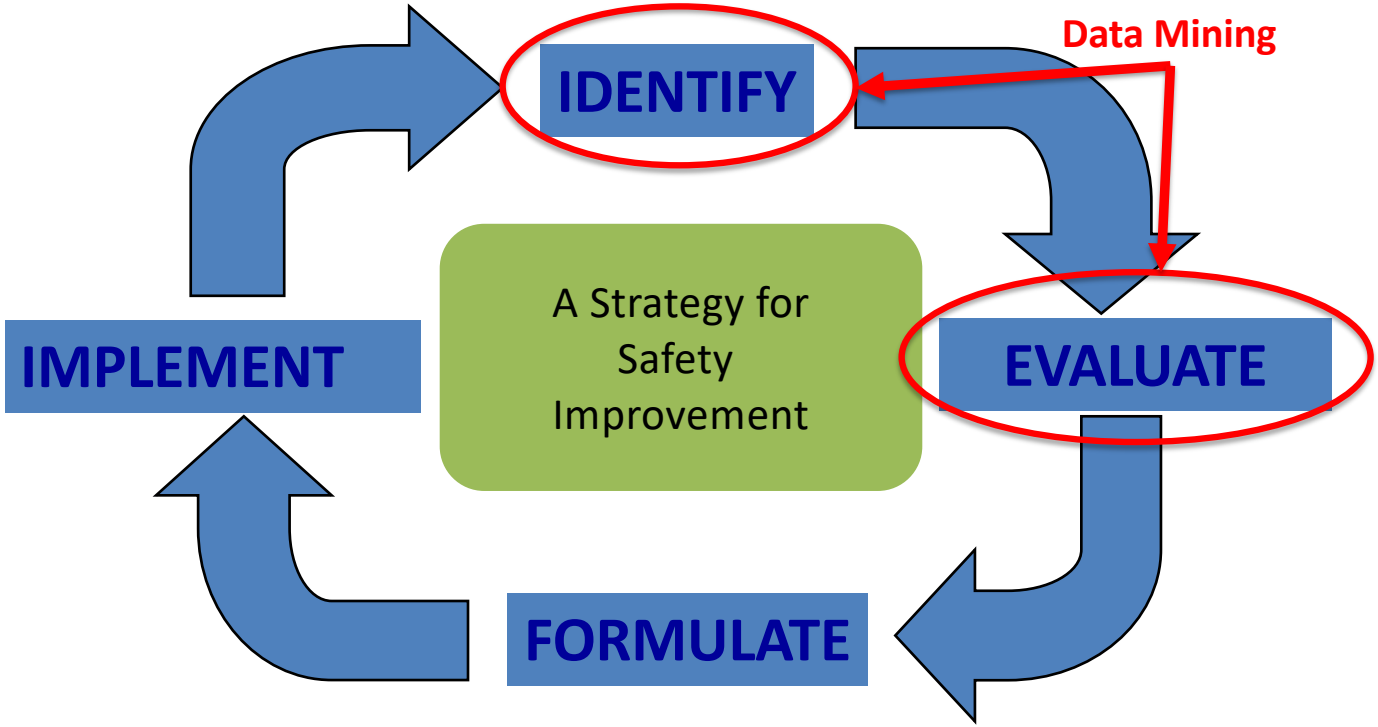
Automatic Anomaly Detection With Machine Learning

Bryan Matthews (KBR)
NASA Ames Research Center
World Aviation Festival
2019 September 4th

