Maintenance Check Flight Accidents

Author: Z. Jay Nagy (PhD Candidate, UniSQ) ISASI#MO6955

Jay is an aerospace engineer with over 25 years experience in commercial airline and state aircraft operations. As a Technical Director at Lufthansa Technik, he was responsible for building and operating a new airframe overhaul base. A few years later, Jay was the regulatory post-holder for introducing the new RAAF A330-MRTT fleet into service. His airframe and aircraft systems development experience includes expert positions held at Boeing and Airbus design offices. Jay is in the final stages of completing his doctoral thesis that explores common patterns in airline test flight accidents. He is a member of The Australian Aviation Psychology Association (AAvPA) and the Australian Society of Air Safety Investigators (ASASI).

Abstract

In the wake of serious maintenance check flight accidents and incidents, the airline industry adopted various new regulatory measures and safety improvements. Our research study found common causal patterns across multiple events, enabling a comparison between investigation outcomes and the global safety response. The results highlighted a misalignment between learnings from past investigations and the in-service risk management framework, a conclusion which raises further questions about the scope and suitability of the current safety approach. This paper provides a brief overview of a five-year doctoral research project, offering a novel approach to non-routine safety investigations.

Introduction

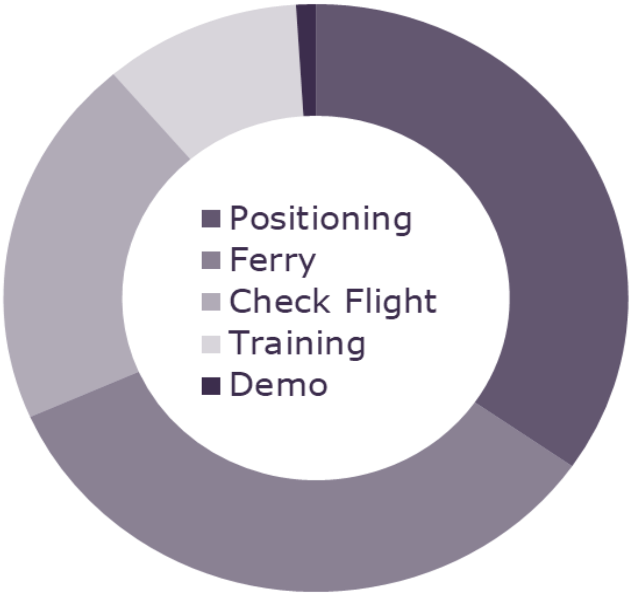
During recent decades, several high-profile accidents highlighted the elevated risk profile of non-routine flight operations (NRFO). A spike in the number of serious accidents (CRJ-200 Jefferson City, A320 Perpignan, B737 Norwich, G650 Roswell) has drastically changed the airline industry's perception of the safety risk associated with non-routine flying (AAIB, 2009; BEA, 2010; NTSB, 2007, 2012).

Non-routine flying is a broad operational category, including positioning and ferry flights, instructional (training) flights, airshow demonstrations, and post-maintenance test flights. Other than the number of non-routine accidents published by Boeing in their annual commercial aviation safety statistics, non-routine flying safety is excluded from regular industry updates (Boeing, 2020). As a result, the research project had to start with building an independent library of non-routine occurrences. A systematic review served as a reliable and repeatable method when sourcing non-routine flying events from air safety investigation reports issued worldwide.

The review identified 99 NRFO accident and incident investigation reports for Western-built commercial jets between 1988 and 2021. Starting the review from 1988 ensured that a complete in-service history is captured for the latest airliner generation. In line with stated research objectives, aircraft with a Maximum Take-off Weight (MTOW) < 60,000 lbs were excluded from the systematic review, limiting the event library to commercial airliners in mainline service.

Figure 1.

Non-routine accidents and incident investigations (1988-2021)



As illustrated in Figure 1., positioning and ferry flight occurrences generated most non-routine investigations over the review period. Compared to the relatively small number of maintenance check flight (MCF) sectors required in support of safely returning to routine airline operations, the number of check flight occurrences was high. Similarly, training flights generated a sizable proportion of non-routine investigations, however the risk profile associated with a less experienced operator at the controls is expected to be higher and may explain that finding.

