LESSONS LEARNED FROM HUMAN OPERATOR INTERVENTION FOR AI NAVIGATION AND FLIGHT MANAGEMENT SYSTEMS

Jomana Bashatah, George Mason University, Fairfax, VA

Lance Sherry, George Mason University, Fairfax, VA

Abstract

The advent of Machine Learning (ML) has made possible the deployment of AI systems to perform complex decision-making formerly performed by "expert" human operators. Although modern airliner avionics were not developed using ML methods, they exhibit the characteristics of complex AI functions, especially regarding the complex decision making required for Flight Planning, Navigation, Guidance and Control functions. To address operational situations that the AI functions are not designed for, human operators are required to monitor and intervene should the AI functions command the vehicle into a hazardous operating region. Accident investigations of modern airliner accidents have cited the failure of the flight-crew to complete the required intervention Standard Operating Procedure in a timely manner. This will be an important issue that must be addressed in the design of the proposed new eVTOL/eSTOl aircraft.

This paper describes a detailed analysis of the SOPs required for use in the operator intervention of the expert system AI functions involved in Part 121 airliner and Part 135 accidents. The analysis identified two main categories of SOP design issues regarding the time it takes to complete the intervention task: (1) designed SOP takes a longer time than the actual allowable time, i.e., it is humanly impossible to complete the intervention task in time and (2) the designed SOP ToP takes reasonable time, but in rare cases the actual time to complete the intervention task is greater than the actual allowable time due to execution problems. The implications of these results are discussed.

Section 1 Introduction

The advent of Machine Learning (ML) has made possible the deployment of AI systems performing complex decision-making formerly performed by human operators. Artificial Intelligence (AI) systems are defined as the "simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision" [1].

Next generation aircraft, such as eVTOL and eSTOL aircraft, have specifically proposed using ML to develop expert system AI Flight Planning, Navigation, Guidance and Control functions.

Despite their advanced autonomous capability, these AI functions will need to be monitored by human operators. The monitoring activities aid with intervention should the AI functions command the vehicle into a hazardous operating region. Initially, the monitoring and intervention may take place by pilots on the aircraft, but eventually, it is feasible for the pilot to be located remotely in a central command center.

Modern airliners are equipped with sophisticated Navigation and Flight Management avionics functions. Although these functions were not developed using ML algorithms, they fit the criteria for AI systems, given the complex decision-making tasks they perform. For example, the standard FMS Vertical Navigation (VNAV) function is an "expert system" that evaluates the state of over 100 parameters to determine which of 360 operational situations the vehicle is in to select the appropriate airspeed, altitude and vertical speed targets, as well as to select the pitch and thrust control models. In fact, the modern FMS VNAV can safely "fly" the aircraft from takeoff to touch-down with only minor human operator intervention for flaps, slats, and landing gear.

Over the past twenty years (2000 – 2020), U.S. airliners have experienced 614 accidents (19 with fatalities) with a fatality rate of < 0.01 fatalities per departure [2]. In a subset of the accidents (61 accident) that have occurred, the accident investigation identified a required intervention by the flight crew during the sequence of events leading up to the hazardous event. The intervention required the flight crew to override the automation when it commanded an inappropriate command by using one or more of the required Standard Operating Procedures (SOPs). The accidents' reports all cite the flight crews with failure to perform the intervention in a timely manner. This raises the question whether the failure was a result of "human error" or the design of the SOP.

This paper describes a detailed analysis of the performance of the SOPs that was required for use in the intervention of the expert system AI functions involved in ten accidents categorized as Part 121 airliner and Part 135. The performance-based analysis identified that all intervention tasks took a longer time to complete before the hazardous event occurred. This was due to two main categories of issues (Figure 1): (1) The designed SOP takes a longer time than the actual allowable time, i.e., it is humanly impossible to complete the intervention task in time and (2) the designed SOP takes a reasonable time, but the actual time to complete the intervention task was greater than the actual allowable time because the cues to initiate the SOP/SOP steps of the intervention task were not salient and/or ambiguous.

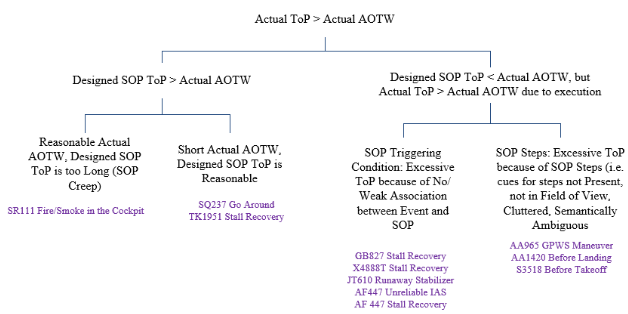

Figure 1. Taxonomy of SOP Issues

The results of the analysis highlight several problems when designing SOPs. The first problem identifies rare scenarios with a low probability of completing the SOP within the AOTW. This is because of the known variability in human performance as well as the variability in the AOTW which prohibits the crew from completing the SOP in some cases, resulting in a significant likelihood of the Time on Procedure (ToP) exceeding the Allowable Operational Time Window (AOTW). The second problem arises from the ambiguous triggering events in the SOP i.e., the starting conditions, that result in a weak association between the event and the required SOP for intervention, as well as ambiguous cues for some SOP steps.

The paper is organized as follows: Section 2 provides an overview of SOPs and SOP performance measures. Section 3 describes the method of analysis of the SOPs in part121/135accidents. Section 4 describes the results of the analysis generating a taxonomy of SOP shortfalls in intervention tasks. Section 5 discusses the implications of these results and future work.

Section 2 SOPs and SOP Performance Measures

What are Standard Operating Procedures (SOPs)

SOPs are universally recognized as essentials to safe aviation operations. The SOPs define flight-crew step-by-step actions to perform the SOP appropriate for the mission event. Mission events include normal operations, abnormal operations, and emergency operations. Abnormal and emergency operations include intervention when the automation commands are not appropriate for the operational situation.

Why are SOPs Important

SOPs define the best response to each situation in each phase of a flight, especially in abnormal, non-normal and emergency conditions. The flight-crew do not have to invent solutions to problems posed by the operational situation but can follow the prescribed SOP.

SOPs also serve to provide a consistent, standardized model of each task that must be performed by each crew member. Effective crew coordination and crew performance depends on the crew sharing a mental model of each task, essentially established by SOPs. They also reflect the airline companies' philosophies and best practices to meet their organizational and operational needs.

