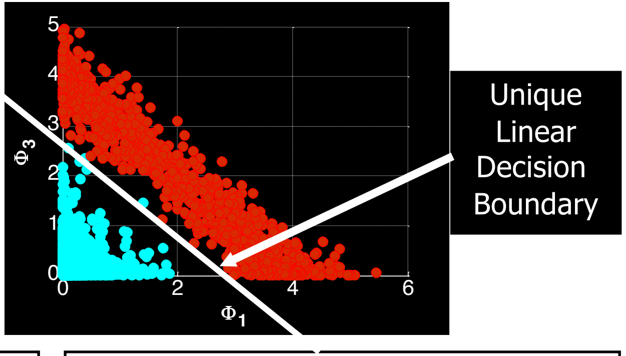
Multiple Kernel Anomaly Detection (MKAD)

Advantages:

- Data can be heterogeneous
- Kernels can be combined and tuned for different data types.
- Take advantage of off-the shelf
anomaly detection algorithms like One-class Support Vector Machines.



Optimization problem



Discrete kernel

Continuous kernel

Text kernel

$$\text{Minimize } Q = \frac{1}{2}\sum_{i=1,j=1}^{l} \alpha_i\alpha_j \left(\beta_d K_d(x_i,x_j) + \beta_c K_c(x_i,x_j) + \beta_t K_t(x_i,x_j)\right)$$

$$\text{subject to } 0 \le \alpha_i \le \frac{1}{l\nu}, \nu \in [0,1], \sum_{i=1}^{l}\alpha_i = 1$$

$$\boxed{h(\alpha,\beta,f_z,\rho) = \sum_i \alpha_i \left(\sum_\lambda \beta_\lambda K_{i,z}^{\lambda}\right) - \rho}$$

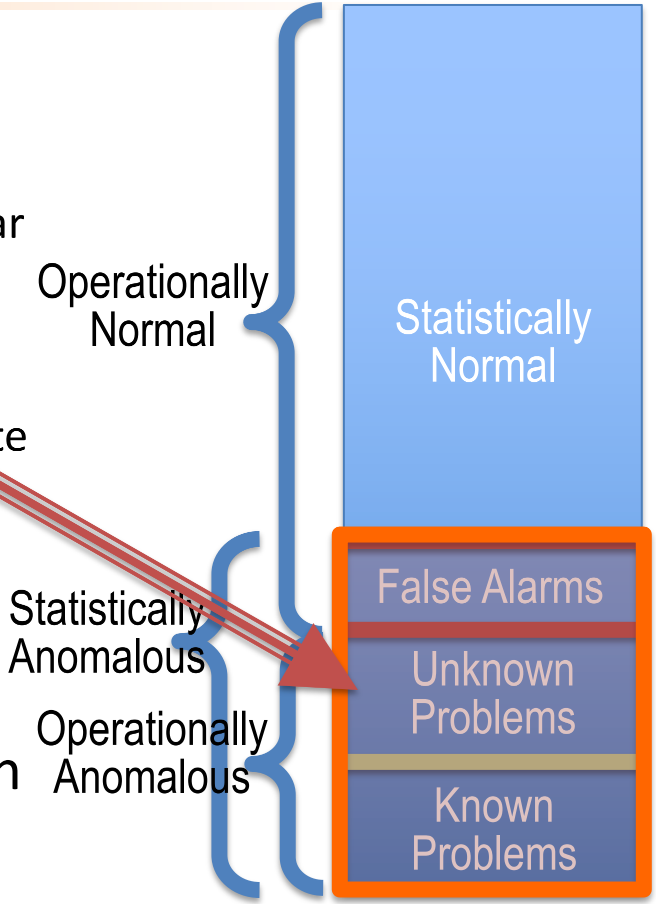
Discrete Kernel: normalized Longest Common Subsequence (nLCS)

