BENEDIKT BADANIK, Ph.D.[1]
(Corresponding author)
E-mail: benedikt.badanik@fpedas.uniza.sk
MARTIN JANOSSY, Ing.[2]
E-mail: martin.janossy@gmail.com
ARTHUR DIJKSTRA, M.Sc.[3]
E-mail: Arthur@admc.pro

[1] Air Transport Department
   Faculty of Operation and Economics of Transport
   and Communications, University of Zilina
   Univerzitna 1, 010 26 Zilina, Slovak Republic

[2] Air Navigation Services of the Czech Republic
   Navigacni 787, 252 61 Jenec u Prahy, Czech Republic

[3] Kooikerboog 7, 1394ND Nederhorst den Berg
   Netherlands

THE USE OF EXPERT JUDGEMENT METHODS FOR DERIVING ACCIDENT PROBABILITIES IN AVIATION

ABSTRACT

Improving safety has always been the top interest in the aviation industry. The outcomes of safety and risk analyses have become much more thorough and sophisticated. They have become an industry standard of safety investigations in many airlines nowadays. In the past, airlines were much more limited in answering the questions about hazardous situations, accident probabilities, and accident rates. Airlines try hard to cope with stricter safety standards. The objective of this paper is to find out and quantify the extent of the expert judgment in helping airlines in the evaluation of the Flight Data Monitoring (FDM) events. On top of that, the paper reveals the method for a careful choice of experts, so that their estimations will maximize the potential of an accurate and useful outcome. Also, the paper provides details of implementation of the classical model into this research, then continues with the calculations and visualization of the outcomes. The outcomes are probability distributions per each aircraft type, then per IATA accident type and finally per FDM event.

KEYWORDS

expert judgment; classical model; Excalibur software; accident probability.

1. INTRODUCTION

Expert judgments have been used during recent decades to gather many expert opinions about different subjects. They have earned an increased attention in risk assessment across various industries. For instance, seeking experts' opinion has played a vital role in maritime, nuclear, aerospace, chemical, economical, meteorological, and technical industries to provide estimations on the desired subjects of interest [1][3]. These include, inter alia, risk assessments and their influence on safety in the area of interest [4]. Expert judgment has recently become recognized as scientific data (i.e., formal elicitation with external validation of expert probability assessments) rather than judgment itself. Data can be collected through observations, programmed trials, surveys, reporting methods, and even expert judgments [5]. There have been several scientific techniques developed that help to treat it as scientific data [2]. For instance, scientific data help synthesize different criteria of the ergatic base complex with the focus on its reliability [6]. Also, scientific data is frequently used in the field of maintenance, repair, and overhaul [7]. Although numerous human error assessment techniques are available for human reliability derivation, they have not been applied in flight safety assessment [8]. The risk assessment and mitigation process were presented through steps logically excluded from the multi-stage process [9].

Generally, the word 'expert' means the person whose judgments are to be elicited regardless of their actual degree of expertise [9]. Another reference describes an expert as someone who has knowledge of an issue at an appropriate level of detail and who is capable of communicating their knowledge [10], [12]. In either way, the statements outline an individual human being with a different personality, experience, and technical background. For this reason, multiple expert opinions about the

targeted point of interest are usually required. But to distinguish an expert from a group of non-experts, one needs to show one's significant knowledge of the subject matter. Therefore, a careful choice of experts has to be assured, so that their estimations will maximize the potential of an accurate and useful outcome.

The first step in expert judgment is the elicitation. The key is to determine, establish, and conduct the process of gathering the experts' opinions. Several methods on the approach were described by O'Hagan et al. [9]. The choice of the number of experts can vary from one to multiple, whereas the analyst often wishes to synthesize the knowledge of more than one expert. This can limit them to the amount of available and valuable experts. The algorithm for combining single probability distributions is known as mathematical aggregation. If the views of a group of experts were gathered and the whole group is treated as a single expert, then the process called behavioral aggregation is used.

This paper uses the classical model that formulates guidelines for using expert opinion in science. Science needs to aim at rational consensus, otherwise the scientific contribution to rational decision making would be compromised [13]. Four principles are created to connect the classical model with the rational consensus, which underlie the model and are described by Cooke [13] and summarized by Aspinall [14]:

1) Accountability/scrutability – individual assessments, realizations, and scores of the experts can be recorded. This allows any future reviewer to analyze the application of the method. Sometimes the true identity of the experts is required, in order to prevent controversial assessments.

2) Empirical control – from scientific point of view, it is important that the expert probability assessments are in principle susceptible to empirical control.

3) Neutrality – before processing the observation results, all experts should be treated equally and their competencies should not be pre-judged [14], [15].

