Limitations of Commercial Aviation Safety Assessment Standards

Rodrigo Rose

STAMP Workshop 2024

Massachusetts Institute of Technology

Introduction

- The processes used to assess the safety of commercial aircraft were developed throughout the 20th century and formalized into standards in the 1990s

- Modern commercial aircraft are highly automated and rely on complex interactions between hardware, software and humans

- The Boeing 737 MAX accidents have highlighted that commercial aircraft are not immune to severe design flaws
  - Government agencies, academics and the standards community were aware of this before the accidents
  - Impetus to address these deficiencies before another major accident

Common View of the Limitations in Boeing 737 MAX Safety Assessment

"When all flight deck effects are considered, the introduction of the MCAS function **invalidated aircraft-level assumptions** for flight crew responses related to erroneous AOA failures under certain conditions"

– *Joint Authorities Technical Review Report*

"Based on the **incorrect assumptions** about flight crew response and an incomplete review of associated multiple flight deck effects, MCAS's reliance on a single sensor was deemed appropriate and met all certification requirements"

– *Lion Air 610 Final Report*

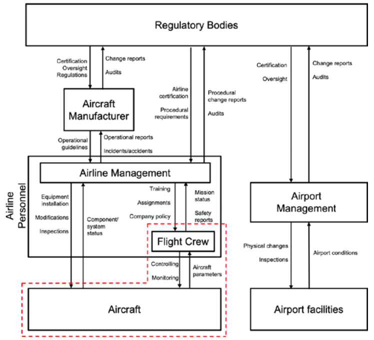
"Boeing made **fundamentally faulty assumptions** about critical technologies on the 737 MAX, most notably with MCAS"

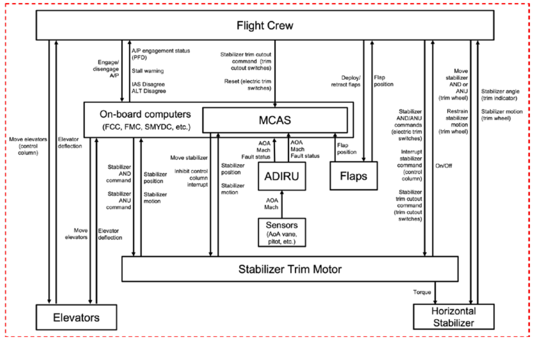
– *House Committee on Transportation & Infrastructure Report*

Most analyses identify the flawed assumptions, but don't systematically question the safety assessment methods that allowed the assumptions to slip through

Limitations in Safety Assessment Standards

- CAST analysis was performed on JT610 and ET302





Beliefs held during the design and safety assessment ⟷ Beliefs held and actions taken during the operation

- Four main limitations identified:

Limited integration of human considerations in the safety assessment process	Absence of a systematic methodology that supports the identification of unrecognized assumptions	Limited guidance for uncovering and managing non-failure based causal scenarios leading to losses	Limited framework for understanding non-linear (e.g., circular or balancing) causal relationships

Limitation 1: Human Considerations

3.7 Aircraft Safety Assessment

The ASA is a systematic, comprehensive evaluation of the complete aircraft to show that safety objectives from the AFHA/PASA and safety requirements from the PASA are satisfied. The difference between a PASA and an ASA is that a PASA is a method to evaluate proposed architectures and derive safety requirements; whereas the ASA is verification that the implemented design meets both the qualitative and quantitative safety objectives and requirements as defined in the AFHA and PASA.

The ASA integrates the results of the various analyses to verify the safety of the overall aircraft and systems. This aircraft safety assessment is refined and updated throughout the development process to reflect the updated design.

The ASA uses the results obtained from the PASA and SSAs and ensures assessment of interdependencies between the aircraft functions and systems. The ASA ensures that system failure modes are considered for inclusion. The ASA also includes applicable common cause consideration results.

For details in performing the ASA, refer to Appendix F.

3.8 Determining Depth of Analysis for Failure Conditions

Failure conditions for the aircraft/system function should be evaluated to determine how the aircraft/system will satisfy safety objectives. The depth of analysis that should be employed in the assessment of the failure conditions is typically based on the failure condition classification, and in some cases, other aircraft/system characteristics. This evaluation generally follows a course to determine what type(s) of analysis/assessment should be employed in analyzing the failure condition, e.g., design or installation appraisal, verification analysis, or qualitative and/or quantitative assessment. While the determination of the course of analysis is straightforward for most categories, additional criteria are usually required to determine the course of analysis for "Major" failure conditions. The Safety Analyst should consult Depth of Analysis Flow Charts and associated text in advisory circular material for the current guidance to be used in determining depth of analysis of failure conditions, e.g., AC 25.1309 draft ARSENAL revised / AMC 25.1309 and AC29.2C.

3.9 Function Development Assurance Level and Item Development Assurance Level Assignment.

Safety process activities within the PASA and PSSA processes include the assignment of FDALs and IDALs which define the level of rigor of development assurance activities. These levels of rigor are used to substantiate, to an adequate level of confidence, that development errors have been identified and corrected. The activities are associated with the process of establishing the characteristics of how potential errors contribute to failure conditions so that FDALs and IDALs may be assigned in accordance with development process principles.

