Cognitive Support for Airline Operations in Complex Environments

Dr. Robert Mauro,
- Decision Research &
- University of Oregon

Dr. Lance Sherry,
- George Mason University Center for Air Transportation Systems Research
Outline

- Nowcasting Uncertain Events
  - Using Big Data and Machine Learning
- Risk Assessment
- Procedure Design
World View

Stochastic

Sensors

Deterministic Displays

Deterministic Procedures
Command & Control

1. Outside world is stochastic; captured by analog sensors
2. Sensor data is converted to deterministic representations
3. Operators match deterministic procedures with deterministic displays to execute the mission
Very Effective Approach …

- Stochastic world view → Deterministic procedure view is very, very successful

- In past 30 years of digital automation, the process has improved even as layers of complex procedures/functions have been added.
  - Simple → Complicated
…Except for Rare Events

- … erratic airspeed sensors (AF 447)
  - Automation to Operator: “your airplane, sir”
- … frozen A-o-A sensors (XL Germany)
  - Automation to Operator: “everything is fine”
- … failed Radio Altimeter coupled to Master Autothrottle (TK 1951)
  - Automation to Operator: “there are discrepancies, but we are going to ignore them.”

Issues:
- Sensors and Sensor Checking logic cannot make sense of the stochastic data
  - Tolerances/Thresholds
  - Erratic data
  - Noise
Stable Approach:
- On glide-path
- On runway center-line
- +/-10 knots of desired speed
- Not excessive Rate-of-Descent (ROD)

Issues:
- Set-up is critical to downpath events/state
- What is tolerance?
- Abort is a complex decision (with significant workload consequences)
Simple/Complicated → Complex

- **Snowden (2005)**
  - **Simple/Obvious**
    - Tightly constrained/No degrees of freedom (Linear System)
    - Procedures: Sense → Categorize → Respond
      - Decisions based on
    - Best Practice Procedure (i.e. optimized)
  - **Complicated**
    - Governed by constraints/Tightly Coupled (Linear System with high combinatorics)
    - Procedures: Sense → Analyze → Respond
      - Decisions based on ruled
    - Good practice (i.e. not always best, but sufficient)

- **Complex**
  - Not governed by constraints/Tightly Coupled (non-linear, some uncertainty)
  - Procedures: Probe → Sense → Analyze → Respond
    - Trial-and-error
  - Emergent/Novel practice

**Transition**
- Instantaneous
- Silent
- Significant Consequences
Challenge

- Provide operators solutions to execute missions in environment in which Stochastic world instantaneously is no longer compatible with deterministic procedures:
  - Rare events
  - Complex dynamic emergent procedures
Challenge

- How to deal with the residual unpredictability
- Our work in flightdeck automation, big data analysis, risk assessment, and procedure development is aimed at providing workable solutions
Nowcasting Uncertain Events Using Big Data and Machine Learning

Dr. Lance Sherry,
  - George Mason University Center for Air Transportation Systems Research

Dr. Robert Mauro,
  - Decision Research &
  - University of Oregon
Airline operations are increasingly captured by massive amounts of data:
- Trajectory
- Atmospheric
- Vehicle system states
- ATC/ATM/AOC system states
- Operator Performance
Bad News

- Massive amounts of data
- Data sets not connected
- Data is noisy/variation
- Data from one operation (e.g. ILS approach) is not applicable same operations at another location
  - ILS approach at ABC is not the same as the ILS approach at XYZ
Magic Genie

- Machine Learning & Data Storage
  - Process massive amounts of data (from same location)
  - Tease out correlations and patterns
Nowcasting

- Use data from a specific procedure to *nowcast* down path performance

- *Nowcast* – forecast based on real-time data for events in the near future (minutes)
Example – Stable Approach

![Graph showing Example – Stable Approach](image-url)
Example – Stable Approach

Stable Approach at 1000’ AGL

Can Nowcast here ...

