Aviation Safety Risk Modeling: Lessons Learned from Multiple Knowledge Elicitation Sessions

J.T. Luxhøj, Ph.D.; LCR, Somerset NJ USA
E. Ancel, Ph.D.; National Institute of Aerospace, Hampton, VA USA
L.L. Green, A.T. Shih, Ph.D., S.M. Jones, Ph.D., NASA Langley Research Center, Hampton, VA USA
M.S. Reveley, NASA Glenn Research Center, Cleveland, OH USA

Keywords: system safety, hazard and risk analysis, knowledge elicitation

Abstract

Aviation safety risk modeling has elements of both art and science. In a complex domain, such as the National Airspace System (NAS), it is essential that knowledge elicitation (KE) sessions with domain experts be performed to facilitate the making of plausible inferences about the possible impacts of future technologies and procedures. This study discusses lessons learned throughout the multiple KE sessions held with domain experts to construct probabilistic safety risk models for a Loss of Control Accident Framework (LOCAF), FLightdeck Automation Problems (FLAP), and Runway Incursion (RI) mishap scenarios. The intent of these safety risk models is to support a portfolio analysis of NASA's Aviation Safety Program (AvSP). These models use the flexible, probabilistic approach of Bayesian Belief Networks (BBNs) and influence diagrams to model the *complex interactions* of aviation system risk factors. Each KE session had a different set of experts with diverse expertise, such as pilot, air traffic controller, certification, and/or human factors knowledge that was elicited to construct a composite, systems-level risk model. There were numerous "lessons learned" from these KE sessions that deal with behavioral aggregation, conditional probability modeling, object-oriented construction, interpretation of the safety risk results, and model verification/validation that are presented in this paper.

Introduction

As a complex domain with numerous interacting elements, aviation presents significant challenges to both researchers and practitioners in modeling system safety risk. With the creation of probabilistic safety risk models, inferences about changes to the states of the accident/incident shaping or causal factors can be ascertained. These predictive safety inferences derive from qualitative reasoning to conclusions based on data, assumptions, and/or premises and enable an analyst to identify the most prominent causal factors leading to a risk factor prioritization. Such an approach facilitates a mitigation portfolio study and assessment. The resulting risk model also facilitates the computation of sensitivity values based on perturbations to the estimates in the conditional probability values. Such computations lead to identifying the most sensitive causal factors with respect to an accident/incident probability. This approach may lead to vulnerability discovery of emerging causal factors for which mitigations do not yet exist that then informs possible future Research & Development (R&D) efforts.

The safety risk methodology in this study uses the flexible, probabilistic approach of Bayesian Belief Networks (BBNs) and influence diagrams to model the *complex interactions* of aviation system risk factors. Accidents are seldom, if ever, the result of a single hazard. In some instances, a shortcoming in the typical hazard analysis approach is to focus on a single hazard and risk assessment. Combining the individual hazard assessments inherent in a complex system to arrive at an overall level of system risk is a difficult challenge. Safety practitioners need to deal with numerous inherent hazard scenarios that a complex system operation can generate. The approach achieves a better understanding of the dynamics and inherent uncertainties in these scenarios. It permits robust inductive reasoning on the hypothesized accident/incident scenarios, ideal for addressing emergent National Airspace System (NAS) operations where there may be obvious data and experience limitations.

This paper focuses on the experience of knowledge elicitation (KE) sessions for the development of three different safety risk models. These models are developed by NASA Aviation Safety Systems Analysis Team in support of the NASA Aviation Safety Program (AvSP). Each KE session had a different set of experts with varying expertise, such as pilot, air traffic controller, certification, and/or human factors knowledge that needed to be elicited to construct a composite, systems-level risk model. In particular, the artful building of the relationships among the risk factors and their associated conditional probabilities are a unique, original contribution in each of the models. There were numerous "lessons learned" from these KE sessions that deal with behavioral aggregation, probabilistic risk modeling, object-oriented construction, interpretation of the risk results, and verification/validation. The "lessons learned" and recommendations from these "lessons" are presented in this paper.

Methodology

When constructing causal models, one of the most important factors that should be considered is the impact of uncertainty. In essence, probability theory derives solutions to reasoning under uncertainty in the face of limited information. Ideally there may be nominal or non-nominal statistical data about the operation under study. However, in novel situations, such as a study on new flight deck automation, these statistical data are scarce. In recent years, Bayesian Belief Networks (BBNs) have emerged as a principal methodology for numerous problems that involve reasoning under uncertainty in complex decision making arenas (see ref. 1). In some applications there may well be an abundance of information available during the development process. However, in other cases, while incident and accident data may exist, these data are not easily searched due to the lack of clear definitions/categories in the search engine. For emergent operations, such as civil uses for unmanned aircraft systems (UAS), the data are sparse.

Belief networks provide symbolic representations of probability models combined with efficient inference algorithms for probabilistic reasoning under uncertainty (refs. 2-5). Undesired events are not deterministic so any modeling effort needs to capture the probabilistic nature of multiple causalities. A BBN is a graphical approach that allows the "quantification" of safety risk models by using conditional probability theory. The next sub-section presents a brief overview of the basic BBN theory and inferencing principles.