Appendix P provides details in performing development assurance level assignment.

3.10 Considerations of Human Error in the Safety Assessment Process

The safety assessment process described in this document assumes that flight crews, cabin crews, maintenance crews, and other individuals participating in the operation of the aircraft follow documented procedures in foreseeable operating conditions (normal, malfunction or abnormal, and emergency). Intentional or unintentional deviation from these procedures is not considered in the safety assessment process described herein.

With the exception of some aspects of the common mode analysis and the zonal safety analysis, the safety effects of potential flight crew and maintenance errors are evaluated using different analysis techniques. See the appropriate certification advisory material on human factors for accomplishing human factor safety evaluations.

4. SAFETY ANALYSIS METHODS

4.1 Fault Tree Analysis/Dependence Diagram/Markov Analysis/Model Based Safety Analysis

FTA, DD, and MA are top-down analysis techniques. These analyses proceed down through successively more detailed (i.e., lower) levels of the design. The MBSA is a technique which models system content and behavior to provide safety analysis results. A reminder that when FTA is presented herein, the DD, MA and/or MBSA analysis techniques may be applicable/selected depending on the circumstances and the types of data desired.

"The safety assessment process described in this document assumes that flight crews, cabin crews, maintenance crews, and other individuals participating in the operation of the aircraft follow documented procedures in foreseeable operating conditions…"

Limitation 1: Human Considerations

Table A-7 - AFHA Format Example

1	2	3	4	5	6
ID #	**Failure Condition**	**Flight Phase**	**Effects of Failure Condition on Aircraft, Crew, Occupants**	**Severity Classification**	**Assumptions, Comments, Rationale or Reference to Supporting Material**
Aircraft Function: (4) Provide Survivable Environment			Sub-Function: (4.1) Provide breathable atmosphere		
Sub-Function: (4.1.1) Provide oxygenated atmosphere					
4.1.1.T1	Unannunciated total loss of oxygenated air to crew or passengers	Climb Cruise Descent	Aircraft: No effect. Crew: Unaware or unable to counter the effects of the condition, the crew may be incapacitated by hypoxia or unable to restore sufficient levels of oxygen to the occupants in time to prevent permanent physiological harm. Occupants: Multiple occupant fatalities or severe injuries are possible due to the direct effects of hypoxia or due to crew incapacitation and subsequent loss of aircraft control.	Catastrophic	14CFR/CS 25.841(a)(2)(ii) "Pressurized Cabins" 14CFR /CS 25.1441(d) "Oxygen equipment and supply" 14CFR /CS 25.1443(c)(2) "Minimum mass flow of supplemented oxygen" AC 25-20 (6)(e)&(7) "Pressurized Ventilation and Oxygen System Assessment for Subsonic Flight Including High Altitude Operations" EASA Certification Review Item "Airworthiness Standards for Subsonic Transport Aeroplanes to be operated above 41,000 ft."