This paper adopts "*maintenance check flights*" as an inclusive term, covering all non-routine operations with a similar objective, like post-maintenance test flights, functional or operational check flights, post-modification evaluation flights, end of lease, or acceptance flights. It is important to highlight that the term "*check flight*" may also refer to verifying flight crew competency, however those flight operations are not considered in this study.

The European Union Aviation Safety Agency (EASA) offers a comprehensive definition for maintenance check flights as:

"a flight of an aircraft with an airworthiness certificate or with a permit to fly which is carried out for troubleshooting purposes or to check the functioning of one or more systems, parts or appliances after maintenance, if the functioning of the systems, parts or appliances cannot be established during ground checks..."

EASA rules require a maintenance check flight, whenever prescribed in the aircraft maintenance manual (AMM) or other approved data; in the operator's continuing airworthiness management system; or requested by the maintenance organisation for fault isolation, troubleshooting, or verification of successful defect rectification (EASA, 2019).

Federal Aviation Administration (FAA) guidance differentiates between two levels of complexity: functional and operational check flights. In FAA terminology, a functional check flight refers to an in-flight functional evaluation of the aircraft and its systems to a test standard, whereas an operational check flight reflects an in-flight verification of prior maintenance action or ongoing troubleshooting steps. The study incorporates learnings from both functional and operational check flight investigations under the inclusive MCF banner (FAA, 2002).

Whilst the airline industry recognizes that non-routine flying carries an elevated risk level, the actual magnitude associated with maintenance check flights is not that well understood by all stakeholders. Available literature suggests a considerable number of threats unique to check flying which will be explored in the following sections.

In-service experience

As highlighted earlier, assembling a representative catalogue of major non-routine events was the initial step in better defining the non-routine safety problem. The NRFO catalogue, for want of a better term, was a necessary precursor before common themes and characteristics could be identified for maintenance check flight events.

Once the NRFO catalogue was broken down into various non-routine flight operation types, the review identified 7 check flight accidents and 13 check flight incident reports for the Western-built commercial airliner fleet. A summary is provided in Table 1. It is beyond the scope of this article to revisit each MCF occurrence in detail, and as such, those landmark events are highlighted here which were later included in a novel case study.

Table 1.

Maintenance Check Flight accidents and incidents (1988-2021)

ID	EVENT DATE	LOCATION	COUNTRY	MODEL
A01	02-April-1993	Margarita Island *	Venezuela	DC-9-15
I01	29-April-1994	Heathrow Airport	United Kingdom	Concorde
I02	22-October-1995	Bournemouth	United Kingdom	B737-236
I03	29-October-1995	San Francisco, CA	United States	B737-500
A02	22-December-1996	Narrows, VA	United States	DC-8-63F
I04	19-December-1997	Shannon	Ireland	MD-82
I05	12-May-2000	Dublin	Ireland	B747-212B
I06	25-November-2000	Newark, NJ	United States	MD-11
A03	08-November-2002	Salamanca	Spain	A340-313
I07	03-December-2002	Munich	Germany	A300-600
I08	11-March-2004	Fort Lauderdale, FL	United States	A300F4-605R
I09	22-October-2006	London Stansted	United Kingdom	B757-204
I10	21-November-2007	South of France	France	A330-202
A04	27-November-2008	Perpignan	France	A320-232
I11	12-January-2009	Norwich	United Kingdom	B737-73V
I12	07-August-2012	North Sea	United Kingdom	B757-2K2
A05	06-December-2013	Tripoli	Libya	E170
A06	20-November-2014	Dallas, TX	United States	B737-7H4
A07	11-November-2018	Alverca **	Portugal	E190
I13	13-July-2021	Luton	United Kingdom	A319-111

Notes: * No investigation report available. **Validation during ferry flight.

