Systematic approach for the design of flight simulator studies

Klaas Fiete Krutein[1] and Linda Ng Boyle[1]
[1]Department of Industrial & Systems Engineering, University of Washington, Seattle, WA

The examination of commercial pilot workload often requires the use of controlled simulated studies to identify causal effects. The specific scenarios to consider within a simulator study require an extensive understanding of the safety situations that can occur in flight while also considering the specific training that pilots are provided within a simulated environment. The purpose of this paper is to provide a more systematic approach to scenario identification based on historical data, feasibility of capturing behavioral changes, simulator constraints, and training curricula.

INTRODUCTION

Previous research shows that the design and layout of cockpit interfaces and the corresponding information displayed can impact the workload of commercial pilots (Harris, 2007; Orlady, 2019). Research in human-automation interfaces show potential to further reduce workload (Schutte et al., 2007). This has been coupled with research on reduced crew operations (RCO), single pilot operations (SPO) and ground-based operational back-up (Harris, 2007; Graham et al., 2014; Bailey et al., 2017; Vu et al., 2018).

Approaches to assess pilot workload can range from interviews with subject matter experts to flight simulator studies. Wolter and Gore (2015) used a task analysis, validated with subject matter experts (SMEs), to compare current operations with two alternative SPO concepts that they identified as nominal and off-nominal scenarios. Off-nominal scenarios were defined as those that do not regularly occur in daily operations. Graham et al. (2014) examined workload for different SPO concepts by testing variations in an automated computer simulation model. Interviews with subject matter expects have also been used to examine pilot workload (Cummings et al., 2016). Lastly, Bailey et al. (2017) conducted a flight simulator study with human pilots. They measured the cognitive workload of pilots under three different operational concepts (current, RCO, SPO) given the current flight deck layout for six different scenarios. The study showed that existing technology in RCO and SPO were not yet safe for implementation. They underscored the need to have research directed towards assistance technologies and robust workload reduction for unique circumstances outside of regular operations.

While all studies mentioned noted the potential safety benefits of autonomous technology, the role of the human pilot and their ability to take over in case of autonomous failures are still not clear. There are several cases in which pilots have successfully intervened for errors caused by technology (Bailey et al., 2017; Helmreich & Foushee, 2019). In the process to support the safe and efficient introduction of alternative flight crew configurations, flight simulators can be used to identify the underlying factors related to high pilot workload. They are particularly useful for identifying elements of flight management that would be beneficial to automate without unnecessary harm to the pilot.

In a flight simulator study, the scenarios to be tested should be carefully selected to better inform policies, training, and design. As simulator availability is often limited and costly, it is beneficial to design experiments that maximize the likelihood of significant outcomes while also minimizing variability and noise. Hence,a systematic approach would help identify and distinguish scenarios prior to actual data collection.

Carefully selected scenarios can help better identify the context where higher pilot workload can occur. An important factor associated with pilot workload is training (Hart, 1986) and task processing time (Helmreich & Foushee, 2019; Salas et al., 2006). Task processing time is also correlated with the distribution of tasks in a crew, also called Crew Resource Management (CRM) (O'Connor et al., 2008; Orlady, 2019). Hence, understanding what is the most efficient distribution of tasks in a crew is important to minimize pilot workload.

Another factor related to pilot workload is the probability of occurrence of a specific scenario, which is independent of the training that the pilots may receive on taking action for that scenario. That is, despite training, frequently occurring scenarios might show lower workload peaks than scenarios that occur less frequently due to experience (Orasanu-Engel & Mosier, 2019). A selection process also needs to account for what can actually be simulated, under what circumstances the scenarios usually occur, and assess the relevance of the inspected scenarios. This is particularly important as safety priorities can change over time.

The goal of this study is to provide a systematic approach for identifying and selecting scenarios for studying a pilot's response to various safety critical events. Based on flight training guidelines and flight incident data, this paper proposes a schema that correlates the probability of occurrence of the safety critical event with the level of preparation through flight training. This helps to account for a sufficient spread of probability and training frequency properties in between scenarios.