How are SOPs Developed and Tested

Airline operators are required to provide flight crews with operating procedures per the applicable regulations as identified in the related regulations section [3]. Original Equipment Manufacturers (OEMs) typically supply airline customers with suggested procedures and checklists that meet the engineering demand of the technology. The OEM supplied procedures reflect the general design of the flight deck and promote the ideal use of the technology [4].

Airlines customize the procedures to meet their company's policies and philosophies [5], or to solve observed incidents and accidents [3].

Modifications of the SOP are approved by the OEM and the Principal Operations Inspector (POI). The POI is an FAA employee assigned to an airline to oversee airline operations, is required to evaluate the procedures with respect to FAA regulations and Advisory Circulars.

FAA Advisory Circular AC 120-92B dictates that all carriers authorized to operate under a Part 121 certificate must have a Safety Management System (SMS). SMSs are a formal, top-down approach to manage risk and assure the effectiveness of risk controls in place. SOPs modified by operators must go through the SMS to be risked out. POIs oversee the process to approve and accept modifications to an SOP.

FAA advisory circular AC 120-71B provides guidance for designing, implementing, and updating SOPs. AC 120-71B describes the characteristics that make for a good SOP (e.g. write steps as imperative, avoid visual clutter, use active verbs). The assumption is that if the SOP meets these characteristics, it will yield a safe and efficient SOP in operational use. It does not take into account the operational environment such as the time available to perform an SOP, the location and salience of visual cues, workload, attention, or fatigue.

Typical industry methods for evaluating SOPs in the SMS include: (i) review by Subject Matter Experts (SMEs) and (ii) human-in-the-loop simulator testing [5]. SMEs tend to be experts in the field and may not represent novice or intermediate pilots. Simulator testing is expensive and requires simulator time as well as flight-crew time. Due to these restrictions, it is cost and time prohibitive to test the procedures for nominal operations as well as off-nominal operations (e.g. wind, aircraft performance, cockpit disruptions/distractions, crew experience, etc.).

It should be noted that AC 120-71B and SMSs does **not** provide instructions on how to test or evaluate the SOP. Most notably, they do **not** mention the role of time as a measure of SOP performance in particular with regard to the Allowable Operational Time Window (AOTW) to complete a procedure before a hazardous event.

Measures of SOP Performance

One way to measure SOP performance is based on time. Time-based measures provide an objective assessment of the performance of the SOP and provide a means for simulation of human performance while taking into consideration all human/environment conditions [6]. The performance of the human operator is captured though the time it takes to complete all operator actions, known as the Time on Procedure (ToP). The ToP is compared to the time in which the SOP must be completed to avoid operational hazards, known as the Allowable Operational Time Window (AOTW). Both AOTW and ToP exhibit variability, due to varying factors of environment/operator conditions.

The ToP and AOTW are related through the Procedure Buffer Time (PBT), which is defined as the simple arithmetic relation of the difference in AOTW and ToP. The PBT is governed by a distribution, and the left tail of this distribution, where $PBT < 0$, is the Probability of Failure to Complete.

The simulation of human performance results in the distribution shown in blue in the upper left of Figure 2. The AOTW distribution, reflecting variability in the operational environment is shown in orange in Figure 2 upper left. The overlap between the ToP and AOTW is summarized by the distribution in the PBT shown in the upper right in Figure 2. The percentage of scenarios where the ToP is greater than the AOTW is the red bar in Figure 2. A big picture view of the ToP vs AOTW is shown in the bottom of Figure 2.

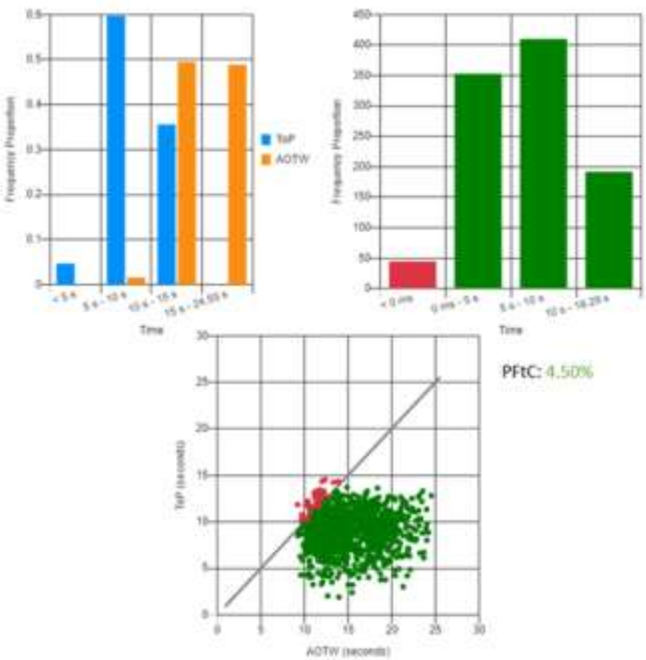

Figure 2. An Example Set of Data for a Given SOP. Human Performance (Top Left: Distributions of ToP & AOTW, Top Right: Distribution of PBT, Bottom: Scatterplot of ToP and AOTW)

Section 3 Method of Analysis

Ten (10) published accident reports were analyzed. For each accident, the following information was identified:

(1) SOPs used for Human Operator Intervention

(2) The AOTW was calculated by examining the timeline of events from the first cue of the triggering event until the hazardous event

(3) The SOP was modeled in an MBSE tool and simulated to obtain the designed SOP ToP.

(4) The PFtC was calculated for the accident scenario.

(5) The accidents were categorized by the relationship between accident AOTW, the designed SOP ToP and the actual ToP.

Modelling Accident SOPs

For step 3 and 4, the SOP performance was modeled using LML/SysML Action Diagrams (ADs). ADs are used to model dynamic behavior and carry the added benefit of being executable. SOP Action Diagrams (SOP ADs) consist of Actors, Actions, and I/Os. Actors are depicted by branches in the diagram (Figure 3). Actions are depicted as squares on each branch. I/Os are depicted by parallelograms and represent the information flow between two actions.

SOP ADs can be simulated in a Model-Based Systems Engineering (MBSE) tool, such as Cameo, Capella, Core, or Innoslate, to provide a means for quantifying the time to complete a procedure. [7] provide a detailed step by step process to model and simulate the SOP using LML Action Diagrams.