Basic concepts in belief networks

Let $X = \{X_1, X_2, \ldots, X_n\}$ be a set of n variables. A belief network consists of a set of variables and a set of directed links between variables. This graphical structure is referred to as a directed acyclic graph (DAG) as in Figure 1. A variable or node represents a set of possible states, and the variable is in exactly one of its states. A DAG is considered acyclic if there is no directed path $X_1, X_{2, \ldots,} X_n$ such that $X_1 = X_n$. Let D be a DAG with one node for each variable in X. Every link from X_i to X_j in the graph indicates a direct dependence between the variables X_i and X_j. The node X_i is called a parent of X_j and X_j is referred to as a child of $X_{i.}$ The set of all parents of a node X_i is denoted as $\prod_i$. For example, in Figure 1 the nodes X_3 and X_6 are the parents of X_5. Therefore, $\prod_5 = \{X_{3,} X_6\}$.

A crucial concept in belief networks is the idea of conditional independence. In general, the variables A and C are independent given the variable B if P (A|B) = P (A|B,C). This statement implies that if the state of B is known then no knowledge of C will alter the probability of A. The notion of conditional independence facilitates the construction of the belief network and leads to efficient algorithms for the Bayesian network computations. In general, a Conditional Probability Table (CPT) given as $P(A|\ B_1, \ldots, B_n)$ is attached to each variable A with parents $B_1, \ldots, B_{n.}$ Inference algorithms that essentially use Bayesian inference methods as in refs (2-5) are employed on a BBN based on the probability tables associated with each node. The fundamental concept in the Bayesian treatment of uncertainties in probabilistic network models is conditional probability. The basic notation P (A|B) = x is read as "given the event B, the probability of the event A is x", and it should be noted that the statement does not mean that whenever B is true then the probability is x. Rather, this statement means that if B is true, and everything else known is irrelevant for A, then P (A|B) = x.

The fundamental rule for probability calculus is given as: P (A|B) $\cdot$ P (B) = P (A, B) where P (A,B) is the joint probability of the event $A \cap B$. Remembering that probabilities should always be conditioned on a context C, the formula should read P (A|B, C) $\cdot$ P (B|C) = P (A, B|C). This leads to the well-known Bayes' rule:

$$P(B|A) = \frac{P(B|A)P(B)}{P(A)} \tag{1}$$

Bayes' rule conditioned on C is:

$$P(B|A,C) = \frac{P(A|B,C)P(B|C)}{P(A)|C} \tag{2}$$

Bayes' formula, which constructs the foundation for inference algorithms in belief networks, can be interpreted as follows. Suppose we are interested in B and we begin with a prior probability P (B), representing our belief about B before observing any relevant evidence. Suppose we then observe A. By (1), our revised belief for B, the posterior probability P (B|A), is obtained by multiplying the prior probability P (B) by the ratio P (A|B) / P (A).

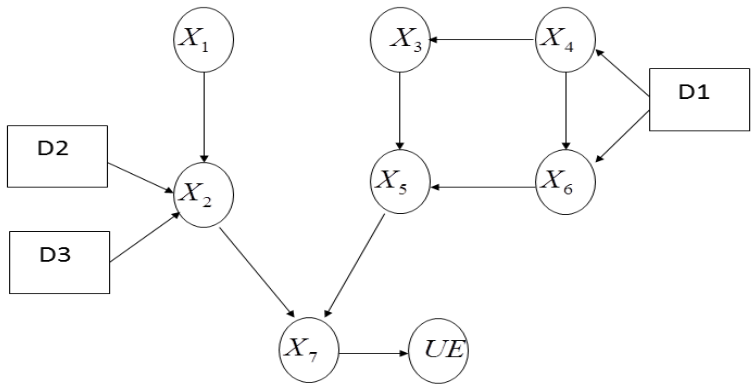

Figure 1 – Influence diagram with chance and decision nodes (UE is the Undesired Event)

Figure 1 is an influence diagram that is an extension of a BBN containing both chance nodes and decision nodes. Chance nodes (i.e., the random variables) are represented as circles and decision nodes (see D1, D2, and D3) are depicted as rectangles. The chance nodes are typically the causal nodes, whereas a decision node represents choices available to the decision-maker. The set of actions available to an individual in any given situation can be represented by a variable or a group of variables that are under the control of the decision-maker, unlike the chance variables. A decision variable can be related to one or multiple chance variables or multiple decision variables can be related to one particular chance variable. Choosing an action amounts to selecting a set of decision variables in an influence diagram, and then fixing their values unambiguously. Such a choice normally alters the probability distribution of other sets of variables that are judged to be consequences of the decision variables.

The conditional probability of one causal factor given the presence of other factor(s) may be estimated using the "beliefs" of subject matter experts. Aviation accidents are rare events so it is challenging to obtain hard data to quantify the models. An event tree could possibly be used to obtain some numerical "seeds" for the model, but an event tree is not an influence diagram and the interpretation of the numbers is not the same. With the BBN approach, the numbers in the conditional probability tables essentially represent the strength of the belief in the conditional causality as assessed by the expert for the scenario under study. A degree of belief approach was used by Ang and Buttery (ref. 6) in their risk assessment study of nuclear power plants. The approach involves moving up the systems ladder a bit and necessitates that the subject matter experts rely upon their mental model repository of similar cases. With a systems expansion viewpoint, the experts establish some basic boundary conditions, such as a towered airport, moderate traffic density, time period, etc. to set the conditioning context. This systems interpretation is consistent with the conceptual notion of "analytic generalization". These conditional probabilities serve to baseline the safety risk model.