Apr 1993 - DC-9-15, Margarita Island (Caracas), Venezuela: The aircraft departed Caracas on a post-maintenance test flight, carrying eight engineers and three crew members. Twenty-eight minutes into the flight the crew started the test program. A few minutes later, the pilot declared a brief mayday, the aircraft entered an uncontrolled descent and crashed into the sea 16 km off Margarita Island. The aircraft disintegrated on impact and sank to a significant depth. The wreckage was not recovered. To this day, the 1993 DC-9 crash remains the worst MCF accident in terms of fatal injuries (FSF, n.d.).

Dec 1996 - DC-8-63, Narrows, United States: A DC-8-63 freighter was destroyed during a post-modification evaluation flight (check flight), fatally injuring all three crew members and three observers on board. The NTSB determined that the probable causes of the accident were: pilot error and the failure of the airline to establish a formal check flight program (NTSB, 1997).

Nov 2008 - A320-232, Perpignan, France: The airplane operated a post-maintenance check flight, in the context of ending a lease agreement, when it was destroyed upon impact off the coast of Canet-Plage. This major accident involved two airlines and a number of aviation authorities and safety boards worldwide. The Perpignan tragedy created significant interest and concern within the airline industry and the final BEA investigation report played a key role in shaping the current safety response (BEA, 2010).

Jan 2009 - B737-73V, Norwich, United Kingdom: The airplane operated a combined check flight and customer demonstration program, having just completed a maintenance visit, when it experienced a serious in-flight upset and loss of control incident. The airplane violently pitched down and lost approximately 9,000ft before the pilot was able to recover and landed it safely. This serious incident served as another major wake up call to the industry, confirming that the check flight safety problem is not limited to a particular manufacturer or design philosophy (AAIB, 2009).

Industry response

Major investigations highlighted the essential role airframe manufacturers had in preventing check flight safety issues. Findings and safety recommendations primarily focused on ensuring that non-routine operations are supported by appropriate instructions for continued airworthiness, including engineering and operations manuals, flight test schedules, and flight operations advice on-demand.

Some findings also highlighted the manufacturers' reluctance in releasing flight test schedules, unless safeguards were in place to ensure that they would not be held accountable for any in-service accidents and loss. Additionally, the AAIB noted an "*attitude conflict*" between aircraft manufacturers and the airlines, highlighting that while the assessed level of operational risk is similar during check flights and production acceptance test flights, the pilot training and qualification standards applied are not comparable (AAIB, 2009).

In 2011, the Flight Safety Foundation organised a Functional Check Flight Symposium for airline operators to revisit check flight challenges faced by the industry. The forum's steering committee included representatives from all transport category airframe manufacturers. Most manufacturers highlighted their own internal risk control plans for mitigating the risk of experimental test flights and customer acceptance flights, however the forum's key message was firmly focused on reassuring the operators that check flights pose no additional safety risk, if only prepared and properly managed by the airlines (FSF, 2011).

Having accepted a key safety recommendation from BEA and the AAIB, Airbus and Boeing released generic check flight schedules for their in-service models, with the provision that airlines remain fully responsible for adapting flight test schedules to local airworthiness requirements and accounting for any configuration differences.

As highlighted in the Norwich investigation report, and later reflected in rule-making proposals, aviation authorities appear to accept the underlying assumption that the overall safety risk associated with check flying is not comparable to the risk of production test flying activity (AAIB, 2009). Unfortunately, that underlying assumption may not fully take into account that during some check flights airline pilots deliberately need to degrade aircraft systems, or test emergency configurations, which is no different from production acceptance test flight objectives.

Whilst it is true that during normal line operations airline pilots fly in the *"middle-of-the-envelope"* and rely on automation to a large degree, check flying is probably more appropriately described as flying *"near-the-edges"* of the flight envelope. As major investigations listed in Table 1. attest, a combination of unexpected or hidden faults, an inappropriate crew response, startle and confusion, or other systemic factors may push the airplane *"beyond-the-edge"*, an uncharted area for most airline Technical Pilots.