What has failed

What phase of flight it failed in

What are the effects of the failure

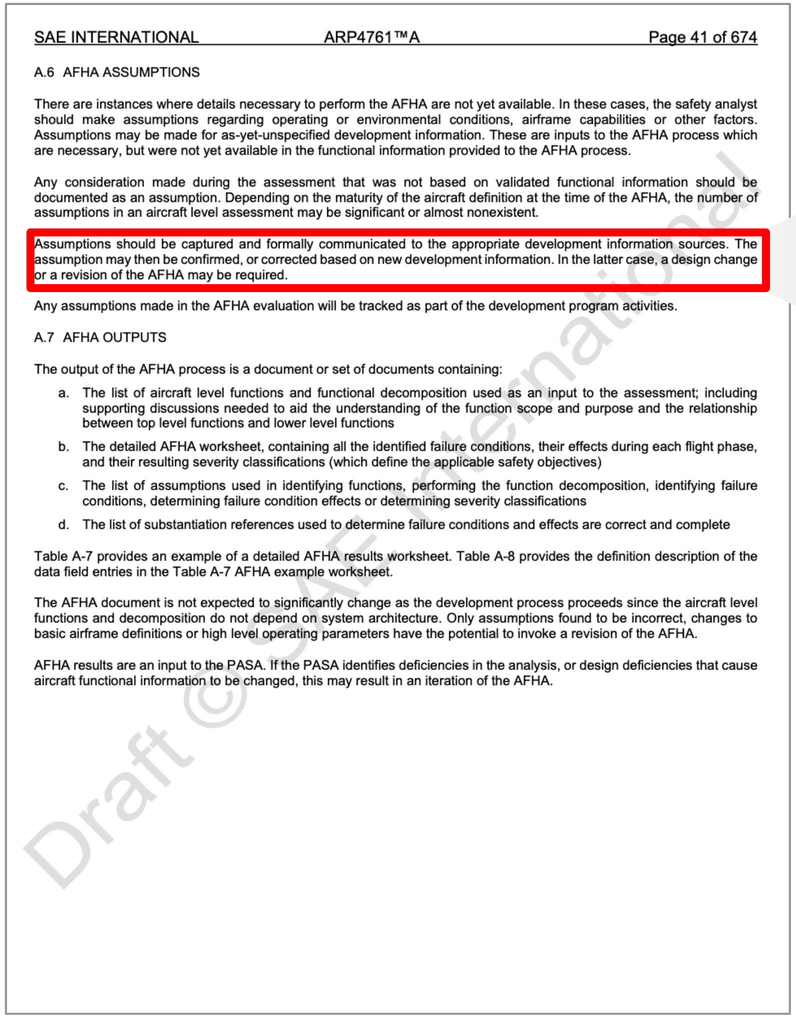
What is the severity of the failure

What assumptions have been made about the failure

- Assumptions about flight crew response are used to make decisions about severity classifications
- Severity classifications are used to make design decisions

Limitation 2: Identification of Assumptions

- In traditional safety assessments, assumptions are listed because there is some level of **doubt** about their validity



A.6 AFHA ASSUMPTIONS

There are instances where details necessary to perform the AFHA are not yet available. In these cases, the safety analyst should make assumptions regarding operating or environmental conditions, airframe capabilities or other factors. Assumptions may be made for as-yet-unspecified development information. These are inputs to the AFHA process which are necessary, but were not yet available in the functional information provided to the AFHA process.

Any consideration made during the assessment that was not based on validated functional information should be documented as an assumption. Depending on the maturity of the aircraft definition at the time of the AFHA, the number of assumptions in an aircraft level assessment may be significant or almost nonexistent.

Assumptions should be captured and formally communicated to the appropriate development information sources. The assumption may then be confirmed, or corrected based on new development information. In the latter case, a design change or a revision of the AFHA may be required.

Any assumptions made in the AFHA evaluation will be tracked as part of the development program activities.

A.7 AFHA OUTPUTS

The output of the AFHA process is a document or set of documents containing:

a. The list of aircraft level functions and functional decomposition used as an input to the assessment; including supporting discussions needed to aid the understanding of the function scope and purpose and the relationship between top level functions and lower level functions
b. The detailed AFHA worksheet, containing all the identified failure conditions, their effects during each flight phase, and their resulting severity classifications (which define the applicable safety objectives)
c. The list of assumptions used in identifying functions, performing the function decomposition, identifying failure conditions, determining failure condition effects or determining severity classifications
d. The list of substantiation references used to determine failure conditions and effects are correct and complete

Table A-7 provides an example of a detailed AFHA results worksheet. Table A-8 provides the definition description of the data field entries in the Table A-7 AFHA example worksheet.

The AFHA document is not expected to significantly change as the development process proceeds since the aircraft level functions and decomposition do not depend on system architecture. Only assumptions found to be incorrect, changes to basic airframe definitions or high level operating parameters have the potential to invoke a revision of the AFHA.

AFHA results are an input to the PASA. If the PASA identifies deficiencies in the analysis, or design deficiencies that cause aircraft functional information to be changed, this may result in an iteration of the AFHA.

"Assumptions should be captured and formally communicated to the appropriate development information sources. The assumption may then be confirmed, or corrected based on new development information. In the latter case, a design change or a revision of the AFHA may be required."

Limitation 2: Identification of Assumptions

Documented Assumption: Continuous unintended nose down stabilizer trim inputs would be recognized as a Stab Trim or Stab Runaway failure and procedure for Stab Runaway would be followed

Runaway Stabilizer

Condition: Uncommanded stabilizer trim movement occurs continuously.

1 Control column. . . . . Hold firmly

2 Autopilot (if engaged) . . . . .Disengage

Do **not** re-engage the autopilot.

Control airplane pitch attitude manually with control column and main electric trim as needed.

3 Autothrottle (if engaged). . . . . .Disengage

Do **not** re-engage the autothrottle.

4 **If** the runaway **stops** after the autopilot is disengaged:

■ ■ ■ ■

5 **If** the runaway **continues** after the autopilot is disengaged:

STAB TRIM CUTOUT switches (both) . . . . CUTOUT

If the runaway **continues**:

Stabilizer trim wheel. . . . . Grasp and hold

- -

6 Stabilizer. . . . . Trim manually

7 Anticipate trim requirements.

"Condition: Uncommanded stabilizer trim movement occurs continuously."