Continuous Kernel: nLCS of SAX representations

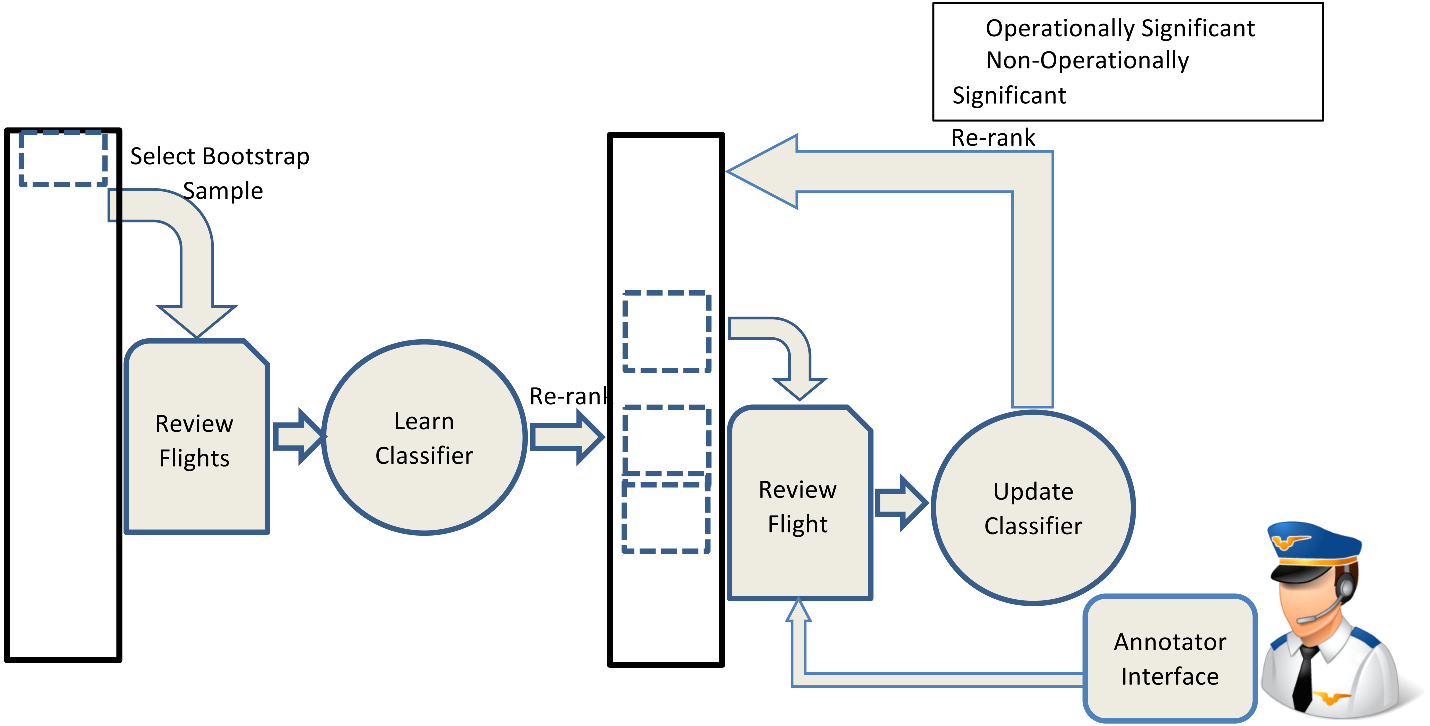
Text Kernel: Euclidian distances between projections onto LDA topic vectors

Data-Driven Methods

- DISCOVER anomalies by
  - learning statistical properties of the data
  - finding which data points do not fit (e.g., far away, low probability)
- Complementary to existing methods
  - Lower false negative (missed detection) rate
  - Higher false positive rate (identified points/flights unusual, but not always operationally significant)
- Data-driven methods -> insights -> modification of exceedance detection