In summary, the research study identified the following problems with the safety response implemented by the airline industry:

- Narrow focus on airline operations and crew competency: improved planning, preparation, and crew training initiatives are a welcome change and proven risk control measures, however there is no objective evidence that the elevated safety risk is exclusively an in-service operational problem.

- Disjointed regulatory development: most safety recommendations were adopted by relevant airworthiness authorities, however extensive delays in the rulemaking and implementation process are symptomatic of a low priority afforded to non-routine flight operations. Despite promising developments, the regulatory framework remains disjointed. To this day, the industry struggles to produce a common set of rules, or even with a common terminology.

- Chasing elusive root causes: over the last 30 years, most member states have markedly improved their compliance with ICAO Annex 13 requirements and recommended practices. Within the scope of the research study, though, it is unfortunate that most MCF investigations were limited to simple root cause analysis techniques or a a linear sequence of events described in a factual account.

- Evidence of hindsight bias: the study also uncovered repeated examples of hindsight bias in MCF investigation reports. With reference to Fischhoff's landmark experiments, knowledge of the outcome was found to change the perceived relevance of event-descriptive data, regardless of the likelihood of the outcome, and the truth of the report (Fischhoff, 1975). Recognising the role of hindsight bias in retrospective accident investigations, it can be seen why it is so tempting for investigators to accept the known outcome as inevitable,

> which also *"simplifies the task of pointing out where people went wrong"*. This approach, however, does not help prevent recurrence (Dekker, 2002).

Our study confirmed earlier research findings that an accident remains (10x-100x) more likely during a check flight, when compared to ultra-safe routine airline operations (Poprawa, 2015).

In an attempt to find a revised approach to resolving the safety gap between routine and non-routine operations, the research project introduced a novel accident analysis framework. The framework can be explained in five main steps, as outlined in the following section.

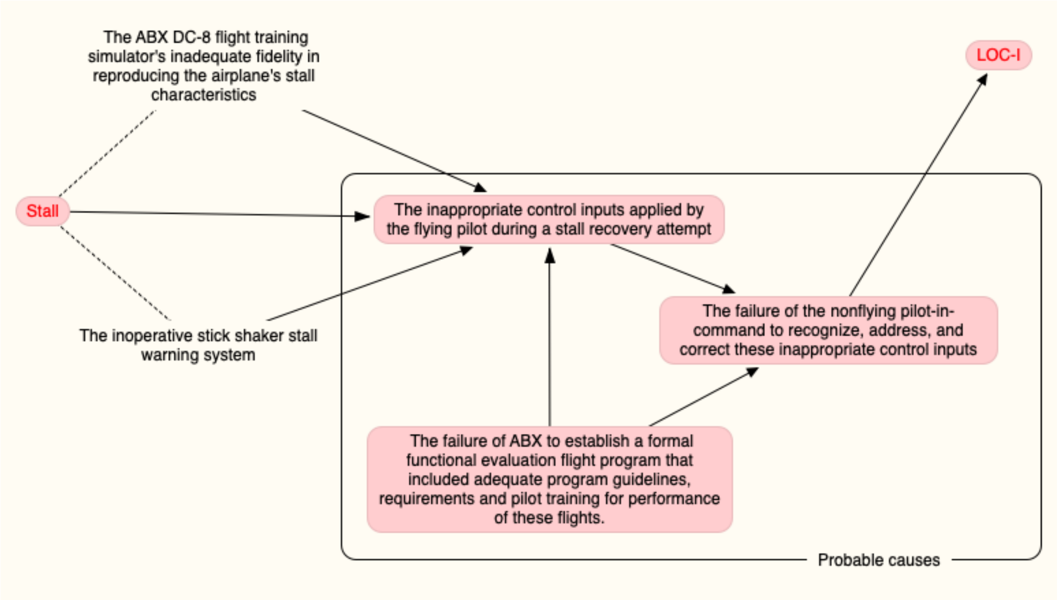
Research study

Step 1. Text analysis - causes and factors extracted from Investigation Reports

While the preferred safety investigation methods may be different, investigation reports produced by transport safety agencies conform to ICAO Annex 13 standards. A common feature of these reports is a heavy reliance on natural text when describing the events leading up to the safety loss, including the investigation team's analysis, findings, annotated causes and contributing factors (ICAO, 2003).