Understanding the Benefits of Automatic Anomaly Detection

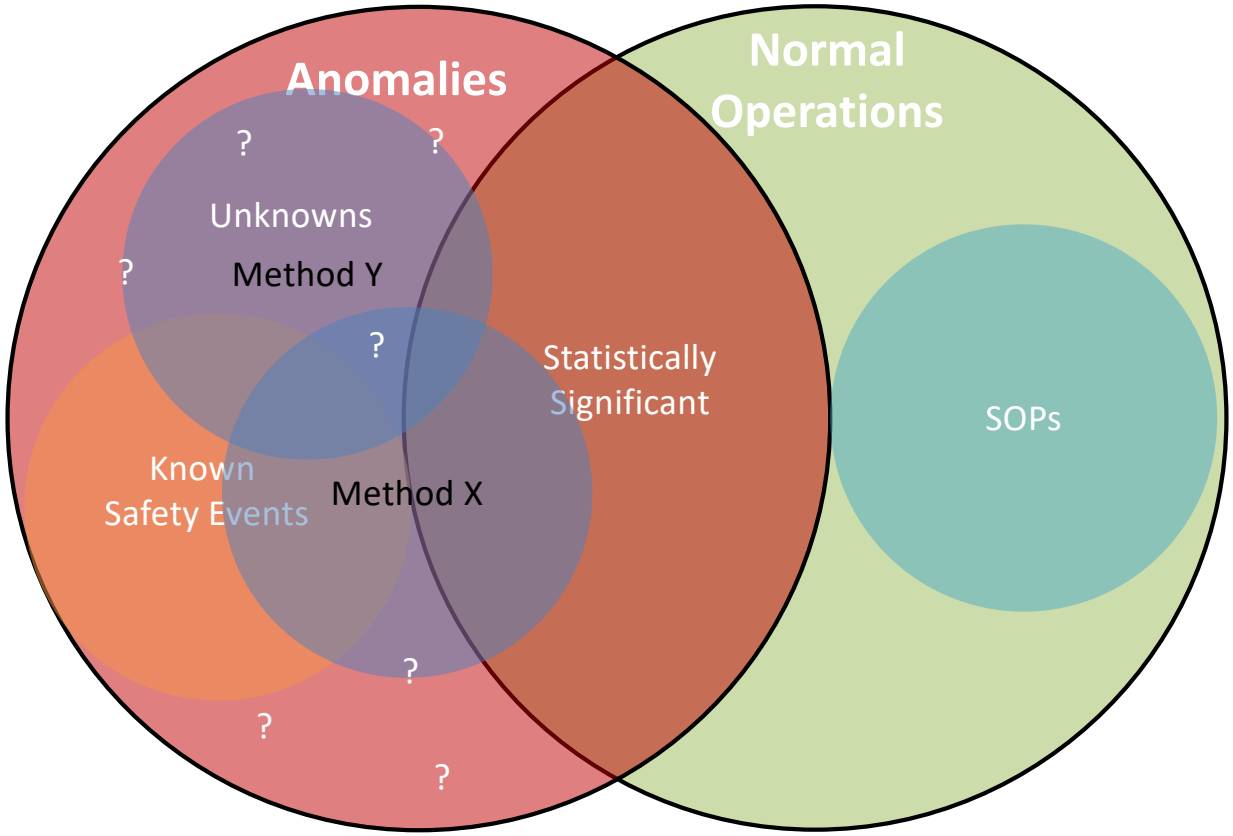
- **Benefits to Safety**
  - Find the unknown unknowns:
    - Catch operationally significant events that may not trigger safety thresholds but could hold high risk safety implications.
  - Capture differences between work imagined vs work performed:
    - Formalize common good practices into procedures.
- **Benefits to Efficiency**
  - Reduce delays:
    - Identify flaws in schedules or operations that result in delays.
  - Save fuel:
    - Identify anomalies in maintenance, flying techniques, OPD adherence, etc. that result in higher fuel usage.

Safety Life Cycle



Reduce the accident/incident rate by identifying and responding to precursor events *before* accidents/incidents occur.

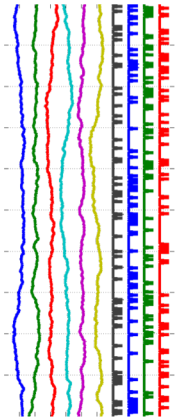
Operational Significance Vs Statistically Significant