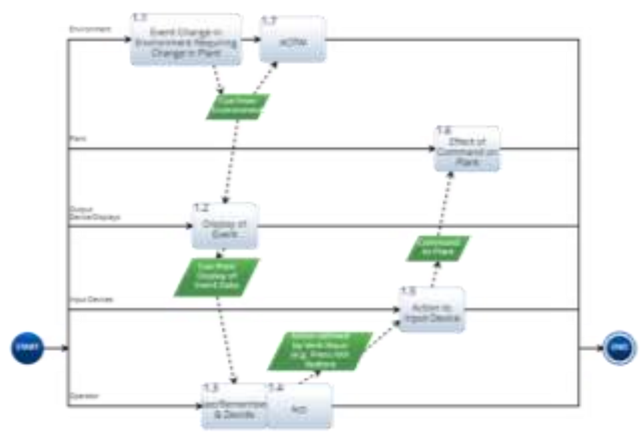

Figure 3. A Generic SOP Action Diagram Modeled in Innoslate

To obtain the Designed SOP ToP, the SOP was modelled using SOP Action Diagrams (SOP ADs). The SOP AD model was then simulated for 10,000 runs, and the resulting metrics were captured. Figure 4 shows a complete SOP AD model for the Asiana 214 (OZ 214) accident. Highlighted in red is the accident AOTW. The blue "actions" show the SOP actions modeled, and the light blue actions show the steps leading to the accident.

The time distributions used in the simulation were obtained from the Human Performance Time Distribution (HPTD) database [6][7][8].

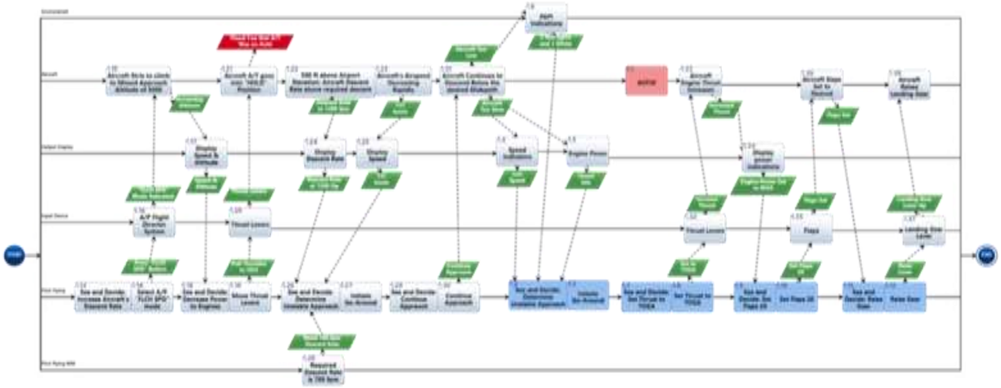

Figure 4. SOP AD Model of Asiana Airlines Flight 214 Accident

Section 4 A Taxonomy of Intervention SOP Issues

Analysis of Ten (10) Part 121/Part 135 modern airliner accidents identified the taxonomy of SOP issues based on the relationship between AOTW, the designed SOP ToP and the actual ToP in Figure 5.

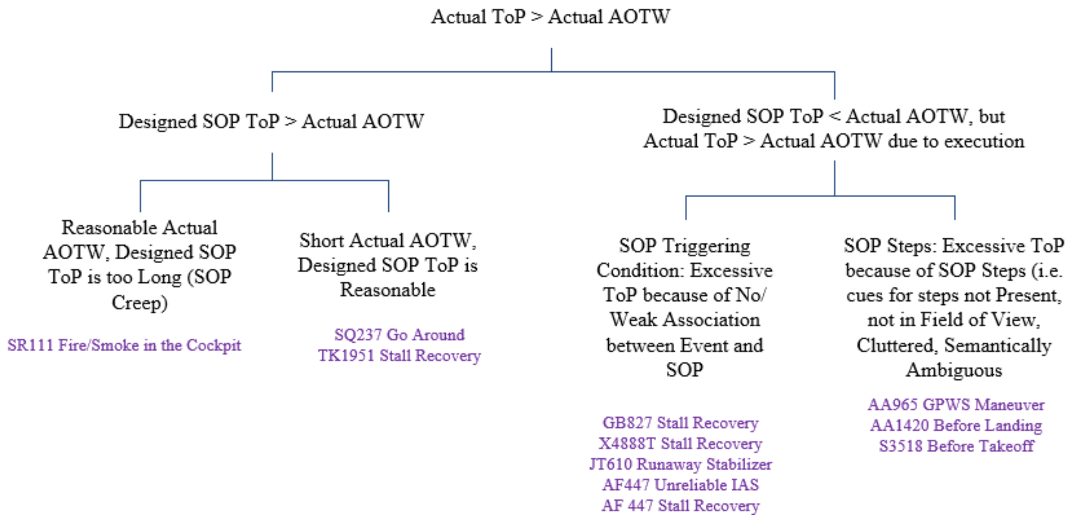

Figure 5. Taxonomy of SOP Design Issues as Identified from Human Operator Intervention Tasks in Accident

For all accidents, the time to complete the intervention task was greater than the actual time the crew had to complete the task (Actual ToP < Actual AOTW). This was due to two main categories:

1. Exogenous Factors (unrelated to the design of the SOP): Accidents where the **designed SOP ToP > Actual AOTW** occurred when there was no time to complete the intervention SOP because of:
    1. The paradox of time allowed vs time needed: An SOP that is impossible to complete in a timely manner i.e. smoke/fire in the cockpit.
    2. Very short AOTW even for a one-step SOP i.e. push go-around button.
2. Endogenous Factors related to the design of the SOP: Accidents where the **designed SOP < Actual AOTW** occurred when there was enough time to complete the intervention SOP, but the crew was unable to complete the intervention because:
    1. There was a weak or no association at all between the SOP required and the

event i.e. pre-stall buffet recognized as clear air turbulence.

2. The cues for an SOP step were not present, not in the Field of View (FoV), lost in the clutter, or were semantically ambiguous.

1. Designed SOP ToP > Actual AOTW

Three (3) of the accidents were identified in this category (Table 1). The flight crew had between 10 seconds (short actual AOTW) to 18 minutes (reasonable actual AOTW) to complete the SOP. By modeling the SOP using time distributions from the HPTD database, the simulations showed that in each accident scenario, the actual AOTW occurred within the 10th %-tile of the designed SOP ToP distribution, meaning that all accidents in this category fell in the left tail of the designed SOP ToP distribution.