Figure 2.

Text analysis (example)



Providing a description in the investigator's primary language, especially when conveyed to a person who understands that same language, is a powerful tool in communicating the proposed causal structure embedded in the report. Some of that context is lost when the report needs to be translated to English, the common language of the international aviation safety community.

More importantly, as ICAO standards do not prescribe a common terminology for annotating probable causes and contributing factors, dissimilar labels can be assigned to seemingly identical causal linkages in safety investigation reports. The same label may record a cause in one report and refer to a systemic causal factor in another investigation. Therefore, there is only limited value in directly comparing causal labels annotated in MCF investigation reports.

In the first instance, full sentences describing causes and contributing factors were extracted word-for-word from the relevant section of the investigation report. If the report did not have a dedicated section, a simple text search was utilised for collecting causal factors from the document. The text for any safety issues, safety recommendations, and safety actions was also extracted from the report, then compared with already identified causes and factors. This additional step was necessary to obtain a more accurate causal description from the contextual information contained in the report, especially when the relevant safety agency prefers not to nominate immediate causes. Figure 2 illustrates an extract from a text analysis result.

Step 2. Causal groups - causes and factors progressively arranged

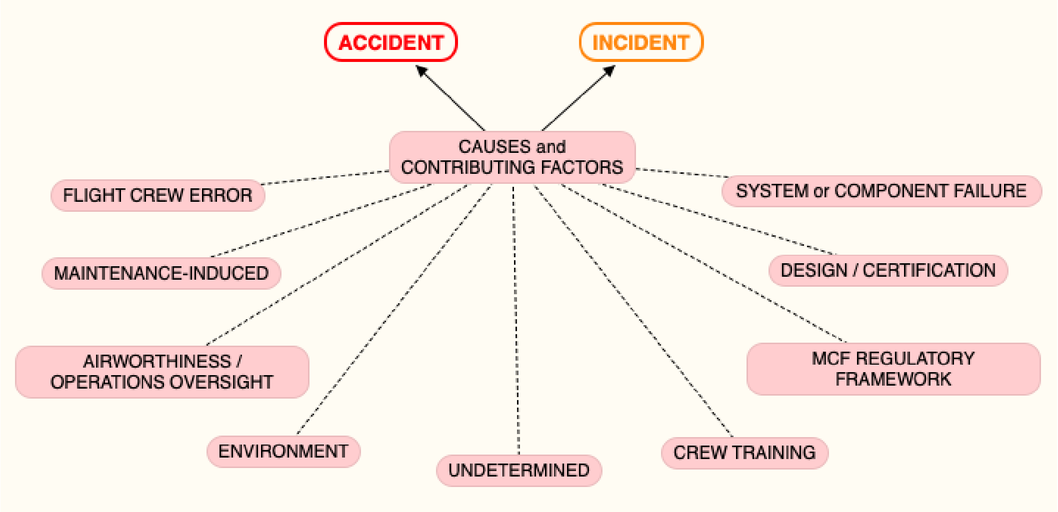
Next, the long-form text was replaced by a short-form causal label. The labels were limited to a few keywords which would be sufficient to describe the investigator's original intent for other safety investigators. Substantially similar causes and factors were assigned a causal group heading based on these short-form labels. Holloway and Johnson describe a similar step in their independent analysis of probable and contributory causes in selected NTSB aviation reports (Holloway & Johnson, 2004).

When the original natural language descriptors are replaced by causal labels, then similar labels are replaced by primary group headings, it is inevitable that some of the original context and meaning is lost during the process. The study identified nine primary causal groups in MCF investigation reports, as seen in Figure 3. The two most frequent causal factors annotated in the reports were *“incorrect pilot action”* and *“component failure”* allocated to *the “Flight Crew Error”* and *“System or Component Failure”* headings, respectively.