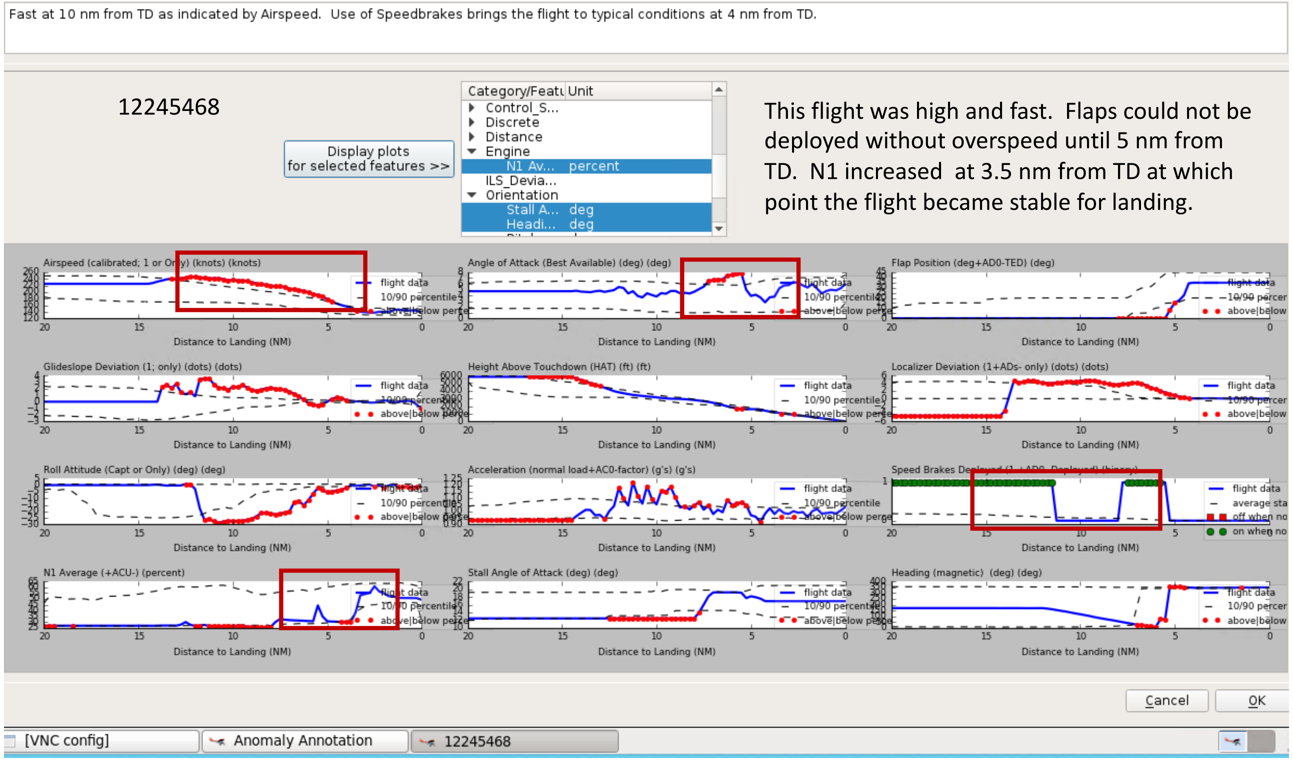


Active Learning Approach

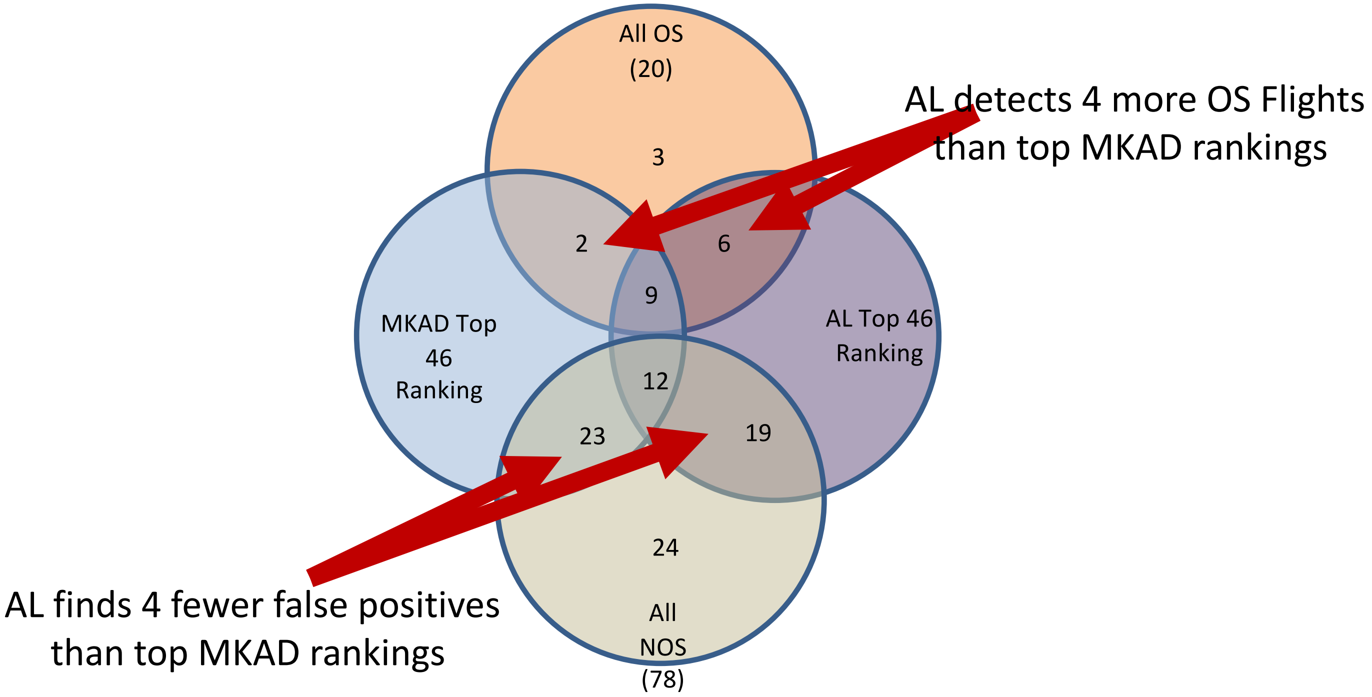


Screen Shot of Annotator

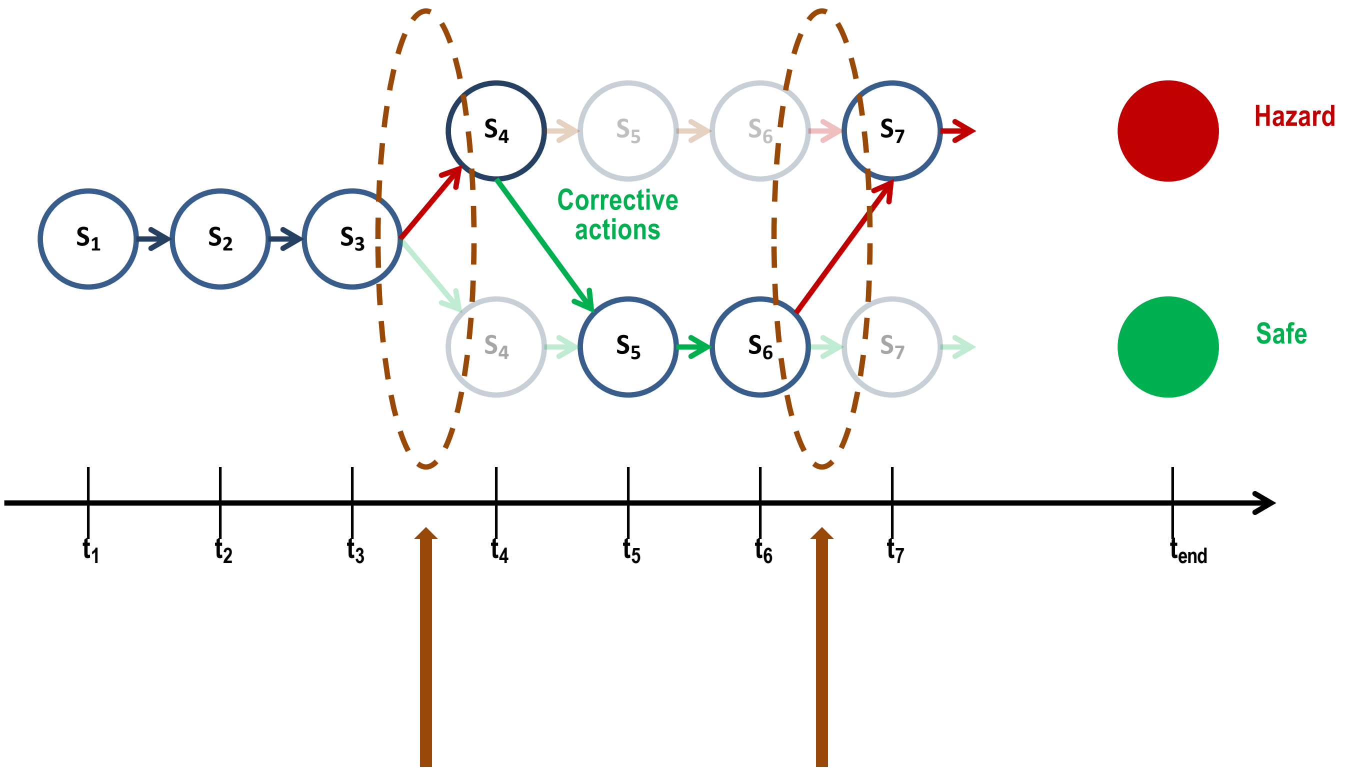
(MKAD Rank 19) Fast at 10 nm from TD as indicated by Airspeed. Use of Speedbrakes brings the flight to typical conditions at 4 nm from TD.