Model quantification occurs by developing the Conditional Probability Tables (CPTs) using data when it is available. In the absence of data, experts' "beliefs" are used. Typically, model quantification involves the fusion of both hard data and "beliefs" and BBNs are ideally suited to such a hybrid or mixed modeling approach. Various elicitation methods of expert beliefs are provided in (refs. 7-9). There are a number of issues concerning human capabilities to consider when eliciting beliefs from the experts. Benson et al. (ref. 9) distinguish between "belief assessment" and "response assessment." Benson et al. (ref. 9) contend that "belief assessment" includes the structuring and conditioning steps in which target propositions are identified and defined and relevant knowledge is evoked from the domain expert." They state that "response assessment" encompasses the encoding and verifying steps in which numerical or verbal qualifiers are attached to the proposition" (ref. 9, p. 1641). Quantitative or semi-quantitative knowledge involves providing numerous conditional probabilities for the BBN. Benson et al. (ref. 9) state "a probability qualifies an individual's belief concerning a target proposition" (ref. 9, p. 1641). The elicitations of numerous probabilities are typically considered the bottleneck in BBN construction and are prone to a number of expert cognitive and motivational biases. Renooji (ref. 10) and (refs. 11-13) present an approach to facilitate probability elicitation in BBNs. This approach involves the use of fragments of text to provide a conditioning context that are derived from the graphical BBN structure. Then the fragments of text are placed adjacent to a probability scale that contains both verbal probability expressions and numerical values. The verbal expressions are of the form "(almost) certain, probable, expected, fifty-fifty, uncertain, improbable, and (almost) impossible" (ref. 11) as shown in Table 1. The authors contend that the combined approach of both verbal and numerical anchors accelerate conditional probability assessments in BBN when used in conjunction with the fragments of text. Ang and Buttery (ref. 6) also used such an approach in their elicitation of subjective probabilities for Probabilistic Safety Assessments (PSAs) in the nuclear power plant industry. However, the Ang and Buttery scale differs from the Renooji and Witteman scale especially in the mappings of numerical probabilities to the word "probable". For the LOCAF, FLAP, and RI safety risk models developed in these studies, the SMEs in each session adapted the scales to the unique boundary considerations of the problem domain under consideration and reached concurrence on the verbal-numerical probability scale to be used in each session to ensure consistency "within" a session. So, for example, the same scale was used for all LOCAF KE sessions. The Hugin BBN software (*http://www.hugin.dk*; Jensen (ref. 3)) is used in this research to perform the Bayesian propagations and calculus.

Table 1– Comparison of Numerical Probabilities and Verbal Descriptors

Renooij & Witteman (ref. 11) Verbal Descriptor	Probability	Ang & Buttery (ref. 6) Verbal Descriptor
certain (almost)	1	
	0.9999	extremely likely (i.e. almost certain)
	0.9	very likely
probable	0.85	
expected	0.75	
	0.7	likely
fifty-fifty	0.5	indeterminate
uncertain	0.25	
improbable	0.15	
	0.1	probable (i.e. credible)
	0.01	unlikely
	0.001	very unlikely
	0.0001	extremely unlikely
(almost) impossible	0	

Inserting Mitigations.

This step involves expert assessments of what causal factors may be impacted by mitigations, such as new technologies/procedures and the possible extent of that impact. These assessments are in the form of conditional probabilities. The BBN modeling approach enables an assessment of single and/or multiple technologies/procedures impacting either single and/or multiple causal factors. At this point in the ongoing research, only the LOCAF model has had mitigations inserted by the SMEs.

Evaluating the risk associated with the mitigations.

An original contribution of the safety risk modeling process is the projection of the relative safety risk of the undesired event. Once a model is constructed, the decision support tool may be used to evaluate various scenario combinations of mitigations. The "best" sub-set of mitigations is identified from a larger set of all safety products and the analyst may drill down to evaluate the relative risk reduction of mitigations not only upon the final consequence or undesired event, but also upon the causal factors comprising the model. The relative risk reductions may be used to evaluate the projected effects that a mitigation portfolio may have, and collectively, these risk reductions by undesired event type paint a mosaic of system safety risk. The introduction of mitigating measures often leads to the introduction of novel hazards, so the safety risk assessment needs to be expanded with these hazards and then the assessment process repeated as the model's ontology may change.

Verification and Validation (V&V) of a BBN is complex since there is a blend of quantitative and qualitative data. The V&V of a composite case may be decomposed into construct validity, internal validity, external validity and repeatability (ref. 14). The V&V decomposition of a BBN system risk modeling approach is discussed in Bareither and Luxhøj (ref. 15). The BBN system risk approach has been vetted in a number of industry/government studies and presentations. The original BBN aviation system safety approach was developed with funding from the FAA and NASA. The BBN approach was used in a previous NASA Research Announcement (NRA) (ref. 16) to support an industry team in completing a study on "System Safety." Currently, the BBN approach is being used to support another NASA NRA (ref. 17) on "Unmanned Aircraft Systems (UAS) in the NAS". In this UAS study, the BBN approach is being used for hazard and risk modeling of small UAS. As previously noted, the NASA Aviation Safety Systems Analysis Team is currently using the BBN approach to support safety risk modeling of accidents/incidents for commercial aircraft. Safety risk models are being constructed to understand the interactions of hazards associated with in-flight Loss of Control, Flightdeck Automation Problems and Runway Incursions. These risk-based causal models will then be used to support a portfolio assessment of the safety technologies (or products) being developed through NASA's AvSP.

Lessons Learned from the Multiple KE Sessions

In this study, multiple KE sessions were held with domain experts to construct probabilistic safety risk models for LOCAF, FLAP and RI mishap scenarios. As previously noted, the intent of these safety risk models is to support a portfolio analysis of NASA's AvSP. Each KE session had a different set of experts with diverse expertise, such as pilot, air traffic controller, certification, and/or human factors knowledge that was elicited to construct a composite, systems-level risk model. There were numerous "lessons learned" from these KE sessions that deal with behavioral aggregation, conditional probability modeling, object-oriented construction, interpretation of the safety risk results, and model verification/validation. The "lessons learned" are summarized below with the intent of suggesting "good practices" for future KE sessions.