... for event here
Data

- 28 days of “radar” surveillance track data
- 8237 flights
- Runway 22R EWR
  - 4-5 second update rate
  - track index
  - aircraft type
  - destination airport
  - seconds past midnight
  - latitude/longitude
  - Altitude
  + weather data
  + nav procedure data

• Results improved by use of FOQA/FDR data
  • 1 sec update rate
  • Aircraft configuration (slats, flaps)
  • Automation targets and modes
  • Aircraft maintenance log

Lateral/Vertical Deviation at 6 nm to Runway Threshold
Logistic Regression Model

Logistic Regression:

\[ h_\theta(x) = \frac{1}{1 + e^{-\theta x}}, \]

\[ x = \text{column vector containing all the feature values} \]
\[ \theta(x) = \text{row vector containing all the regression coefficients} \]
\[ h_\theta(x) = \text{predicted probability that a flight with feature-vector } x \text{ experiences an unstable approach after reaching 1000’ AGL.} \]

Cost Function:

\[ J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] \]

\[ \theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \]
Nowcast at 6nm/2000’ AGL for Nowcast 1000’ AGL

<table>
<thead>
<tr>
<th></th>
<th>Actual Unstable</th>
<th>Actual Stable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Unstable</td>
<td>962</td>
<td>281</td>
</tr>
<tr>
<td>Predicted Stable</td>
<td>518</td>
<td>1397</td>
</tr>
</tbody>
</table>

Recall 65%
% Predicated Unstable of Total Actually Unstable

Precision 77%
% Actually Unstable from Total Predicted Unstable

Accuracy 74%
Concept of Operations: Wise Associate

Unstable Approach Nowcasting System

- Navigation procedures
- Model parameters
- Aircraft type

Actual State

Historical Track Data

Historical Weather Data

Aircraft Parameters

Navigation Procedures

Nowcasting Algorithms

Trained Model Parameters

Upload algorithm/coefficients

Nowcast Results

UNSTABLE - speed
Nowcast Interpretation

- “Hey we got a strong crosswind today. When ATC vectors to localizer intercept late and there is this crosswind (tailwind on the base leg), flights tend to overshoot the runway centerline and have to fight to (over) correct”
  - So, anticipate to account for the tailwind in the intercept course

- “in peak arrival push, Small category aircraft have to keep their speed up in the initial approach and have to bleed-off speed in short distance while descending”
  - So, be prepared to add drag rapidly to avoid overspeed
Proposal: Wise Associate

- Automation that acts like a “back seat driver”
  - Always warning about potentially dangerous events
    - Events may not occur
    - Warnings may not always be accurate (< 25%)

- Benefits:
  - Allows operator to provide some attention to factors/events that have historically

- Applications:
  - Flight deck operations
  - Dispatch
  - Traffic Flow Management
  - Air Traffic Control
Research Questions

- Will human operators accept automation warnings?
- How accurate does it need to be to avoid “nuisance” alerts and be turned-off?
- What kind of display/aural alert would work best?
End Part 1
Risk Assessment

Dr. Robert Mauro
Decision Research &
Institute of Cognitive & Decision Sciences
University of Oregon

Dr. Immanuel Barshi
NASA Ames Research Center
Aviation Operations
Outline

- Topics
  - Thinking about Risk
  - Problems and Issues in Risk Assessment
  - Tool for Risk Identification Assessment & Display (TRIAD)
What is risk?

- Which poses the greatest risk?
  - Meteor strike on the Ops Center
  - W&B program error causes 2% decrease in fuel efficiency
  - All aircraft in a fleet grounded
  - Loss of an aircraft and crew

Dayton, Ohio
What is risk?

- Combination of:
  - probability (likelihood)
  - consequences (threat value)
- Risk = f(p, c)
Common Methods of Risk Assessment

- Informal “seat-of-the-pants” approach
- Probabilistic Risk Analysis (PRA)
Probabilistic Risk Analysis

- **Typical steps**
  - Decomposition of paths leading to undesired state (e.g. FMEA)
  - Evaluation of probability functions associated with each path
  - Determination of costs associated with each possible undesired state
  - Integrated risk function produces risk value

- **Problems**
  - Complex process
    - Difficult to understand and communicate, so people avoid it
    - Conducted by decision analysts using input from domain experts
  - Costly and time consuming
    - So reserved for occasional “big” issues
  - Accuracy depends on values that are difficult to estimate and identification of paths that are difficult to explicate
    - Calculations may give misleading impression of precision
  - Difficult to obtain consistent cost metrics
    - Different types of consequences are hard to equate
  - Simplified versions don’t fit all situations
Informal risk assessment

- **Process**
  - Managers consider problem and produce global judgment based on past experience and logical analysis

- **Problems**
  - Subject to numerous biases
    - Availability – whatever comes to mind
    - Representativeness – whatever fits expectations
  - Process is obscure
    - Leads to argument not discussion
    - Difficult to document
    - Difficult to improve
The Risk Matrix

- Semantics matter
  - Why use particular labels?
  - People often disagree on what the labels mean.