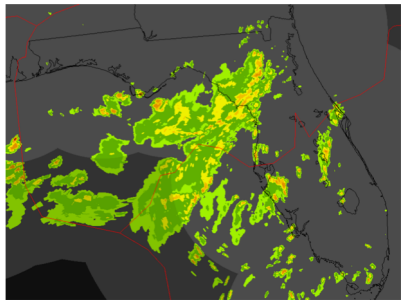


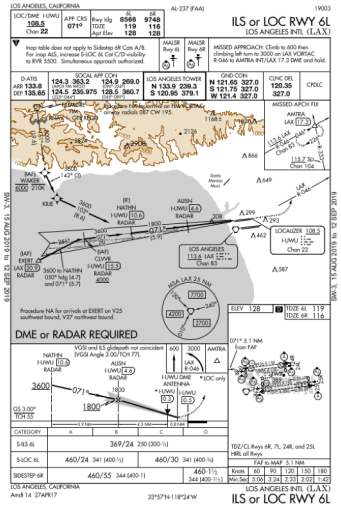
Overcoming the Challenges of Data Integration and Automation

- **Leveraging and fusing heterogeneous data**
  - Flight Operational Quality Assurance Data (FOQA)
    - Aircraft centric view (usually de-identified)
  - Weather (current/forecast)
    - Identify levels of severity that impact traveled routes and planning.
  - Radar Surveillance
    - Captures interactions with other flights
  - Procedural Information
    - Standard Operating Procedures
    - RNAV Arrivals/Departures
  - NOTAMs
    - Changes that may disrupt normal planned operations.
  - Pilot/Controller Safety Reports
    - Captures context and trends of safety incidents.

CIVIL AIRCRAFT OPERATIONS WITHIN OR TRANSITING THE TERRITORIAL AIRSPACE OF THE UNITED STATES (U.S.). THE FOLLOWING SPECIAL SECURITY REQUIREMENTS FOR CIVIL AIRCRAFT OPERATIONS WITHIN OR TRANSITING U.S. TERRITORIAL AIRSPACE ARE IN EFFECT PURSUANT TO 14 CFR 99.7, SPECIAL SECURITY INSTRUCTIONS, AND 49 USC 40103 AND 41703, IN ADDITION TO THE REQUIREMENTS PRESCRIBED IN 14 CFR PART 99, SECURITY CONTROL OF AIR TRAFFIC. THE REQUIREMENTS IN THIS NOTAM SUPERSEDE THE REQUIREMENTS FOR CIVIL AIRCRAFT OPERATIONS WITHIN OR TRANSITING U.S. TERRITORIAL AIRSPACE FORMERLY PUBLISHED IN FDC NOTAMS 3/2735, 3/2768, AND 3/3013. SECURITY REQUIREMENTS FOR CIVIL AIRCRAFT OPERATIONS TO OR FROM U.S. TERRITORIAL AIRSPACE ARE PROVIDED IN A SEPARATE NOTAM. IN ADDITION TO THE REQUIREMENTS OF THIS NOTAM, CIVIL AIRCRAFT OPERATIONS WITHIN OR TRANSITING THE TERRITORIAL AIRSPACE OF THE U.S. MUST ALSO COMPLY WITH ALL OTHER APPLICABLE REGULATIONS PUBLISHED IN TITLE 14, CODE OF FEDERAL REGULATIONS (CFR). PART I. CIVIL AIRCRAFT OPERATIONS WITHIN U.S. TERRITORIAL AIRSPACE A. CIVIL AIRCRAFT WITH A MAXIMUM CERTIFICATED TAKEOFF GROSS WEIGHT LESS THAN OR EQUAL TO 100,309 POUNDS (45,500