LOCAF Model
The LOCAF model was constructed as an Object-Oriented Bayesian Network (OOBN) by NASA AvSP systems analysis personnel and the same three SMEs for all KE sessions. Luxhøj et al. (ref. 18) present a case study on the technical details of the LOCAF model. The LOCAF model is the most developed to date of the three BBN models. The baseline model has been constructed, CPTs elicited and NASA AvSP products inserted. In addition, the model has been through several levels of external review. A summary of the multiple model building and review sessions is shown in Figure 2. A high level depiction of the LOCAF model with its three sub-nets of Flight Crew Conditions, Environment, and System Component Failure is shown in Figure 3. Each sub-net has its own set of causal factors and their interactions.

The systematic development of the LOCAF model initiated with a data analysis of numerous LOC accidents, followed by the identification of common and principal causal factors. Three operational SMEs participated in model development in three KE sessions. These SMEs included a pilot, air traffic controller and a certification expert. The terminology and definitions of all nodes in the BBN were reviewed with the SMEs. The SMEs offered suggestions to improve the clarity of the node definitions. It was important to preserve the participation of the same three SMEs for all KE sessions for the continuity and consistency of model development and conditional probabilities. The OOBN structure with its inherent use of sub-nets facilitated model construction as well as CPT elicitation. The top-level model was able to be decomposed into smaller "chunks" and later synthesized into a cohesive whole. The multi-day KE session was divided into time slots of not more than 1½ hours in duration with multiple, short breaks given to the SMEs and the team in order to maintain focus. CPT elicitation in BBNs is known as the "bottleneck" and the LOCAF model construction was no exception. Mostly binary state nodes were used; however, there were a few tertiary state nodes that added to modeling complexity. The conditional probabilities of each SME were elicited and recorded. The elicitation proceeded in a randomized sequence with the SMEs. Typically, the facilitator started with an unconditioned node as a "warm-up" to a more involved parent node with children. The development team worked diligently to not have more than five inputs (i.e., causal factors) leading into any one node for human cognitive reasons (ref. 1). During the LOCAF CPT elicitation, it was discovered that a CPT "shortcut" for a large table might involve "blanketing" a section of the table with a numerical modifier for a change of state in one of the variables if the SMEs concurred that the "stability" of their assessments facilitated such a "global" adjustment. As reported in Ancel, et al. (ref. 19), the final developed baseline LOCAF model exhibited close correspondence with two different historical datasets (i.e., in one case, the LOCAF indicated a 15.92% LOC probability vs. 13.81% historical LOC and in another case, the LOCAF probability of 10.11% compared to an historical dataset indicating a 12.84% LOC occurrence).