OS/NOS Flight Overlap Between Methods

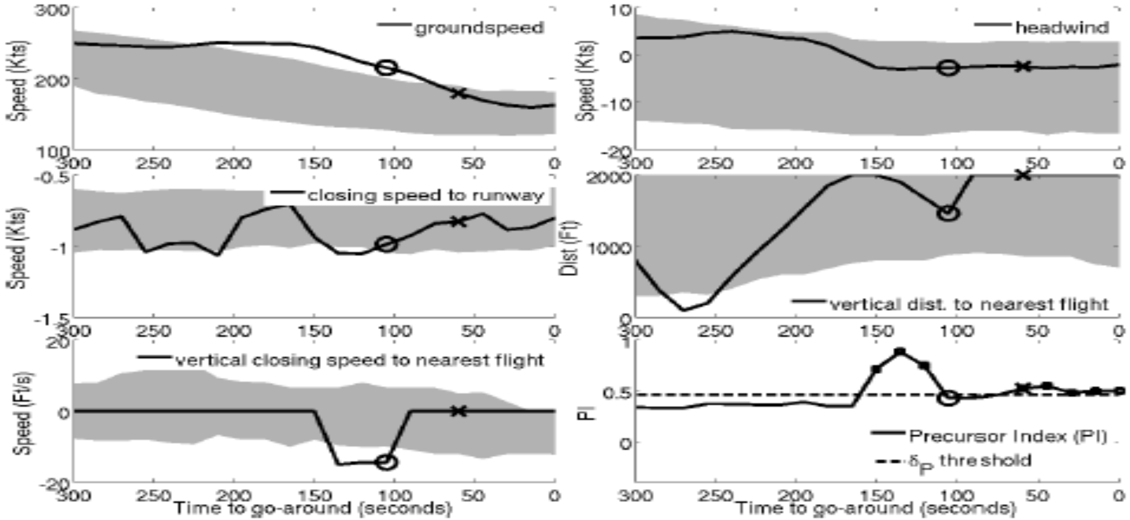


Precursors



S_1
S_2
S_3
S_4
S_5
S_6
S_7
S_4
S_5
S_6
S_7
Corrective actions
Hazard
Safe
t_1
t_2
t_3
t_4
t_5
t_6
t_7
t_{end}