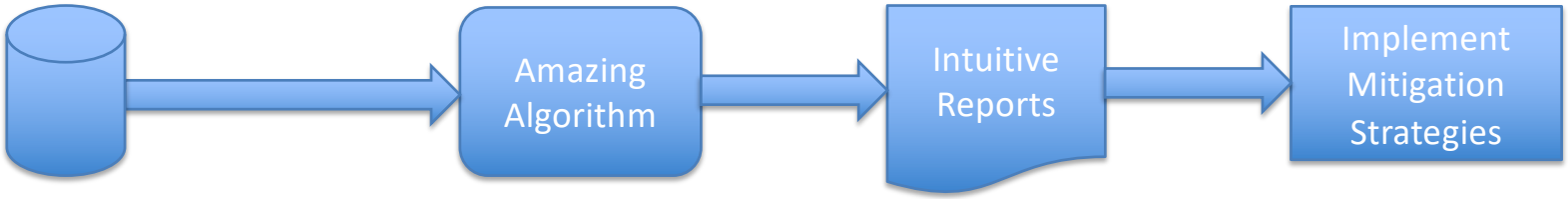




"From a pilot's point of view this arrival could be better, safer, and more user friendly by simply extending the distance from BLUZZ and COPEN intersection. 9.7 miles is simply not enough to lose altitude and slow down. Could it be increased to possibly 20 or 25 miles? The chart states that from COPEN intersection to the airport is 40NM, surely there is enough room to make this possible."

Four V's of Big Data (NASA Ames Sherlock Data Warehouse)

- Volume:
  - Radar Tracks: 47 facilities (1 year)
    - ~423 GB (Compressed)
    - ~3.2 TB (CSV)
  - Weather and Forecast (Entire NAS)
    - CIWS ~2.8 TB
- Velocity
  - Radar Tracks: 47 Facilities
    - ~35 GB/month (compressed).
    - ~268 GB/month (uncompressed)
  - Weather and Forecast (Entire NAS)
    - CIWS ~233 GB/month
- Veracity
  - Data drop outs
  - Duplicate tracks
  - Track ending in mid air
  - Reused flight identifiers
- Variety
  - Numerical (continuous/binary)
  - Weather (forecast/actual)
  - Radar/Airport meta data
  - ATC Voice
  - ASRS text reports (Pilot/Controller)



Additional Challenges

- **Costs**
  - Software/hardware to record and archive historical data.
  - Curating data.
  - Analyzing data and operations.
  - Disseminate proactive safety measures.
- **Sensitive Data**
  - Fear of punitive action.
  - Leaked preliminary analysis may result in loss of public confidence/customers.
- **Research Level Vs Off the Shelf Tools**
  - Off the shelf products:
    - Can provide quick actionable insight.
    - Limited in scope of what they can identify.
  - Research products:
    - Have the potential to find emerging trends
    - Requires investment by the stakeholder in both time and resources to validate findings

Identifying Precursors with Machine Learning

When We Know About a Problem…
How Do We Discover Leading Factors?

Who needs explanations?

- **Airlines**
  - Proactively improve fleet safety
  - Predictive maintenance
  - Improve pilot training

- **Federal Aviation Administration/Eurocontrol**
  - Better airspace design
  - Refine standard operating procedures
  - Hazard identification