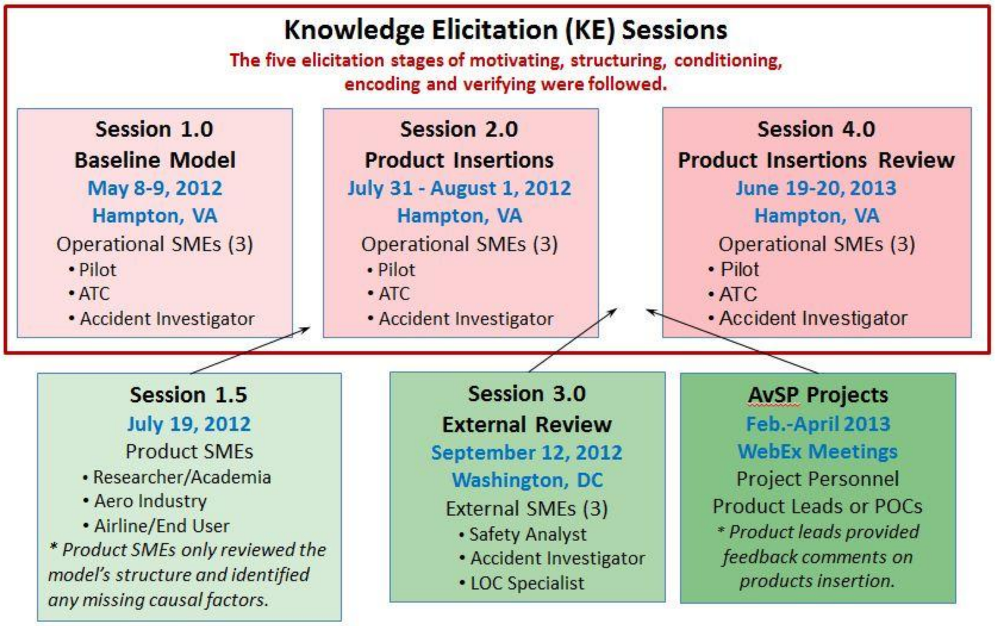

Figure 2 – Multiple KE Sessions for the LOCAF Model

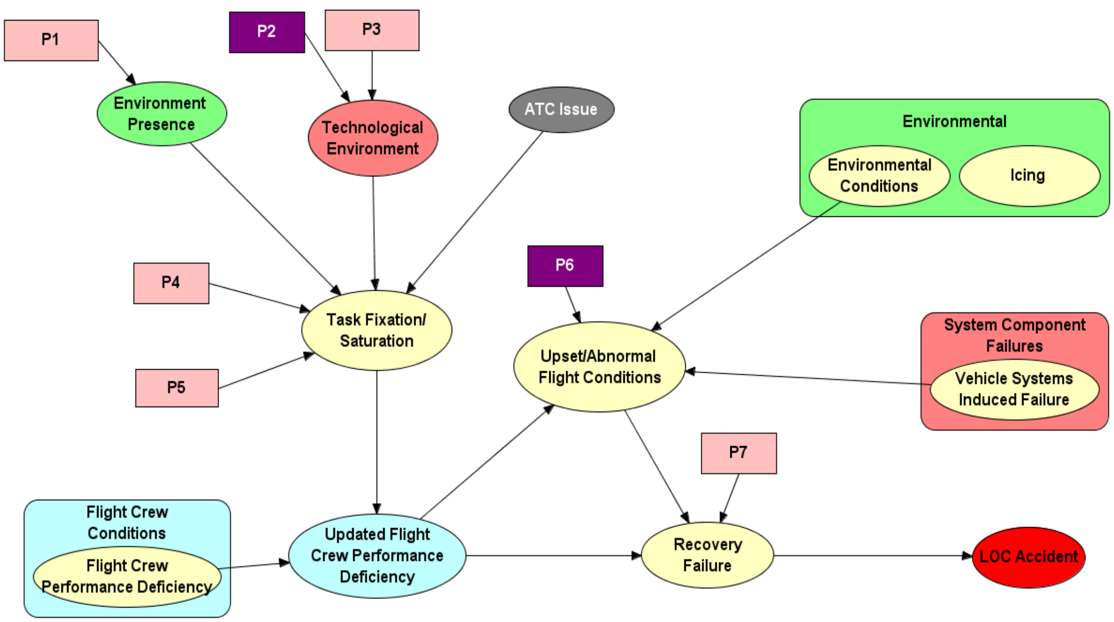

Figure 3 – Top-level LOCAF model

FLight Automation Problems (FLAP) Model

While the LOCAF model construction followed a "structured" process with its OOBN decomposition/synthesis approach, the FLAP model was constructed as a flat network rather than the object-oriented network. The issues surrounding advanced automation in the cockpit are quite complex and integrated and the FLAP model reflected this interplay between automation and the human. It was essential for this model development that a human factors expert participated in the KE session. Initially it was planned to have two 2-day sessions to build the baseline model; however, due to scheduling issues with the SMEs and the NASA team, it was decided to hold one 4-day KE session with 4 SMEs – 2 pilots, 1 human factors expert, and 1 system integration expert. However, due to internal organizational issues, only 3 SMEs actually participated in the KE session. A 4-day meeting proved challenging; however, lessons learned from the LOCAF model building process were insightful. For the FLAP model, it was important to carefully clarify all node terminology and definitions, such as "automation surprise". The human factors expert provided a strong link between the two pilots in communicating their thought processes and rationales to the model building team. During the CPT elicitations, the probability "blanketing" adjustments were again used with SME concurrence to efficiently complete the tables. Usually with the CPT elicitations, the completion of the table moves from the "worse" case (i.e., all given causal factors in their worse state) to the "best case" (i.e., all given casual factors in their "best" state). One of the SMEs suggested completing the CPTs by moving in the opposite direction – from the "best" case to the "worse" case. However, it quickly became apparent to the SMEs and the KE facilitator that the cognitive burden with this approach was significant and did not offer any elicitation procedural efficiencies. At this point in time, the FLAP baseline model awaits for its external review with a different set of SMEs as preparation for the AvSP product insertion phase.

Runway Incursion (RI) Model

The RI model was initially planned to be a smaller model compared to the LOCAF and FLAP models; however, it developed into a fairly large BBN. Some key lessons learned from the RI session include the importance of preparation before the KE session and starting with a clear scope that the SMEs can then refine; simplifying the structure of a very "people intensive" and "communication intensive" modeling situation; and simplifying the event reporting structure imposed by the FAA data source and definitions. As originally constructed, the RI model is not an OOBN. There were 4 SMEs involved in the model construction KE session along with the NASA team. There was a senior pilot with both military and commercial experience, a regional jet pilot, an air traffic controller and a human factors expert. The RI model is very operational in nature and the "scoping" of the model and establishing boundary conditions were quite important in this case. Again, the "blanketing" adjustments during CPT elicitations proved efficient. There was a unique lesson learned from the RI session. One SME fell ill and was not able to appear in person during the KE, so the team held an impromptu telecon/WebEx to finish off the KE session. This adjustment is important because the SME appreciated that individual contributions were acknowledged and the team obtained conditional probabilities from four SMEs rather than three. The lesson is to be flexible and know how to handle an unexpected event. At the present time, the RI model is being expanded to a larger runway safety model (RUNSAFE) that considers both runway incursion and runway excursion events. The RUNSAFE baseline model is under development.