Table 1. Accidents where Designed SOP ToP > Actual AOTW

Accident	SOP	Actual AOTW	Simulated ToP	Classification	Percent of Designed SOP ToP > Accident AOTW
SR 111	Air Conditioning Smoke, Smoke/Fumes of Unknown Origin	18.1 minutes	Triang(15.1, 21.5, 28) minutes	Reasonable Actual AOTW, Designed SOP ToP too Long	96.67%
SQ 237	Go-Around	10 seconds	Triang(6.5, 20.7, 35.3) seconds	Short Actual AOTW, Reasonable SOP ToP	99.78%
TK 1951	Stall Recovery	16 seconds	Triang(15.8, 26.8, 35.6) seconds	Short Actual AOTW, Reasonable SOP ToP	100%

1.1 Paradox of Time Allowed Vs Time Allotted

"Air Conditioning Smoke" checklist and the "Smoke/Fumes of Unknown Origin"

The poster child for the first leaf node of the taxonomy is the SR 111 accident. On September 2nd, 1998, the MD-11 departed New York heading for Geneva, Switzerland. Fifty-two (52) minutes after departure (at 01:10:38), the pilots detected some smoke present in the cockpit. Swiss Air SOPs include two checklists in the case of unusual odor/smoke present in the cockpit: the "Air Conditioning Smoke" checklist and the "Smoke/Fumes of Unknown Origin" checklist.

The "Air Conditioning Smoke" checklist was initiated a little over two minutes after the initial smoke presence (01:12:52). The first checklist concluded with referring the pilots to the "Smoke/Fumes of Unknown Origin" checklist, which was initiated around 01:23:45. Two minutes into the checklist, the Flight Data Recorder (FDR) recorded failure of the instruments. Just eighteen minutes after the initial detection of smoke, the aircraft struck the Atlantic Ocean, killing all 229 people on board (01:31:18).

The "Air Conditioning Smoke" checklist and the "Smoke/Fumes of Unknown Origin" checklist were simulated in sequence starting with the former. During the accident, "The first item on the Smoke/Fumes of Unknown Origin Checklist … was carried out about 13 minutes after the odor first became apparent" [9][1] . The two checklists typically "could take 20 to 30 minutes to complete" [9][2] . The actual AOTW for this accident started when the pilots realized the presence of smoke and ended with the aircraft striking into the water (1084 seconds, or 18.067 minutes). With fire being a strictly random phenomenon, there was an insufficient time to run both checklists, due to the creep of SOP steps over time.

[1] Transportation Safety Board of Canada (2002) p. 232

[2] Transportation Safety Board of Canada (2002) p. 255

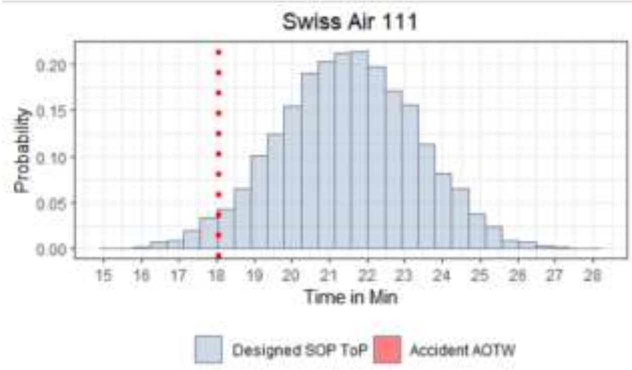

Figure 6. Results of the Swiss Air 111 SOP AD Simulation

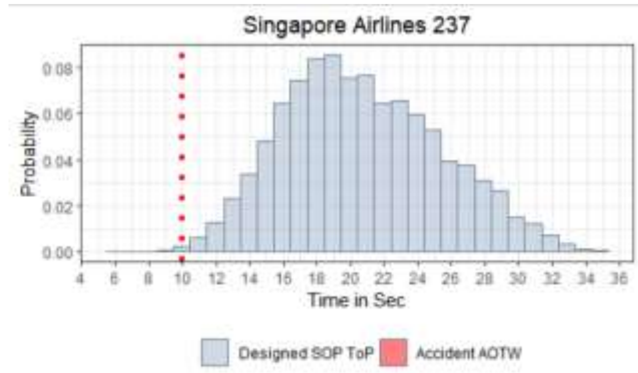

Figure 7. Results of the Singapore Airlines 237 SOP AD Simulation

1.2 Very Short AOTW

Go Around

The crew on the Boeing 777 on approach to Munich at 23:58:42 received landing instructions for the approach to the Instrument Landing System (ILS) for Runway 08R. At 24:08:47, they received clearance to land. Fifty feet above the runway, the airplane started to bank to the left. The aircraft landed 420 m beyond the threshold with the left main landing gear, and the right main landing gear touched down three seconds later. As a result, the aircraft veered to the right, crossed the runway to the other side before coming to a stop on the grass to the left of the runway.

Once the crew noticed that the aircraft banked to the left, the crew initiated a "Go-Around" by pushing the TO/GA buttons on the thrust levers and retracting the spoilers. The crew made the discovery only 30 ft above the ground, leaving them no time to complete a go-around. The crew didn't notice that the TO/GA button was disabled, since the left main landing gear has already touched down.

The AOTW for the SQ 237 accident started when the aircraft banked the left and ended when the aircraft's left main landing gear touched the ground (about 10 seconds). With only ten seconds, there was not enough time to initiate a "Go-Around" before the Auto Flight System (AFS) switched to rollout mode.

1.2 Stall Recovery

The crew on the Boeing 737 initiated the pre-discussed landing approach. As a result of a faulty radio altimeter, the crew heard multiple aural warnings during their approach. The warnings were attributed to the landing gear, and they were treated as a nuisance as the aircraft was still thousands of feet above the ground. As the crew continued their approach, they had to approach the glideslope signal from above, an uncommon yet possible approach. For this approach, the aircraft must slow down and descend rapidly, so the crew selected the "vertical speed" mode. As a result of this change, the throttles went into "RETARD" mode, a mode in which the engines are idle, and the nose is slightly pitched up. With the low airspeed and the slightly pitched nose, the stick shaker activated to warn the crew of an imminent stall. At only 460 ft above the ground, the crew initiated the stall recovery maneuver, only 16 seconds before the crash. At that point, the aircraft was too low and hit the ground killing 8 people and injuring 117 more.