In the context of causal reasoning, natural language also offers the ability to express a whole range of potential causal relationships, from almost certain, through neutral, to negative causal effects. This characteristic was essential in building cognitive maps from accident investigation reports, as described in the next step.

Figure 3.

Causal groups in MCF reports



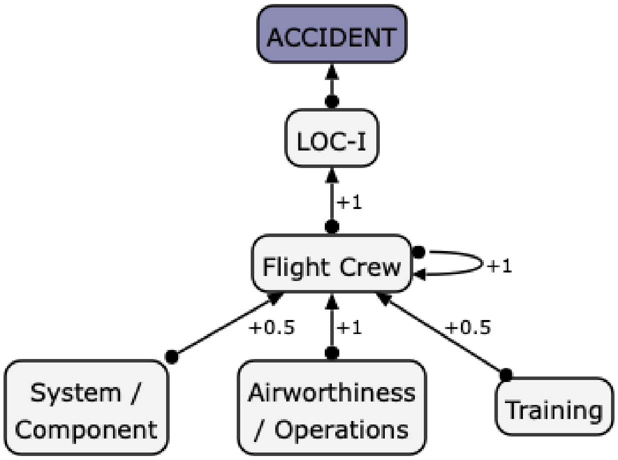
Step 3. Concept maps - causal logic embedded in investigation reports

Concept maps are directed graphs, where nodes (concepts) are linked by edges (arrows). The focus is on describing the relationship between the nodes, supported by labels (typically action verbs) assigned to the directed edges. Unlike in a mind map, a concept node can have multiple parent nodes, a flexible hierarchy which allows for illustrating interconnected graphs. Concept maps are ideal for describing and analysing cause and effect relationships, including propagation effects.

A trivalent causal map was drawn up for each check flight event. Causal concepts and unidirectional arrows between concept nodes reflect the causal logic annotated in the corresponding investigation report. The strength and direction of causal effects is mirrored from the investigation reports, i.e., a +1 sign annotates a (probable) cause, while a +0.5 value signs a contributing factor. Neutral causal links are self-explanatory, indicated by the lack of arrows between concept nodes. A causal map example is provided in Figure 4.

Figure 4.

Causal map example - Accident



The trivalent maps were drawn at a causal group level. It must be highlighted that while these causal maps summarise key investigation outcomes, they cannot be considered as complete accident models. Some finer details about systemic causes, contributing factors, and the interactions between those elements remain hidden at this higher abstraction level.

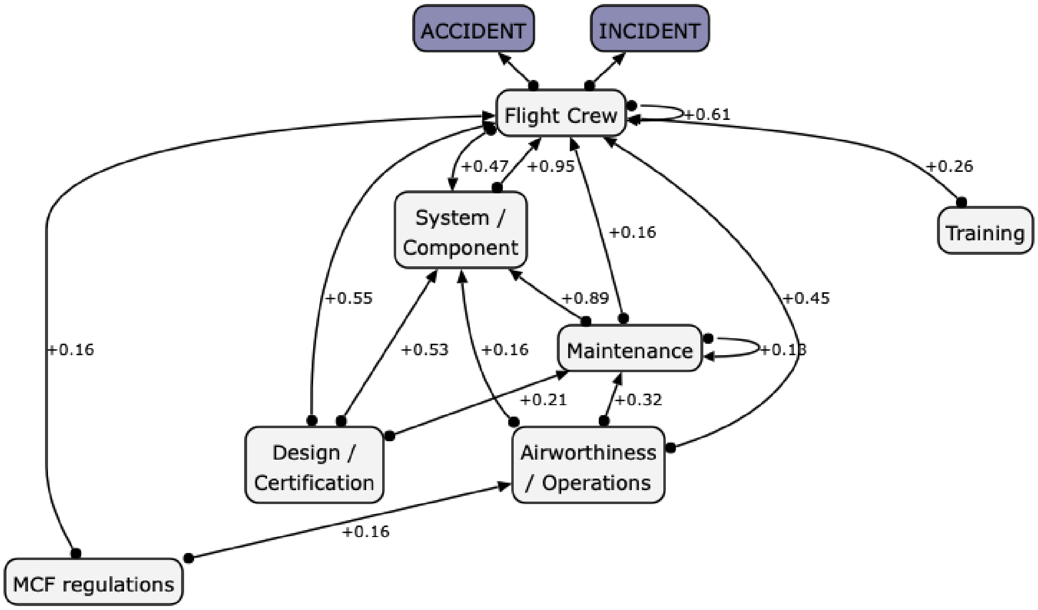
Step 4. Global map - overall knowledge map