DOI 10.1177/1071181319631524

METHOD

A high-level overview of the data wrangling and scenario selection process is shown in Figure 1. Data from the ASN Aviation Safety Database (Flight Safety Foundation, 2018) was used for this process. The justification behind using this database lies within the fact that it provides publicly available data on worldwide incidents and aircraft events. Other databases, such as the National Transportation Safety Board (NTSB), provides data only for the U.S. airspace (National Transportation Safety Board, 2018). If more detailed data with a focus on specific US regions are desired, the NTSB data can be a valuable source. The method presented in this paper also uses official training material to assess the scenarios on training frequency (Civil Aviation Authority, 2017).

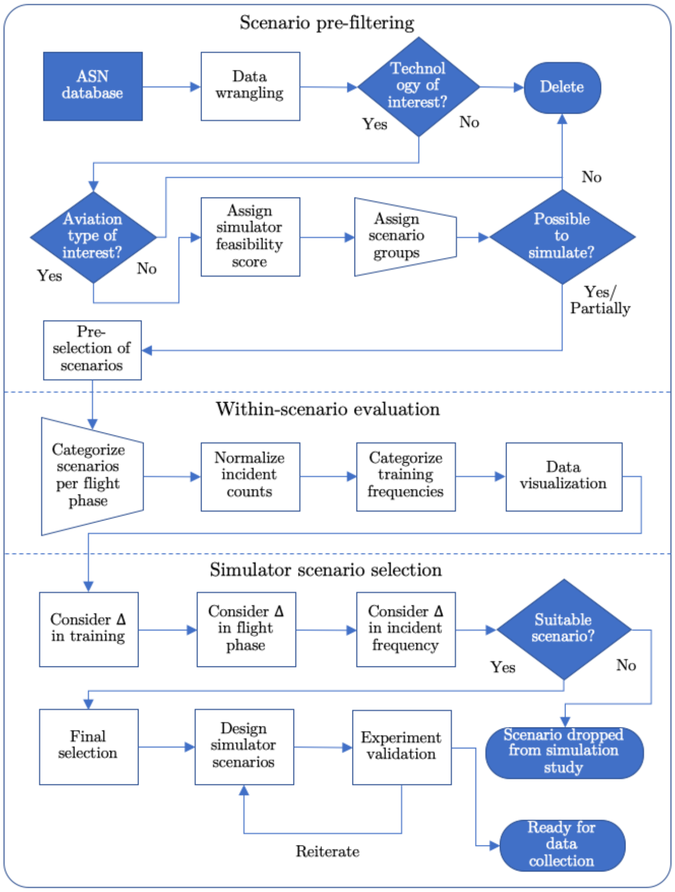

Figure 1: Proposed Data Wrangling and Scenario Selection Process

The following subsections include an example case of the scenario selection for a study investigating peak in-flight workload in non-normal situations in commercial aviation. This example case is geared towards pilots of modern commercial airplanes and therefore utilizes corresponding training material valid for the Boeing 737-800 (Civil Aviation Authority, 2017).

Scenario Pre-Filtering

The scenario pre-filtering phase reduces the number of incidents based on technical relevance for the example case. It aggregates the remaining incidents to higher level scenario groups to facilitate analysis, and assesses the simulation suitability of these scenario groups.

In the example case, the database is filtered to include only incidents that occurred after the introduction of the glass cockpit (Sweet, 1995), which was commercially introduced in the Boeing 767 on September 26, 1981; the Boeing 757 and the Airbus A310 included glass cockpit shortly thereafter (Sweet, 1995). The database is then filtered to include only commercial passenger airplanes. This filter removes all other plane types such as military, leisure and business aircraft that do not necessarily include two pilot operations. The clustering of the data into scenario groups provides a useful visualization of this process step (see Table 1).