As previously noted, V&V of a BBN is complex due to the blending of quantitative and qualitative data. The LOCAF model was able to represent a generic loss of control environment very effectively (ref. 19). In general, the V&V of a composite case may be decomposed into construct validity, internal validity, external validity and repeatability (ref. 14). Besides the numerical validation, the LOCAF model's assumptions, structure, as well as the placement of the products (i.e., construct validity) were reviewed by external SMEs. The construction and the verification of the LOCAF model were conducted in four steps. The first step was the initial review of the links and nodes and collection of the numerical values for CPTs for the baseline LOCAF model, performed by the same three SMEs (i.e., both construct and internal validity). During the second step, the baseline model was reviewed by a different SME group (comprised of experts in various fields in which NASA technologies are being developed) in order to identify any missing links/nodes in representing a LOC environment (i.e., external validity). Following the establishment of the baseline model, the third step involved the placement of the NASA safety products into the model. The initial SME group that participated in the construction of the model was also used in this step in order to achieve consistency in the assumptions. The SMEs not only placed the products in the model, but also they provided numerical values of the prospective impacts of each technology on its attached node. The final step included the review of the entire model with the technology insertions by another external SME panel (i.e., external validity). This panel involved professionals from the NTSB and FAA, as well as individuals experienced in systems engineering and modeling for a comprehensive review. Besides the SME panels, the LOCAF model results were presented at two internal NASA technical forums and three conference paper sessions to seek further discussions and comments. The LOCAF model was thoroughly documented to ensure that the model building process could be replicated (i.e., repeatability). The team plans to follow a similar V&V process for both the FLAP and RI models; however, it is acknowledged that while historical data exist for calibration purposes for the RI model as with the LOCAF model, the historical data for calibration purposes for the FLAP model may be rare. For the FLAP model, it may be that key segments of the model may be compared with data from aviation human factors databases.

As a common factor in all the KE sessions, it was discovered that the arranging the logistics of the meeting was vital for the project. Identification and selection of the qualified SMEs, scheduling of several people with busy schedules, and securing a suitable venue require substantial preparation and time. Due to contractual differences and compensation plans, SMEs involved in government versus private industry require different approaches. For example, the participation of non-government SMEs appears more straightforward because their time and travel were made by the same arrangement. However, the participation of government SMEs in a KE session may become quite complex should compensation for time and travel be involved and subjected to different processes. For that reason, it is advisable to identify alternative experts and contingency plans in cases where the SMEs are unable to attend the meetings due to contractual issues, health reasons, scheduling conflicts or even unexpected events like government shutdown. The KE facilitator needs to consider these aspects when coordinating the arrangements for the KE session.

Recommendations from "Lessons Learned"

Consistent with the U.S. Environmental Protection Agency (EPA) task force white paper (ref. 20), expert elicitation is considered as the "formal, systematic process of obtaining and quantifying expert judgment on the probabilities of events, relationships and parameters." (ref. 20, p. 116). This section summarizes recommendations from the multiple KE sessions that are highly consistent with the U.S. EPA task force white paper (ref. 20) on expert elicitation. SME KE is fundamental to quantitative estimation of uncertain values in the absence or the inadequacy of hard data in complex problem domains and can also assist with model conceptualization. The LOCAF model in particular demonstrates that KE can provide useful, credible results. A KE session requires a significant investment of resources and time in order to provide credible results and several general suggestions for "good practices" are noted above. Some specific recommendations from the "lessons learned" from the LOCAF, FLAP and RI KE sessions are as follows:

1) KE is especially useful when an emerging scientific or engineering challenge lacks a consensus interpretation and database or for a challenge that is dependent on uncertain events. As shown from the LOCAF, FLAP and RI models, expert beliefs about the value and meaning of data can provide valuable insights. As noted by Richard Hamming, considered by many as the "father" of numerical methods, "The purpose of computing is insight, not numbers" (ref. 21). Also, as shown in the LOCAF model with its OOBN constructs, it may be useful to disaggregate or decompose the problem into smaller sub-nets, develop these separately, and then synthesize or aggregate these sub-nets into a top-level model for subsequent analysis.

2) It is important that the KE session be well-designed and implemented to help ensure the model's credibility and endorsement of the results within the organization and by external parties. The NASA Aviation Safety Systems Analysis Team may use the KE results to communicate to a regulatory agency, such as the FAA, to encourage transparency, credibility, objectivity (unbiased and balanced), rigor (control of heuristics and biases), and relevance to the technical problem under study. The KE session and BBN models are generally context-specific so the interpretation of the results should be performed with caution. With each team member working in concert, a KE session has distinct, specialized roles for members of the project team (such as a generalist, modelers/analysts, and SMEs) and each role/member contributes in a unique way to the final risk model.

3) It is concluded that a KE is appropriate for research challenges with complex technical problems and unobtainable data characterized by a high level of uncertainty. As per the EPA report, it is also noted that a KE session can serve as a proxy for traditional data but should not be used as a substitute for conventional research when empirical data can be obtained within the available time and resources.

4) Some issues to be addressed in the design and execution of a KE session include (U.S. EPA, ref. 20):
    - Standards of quality for KE that are a function of their intended use (e.g., to inform research needs, to inform regulatory decisions) and a minimum set of best practices.
    - How to interpret the quality of the results and the KE process?
    - How does acceptability of the results depend on context?
    - The role of stakeholders (e.g., product designers) in the KE process to provide input on relevant questions or issues.

- Appropriateness of secondary application of KE results (i.e., the use of results beyond the intended purpose of the initial study).
- If and how to combine experts' judgments?
- Comparison of alternative types of research findings: empirical data, external expert recommendations and KE results.
- How the KE results should be used and communicated to decision-makers?

5) The development of training materials for the KE sessions could be helpful for future sessions. For example, in the KE session for the RI model, a small BBN of sub-set of risk factors was developed to illustrate to the SMEs the numerical Bayesian computations. The development of additional, similar material should improve the efficiency of the KE session. The KE session is quite dependent on the skill of the KE facilitator and it would be beneficial to develop an informal community of practice where this knowledge could be shared with other potential KE facilitators.