Next, the individual causal maps were combined into a global map. Prior to combining the reports, different weights were assigned to corresponding adjacency matrices, accounting for the perceived relative value of causal knowledge elicited from the investigation report. In other words, an extensive systemic investigation report received more credit than a relatively simple factual report or root cause analysis, when compiling the global map. The result is a collective knowledge map about lessons learnt from past MCF occurrences. The global map depicted in Figure 5, however, does not make every single causal linkage visible on the overall graph, only weak and strong patterns are revealed across multiple investigation records.

Looking more closely at the result it is not surprising that strong patterns are revealed between maintenance-induced causes and system or component failures, which ultimately link up with the flight crew cluster, including pilot decision-making, action, or inaction. The map also reveals the role of design- and certification-induced causal factors, which increase the causal effect on system or component failures and factors listed under the flight crew heading.

Figure 5.

Global causal map



Factors in the airworthiness and operations cluster also have an important causal effect on flight crew decision-making and actions, while training has a similar, but somewhat lesser impact. A more surprising finding is the relatively strong internal causal interaction between various causal factors listed under the flight crew heading.

Step 5. Case study - common structure and accident patterns

Once the global map was assembled, it would have been tempting to stop the study and draw some conclusions about the current safety response. That approach would have resulted in identifying a misalignment between what we already know from past MCF investigation reports (i.e., important systemic causes may also emerge from design and certification shortfalls, not only from in-service operational problems) and the “plan and prepare” risk controls implemented by the industry. It is important to highlight that the elevated check flight safety risk should not be portrayed as an in-service risk management problem alone, but that would not have addressed the original research objectives.

Learning from past MCF investigations is very important, however our safety response should not be limited to fixing individual causal trees drawn up for past occurrences. It is extremely unlikely that the same accident or incident pattern would unfold in the same manner, especially when complex systemic causes and factors interact. Furthermore, the global causal map is limited to the knowledge encoded in the original investigation reports. The narrow focus on airline operations and flight crew competency, combined with elusive root causes, unclear stopping rules, and the evident hindsight bias in some MCF investigation reports, heavily influences what we can learn from past occurrences.

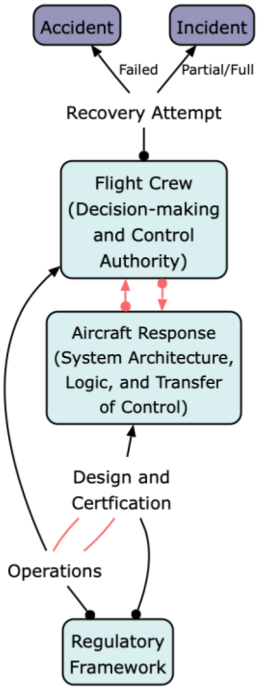
In an attempt to address some of the inherent major shortfalls, the study introduced two new focus questions:

Question 1 - WHY did it make sense for the crew to continue the test program? and

Question 2 - WHY did the recovery attempt work / fail to work?

Figure 6.

Case study results



As a final step, a comparative case study was conducted across the three landmark incidents highlighted in the introduction. Instead of looking for root causes of the safety problem, the case study tried to answer the new focus questions, searching for similarities and dissimilarities between accident patterns. The primary benefit of structuring the case study around the new focus questions is the credit given to both positive and negative causal links, where a negative link refers to a causal decrease, not a negative safety impact.