*"If the runaway stops after the autopilot is disengaged: **DONE**."*

Reality:

- MCAS stabilizer movement not continuous
- MCAS commands bounded by 2.5° authority
- Pilots can counter nose-down movement with manual electric trim inputs
- No MCAS command for 5 seconds after reset

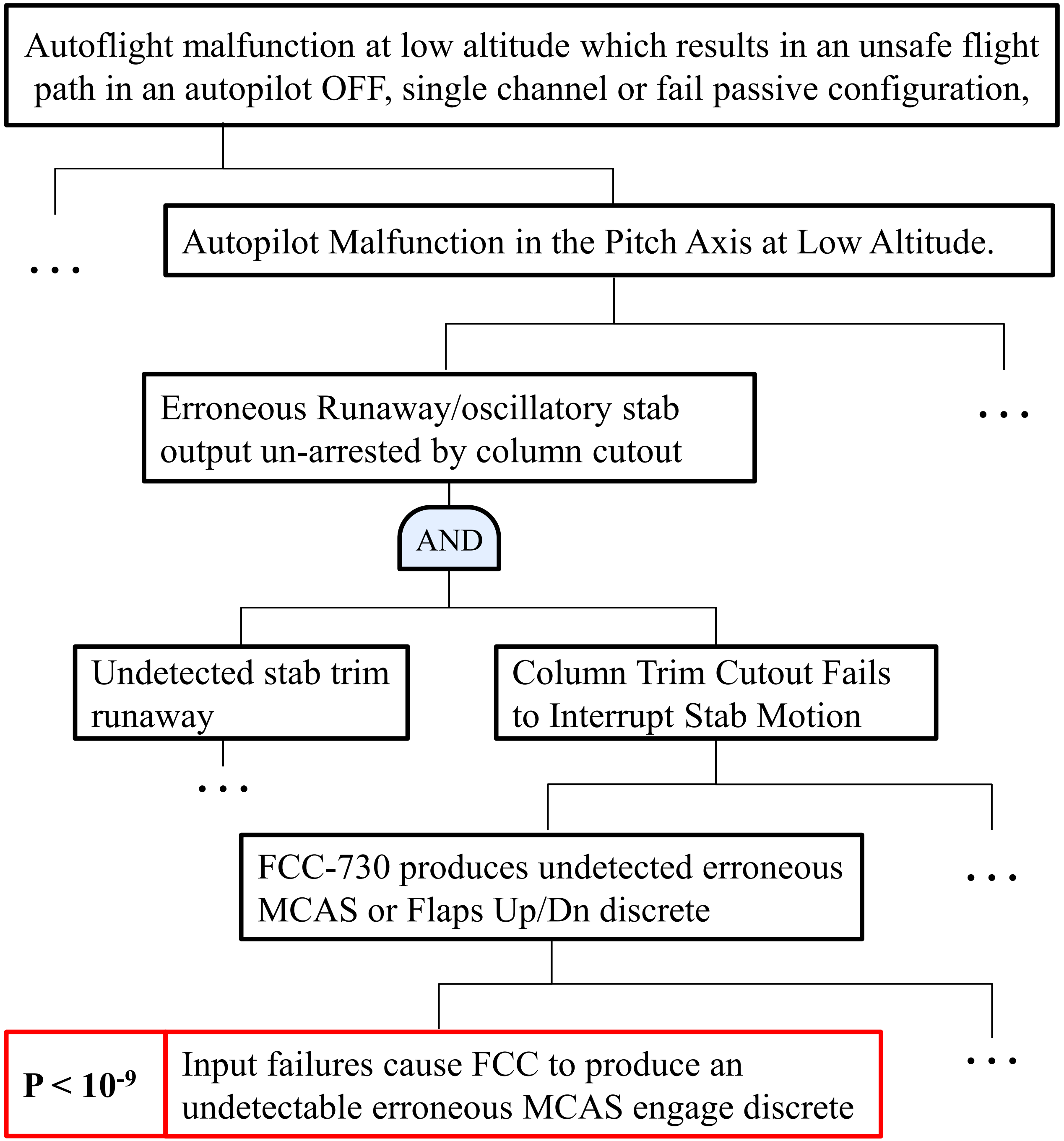
Undocumented Assumption: Erroneous MCAS activations always result in "continuous unintended nose down stabilizer trim inputs"