A special software package was used in this paper to calibrate the experts, generate the scores on information they have provided, and calculate global weights, item weights, etc. It implements Cooke's classical model and it was developed by TU Delft with the support of the European Union. Its name, Excalibur, stands for 'EXpert CALIBRation'.

Moreover, the user is able to adjust the settings for their own purposes, such as the calibration power, significance level, scale used (uniform or log-uniform) to achieve the desirable outcome. However, the key to understanding Excalibur is first to understand the classical model and the core statistics it entails. The analyst has to make the right decision to choose the right scale for each variable, which requires certain preliminary knowledge of scaling in probability distribution and statistics.

This paper is structured as follows: chapter 2 provides information on methods, results are outlined in chapter 3, chapter 4 discusses different results per Flight Data Monitoring (FDM) event, observed in the form of every-day data from each aircraft. The investigators in the department are collecting the Aircraft Condition Monitoring System (ACMS) data, and based on their observations, a certain FDM event is sometimes derived. Most of the time the FDM events are the actual exceedances of the limits one aircraft is able to handle. Based on the severity and limitations of exceedances, the FDMs are divided into three classes. Generally, Class 3 events are the most undesirable with the biggest future risk implication possible. However, there is no general agreement in the aviation on the definitions of these events, yet each airline has to decide how it will classify these events so that it suits their needs. Chapter 4 is followed by chapter 5 with conclusions and ideas for further research.

2. METHODS

This paper shows the extent to which a particular airline is willing to apply expert judgment into its everyday business. That is why the flight safety department (of a particular airline) started a research project with the aim to classify the FDM events according to accident risk. If the data are not available, one of the best methods to collect the missing data is the expert judgment elicitation method [16]. We have asked the flight crews to become experts for the purpose of this research. The pilots were asked to perform flight checks of their flying colleagues and to record any flight route issues. The expert judgment method was already applied to gather the pilots' opinions on accident probabilities based on the FDM events. It is important to note that these are their subjective opinions. A group of 20 company pilots were asked to give their probability distributions regarding 12 IATA accident types, namely: controlled flight into terrain, loss of control in

flight, runway incursion/collision, mid-air collision, runway excursion, in-flight damage, ground damage, undershoot, hard landing, gear-up landing, tail strike, and off-airport landing. The description of the IATA accident types is not part of this paper. Only 14 experts from the group were processed (the number of experts included in the study depends on their trustworthiness measured by the EXCAL-IBUR software). This paper presents the data for 2010 and 2011.

Each month several FDM events that occurred and compromised the safety of the airline were presented to the higher management. In order to expand the view on the usefulness of these events, there is a desire to look for new methods of how to utilize them.

A set of Class 3 FDM events was chosen based on the opinions of the people involved in the process of their analyses. Roughly 160 FDM events were ranked by data analysts and the top 40 were used for the elicitation. The reason for this is that there is a big amount of data coming into the investigators' hands each day, and the circumstances needed to be simplified to use expert judgment.

2.1 Elicitation

The elicitation process was intended to be done with 20 different experts, five per each aircraft type – Boeing 737, 747, 777, and Airbus 330. This paper focuses on the presentation of the B747, B777, and A330 results. In the end, only 14 of them gave their judgments. Every single expert was given a questionnaire with the information about the task. Since the elicitation process can be quite time consuming, the experts had enough time to carefully read and understand their role. A phone call was later conducted with each expert to clarify any parts of the questionnaire that might not have been clear enough. If this was the case, the questions were then answered by the analyst. They were forbidden to use any external sources, including the internet or other colleagues to estimate their answers. The goal of the expert judgment is also to gather their current knowledge, so any possible bias has to be avoided. This part of the elicitation thus remains the one with the least trustworthiness, as the best way of performing the elicitation is a direct meeting with the experts to reach adequate control of the process.

The first step for them was to answer eight seed questions (or seed variables) which serve as a foundation for their weights. The better each expert answers all of the eight seed questions, the higher the weight he or she is attributed. The answers of the seed questions were only known to the analyst and not to the experts. Based on their predictions, the weights were given as a measure of how far away the experts were with their answers from the true values. These questions, according to the literature, should be within their field of knowledge.

The questions used for weighting were formulated as follows:

– How many IATA flights were conducted worldwide in the year 2010?
– How many aircraft hull losses (accidents in which the aircraft is destroyed or substantially damaged and is not subsequently repaired) were there per 10 million flights in 2010?
– How many class 3 high-speed events at 500 feet height above touchdown (HAT) ($>V_{approach}+30$ knots) were there per 10,000 (1000) flights with the ABC AIRLINE in 2011 (ABC AIRLINE is used to protect the privacy of the real airline that offered the data? The authors processed real data, but the source of the data is not disclosed because the airline wished to stay anonymous).
– How many Air Safety Reports (ASR) were written in 2011 by the ABC AIRLINE pilots.
– How many ABC AIRLINE ASRs were classified as High Risk in 2011?
– How many ABC AIRLINE ASRs were classified as Medium Risk in 2011?
– How many take-off configuration warnings were there at the ABC AIRLINE in 2011?
– How many rejected take-offs at a speed rate higher than 80 knots were there at the ABC AIRLINE in 2011?