- Categories imply precision
  - E.g., what if an issue can range from “unlikely” to “possible”?

- Colors imply decisions
  - Who made the decision rules?

- Why 5 X 5 matrix?
  - Categories imply that one should treat everything in a “box” the same.
The Risk Matrix

- Risk scores (numbers in the category boxes) imply “real” values (and ratio level measurement)
- This implies symmetry
  - E.g., all boxes with the same value are the same.
- But the scales are (at best) ordinal.
- This can result in reversals in which riskier combinations receive lower scores.
Risk Assessment Wish List

- Structure discussion
- Document the process followed
- Easy path to improve assessment when desired
- Relatively easy to understand and communicate
  - So it will be used
- Relatively inexpensive and time effective
  - So it will be used whenever needed
- Do not want to rely on misleading calculations
- Reveal (not hide) uncertainty
- Obtain consistent consequence metrics
- Valid in all situations
Middle way

- **Goals**
  - Provide decision maker with information needed to make sensible risk assessments and decisions
  - Make bases of assessments and decisions explicit
  - Allow assessments to be made with different levels of precision and effort
  - Provide for different levels of analysis

- **Systematic approach**
  - Requires well-defined problems
  - Components of risk clearly specified
  - Different types of consequences handled separately
  - Uncertainties displayed
  - Risk functions not imposed
General Approach to Risk Assessment

- Current Risk Assessment
  - Define problem
  - Specify possible outcomes
  - Assess probabilities
  - Assess consequences
  - Display risk summaries

- Forecast Risk Assessment
  - Specify possible interventions
  - Assess expected effects of interventions
  - Display risk reduction summaries
Developing Solutions - TRIAD

TRIAD
Tool for Risk Identification, Assessment, & Display

NASA

DECISION RESEARCH
1. Specify Problem & Outcomes
2. Likelihood Estimation

**Likelihood Estimation**

**Instructions:** For each outcome, enter your best estimate of the likelihood. Specify minimum and maximum values. Values may be entered directly in the columns labelled "Min", "Best" and "Max". Likelihood categories (right) may be used to simplify the process: click on a category then hit a button (Min, Max, Best, or All) to enter the values. The values must lie between 0 and 1 and "Min" ≤ "Best Estimate" ≤ "Max". Hit 'Return' to go back to the Recommendations page.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Min</th>
<th>Best</th>
<th>Max</th>
<th>Min</th>
<th>Best</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Aircraft Wt error &gt; threshold</td>
<td>1/1,000,000</td>
<td>1/100,000</td>
<td>1/50,000</td>
<td>once every 2500 days</td>
<td>once every 250 days</td>
<td>once every 125 days</td>
</tr>
<tr>
<td>2) Aircraft Wt error &lt; threshold</td>
<td>1/4000</td>
<td>1/2000</td>
<td>1/400</td>
<td>once every 10 days</td>
<td>once every 5 days</td>
<td>1 times/day</td>
</tr>
<tr>
<td>3) Aircraft Balance error &gt; threshold</td>
<td>0</td>
<td>5/10,000,000</td>
<td>1/1,000,000</td>
<td>once every 0 days</td>
<td>once every 5000 days</td>
<td>once every 2500 days</td>
</tr>
<tr>
<td>4) Aircraft balance error &lt; threshold</td>
<td>1/1,000,000</td>
<td>5/10,000</td>
<td>1/1,000</td>
<td>once every 2500 days</td>
<td>once every 5 days</td>
<td>once every 3 days</td>
</tr>
</tbody>
</table>

**Events/Day Translation**

- To translate frequency/operation to events/day, enter operations/day here: 400

---

[Graphics and logos from NASA and MASON UNIVERSITY]
3. Consequence Evaluation

**LIFE & HEALTH EVALUATION**

This scale measures the threat to the lives and health of humans that may occur as a direct result of the problem. The effects may be immediate or delayed.

**Instructions:** Select an outcome and then rate it on the evaluation scale. Choose the consequence level that best reflects the harm that would occur in the worst reasonable case of that outcome. Once you have rated an outcome, the rating will appear in the column on the right. Continue to select outcomes and rate accordingly until all listed outcomes have been rated.