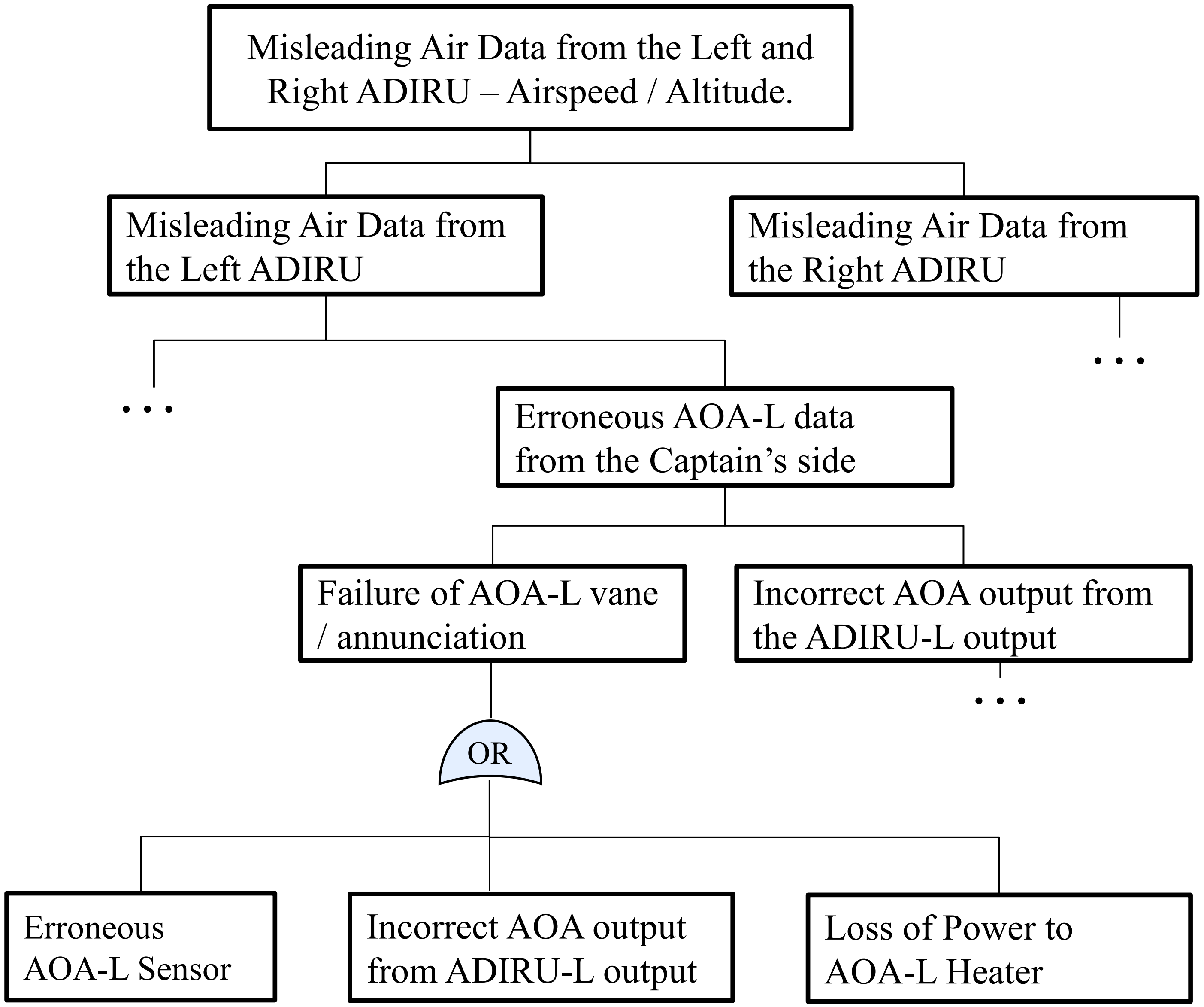
Limitation 3: Capturing Non-Failure Cases

- Developmental factors
  - Unsafe interactions between intended functions/behavior
  - Unsafe combinations of failures and intended behavior

- Non-developmental factors
  - Maintenance error
  - Manufacturing error
  - Operational error
  - etc.

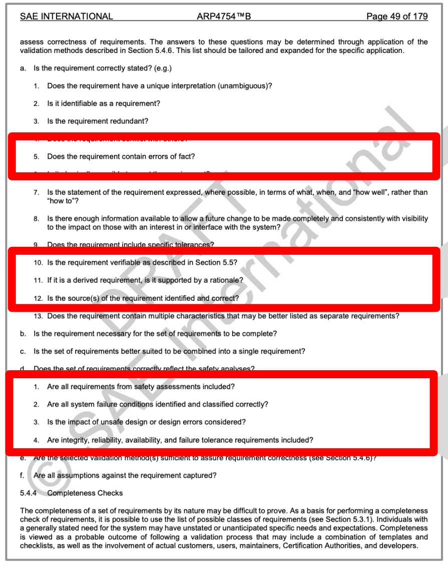
Difficult to obtain meaningful probabilities for


Autoflight malfunction at low altitude which results in an unsafe flight path in an autopilot OFF, single channel or fail passive configuration,
. . .
Autopilot Malfunction in the Pitch Axis at Low Altitude.
Erroneous Runaway/oscillatory stab output un-arrested by column cutout
. . .
AND
Undetected stab trim runaway
. . .
Column Trim Cutout Fails to Interrupt Stab Motion
FCC-730 produces undetected erroneous MCAS or Flaps Up/Dn discrete
. . .
P < 10-9
Input failures cause FCC to produce an undetectable erroneous MCAS engage discrete
. . .


Misleading Air Data from the Left and Right ADIRU – Airspeed / Altitude.
Misleading Air Data from the Left ADIRU
Misleading Air Data from the Right ADIRU
Erroneous AOA-L data from the Captain's side
Failure of AOA-L vane / annunciation
Incorrect AOA output from the ADIRU-L output
OR
Erroneous AOA-L Sensor
Incorrect AOA output from ADIRU-L output
Loss of Power to AOA-L Heater

Limitation 3: Capturing Non-Failure Cases

- ARP4754 seeks to minimize development errors
- ARP4754 requires “Correctness Checks” to be conducted through its recommended “Validation Methods”


SAE INTERNATIONAL ARP4754™B Page 49 of 179

assess correctness of requirements. The answers to these questions may be determined through application of the validation methods described in Section 5.4.6. This list should be tailored and expanded for the specific application.

a. Is the requirement correctly stated? (e.g.)