Table 1: Summary of Accident Causes

Scenario Group	Frequency	Simulation suitable*
Security	488	No
Landing/Take-off	414	Limited
Collision	417	Limited
Engines	262	Yes
Undercarriage	256	Limited
Weather	78	No
Fire	72	Yes
Flight crew	54	No
Systems	42	Yes
Airframe	40	No
Cargo	36	Yes
Flight control surfaces	36	Yes
External factors	25	Yes
Instruments	21	Yes
Maintenance	17	No
ATC & Navigation	15	No
Unknown	12	No
Pressurization	5	Yes

*Assessed based on simulator available for this study

A study on pilot workload requires the scenario selection to further consider whether a scenario is:

1. Related to flight operations
2. Related to pilot interactions with the aircraft
3. Possible to simulate in a simulator

Scenarios from the groups of security, landing/take-off, and collision incidents are not controllable in a fixed-based simulator, but represent a high portion of the incidents. Given that these scenarios were attributed to human errors rather than the technical issues, they are the outcomes of some latent (or hidden) variables. For that reason, they are not considered independent (or controllable) variables. Figure 2 shows the number of incidents for the remaining groups. It can be observed that engine failures appear quite frequently, compared to other incident scenarios. This information can be used for further processing in the next phase, within-scenario evaluation.

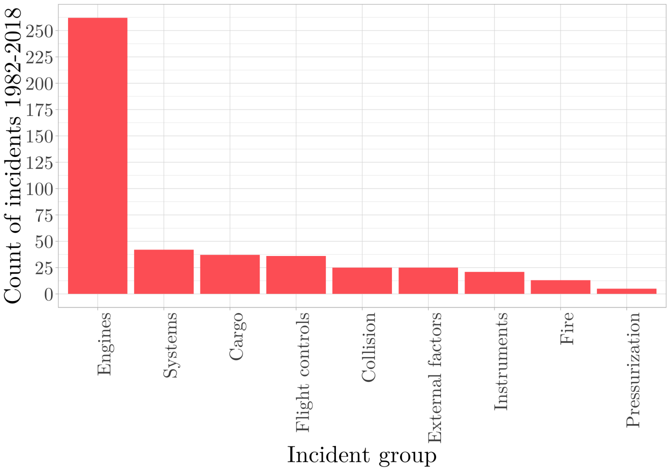

Figure 2: Scenario group occurrences in commercial aircraft since 1982

Within-Scenario Evaluation

The within-scenario evaluation phase aims to identify the distribution of the different scenario groups over the flight phases. This process will make each flight phase comparable by normalizing the incident counts per flight phase time span, and by allocating training frequencies to each scenario per flight phase.

In the context of the example case, the incident scenarios are examined at each flight phase (during take-off, en route, etc.). Nine different phases were observed in the database (see Table 2). The example case focuses on in-flight workload. Hence, the main interest are incidents that happen during the actual flight phases between take-off and landing to assess the pilots workload in-flight (excluding Taxi/Pushback/Towing and Standing). Thus, focusing on the flight phases shown in Figure 3, which was illustrated in the style of Boeing (2018).

With time and resource availability constraints in mind, the number of scenarios in a simulator study should be limited to a manageable scope. In the example case, three scenarios with different incident counts are considered in addition to a baseline scenario. To ensure a sufficient spread and variety of incident characteristics, an engine incident, a flight control surfaces incident and an instrument incident are chosen from the list in Table 2.

Table 2: Simulation Suitable Accident Cause Group Statistics per Flight Phase

			In Flight (S_i)					On Ground	
Scenario Group (S_j)	SF*	Sum	Take-off	Initial climb	En route	Approach	Landing	Taxi/Pushback/Towing	Standing
Undercarriage	L	256	21	0	6	0	209	12	8
Engines	Y	255	49	34	89	50	18	7	8
Fire	Y	71	4	1	22	7	7	4	26
Systems	Y	42	2	0	2	2	26	2	8
Flight controls	Y	36	10	6	4	14	2	0	0
Cargo	Y	36	15	5	8	4	3	0	1
External factors	Y	25	1	0	1	0	1	1	21
Instruments	Y	21	2	2	15	2	0	0	0
Pressurization	Y	5	0	0	5	0	0	0	0
Totals		747	104	48	152	79	266	26	72

*SF = Flight simulator feasibility st

The specific flight phase that these incidents can occur is also taken into consideration. Figure 4 provides an overview on the split of the three selected scenarios on all phases from take-off to landing. This figure shows that engine failures occur widely across all flight phases, with a peak during the en-route phase. Flight control surfaces fail mostly during take-off and landing approach, as their use is the highest during these phases. Only flight instruments show a clear peak during en-route operations but are equally distributed otherwise. However, this only reflects the total counts of incidents and not the actual time spent during these phases. Figure 3 shows that the time spent on Take-off/initial climb and approach/landing phases are much lower than for en-route phases (Boeing, 2018).