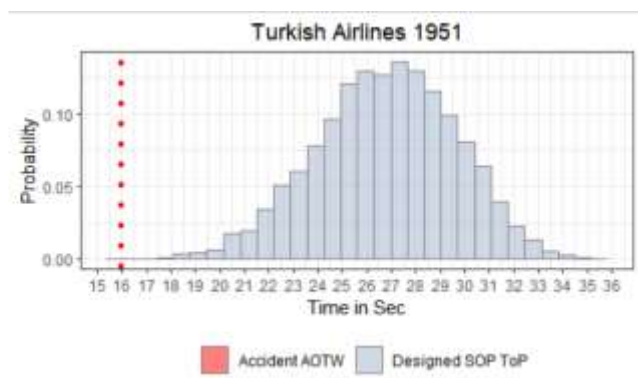

Figure 8. Results of the Turkish Airlines 1951 SOP AD Simulation

1. *Designed SOP ToP < Actual AOTW*

Seven (7) accidents were categorized as having a designed SOP ToP less than the actual AOTW. This was due to (1) the weak association between the triggering event and the SOP and (2) the SOP steps having ambiguous cues. Accidents in this category had the characteristic of having an actual AOTW that occurred within the 90th %-tile of the designed SOP ToP, meaning that the accidents fell into the right tail of the designed SOP ToP distribution.

2.1 SOP Triggering Condition: Weak/No Association Between Event and SOP

Four (4) accidents fell into the category of a weak association between the triggering event and the required SOP for intervention.

Table 2. Accidents with SOP Triggering Problem

Accident	SOP	Actual AOTW	Simulated ToP	Classification	Percent of Designed SOP ToP > Accident AOTW
GB827	Stall Recovery	92 seconds	Triang(28.9, 76.6, 133.9) seconds	SOP Triggering Condition: No/Weak Association between Event and SOP	3.98%
X4888T	Stall Recovery	61 seconds	Triang(26.8, 101.8, 178.5) seconds	SOP Triggering Condition: No/Weak Association between Event and SOP	9.87%
AF447	Stall Recovery	217 seconds	Triang(45.9, 167, 300) seconds	SOP Triggering Condition: No/Weak Association between Event and SOP	9.32%
AF447	Unreliable IAS	46 seconds	Triang(33.3, 87.5, 165.2) seconds	SOP Triggering Condition: No/Weak Association between Event and SOP	1%
JT610	Runaway Stabilizer	7.95 minutes	Triang(2.6, 5.7, 9.7) minutes	SOP Triggering Condition: No/Weak Association between Event and SOP	4.07%

Stall Recovery

The crew on the DC-8 were evaluating the performance of the aircraft after a series of modifications. The stall testing procedure was to first calculate the speed at which the aircraft will stall and note the speed and time at which the stick shaker would activate. With that information on hand, the crew began to slow the aircraft approaching the stall speed. At a speed earlier than what was calculated, the crew acknowledged a buffet sensation which was not recognized as a pre-stall buffet. During the seconds after that, the flight engineer on board recognized that the aircraft was entering a stall, but the stick shaker hadn't activated. Two seconds after recognizing that they were in a stall, the crew of Airborne Express 827 started to perform the stall recovery procedure by setting the engine power to max but kept the nose of the aircraft up. Not having the stick shaker as a trigger, the crew entered the aircraft into an imminent stall which they were not able to recover from. This led to the aircraft striking terrain and killing everyone on board.

The stall recovery procedure in this case should have been triggered by the stick shaker. With the absence of the stick shaker, the high angle of attack and the pre-stall buffet should have acted as the triggering cue. The high angle of attack was obscured due to the absence of the natural horizon, and the pre-stall buffet was not recognized because of confusion. This led to an excessive ToP as a result of the delay in initiating the procedure.

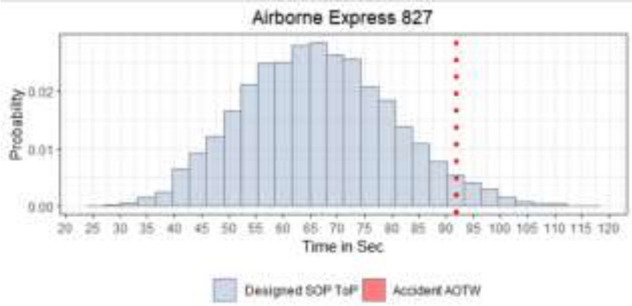

Figure 9. Results of the Airborne Express 827 SOP AD Simulation

The crew on the A320 were on an acceptance flight, as the aircraft was scheduled to return to its owner at the end of a lease agreement. As the crew needed to test the stall protection system, they attempted to deliberately enter the aircraft into a stall by idling the engines and disengaging auto thrust. The crew were sure that the aircraft systems would kick in, and immediately pitch the nose of the aircraft down and increase thrust to the engines. The sophisticated aircraft systems, however, did not activate, and the aircraft entered an impending stall without any recognition from the crew. Two minutes later, the stall warning sounded, and the crew set the thrust levers to TOGA and pitched the nose down. One important cue, however, was masked; the fact that the flight control had passed to direct law, which requires pilot input to the horizontal stabilizer. "The manual use of pitch trim, which is not included as a reminder in the approach-to-stall procedures, only occurs very rarely in operation and occasionally in training" [10][3].

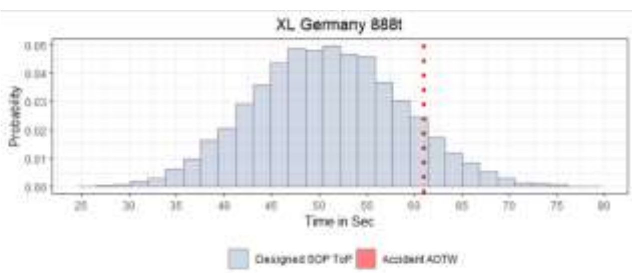

Figure 10. Results of the XL Germany 888t SOP AD Simulation

The crew on the Airbus A330 took off from Rio de Janeiro heading to Paris France. Entering stormy weather, the pilot reduced the speed in anticipation of turbulence. Two hours into the flight, the auto pilot and auto thrust disengaged as a result of the pitot tubes freezing. Of the two co-pilots, the less experienced pilot took command of the aircraft. In an attempt to maintain speed, the pilot pulled back on his side stick instructing the aircraft to climb and raised the nose of the aircraft. Because the power to the engines had been decreased, the pilot's actions entered the aircraft into a deep stall position and triggered the aural message "Stall" and warning alarm 75 times. This situation called for a "stall recovery" procedure, by initiating full thrust on the engines and pitching the nose of the airplane down. Confused about the airplane's trajectory, the pilot continued to pitch the nose up. Shortly after, the aircraft struck the Atlantic Ocean killing all 228 on board. The stall warning "did not elicit any response from the crew" [11][4]. After the warning, any nose down input on the side stick could have resulted in the aircraft's recovery from the stall. The un-linked side stick of the Airbus made it impossible for the pilot not flying to know of the pilot flying's nose-up inputs.