1. Does the requirement have a unique interpretation (unambiguous)?
2. Is it identifiable as a requirement?
3. Is the requirement redundant?
5. Does the requirement contain errors of fact?
7. Is the statement of the requirement expressed, where possible, in terms of what, when, and “how well”, rather than “how to”?
8. Is there enough information available to allow a future change to be made completely and consistently with visibility to the impact on those with an interest in or interface with the system?
9. Does the requirement include specific tolerances?
10. Is the requirement verifiable as described in Section 5.5?
11. If it is a derived requirement, is it supported by a rationale?
12. Is the source(s) of the requirement identified and correct?
13. Does the requirement contain multiple characteristics that may be better listed as separate requirements?

b. Is the requirement necessary for the set of requirements to be complete?

c. Is the set of requirements better suited to be combined into a single requirement?

d. Does the set of requirements correctly reflect the safety analyses?

1. Are all requirements from safety assessments included?
2. Are all system failure conditions identified and classified correctly?
3. Is the impact of unsafe design or design errors considered?
4. Are integrity, reliability, availability, and failure tolerance requirements included?

e. Are the selected validation method(s) sufficient to assure requirement correctness (see Section 5.4.6)?

f. Are all assumptions against the requirement captured?

5.4.4 Completeness Checks

The completeness of a set of requirements by its nature may be difficult to prove. As a basis for performing a completeness check of requirements, it is possible to use the list of possible classes of requirements (see Section 5.3.1). Individuals with a generally stated need for the system may have unstated or unanticipated specific needs and expectations. Completeness is viewed as a probable outcome of following a validation process that may include a combination of templates and checklists, as well as the involvement of actual customers, users, maintainers, Certification Authorities, and developers.

“Does the requirement contain errors of fact?”

“Is the requirement verifiable?”

“Is the source of the requirement identified and correct?”

“Are all requirements from safety assessments included?

Are all system failure conditions identified and classified correctly?

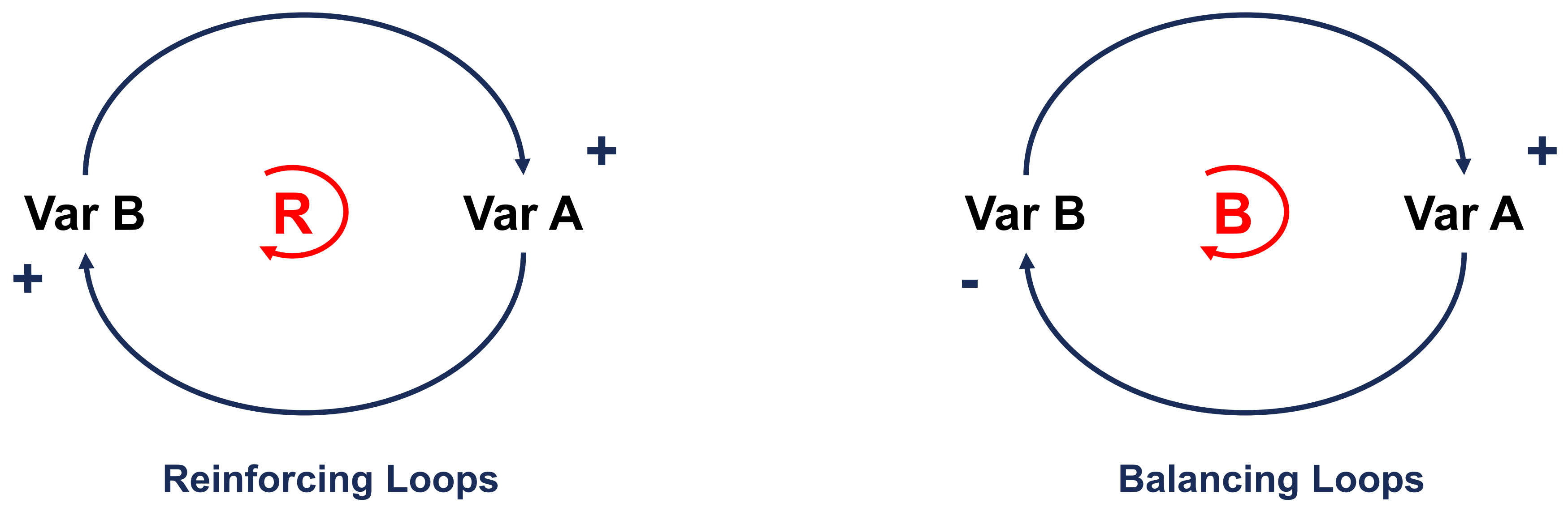
Is the impact of unsafe design or design errors considered?”