6) It is important for credibility and defensibility to select SMEs who do not have conflicts of interests and who can be impartial. In general there should be a process of nominating and selecting SMEs on a case-by-case basis. In our study, the sponsor was involved in the nomination and final selection of the SMEs to ensure quality and credibility of the process. The SMEs were identified by name and organizational affiliations; however, their responses and judgments remained anonymous. A record of all judgments was maintained and if necessary, could be provided for any auditing purposes or if requested during a peer review. Peer review of the results is an essential component of the KE process. The peer review focused on the KE process, such as how the experts were selected, what information they provided, how the KE session protocols were conducted, how biases were controlled, and how the modeling results were analyzed and presented. The purpose of the peer review was to review the KE process, not necessarily to question the details of the modeling results; however, order of magnitude comparative results are important to review. For example, in the LOCAF model development, a formal peer review was conducted by external stakeholders and changes about modeling terminology and structure were collected and then provided to the original SMEs for their considerations. Moreover, placements of NASA's AvSP products in the LOCAF model were briefed to and reviewed by the project managers and product technology leads. Suggestions and/or clarifications about the capabilities of their products were provided to the SMEs post-KE session for their consideration; however, the operational SMEs retained the final control of the AvSP product placements in the LOCAF model.

Finally, descriptions of the KE process and its protocols, model development and analysis, SME selection and probability elicitation processes, peer review procedures and BBN modeling results are being planned for presentation in workshops, colloquia, and professional society meetings to promote dialogue, encourage innovation and improve the quality and appropriate use of KE assessments.

Acknowledgement

This research is supported by funding from the NASA Aviation Safety Program.

References

1. Fenton, N. and Martin, N., *Risk Assessment and Decision Analysis with Bayesian Networks*, CRC Press, Boca Raton, FL, 2013.

2. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, 1988.
3. F.V. Jensen, *Introduction to Bayesian Networks*, Springer Verlag, New York, 1996.

4. Huang, C. and Darwiche, A., "Inference in Belief Networks: A Procedural Guide," *International Journal of Approximate Reasoning*, Vol. 11, No. 1 (1994), pp. 1-45.

5. Lauritzen, S.L. and Spiegelhalter, D.J., “Local Computations With Probabilities on Graphical Structures and Their Application to Expert Systems,” *Journal of the Royal Statistical Society, Series B (Methodological),* Vol. 50, No. 2 (1988), pp. 157-224.

6. Ang, M.L. and Buttery, N.E., “An Approach to the Application of Subjective Probabilities in Level 2 PSAs,” *Reliability Engineering and System Safety*, Vol. 58 (1997), pp. 145-156.

7. Ayyub, B.M., *Elicitation of Expert Opinions for Uncertainty and Risks*, CRC Press, New York, 2001.

8. Vick, S.G., *Degrees of Belief*, American Society of Civil Engineers, Reston, VA, 2002.

9. Benson, P.G., Curley, S.P. and Smith, G.F., “Belief Assessment: An Underdeveloped Phase of Probability Elicitation,” *Management Science*, Vol. 41, No. 10 (1995), pp. 1639-1665.

10. Renooji, S.,“Probability Elicitation for Belief Networks: Issues to Consider,” *The Knowledge Engineering Review*, Vol. 16, No. (3) (2001), pp. 255-269.

11. Renooji, S., and Witteman, C.L.M., “Talking Probabilities: Communicating Probabilistic Information With Words and Numbers,” *International Journal of Approximate Reasoning*, Vol. 22 (1999), pp. 169-194.

12. van der Gaag, L.C., Rennoji, S., Witteman, C.L.M., Aleman, B.M.P. and Taal, B.G., “How to Elicit Many Probabilities,” *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 1999, pp. 647-654.

13. Druzdel, M.J., and van der Gaag, L.C., “Elicitation of Probabilities for Belief Networks: Combining Qualitative and Quantitative Information,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 12, No. 4 (2000), pp. 481-486.

14. Yin, R.K., *Case Study Research: Design and Methods*, 3rd ed., Sage Publications, London, 2003.

15. Bareither, C., and Luxhøj, J.T., “Uncertainty and Sensitivity Analysis in Bayesian Belief Networks: Application to Safety Risk Assessment,” *International Journal of Industrial and Systems Engineering*, Vol. 2, No. 2 (2007), pp. 137-158.

16. Miller, M.E., and Arkind, K., “NextGen Advanced Concepts and Vehicles: System Level Assessment,” NASA Technical Interchange Meeting, San Antonio, TX, October 15, 2009 (NASA NRA #NNH06ZEA001N-IAC1).

17. Luxhøj, J.T., “Predictive Analytics for Modeling UAS Safety Risk,” *SAE International Journal of Aerospace*, Vol. 6, No. 1 (2013), pp. 128-138 (NASA NRA #NNH10ZEA001N).

18. Luxhøj, J.T., Shih, A.T., Ancel, E., Jones, S.M. and Reveley, M.S., “Safety Risk Knowledge Elicitation in Support of Aeronautical R&D Portfolio Management: A Case Study,” *Proceedings of the International Conference of the American Society for Engineering Management*, Hilton Virginia Beach Oceanfront, Virginia Beach, VA, October 17-20, 2012.

19. Ancel, E., Shih, A.T., Jones, S.M., Reveley, M.S. and Luxhøj, J.T., “Predictive Safety Analytics: Inferring Aviation Accident Shaping Factors and Causation,” *Journal of Risk Research*, to appear 2014.

20. U.S. Environmental Protection Agency (EPA), “Expert Elicitation Task Force White Paper,” prepared for the Science and Technology Council, U.S. EPA, Washington, SC, August, 2011.