EUROCONTROL

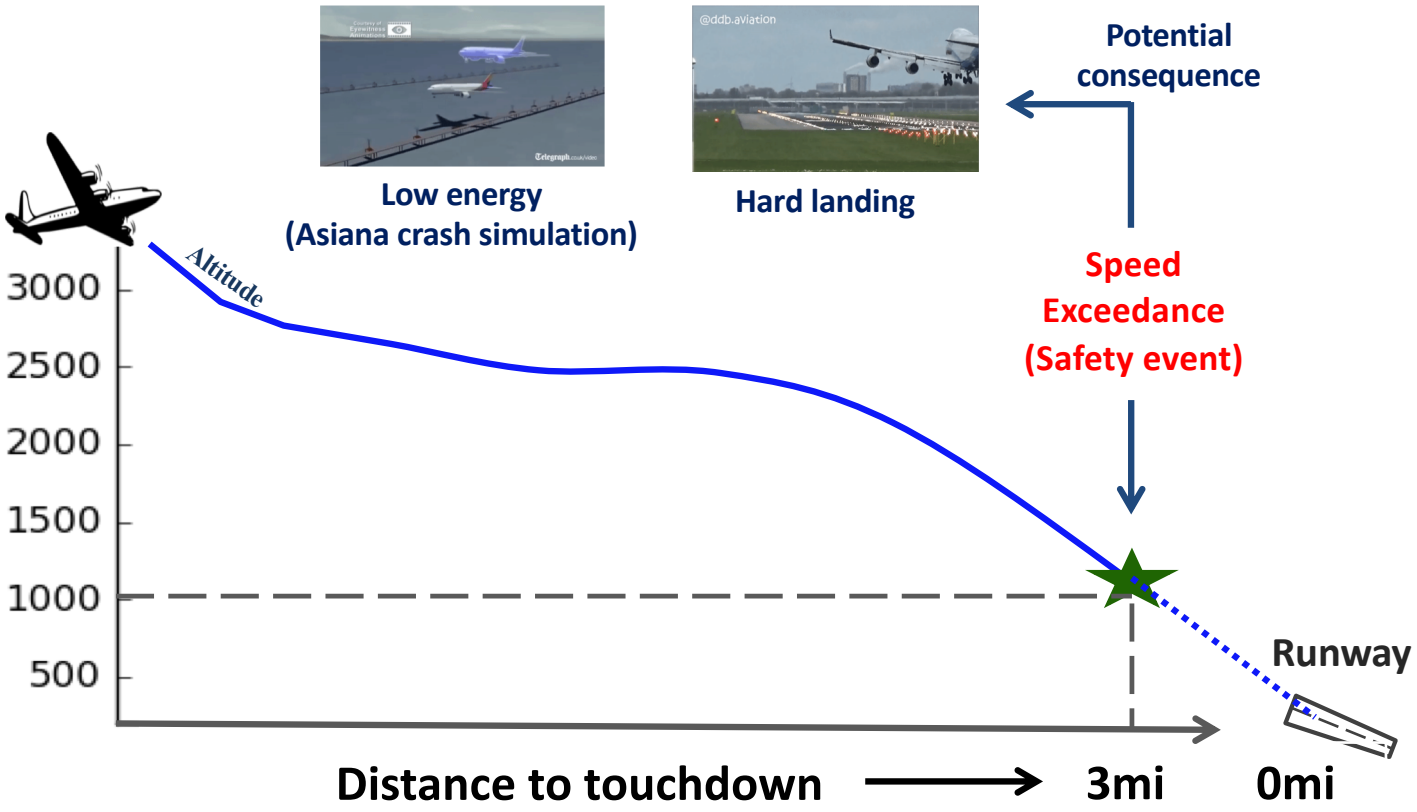
Proactive Safety

- When do degraded states begin to appear?
- What are the degraded states?
- Are there corrective actions?
- What is the likelihood?

Precursor Definition:

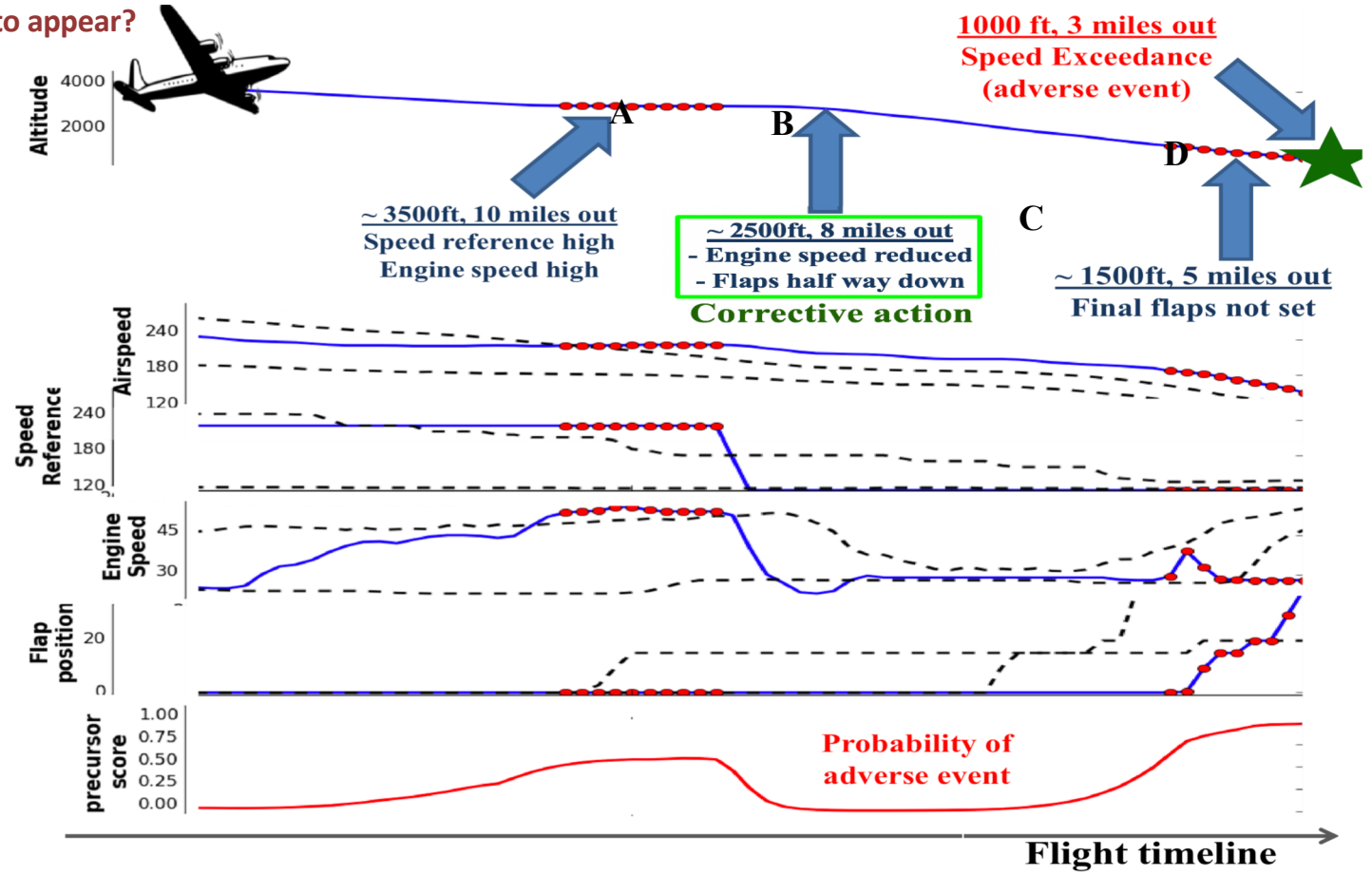
Given a sequence of events: an action that leads to an unrecoverable state is defined as a precursor.

Safety Event Explanation



Precursors – Incident flight

- When do degraded states begin to appear?
- What are the degraded states?
- Are there corrective actions?
- What is the likelihood?



Characteristics and Advantages of the Method

- Any event of interest
  - **Adverse event**
    - Low level events such as vehicle (flight) related events
    - High level events such as airspace congestion, nation-wide delays
  - **Favorable events**
  - high throughput, mission success
  - **Non-aviation**
    - Stock market crash, financial events
    - Ground transportation, activity recognition, health
- Input
  - Time series data
  - Knowledge of the safety event (high-level labels)
- Output
  - Precursor time instances
  - Corrective actions
  - Precursor variables
  - Probability score

Planned open source code release: Fall 2019

Summary

- Benefits of automatic anomaly detection
  - Can discover patterns and insights into operations that may not be initially visible to humans.
  - Leveraging these insights can proactively increase safety and efficiency.

- Overcoming the challenges of data issues in automation
  - Challenges exist both technically mining “Big Data” and logistically in implementing data science programs.

- Identifying precursors with algorithms
  - Discovering precursors to adverse events brings to light possible vulnerabilities that when identified can be mitigated to lower safety risk.

Thank you

ACKNOWLEDGEMENT

This research is supported by the NASA Airspace Operation and Safety Program and the NASA System-wide Safety Program.

CONTRIBUTORS

Vijay Janakiraman (Intuit*)
Daniel Weckler (KBR)

CONTACT

Bryan Matthews (Bryan.L.Matthews@ nasa.gov) (KBR)
Nikunj Oza (nikunj.c.oza) (NASA)
NASA Ames Research Center

*Previously USRA