<table>
<thead>
<tr>
<th>Life &amp; Health - Consequence Evaluation Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>None</td>
</tr>
</tbody>
</table>

1) Aircraft Wt error > threshold | Very Low |
2) Aircraft Wt error < threshold | None |
3) Aircraft Balance error > threshold | Very Low |
4) Aircraft balance error < threshold | None |
5)
3. Consequence Evaluation

**PROPERTY DAMAGE EVALUATION**

This scale measures the financial costs that would be associated with each outcome. This includes the costs of repairing or rebuilding company equipment and vehicles, the costs of repairing or rebuilding structures belonging to other entities, and the cost of recovering from environmental damage, compensating individuals for related losses, and relocating people as a result of that damage.

**Instructions:** Select an outcome and then rate it on the evaluation scale. Choose the financial cost value that would be associated with the worst reasonable consequence of the outcome. Once you have selected a category, the rating will appear to the right of the outcome. Continue to select outcomes and rate accordingly until all listed outcomes have been rated.

<table>
<thead>
<tr>
<th>Property Damage - Consequence Evaluation Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>None</strong></td>
</tr>
<tr>
<td><strong>None</strong></td>
</tr>
<tr>
<td>1)</td>
</tr>
<tr>
<td>2)</td>
</tr>
<tr>
<td>3)</td>
</tr>
<tr>
<td>4)</td>
</tr>
<tr>
<td>5)</td>
</tr>
</tbody>
</table>
3. Consequence Evaluation

MISSION SUCCESS EVALUATION

This scale measures the threat to the company’s mission posed by the problem under consideration. Direct financial costs for repairing or replacing equipment are included in the Property Damage scale. This scale seeks to measure the harm to the company that would result from mission objectives not being met. This includes direct costs of interruptions in service (e.g., lost baggage, delays, cancellations) caused by a problem and indirect costs caused by regulatory actions.

Instructions: Select an outcome and then rate it on the evaluation scale. Select the mission success consequence that would be associated with the worst reasonable case associated with the outcome. Once you have selected a category, the rating will appear to the right of the outcome. Continue to select outcomes and rate accordingly until all listed outcomes have been rated.

**Mission Success - Consequence Evaluation Scale**

<table>
<thead>
<tr>
<th>None</th>
<th>Very Low</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Disruption Single/Aircraft Flight</td>
<td>Disruption Multiple Aircraft/Flights</td>
<td>Disruption Station/Many Operations</td>
<td>Disruption Multiple Stations</td>
<td>Company Threatened</td>
</tr>
</tbody>
</table>

1) Aircraft Wt error > threshold | None
2) Aircraft Wt error < threshold | Very Low
3) Aircraft Balance error > threshold
4) Aircraft balance error < threshold
3. Consequence Evaluation

### SOCIAL AMPLIFICATION EVALUATION

The damage caused by people’s reaction to a failure can far outweigh the damage caused by the failure itself. This “social amplification” effect is a complex function of the physical consequences of the event, public perceptions, media effects, and political activity. The scale below is designed to estimate the size of these effects.

**Instructions:** Select an outcome and then rate it by checking all of the boxes on the evaluation scale that apply. Be sure that the outcome is highlighted. The rating will appear to the right of the outcome. Continue to select outcomes and rate accordingly until all listed outcomes have been rated.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How much of a social amplification effect would be expected if:

1) Aircraft Wt error > threshold
   - Very Low
2) Aircraft Wt error < threshold
   - Moderate
3) Aircraft Balance error > threshold
4. Risk Summary
III. LOGARITHMIC RISK ASSESSMENT DISPLAY

These graphs present a visual summary of the risk profile associated with this issue. The values displayed are collapsed across outcomes but displayed separately for each level of each type of consequence. The error bars display the minimum and maximum probabilities.
V. OUTCOME SUMMARY TABLE

This table provides you with the probability and consequence scale ratings by outcome. The best estimate of the likelihood of each outcome is presented in digital form and coded by increasingly dense hatchings from white to black. The information in this table should not be changed or manipulated. To sort the outcomes in order of increasing probability, click on "Sort by Increasing Probability". To sort the outcomes in order of decreasing probability, click on "Sort by Decreasing Probability". To sort the outcomes alphabetically by label, click on "Sort by Outcome Label".