21. http://en.wikipedia.org/wiki/Richard_Hamming

Biographies

J.T. Luxhøj, Ph.D., LCR, 40 MacAfee Road, Somerset, NJ 08873, USA, telephone – (732) 259-9623, e-mail – jtluxhoj@gmail.com

Dr. Luxhøj has led and participated in FAA- and NASA-sponsored research teams over the past 28 years in the areas of system safety and risk analysis. He served as an Office of Naval Research (ONR) Distinguished Summer Faculty Fellow for NAVAIR System Safety co-located at Patuxent River, MD and Lakehurst, NJ during the summers of 2012 and 2013. He has over 150 technical publications and is the co-author of *Engineering Economy, 13th edition* published by Prentice-Hall. He is a Fellow in the Institute of Industrial Engineers (IIE).

E. Ancel, Ph.D., Research Engineer, National Institute of Aerospace, Located at NASA Langley Research Center Building 1209, Room 190-22, Mail Stop 442, Hampton, VA 23681, USA, telephone – (757) 864-6082, e-mail – Ersin.Ancel@nasa.gov

Dr. Ancel received his B.S. from Istanbul Technical University, Turkey, and M.S. from Old Dominion University, Norfolk, VA in Aerospace Engineering in 2005 and 2007, respectively. He received his Ph.D. from Old Dominion University in Engineering Management and Systems Engineering in 2011. He is currently working as a Research Engineer in aviation safety at National Institute of Aerospace (NIA) as a NASA contractor. He is in charge of developing aviation accident models within the National Airspace System (NAS) and evaluating the effects of various NASA technologies over the implementation of Next Generation Air Transportation System (NextGen). His past experiences include probabilistic safety assessment (PSA) applications in Surry Nuclear Power Plant as well as modeling of complex socio-technical systems and developing serious gaming solutions involving the NAS.

L.L. Green, Vehicle Analysis Branch (E401), Systems Analysis and Concepts Directorate, NASA Langley Research Center, B1209, Rm128H, Mail Stop 451, Hampton, VA 23681, USA, telephone – (757) 864-2228, e-mail – Lawrence.L.Green@nasa.gov

Lawrence Green has been a civil servant at NASA Langley Research Center (LaRC) for over 38 years. Mr. Green holds a Bachelor's Degree in Aerospace Engineering from the University of Cincinnati and a Master's Degree in Mechanical Engineering from the George Washington University. Mr. Green was a member of the NASA LaRC aeronautics community until 2005. Since then Mr. Green has worked within the NASA LaRC space community. Mr. Green’s aeronautics work spans a wide variety of topics in the development and application of computational fluid dynamics (CFD) codes, multidisciplinary analysis and design, aircraft stability and control and design under uncertainty. Beginning in about 1999, Mr. Green formally began working in uncertainty quantification and management and has continued in this area since then, supporting the NASA Space Shuttle, Ares I-X launch vehicle, Orion Crew Capsule and Aviation Safety Programs. Mr. Green was a co-author of NASA Standard 7009 for Models and Simulations and has published 52 journal and conference papers discussing a wide range of topics related to the development and application CFD methods, multidisciplinary design, uncertainty quantification and propagation, reliability assessment, probabilistic methods, Bayesian techniques and runway safety.

A.T. Shih, Ph.D, Vehicle Analysis Branch, Systems Analysis and Concepts Directorate, NASA Langley Research Center, Hampton, VA 23681, USA, telephone – (757) 864-4450, e-mail – Ann.T.Shih@nasa.gov

Dr. Shih is an Aerospace Engineer at NASA Langley Research Center. Currently, she is the lead of Systems Level Integrated Modeling in Systems Analysis and Methods (SAM) within the Safety Analysis and Integration Team (SAIT) of the Aviation Safety Program (AvSP). Her other NASA experience includes the propulsion flowpath analysis for the X-43A vehicle in the Hyper-X program, and system integration failure analysis for the Ares I Launch Vehicle developed as part of the Constellation Program. She received her B.S. in Aerospace Engineering from the Georgia Institute of Technology, her M.S. in Aerospace Engineering from the University of Illinois at Urbana-Champaign, and a Ph.D. in Mechanical Engineering from the University of Illinois at Chicago.

S.M. Jones, Ph.D., NASA Langley Research Center, Hampton, Virginia 23681,USA, telephone – (757) 864-7642, e-mail – Sharon.M.Jones@nasa.gov

Dr. Jones received her B.A. in Mathematics from Hampton University, M.E. in Systems Engineering from the University of Virginia, and Ph.D. in Engineering Management from Old Dominion University. Her current research interests include decision analysis methods for aerospace technology portfolio assessments and risk analysis. Some of her prior NASA research activities included conducting numerous studies to compute the expected impact of aeronautics research on future aircraft direct operating costs and the development and evaluation of computer vision algorithms for space telerobotic tasks. She is currently on a Programming Analysis detail assignment in NASA's Independent Program Assessment Office (IPAO), where she conducts cost, schedule and risk analyses for the agency. She is a Member of INFORMS and is a Senior Member of AIAA.

M.S. Reveley, NASA Glenn Research Center, 2100 Brookpark Road, Mailstop 5-11, Cleveland, OH, USA, telephone – (216) 433-6511, e-mail – Mary.S.Reveley@nasa.gov

Mary Reveley is an Aerospace Engineer at the NASA Glenn Research Center. Currently, she is on the Aviation Safety Analysis Team (ASAT), supporting systems and portfolio analysis for the NASA Aviation Safety Program. She received her B.S. in Aerospace Engineering from the Ohio State University, and her MBA from Cleveland State University. Mary has worked in systems analysis for NASA's Aviation Safety Program since 1998 performing cost analysis, safety risk analysis, portfolio assessments, state of the art assessments, and gap analyses. Mary is a Senior Member of the AIAA and the FAA's Integrated Safety Risk Assessment Advisory Committee.