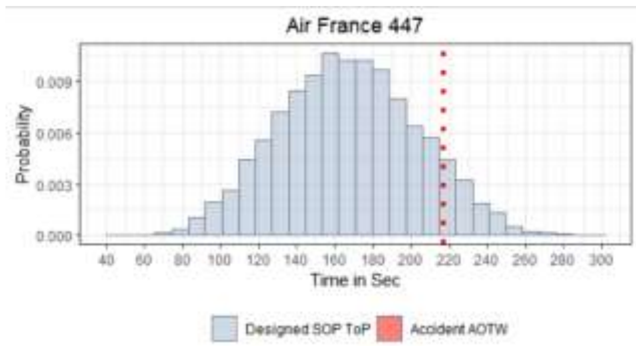

Figure 11. Results of the Air France 447 Stall Recovery SOP AD Simulation

[3] BEA (2010) p. 101

[4] BEA (2012) p. 179

Figure 12. Results of the Air France 447 IAS Disagree SOP AD Simulation

Runaway Stabilizer

As the Boeing 737 Max was taking off from Jakarta Indonesia, the pilot's stick-shaker activated, indicating that their aircraft was about to stall as a result of an uncalibrated Angle of Attack sensor. As they continued to climb, the pilots received a "IAS DISAGREE" warning on the Primary Flight Display (PFD). Less than a minute later, the pilots acknowledged that there was an altitude disagree as well. While attempting to complete the "IAS Disagree" procedure, the aircraft entered a series of nose dives (26 automatic trim commands), and manual nose-up commands (34 manual trim) by the crew. The automatic nose-dives were a result of the Maneuvering Characteristics Augmentation System (MCAS) activation, a system that the pilots were not aware of. To offset the commands of MCAS, the pilots would have to recognize that the trim wheel was moving erroneously, and that they would disengage the auto-trim in the "RUNAWAY STABILIZER" non-normal procedure. At no point during the chaos did any of the pilots recognize an abnormal trim wheel movement, though the movement of the wheel could be heard in the CVR and initiate a "RUNAWAY STABILIZER" procedure.

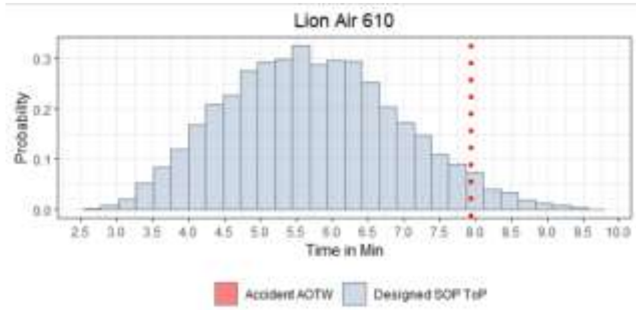

Figure 13. Results of the Lion Air 610 SOP AD Simulation

2.2 SOP Steps: Cues not Present, not in Field of View, Cluttered, or Semantically Ambiguous

Three (3) accidents were categorized as having unclear cues for the steps in the SOP required for intervention. In all the scenarios, the reports cited that the crew "forgot" to complete a step in the procedure. The analysis identified that in all these scenarios, there was not a clear, unambiguous cue to remind the crew to complete the steps.

Table 3. Accidents with Ambiguous Cues for SOP Steps

Accident	SOP	Actual AOTW	Simulated ToP	Classification	Percent of Designed SOP ToP > Accident AOTW
AA965	GPWS Maneuver	13 seconds	Triang(10.9, 54.1, 126.4) seconds	SOP Steps: Cues not Present, not in FoV, Cluttered, or Semantically Ambiguous.	37.3%
AA1420	Before Landing	206 seconds	Triang(51, 131, 259.4) seconds	SOP Steps: Cues not Present, not in FoV, Cluttered, or Semantically Ambiguous.	2.87%
S3518	Before Takeoff	160 seconds	Triang(60.7, 109.8, 186.6) seconds	SOP Steps: Cues not Present, not in FoV,	4.13%

				Cluttered, or Semantically Ambiguous.	

GPWS Maneuver

The crew on the Boeing 757 were cleared for a VOR DME approach to Runway 19 in Cali, Columbia. Without a radar, ATC relied on aircrafts to report their distance to the runway each time they pass one of the waypoints. After erasing all waypoint information from the computerized systems, the pilots attempted to re-program one of the waypoints. However, the pilot entered the wrong waypoint information, and the aircraft started to veer off the programmed course, and into a deadly terrain course. As the aircraft continued to descend, the crew was unaware that the aircraft was heading straight to the mountains. Thirteen seconds before impact, the Ground Proximity Warning System (GPWS) sounded with the infamous "terrain, terrain, whoop, whoop" sound. As a result of the warning, the crew added thrust and pitched the node of the aircraft up, however, they forgot to retract the already extended speed brakes. The GPWS maneuver was unsuccessful, as the crucial step to lift the aircraft was not completed. The aircraft struck terrain and killed all but 4 people on the aircraft. The missing step to retract the speed brakes led to a long ToP as there was an assumption that "the flightcrew should have recognized that the spoilers were still extended during the attempt to avoid the terrain and should have retracted them early in the escape maneuver" [12][5].

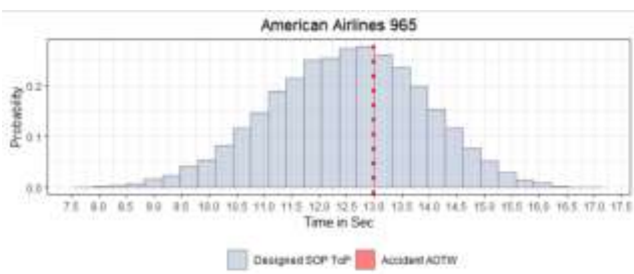

Figure 14. Results of the American Airlines 965 SOP AD Simulation

Before Landing

The crew on the MD-82 took off and was expected to land in 4 minutes in Little Rock, Arkansas, where the weather was progressively getting worse. ATC cleared the aircraft for an ILS approach, since visibility at the airport was very low. Visibility was less than a mile, and there was heavy rain at the airport, yet the crew decided to continue their landing approach. As the crew were completing the "Before Landing" checklist, they forgot to deploy the spoilers, a step that that is crucial to help the aircraft come to a stop. At the time of the accident, the checklist did not require any call-outs or visual confirmation of the spoilers being deployed and "the captain failed to manually extend the spoilers when they did not deploy" [13][6]. The report stated that "the flight crew's failure to detect that the spoilers had not deployed might have been avoided if a procedural requirement similar to the one in Boeing's MD-80 FCOM had been in place at the time" [13][6].