The seed question number three was designed differently for two out of five of each expert per aircraft. Whereas three experts were asked to make their estimations per 10,000 flights, two were asked to make the same per 1000 flights. This question can later lead to a bias in the weighting results and will therefore be left out of the probability analysis.

The second step for each expert was to make their estimations on the variables of interest (or the target questions, or target variables). These were the actual variables for which the expertise was required. The elicitation sheet consisted of 50 Class 3 FDM events in rows and 14 IATA accident types in columns. There was a different elicitation structure used for Boeing 747 pilots with 12 IATA accident types, whereas there were 14 in the case of Boeing

777, and Airbus 330 (in addition to the IATA (2013) accident types, two more accident scenarios were added to the elicitation, namely in-flight injuries and ground injuries). Also, the number of FDM events used for the aircraft was different for each aircraft type. The objective was to quantify the probabilities of a single FDM event contribution to the IATA accident types. Since the airline already operates on a high-safety level, and the FDM Class 3 events are quite rare, there was a need to multiply the severity by stating at the beginning of the elicitation that these Class 3 events must happen under unfavorable conditions. This statement has to provide the experts with a mindset of thinking about the worst possible circumstances in which one of them can find themselves in a certain situation. It was left to the subjective opinion of each expert what he or she considers as an unfavorable condition. In general, these should include conditions like bad weather, fatigue, jet lag, small technical failure, disease, time pressure, and so on.

To cope with the uncertainty, the answers were provided as probability distributions with a span of three quantiles. The quantiles chosen were the 5%, 50%, and 95% values for each seed question, and for all the target questions as well. It is important for the experts in this phase to completely understand what their task is going to be.

With the elicitation data available, the next step is to turn the expert judgment into a representative effigy. All the information from the previous paragraphs is combined together and used to deliver the desired results, which are accident probabilities and their distributions per aircraft, per FDM event, and per accident type. The classical model implemented in the Excalibur is applied to the gathered data. Afterwards the results are presented and visualized. In the end, a comparison of the aircraft types is made, together with the FDM probabilities and accident types.

This method helps to provide answers to the following questions: What are the results of the experts' weighting and who is the best expert? How are the accident probabilities derived and what are the differences between the different aircraft types? What are the probability distributions per accident, per FDM event and per aircraft type?

The first step is the calculation of the experts' weights with the classical model through the Excalibur. The weights are calculated based on the experts' performances in the seed questions. To obtain the weights, the seed questions are put into the Excalibur together with their realization values. The weights are calculated, and the best expert is given the highest weight. This process is repeated four times for each aircraft and with five different experts every time, as one of the goals is to create probability distributions per aircraft.

2.2 Excalibur software setup

The Excalibur setup for the calculations of the Boeing 777 experts is shown in *Figure 1*. The layout of the software in this figure consists of three windows: expert data, realizations data, and assessments for the experts.

The first window is the input of the number of experts with their names. The real names of the experts are replaced by the abbreviations B777-1 – B777-5 due to confidentiality. The second window, realizations data, is the input for the seed questions with their scaling and true realization values. The IDs column represents the seed questions in the same order in which they were written in the questionnaire. Scaling is chosen to be uniform for all the seed questions according to Aspinall and Cooke (personal communication) [17]. The realization column assigns the known true values of each question. Seed question 3 is completely removed from the
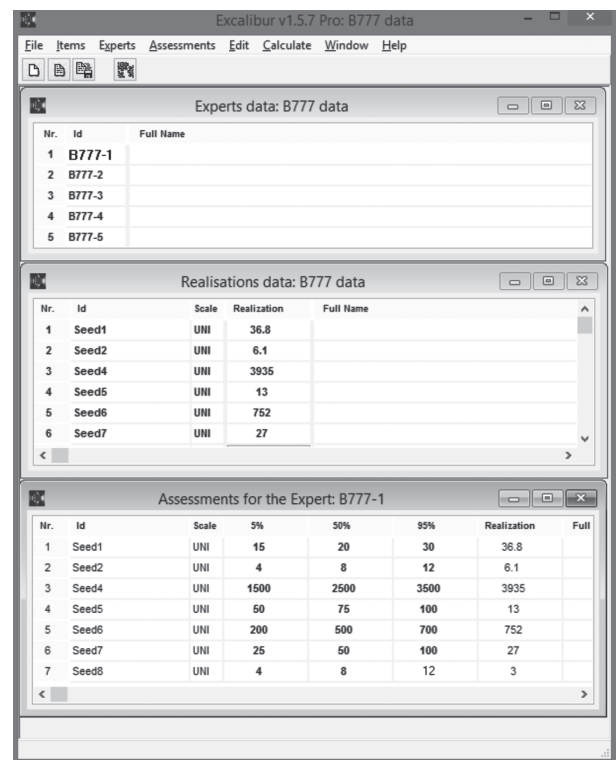