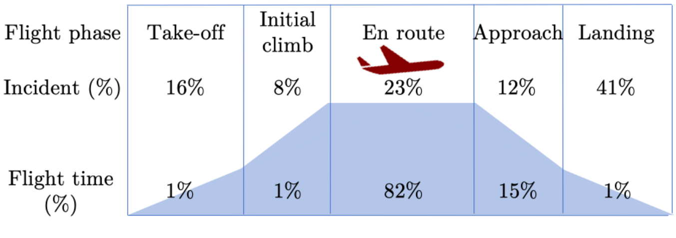

Figure 3: Proportions of incidents per flight phase in commercial aircraft since 1982 (Percentage of flight time sourced from Boeing (2018))

Comparing this to the time actually spent during each flight phase, the probability of these incidents over time is much higher during flight phases outside of en-route operations (Figure 4).

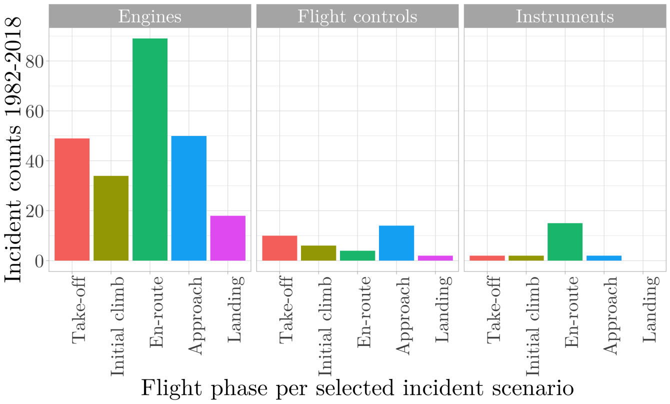

Figure 4: Selected scenario occurrences per flight phase in commercial aircraft since 1982

A scenario score was computed to account for the time each scenario group is observed in each flight phase. This score provides a value that is comparable across the different flight phases (illustrated in Figure 5). The scenario score, Sc_{ij}, is computed as:

$$Sc_{ij} = \frac{\frac{1}{\text{Flight time}} \times S_{ij}}{\sum_{i=1}^{5} \frac{1}{\text{Flight time}} \times S_{ij}} \tag{1}$$

where i = flight phase, from i=1 (take-off), 2 (initial climb) to 5 (landing), j = scenario group, and S_{ij} = frequency of scenario j at flight phase i.

As an example, let's say we are interested in engine failures during the take-off phase. In this case, i = Engines and j=Take-off. The scenario score will then be computed as:

$$\begin{aligned} Sc_{ij} &= \frac{1 \times 49}{1 \times 49 + 1 \times 34 + \frac{1}{82} \times 89 + \frac{1}{15} \times 50 + 1 \times 18} \\ &= \frac{49}{105.42} = 0.46 \end{aligned} \tag{2}$$

Figure 5 displays the scenario score over the investigated scenarios (the point size and numbers in the graph represent the absolute counts). In relative terms, these graphs show that take-off procedures and landings have higher frequencies per percent of time than en-route operations. They are therefore, more prone to failures (with the exception of instrument failures). Training schedules for Boeing 737 pilots are available from the Civil Aviation Authority (2017). Interviews with a subject matter expert was also conducted. This graph is then used for the simulator scenario selection phase.