These processes are not step-by-step methodologies to interrogate and challenge what you think is true about the system

- ARP4761’s Common Mode Analysis (CMA) qualitatively considers how aspects like software error, pilot training, or manufacturing defects can invalidate logic in FTA

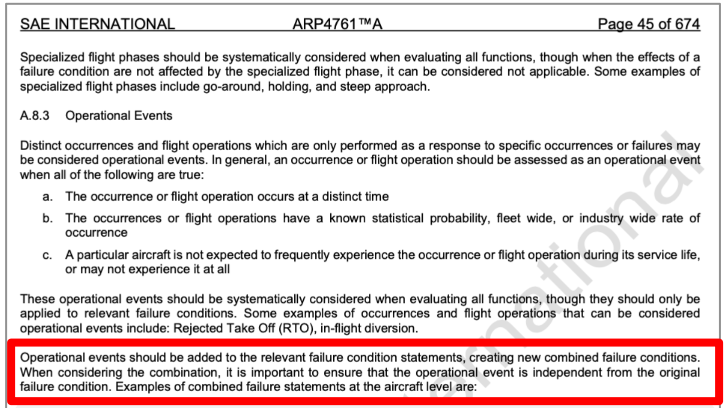
Limitation 4: Capturing Complex Non-linear Causality

- Non-linear causality often involves behavior that reinforces itself or cancels itself out



- Capturing non-linear causality requires being able to capture repeated actions, appropriate timing of decisions, sequences of crew and automated actions, etc.

Limitation 4: Capturing Complex Non-linear Behavior



SAE INTERNATIONAL ARP4761™A Page 45 of 674

Specialized flight phases should be systematically considered when evaluating all functions, though when the effects of a failure condition are not affected by the specialized flight phase, it can be considered not applicable. Some examples of specialized flight phases include go-around, holding, and steep approach.

A.8.3 Operational Events

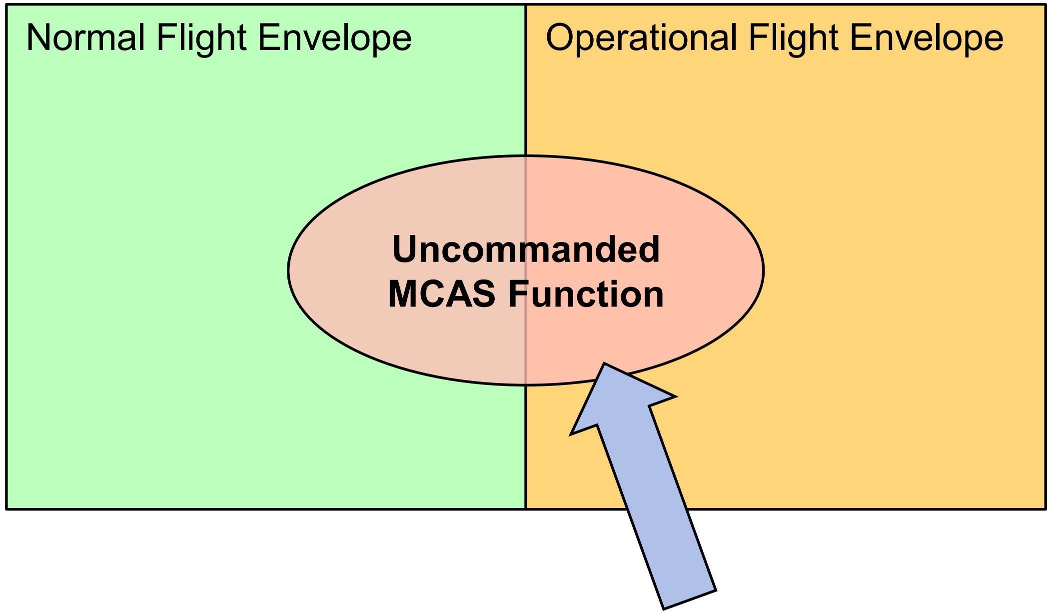
Distinct occurrences and flight operations which are only performed as a response to specific occurrences or failures may be considered operational events. In general, an occurrence or flight operation should be assessed as an operational event when all of the following are true:

a. The occurrence or flight operation occurs at a distinct time
b. The occurrences or flight operations have a known statistical probability, fleet wide, or industry wide rate of occurrence
c. A particular aircraft is not expected to frequently experience the occurrence or flight operation during its service life, or may not experience it at all

These operational events should be systematically considered when evaluating all functions, though they should only be applied to relevant failure conditions. Some examples of occurrences and flight operations that can be considered operational events include: Rejected Take Off (RTO), in-flight diversion.