It is beyond the scope of this paper to fully explain the case study, and as such, only the results are summarised here. Figure 6. provides a graphical aid to the case study outcome:

- Building on the global knowledge map obtained from past investigations, the problem was simplified by accepting that there is a complex interaction of systemic factors between operational and design / certification causal groups.
- The case study also revealed that the regulatory framework has an underlying influence on both the operational and design streams.
- When those systemic factors are all considered, common MCF accident patterns can be best described as a complex interaction between the flight crew's decision making and control authority and the in-flight aircraft response. Aircraft system architecture, the associated system logic, and the nature and timing of transferring control from automated systems to the human pilot appear to be influential in whether the ensuing recovery attempt(s) are successful or not.

Conclusion

Contrary to the current "plan and prepare" approach taken by the airline industry; check flight accidents are not isolated in-service events. MCF accidents and incidents emerge from a systemic interaction of design and operational factors, influenced by underlying shortfalls in the regulatory framework. Check flight accident patterns can be described as a complex interaction between the flight crew's decision making and control authority vs the in-flight aircraft response, further aggravated by any problems when automated systems transfer control to the human operator.

The study also highlighted that current airliner designs continue to rely on the human pilot as a last resort, when an unexpected or previously unknown systemic interaction emerges during a check flight. This finding underlines the need to recognise the human operator's positive safety contribution(s) in accident models, in lieu of simply pointing at the crew when the recovery attempt(s) fail to work.

And finally, the study also confirmed that valuable additional safety lessons can be learnt from combining multiple investigation results, providing the original in-service operational objectives were similar.

REFERENCES

AAIB. (2009). *AAIB Bulletin: 9/2010 SERIOUS INCIDENT Boeing 737-73V, G-EZJK West of Norwich, Norfolk 12 January 2009 at 1545 hrs*. Air Accident Investigation Branch.

BEA. (2010). *Accident on 27 November 2008 off the coast of Canet-Plage to the Airbus A320-232 registered D-AXLA operated by XL Airways Germany*. Bureau d'Enquêtes et d'Analyses.

Boeing. (2020). *Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations 1959-2019*. Boeing Commercial Airplanes.

Dekker, S. (2002). Reconstructing human contributions to accidents: The new view on error and performance. *Journal of Safety Research*, *33*(3), 371–385.

EASA. (2019). Regulations (EU) No 965/2012 and (EU) No 1321/2014. *Official Journal of the European Union*, *L 228/106*.

FAA. (2002). *FSAW 02-12 Guidance Addressing Operational Check Flights Following Maintenance of Air Carrier Aircraft* (Nos. 8300.10; Fourth).

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288.

FSF. (n.d.). *Aviation Safety Net Database*. https://aviation-safety.net/database/databases.php.

FSF. (2011). *Functional Check Flight Compendium*. Flight Safety Foundation.

Holloway, C. M., & Johnson, C. W. (2004). *Distribution of causes in selected US aviation accident reports between 1996 and 2003*.

ICAO. (2003). *DOC9756 AN/965 Manual of Aircraft Accident and Incident Investigation Part IV Reporting*. International Civil Aviation Organization.

NTSB. (1997). *Aircraft Accident Report Uncontrolled Flight Into Terrain ABX AIR (Airborne Express) Douglas DC-8-63, N827AX Narrows, Virginia December 22, 1996* (AAR-97/05).

NTSB. (2007). *AAR 07-01 ACCIDENT REPORT Crash of Pinnacle Airlines Flight 3701 Bombardier CL-600-2B19, N8396A Jefferson City, Missouri October 14, 2004*.

NTSB. (2012). *NTSB/AAR-12/02 ACCIDENT REPORT Crash During Experimental Test Flight Gulfstream Aerospace Corporation GVI (G650), N652GD Roswell, New Mexico April 2, 2011*.

Poprawa, S. (2015). Maintenance test flyingan accident waiting to happen? *The Aeronautical Journal*, *119*(1216), pp 781–790.