Simulator Scenario Selection

In this last process phase, the correlation between training frequency and the scenario score is taken into account. Figure 5 shows the potential in simulating various engine failures during different flight phases. For example, the engine scenario can be used to assess differences in pilot workload given training and incident frequencies. Instrument failures and flight control surface failures do not show the same spread when reviewed in terms of within-scenario discrepancies. These latter two scenarios could still be considered in simulator studies, but may not show as strong a significant difference as would be expected for the engine failure scenarios. A simulator study on engine failures may be able to detect differences in pilot workload for various scenarios and at different flight phases (i.e., Take-off, Initial climb, Approach, Landing flight phases).

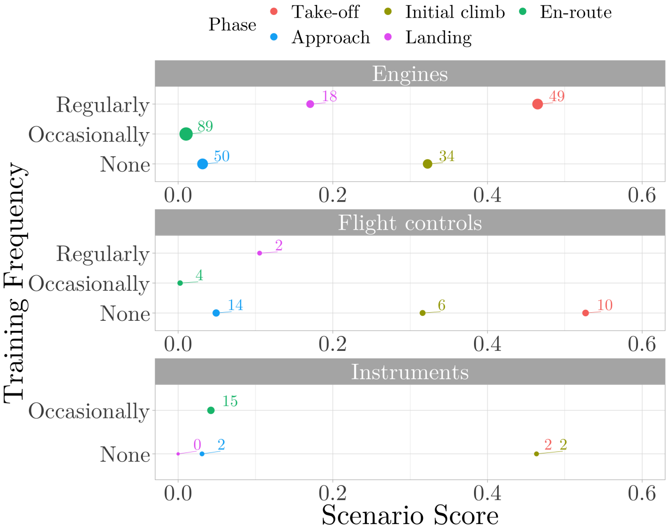

Figure 5: Scenario score vs. Training frequency per scenario group

DISCUSSION

This paper presents a method for selecting a justifiable set of scenarios for flight simulator studies that aim to examine pilot workload. This systematic approach allows the design of a more robust experimental design. The scenarios that were eliminated are not less important, but rather should be examined using other data collection tools that may provide more naturalistic observations.

Depending on sample size, number of independent variables selected, and simulator options, fractional or full factorial designs are examples of experimental designs that can be considered to examine these scenarios. The selection of scenarios can be useful even if one is not designing a controlled study. The framework is sufficiently flexible to allow alternative scales for scenario selection. In addition to applications in human factors, the systematic approach for scenario selections can be used more broadly to inform policy, pilot training and education, and flight deck design.

Regulations. For aviation authorities such as Federal Aviation Agency (FAA), the proposed framework provides context for the mandatory training requirements. For example, there may be failure scenarios that training is not provided for on a regular basis, but have a large safety impact. Or if training is provided, the time that the failure occurs may be of greater consideration. Based on Figure 5, engine failures and flight control surface failures are more prominent during the initial climb, whereas instrument failures are observed more during take-off.

Training. The proposed framework can also be useful to improve the overall quality of training, particularly in improving the effectiveness of CRM methods. Trainers may be interested in designing training curricula that provide more varied practice flights to supplement the mandatory training requirements. This could lead to lower pilot workload on those scenarios that reside on the lower part of the scale, such as failures during initial climb and approach.

Flight Deck Design. Aircraft manufacturers that aim to improve the flight deck could use this method to review checklist set ups and optimize human-machine interfaces in cockpit designs. This can lead to higher level of preparedness and better feedback for the pilots. The information displays can also better prioritize information displayed to support better pilot decision-making during safety critical situations.

These applications show the value of using a systems approach to be more proactive in identifying pilot workload and more specifically, the context and the possible timing of safety critical incidents. As the industry moves forward with automation and co-automation, testing the scenarios that are common as well as rare will be very important as we continue to redefine the supervisory role of the pilot.

ACKNOWLEDGEMENTS

This project was sponsored by the Joint Center for Aerospace Technology Innovation (JCATI), in collaboration with Esterline (Tim Robinson, Michael Utsugi). The authors also acknowledge Brian Bradley of Bradair Aviation Consultancy for subject matter advice.