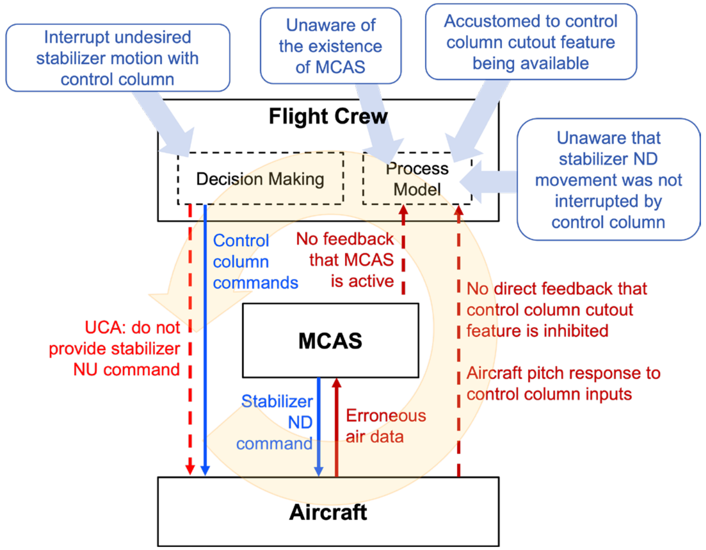
Operational events should be added to the relevant failure condition statements, creating new combined failure conditions. When considering the combination, it is important to ensure that the operational event is independent from the original failure condition. Examples of combined failure statements at the aircraft level are:

"Operational events should be added to the relevant failure condition statements, creating new combined failure conditions. When considering the combination, it is important to ensure that the operational event is independent from the original failure condition."

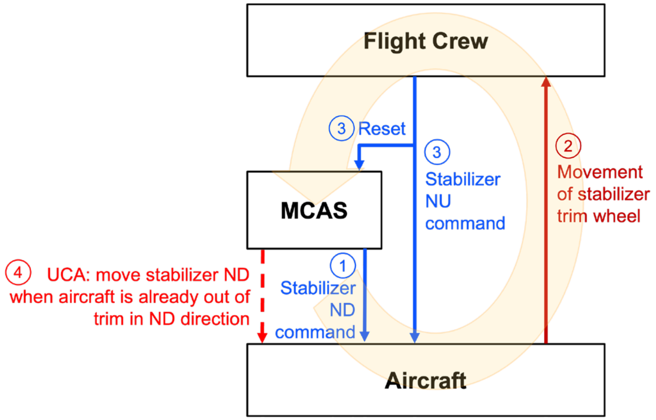


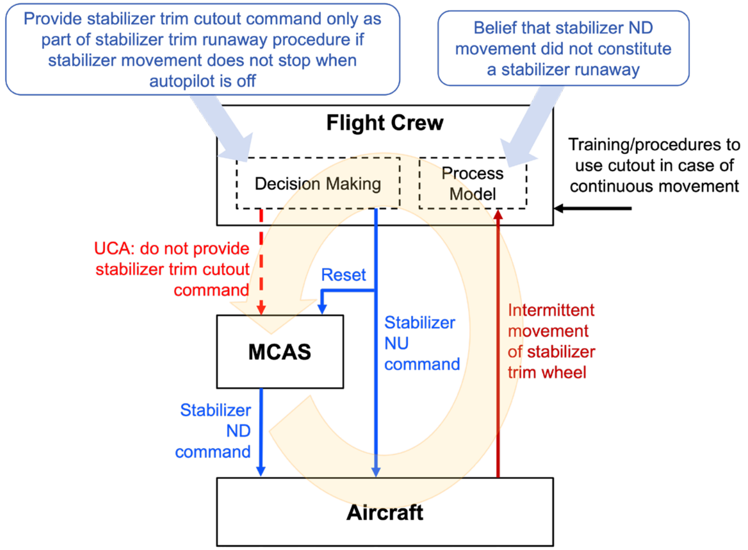
Are these events independent?

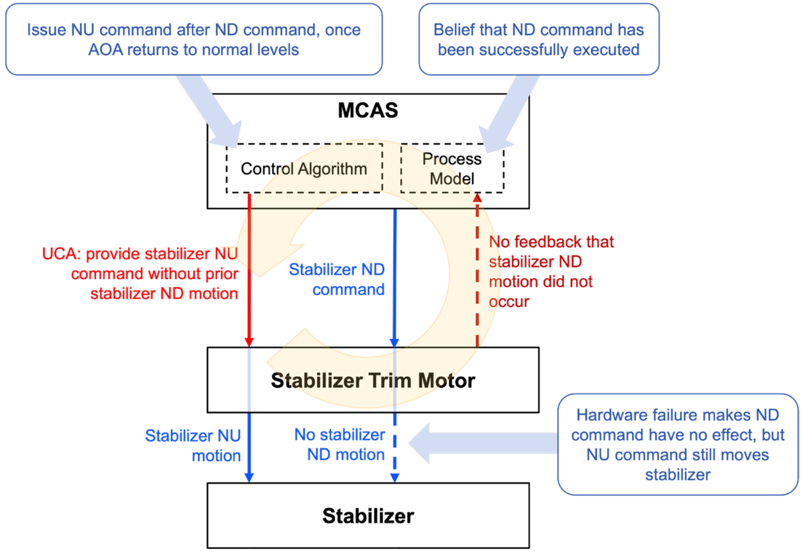
Moving Forward



Can STPA help address some of these gaps?







Thank you!

Questions, Comments, Feedback?

rlrose@mit.edu