The "Before Landing" Checklist requires the crew to arm spoilers only after the landing gear has been down, and all three green lights have been illuminated. This is not an instantaneous step, i.e. variable time to complete this verification step. While the crew was waiting for verification, ATC gave the crew a windshear alert, thereby interrupting the checklist. The crew never went back to complete the checklist. There was no cue to remind the crew that the aircraft was not prepared for landing (no spoilers). Furthermore, the "Before Landing" Checklist did not contain a "Verification" step, where the flight crew visually confirm that the spoilers had been deployed.

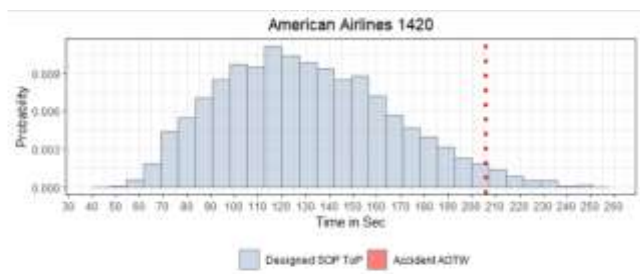

Figure 15. Results of the American Airlines 1420 SOP AD Simulation

Before Takeoff

The crew on the ATR 42 arrived late the aircraft and started to rush through the checklists to takeoff as soon as possible. Typically, the crew on an aircraft

[5] Aeronautical Civil of the Republic of Columbia (1996) p. 47

[6] NTSB (2001) p. 134

would start the engines, then complete all the required checklists, a process would typically leave enough time for the aircraft systems to initialize. Because the crew had rushed through the checklists, and missed some steps, the crew on the ATR 42 took off after 2.5 minutes of starting their engines. Not knowing that the aircraft needs a minimum of 180 seconds to initialize the Attitude and Heading Reference System (AHRS), the crew took off with their systems offline and flew blind. Shortly after takeoff, the aircraft struck terrain without the crew's knowledge of the altitude or where the aircraft was heading.

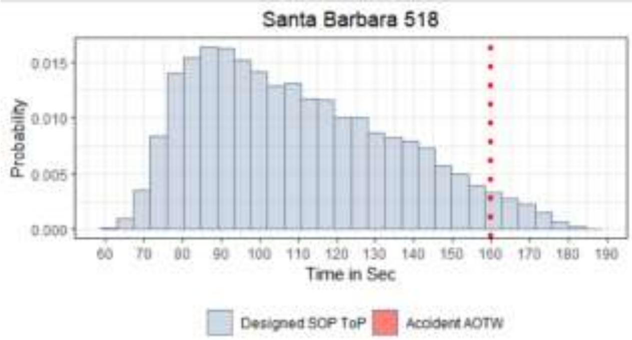

Figure 16. Results of the Santa Barbara 518 SOP AD Simulation

Section 5 Conclusions

This paper reports the results of the analysis of ten accident scenarios identified events that required flight crew intervention. The flight crew intervention was defined by Standard Operating Procedures (SOPs) that defined the situations to initiate and the sequence of steps to mitigate the hazardous operational situation.

The analysis was based on a time-based model of SOP performance that is based on the Available Operational Time Window (AOTW) and Time-on-Procedure (ToP) [6]. The analysis used a model-based simulation of SOPs using empirical human response times [7][8].

The analysis identified two mutually exclusive categories of intervention events. The first category is when the designed SOP of the intervention took more time than the AOTW. This happened because (1) the SOP kept getting modified and over time, the SOP became too long (SOP Creep) or (2) there was simply not enough time for a crew to complete the SOP even in nominal situations. In both cases, the analysis found that it was humanly impossible for a crew to complete the required SOP before impact; i.e., this category of accidents was unpreventable.

The second category identified scenarios where the designed SOP ToP for the intervention took less time than the actual AOTW, but execution conditions specific to the accident led to an increased actual SOP ToP. The conditions included (1) ambiguous triggers to initiate the required SOP or (2) ambiguous cues for SOP steps i.e. no reminders of missing steps. In this category, the accidents could have been prevented, given more clear triggering conditions and/or cues in the SOP.

Impossible Human Operator Intervention: SOP Creep and Unrecoverable Zones

SOP Creep

In the case of smoke in the cockpit, the SOP was too long given the variability in time for smoke to transition to fire (SR 111). This was mainly due to the fact that the SOPs required for the intervention kept getting modified though time, and no testing was conducted to ensure that the steps could be completed in time. The "regulations do not require that checklists for isolating smoke or odours that could be related to an overheating condition be designed to be completed in a time frame that minimizes the possibility of an in-flight fire being ignited or sustained." [9][7].

Unrecoverable Zones for Flight Crew Intervention

The analysis identified zones in which a crew intervention would not be beneficial, and an accident/incident is imminent. These zones are called "Unrecoverable zones", and are characterized by (1) a stall recovery on final approach with or without a stick-shaker or correct V-stall calculation, or (2) Autoland roll due to localizer deviation close to touchdown. For the two accidents in this category, there was at least a 95% probability of failing to complete the procedure according to the human performance simulations of the accidents and their accompanying procedures.

In the case of avoiding a landing off the runway center line, there was not sufficient time to reasonably allow the flight crew to respond with a go-around (SQ

[7] Transportation Safety Board of Canada (2003) p. 253

237). In the case of a stall recovery, the aircraft was simply too low near the ground for any response to be successful (TK 1951).

Possible Human Operator Interventions: Ambiguous Triggering Conditions and Cues for SOP Steps

Ambiguous Triggering Conditions

The analysis also identified accidents where the starting condition of the SOP was unclear. This led to a disassociation between the hazardous event and the required procedure for intervention. The disassociation caused an abnormally long ToP, longer than the assumed ToP of the designed SOP. In the case of unreliable airspeeds, the "crew of flight AF 447 did not associate the disappearance of the speed information and the ECAM messages associated with the "Unreliable IAS" procedure" [11][8].

The disassociation occurs because of two reasons; either because the cues to initiate the SOP are from the external environment (ABX 827), or the cues to initiate the SOP are from a cockpit alert, but there is no association between the cue and the required procedure (XL Germany, Lion Air). Special care must be taken when designing the SOPs to recognize both situations. Furthermore, many SOPs do not define the triggering conditions for the SOP, or if they do, they are vague assertions.

These triggering conditions must be highlighted and trained for especially for rare hazardous scenarios. It may also make sense to develop AI/ML systems to detect these situations. Of course, this is problematic, as the AI/ML needs a large set of data to perform the AI/ML training and testing, and these events are, by definition, rare events.