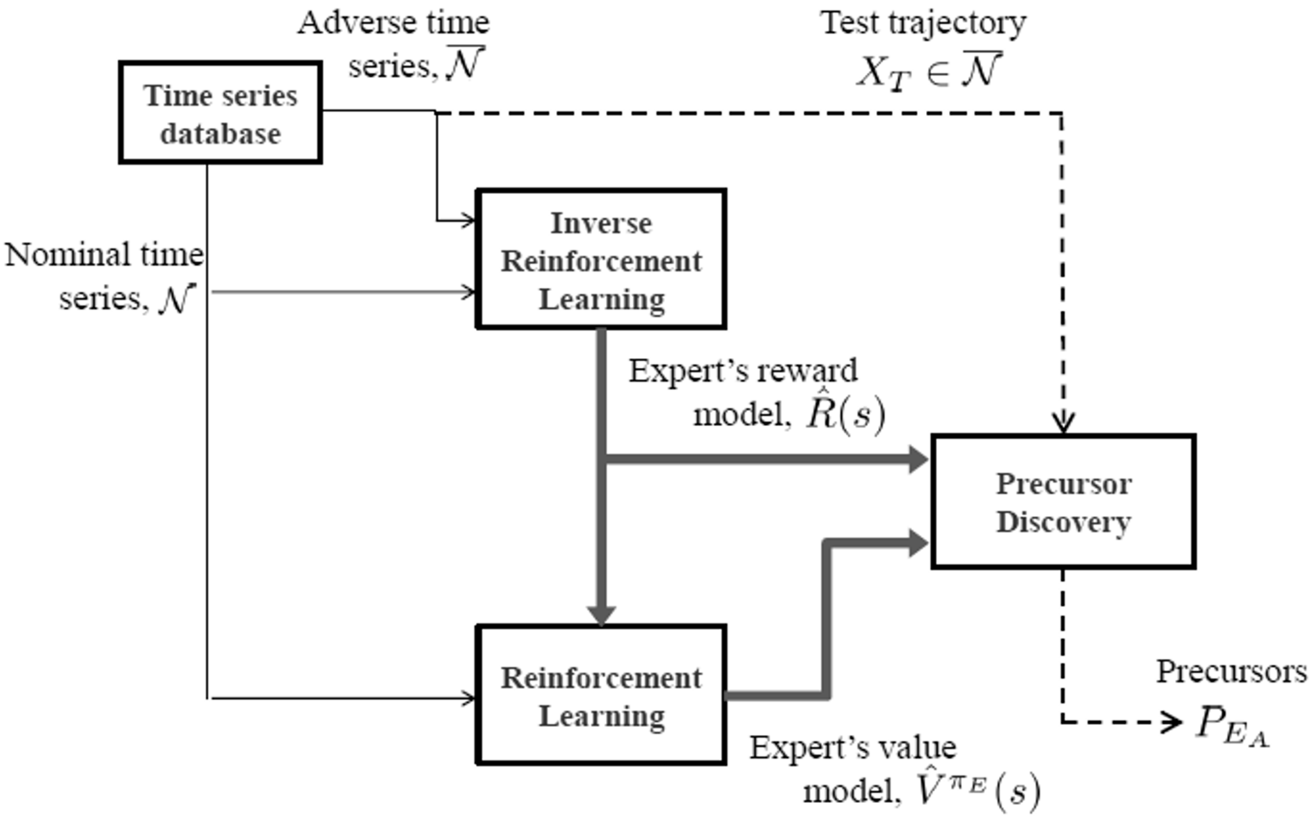
Precursor to go-around—Energy Mismanagement



Upper left plot shows that ground speed is higher than normal. This is responsible for precursor index (lower right) being high, indicating high probability of go-around.

Algorithm automatically identified precursor, given go-around condition as target effect.

Automatic Discovery Of Precursors in Time series (ADOPT)



Upcoming Work

- Anomaly Detection / Active Learning
  - Further testing of algorithm and user interface
  - Incorporate multiple anomaly detection algorithms, active learning strategies, users
- Precursor Identification
  - Ongoing work on autoencoder based algorithm
- Text
  - Sustained effort!
  - Aviation Safety Reporting System: Use unredacted reports
  - Non-traditional data sources: pilot blogs, others

Thank You!

Contact: **nikunj.c.oza@nasa.gov**

Papers

[1] V. Janakiraman, Explaining Aviation Safety Incidents Using Deep Temporal Multiple Instance Learning, *KDD 2018.*

[2] K. Das, I. Avrekh, B. Matthews, M. Sharma, and N. Oza, ASK-the-Expert: Active Learning Based Knowledge Discovery Using the Expert, *ECML-PKDD* 2017.

[3] V. Janakiraman, B. Matthews, and N. Oza, Discovery of Precursors to Adverse Events Using Time Series Data, *SDM* 2016.

[4] B. Matthews, D. Nielsen, J. Schade, K. Chan, and M. Kiniry, Comparative Study of Metroplex Airspace and Procedures Using Machine Learning to Discover Flight Track Anomalies, 34th *DASC*, 2015.

[5] S. Das, B. Matthews, N. Oza, and A. Srivastava, Multiple Kernel Learning for Heterogeneous Anomaly Detection: Algorithm and Aviation Safety Case Study, *KDD* 2010.