REFERENCES

Bailey, R. E., Kramer, L. J., Kennedy, K. D., Stephens, C. L., & Etherington, T. J. (2017, September). An assessment of reduced crew and single pilot operations in commercial transport aircraft operations. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)* (pp. 1–15). St. Petersburg, FL: IEEE. doi: 10.1109/DASC.2017.8101988

Boeing. (2018, October). *Statistical Summary of Commercial Jet Airplane Accidents* (Presentation). Seattle: The Boeing Company. Retrieved 2018-11-07, from `https://www.boeing.com/resources/boeingdotcom/company/about\_bca/pdf/statsum.pdf`

Civil Aviation Authority. (2017, July). *Training, skill test or proficiency check for ATPL, MPL and type ratings MPA and SPHPCA.* Civil Aviation Authority. Retrieved 2018-12-23, from `https://publicapps.caa.co.uk/docs/33/SRG1158Issue05Enabled.pdf`

Cummings, M. L., Stimpson, A., & Clamann, M. (2016, January). Functional Requirements for Onboard Intelligent Automation in Single Pilot Operations. In *AIAA Infotech @ Aerospace.* San Diego, California, USA: American Institute of Aeronautics and Astronautics. doi: 10.2514/6.2016-1652

Flight Safety Foundation. (2018, Oct). *ASN Aviation Safety Database* (Tech. Rep.). Author. Retrieved 2018-11-05, from `https://aviation-safety.net/database/`

Graham, J., Hopkins, C., Loeber, A., & Trivedi, S. (2014, April). Design of a single pilot cockpit for airline operations. In *2014 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 210–215). Charlottesville, VA, USA: IEEE. doi: 10.1109/SIEDS.2014.6829879

Harris, D. (2007, September). A humancentred design agenda for the development of single crew operated commercial aircraft. *Aircraft Engineering and Aerospace Technology*, *79*(5), 518–526. doi: 10.1108/00022660710780650

Hart, S. G. (1986). The relationship between workload and training: An introduction. In *Proceedings of the human factors society annual meeting* (Vol. 30, pp. 1116–1120).

Helmreich, R. L., & Foushee, H. C. (2019). Why CRM? Empirical and Theoretical Bases of Human Factors Training. In *Crew Resource Management* (pp. 3–52). Elsevier. doi: 10.1016/B978-0-12-812995-1.00001-4

National Transportation Safety Board. (2018, December). *Aviation Accident Database & Synopses.* Washington, DC. Retrieved 2018-12-30, from `https://www.ntsb.gov/\_layouts/ntsb.aviation/index.aspx`

O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. A. (2008, October). Crew Resource Management Training Effectiveness: A Meta-Analysis and Some Critical Needs. *The International Journal of Aviation Psychology*, *18*(4), 353–368. doi: 10.1080/10508410802347044

Orasanu-Engel, J., & Mosier, K. L. (2019). Flight Crew Decision Making. In *Crew Resource Management* (pp. 139–183). Elsevier. doi: 10.1016/B978-0-12-812995-1.00005-1

Orlady, L. M. (2019). Airline Pilots, Training, and CRM in Todays Environment. In *Crew Resource Management* (pp. 553–579). Elsevier. doi: 10.1016/B978-0-12-812995-1.00021-X

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006, June). Does Crew Resource Management Training Work? An Update, an Extension, and Some Critical Needs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 392–412. doi: 10.1518/001872006777724444

Schutte, P. C., Goodrich, K. H., Cox, D. E., Jackson, E. B., Palmer, M. T., Pope, A. T., . . . Barry, J. S. (2007). *The Naturalistic Flight Deck System: An Integrated System Concept for Improved Single-Pilot Operations* (Tech. Rep. No. NASA/TM-2007-215090).

Sweet, W. (1995, September). The Glass Cockpit - Pilots who work in glass cockpits throw no stones, despite some deep ambivalences. *IEEE Spectrum*(September 1995).

Vu, K.-P. L., Lachter, J., Battiste, V., & Strybel, T. Z. (2018, September). Single Pilot Operations in Domestic Commercial Aviation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(6), 755–762. doi: 10.1177/0018720818791372

Wolter, C. A., & Gore, B. F. (2015, January). A Validated Task Analysis of the Single Pilot Operations Concept. *National Aeronautics and Space Administration: NASA/TM2015218480*, 1–18.