Ambiguous Cues for SOP Steps

In the case of a GPWS Maneuver (AA 965), the designed SOP assumed that the flight crew would remember that the spoilers were extended during intervention, therefore, the step to "retract spoilers" was omitted from the SOP. The oversight of adding the step led the crew to forget this important step in the heat of the moment.

In the case of deploying spoilers before landing (AA 1420), the SOP requires that the spoilers be deployed only after the landing gear verification has been noted. This verification step takes a variable amount of time. The interruptions during the waiting period led to a missed SOP step. The design of the SOP did not aid the crew in recognizing the fatal error that had occurred, as there was no requirement of callouts or visual confirmation of deploying the spoilers.

In the case of initializing aircraft systems before takeoff (SB 516), the SOP did not highlight the requirement to start the aircraft and wait a minimum of three minutes before takeoff. The SOP design assumed that crew would typically wait this time to complete the required checklists.

Not Enough Time to Complete SOP Due to Variability in AOTW and ToP

This analysis identifies rare scenarios in which the flight crew have low probability of completing the SOP within the AOTW (SQ 237, TK 1951). This is not to say the SOP cannot be completed with a ToP less than the AOTW, but that given the known variability in AOTW as well as the human performance when completing SOPs, there is a significant likelihood (i.e., Probability of Failure to Complete Task – PFtC) of the ToP exceeding the AOTW.

It is imperative that the variability in these scenarios be uncovered during the design process. This cannot be accomplished by evaluating the user-interface or analyzing the SOP, it must be accomplished by analysis of the variability in human performance completing the SOP under varying AOTW operational scenarios. This can be mitigated by modeling and simulating the SOP using MBSE methods that take into account the variability by using time distributions.

Future Work

The problem with triggering cues of an SOP could be solved using Natural Language Processing (NLP) to evaluate SOPs. The evaluation of the SOP includes identifying SOPs with unclear/ambiguous starting conditions and SOPs with unclear/ambiguous cues for steps. The evaluation could also include giving a score to each SOP based on the clarity and ambiguity of the triggering events and cues.

Future work also includes using AI/ML systems on aircraft performance data to detect hazardous

[8] BEA (2012) p. 210

events such as an aerodynamic stall. The ML/AI system could use previously acquired data from accidents/incidents to detect conditions of hazardous events, especially in cases where aircraft systems fail to alert of the hazardous event.

References

[1] S, Sneha, Beschi Raja, 2019, "A conceptual Overview and Systematic Review on Artificial Intelligence and It's Approaches", IJETIE.

[2] Airbus, 2002, Operational Philosophy: Introduction (01.010, July 28/05). A318/A319/A320/A321: Flight Crew Training Manual. France

[3] Federal Aviation Administration, 2017, Advisory Circular 120-71B "Standard Operating Procedures and Pilot Monitoring Duties for Flight Deck Crewmembers", https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-71B.pdf

[4] Airlines for America, 2021, Safety Records of U.S. Air Carriers, https://www.airlines.org/dataset/safety-record-of-u-s-air-carriers/

[5] Barshi, Immanuel, Robert Mauro, Loukia Loukopoulou, 2016, Designing Flightdeck Procedures, NASA/TM-2016-219421.

[6] Kourdali, Houda Kerkoub, Lance Sherry, 2016, A Systems Engineering Method and Simulation of Standard Operating Procedures.

[7] Bashatah, Jomana, Lance Sherry, 2021, A Model-Based Approach for the Qualification of Standard Operating Procedures, IEEE, pp. 1-10.

[8] Bashatah, Jomana, Lance Sherry, Steve Dam, Lauren Flenniken, Patrick Hartmann, and Tom Harold, 2021, Analyzing Standard Operating Procedures Using Model-Based Systems Engineering Diagrams, Wiley, pp. 1130-1144.

[9] Transportation Safety Board of Canada, 2003, In-Flight Fire Leading to Collision with Water Swissair Transport Limited McDonnell Douglas MD-11 HB-IWF Peggy's Cove, Nova Scotia 5 nm SW 2 September 1998.

[10] 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.

[11] BEA, 2012, Final Report on the Accident on 1st June 2009 to the Airbus A330-203 Registered F-GZCP Operated by Air France Flight AF 447 Rio de Janeiro-Paris.

[12] Aeronautical Civil of the Republic of Columbia, 1996, Controlled Flight into Terrain American Airlines Flight 965 Boeing 757-223 N651AAA Near Cali Columbia December 20 1995.

[13] National Transportation Safety Board, 2001, Runway Overrun During Landing American Airlines Flight 1420 McDonnel Douglas MD-82, N215AA Little Rock Arkansas June 1 1999.

[14] Komite Nasional Keselamatan Transportasi, 2019, Aircraft Accident Investigation Report PT. Lion Mentari Airlines Boeing 737-8 (MAX); PK-LQP Tanjung Karawang, West Java Republic of Indonesia 29 October 2018.

[15] Dutsch Safety Board, 2010, Crashed During Approach Boeing 737-800 near Amsterdam Schipol Airport 25 February 2009.

[16] National Transportation Safety Board, 2014, Descent Below Visual Glidepath and Impact With Seawall Asiana Airlines Flight 214 Boeing 777-200ER, HL7742 San Francisco, California July 6, 2013.

[17] Federal Aviation Administration, 2015, Advisory Circular 120-92B, "Safety Management Systems for Aviation Service Providers", https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-92B.pdf

[18] National Transportation Safety Board, 1988, Northwest Airlines, INC. McDonnel Douglas DC-9-82, N312RC Detroit Metropolitan Wayne County Airport Romulus, Michigan August 16, 1987.

[19] German Federal Bureau of Aircraft Accidents Investigation, 2018, erious Incident on 3 November 2001, Munich, Boeing B 777-300 ER.

[20] National Transportation Safety Board, 1997, Uncontrolled Flight Into Terrain ABX Air (Airborne Express) Douglas DC-8-63, N827AX Narrows, Virginia December 22, 1996

Acknowledgements

Houda Kourdali (Mitre), Immanuel Barshi (NASA), Robert Mauro (Decision Research Inc.), Len Adelman, John Shortle, George Donohue, Jon West,

Oleksandra Snisarevska, Amy Rose (GMU/CATSR), Vivian Motti (GMU).

Email Addresses

jbashata@gmu.edu

lsherry@gmu.edu

2022 Integrated Communications Navigation and Surveillance (ICNS) Conference

April 5-7, 2022