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
<th>Life &amp; Health</th>
<th>Property</th>
<th>Operation</th>
<th>Amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Aircraft Wt error &gt; threshold</td>
<td>0.00001</td>
<td>Very Low</td>
<td>Low</td>
<td>None</td>
<td>Very Low</td>
</tr>
<tr>
<td>2) Aircraft Wt error &lt; threshold</td>
<td>0.0005</td>
<td>None</td>
<td>Very Low</td>
<td>Very Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>3) Aircraft Balance error &gt; threshold</td>
<td>0.0000005</td>
<td>Very Low</td>
<td>Moderate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Aircraft balance error &lt; threshold</td>
<td>0.0005</td>
<td>None</td>
<td>Very Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV. LOGARITHMIC OUTCOME DISPLAY

These graphs present a visual summary of the risk profile associated with this issue. The values for each outcome are displayed separately by type of consequence. The error bars display the minimum and maximum probabilities.
5. Forecast Risk

- Devise possible intervention(s)
- Consider previously determined outcomes
- Rate new likelihood given intervention
  - Particular intervention may/may not affect likelihood of particular outcomes
- Rate new consequence given intervention
  - Particular intervention may/may not affect consequences of particular outcomes
- Display Results
Forecast Risk Displays
Conclusion

- TRIAD
  - Easy to use tool for risk assessment under uncertainty
  - Provides information needed for decision without misleading oversimplification
  - Displays help guide group discussion and decision-making, not replace it
Conclusions

- TRIAD
  - Avoids using risk functions that do not match reality
  - Avoids use of arbitrary values to equate different types of consequences
  - Avoids logical / mathematical errors in combining assessments of different outcomes
End Part 2
Procedures that Work
Designing Excellent Procedures

Immanuel Barshi
NASA Ames Research Center

Asaf Degani
General Motors Advanced Technology Center

Loukia Loukopoulou
Swiss International Airlines

Robert Mauro
Decision Research & University of Oregon
Where
Why

- Operate
- Reduce error
- Substitutability
- Operational efficiency
- Organizational efficiency
- Managerial control
- Risk management
When

- Something is new
- Something has changed
- Something isn’t working
Why Procedures Fail

- **Bad procedures**
  - Fail to take into account an important component of the task
  - Too narrow
  - Too broad

- **Good procedures that went bad**
  - Can’t adapt (and aren’t changed)

- Aren’t followed
4P’s
Practices
Procedures
Policies
Philosophy
How

- Understand
  - Goals & Requirements
  - Task Analysis
- Design
- Implement
- Evaluate
How: Understand

- Goals for procedures
- Goals of procedures
  - Correct
  - Reliable
  - Robust
  - Resilient
  - Efficient
- Task Analysis
How: Design

- Task Procedure
  - Domain
    - Technology
    - Human
    - Environment
  - Requirements
  - Conflicts
  - Prohibitions
    - Margins, Barriers, Buffers
  - Sequences
  - Timing
  - Formative Evaluations

- Phase Procedure
  - Sequential Task Procedures
  - Interwoven Task Procedures
How: Implement

- Training
  - How
  - Why
  - Why not
- Change Management
How: Evaluate

- Observations
  - General
  - Targeted
- Surveys
- Automated data collection (FOQA)
- Voluntary reports (ASAP, ASRS)
Results
Example

- **Design Overview**
  - Phase I: Old procedures
    - line pilots observed at random
  - Phase II: New procedure tests
    - a small group of test crews observed flying under old procedures and proposed procedures for a whole month
  - Phase III: New procedures after adoption
    - line pilots observed at random 4-5 months after adoption of new procedures
  - Phase IV: New procedures after stabilization
    - line pilots observed at random 8-10 months after adoption of new procedures (+ winter ops)
Example

- Measurement overview
  - Pilots observed using structured observation log
  - Pilots surveyed using open/closed ended questions returned directly to NASA
  - Observers’ questionnaire
  - “Event” details
  - Observer training
Figure 1. Mean Number of Problems on Target Items per Flight

- Line Crews Original Procedures
- Test Crews Original Procedures
- Test Crews New Procedures Week 1
- Test Crews New Procedures Week 3
- Line Crews New Procedures 4 Months
- Line Crews New Procedures 8 Months
Figure 3. Problems on Target Items by Phase of Flight

Proportion

Before Push, Pushback, Taxi Out, Takeoff, Cruise, TOD-10000, Approach/Landing, Taxi In/Shut Down

Legend:
- Line Crews
  - Original Procedures
- Test Crews
  - Original Procedures
  - New Procedures
  - Week 1
  - Week 3
  - 4 Months
  - 8 Months

Data points show the proportion of problems on target items across different phases of flight for line crews and test crews under various new procedures and time periods.
Thank You!