Figure 1 – The Excalibur setup for the calculation of the weights

elicitation because it can bias the weights. The last window, assessments for the expert: B777-1 shows the answers for just the first expert on the seed questions due to space limitations. All the other experts and their assessments have the same format with different data inputs. The window consists of question IDs, scale, and the 5%, 50% and 95% quantiles of their distributions. The last column, realization, is the true value known only to the analyst.
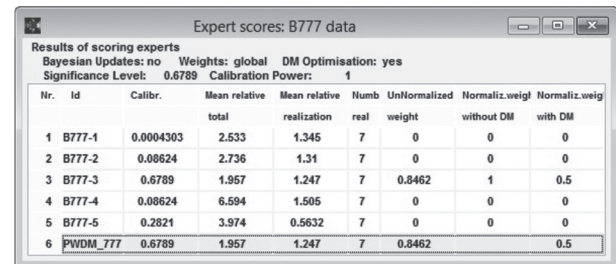
2.3 Running the calculations and gathering weights

Following the correct setup and data input operations in the Excalibur, the weights are calculated. Before this step, a literature review is required on how to arrive at the best weighting method since the classical model can compute four different weighting schemes: global weights, user weights, equal weights, and item weights. Each of the weighting methods results with different calculations of a decision maker. The sum of the experts' weights and the decision maker's weight is always equal to one. If the global weights are chosen, then the global weight decision maker uses performance-based weights determined per expert on his/her calibration and information score. Thus they are also called performance weights. In the case of equal weighting scheme, each expert is given an equal unnormalized weight. This is determined by the weight density for N experts as a 1/N. In the case of five experts, each one is assigned the weight of $1/5 = 0.2$, and the sum is always 1. The third option is to choose the item weighting scheme. The principle is the same as with global weighting, but in global weighting the overall measure of informativeness is used. In item weights, these are determined per expert and per variable as the product of calibration and information for the given item (question). The last option of user weights assigns self-determined weights to each expert [18].

Global weights were chosen with the decision maker optimization turned on, and calibration power 1.00. The resulting decision maker is called PWDM_777. Since the calibration power is the strongest at its highest value 1.00, the highest weight will be assigned to the best calibrated expert.

When it comes to the Boeing 777 data, the best expert rewarded with the weight is the B777-3. They outperformed the rest of the group with a high calibration score (column Calibr.), and their influence on the resulting decision maker is the highest.

After normalizing the weights, they were assigned the weight 0.5 (column "with DM") and the weight 1, respectively (column "without DM"). The next step is to use these weights for the target questions. In this case, only one expert is assigned a weight, and only their estimations are used to answer the target questions. The results are sketched in *Figure 2*.

Results of scoring experts
Bayesian Updates: no   Weights: global   DM Optimisation: yes
Significance Level: 0.6789   Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weight without DM	Normaliz.weight with DM
1	B777-1	0.0004303	2.533	1.345	7	0	0	0
2	B777-2	0.08624	2.736	1.31	7	0	0	0
3	B777-3	0.6789	1.957	1.247	7	0.8462	1	0.5
4	B777-4	0.08624	6.594	1.505	7	0	0	0
5	B777-5	0.2821	3.974	0.5632	7	0	0	0
6	PWDM_777	0.6789	1.957	1.247	7	0.8462		0.5

Figure 2 – Expert scores for the B777 data

3. RESULTS

The same calculation process is repeated three times for each aircraft separately. All the inputs are the same, only the weights change depending on different performances of different experts per aircraft. This paragraph provides the results for the accident probabilities. Under the same circumstances and same seed questions, the best performing expert is B777-3, who achieved a weight of 0.5 – the highest in the expert group of the Boeing 777 aircraft.

3.1 Probabilities of Boeing 777 FDM events per accident type

Figure 3 depicts the probabilities of FDM events contributing to the controlled flight into terrain. The Y axis is the number of accidents that can occur in 10,000 occurrences of a particular FDM event. The X axis represents single FDM events chosen by the experts from the set. The graph area consists of probability distributions with the median values highlighted. The higher the median value lies in relation to the Y axis, the higher the number of potential accidents a particular FDM event can lead to. Therefore, the risk posed by different FDM events is higher or lower depending on particular events.

The accident type with the highest number of contributing FDM events in case of the Boeing 777 is the loss of control with 44 events (the figure is not part of this paper). The second one is the controlled flight into terrain with 41 events (*Figure 3* depicts 21 of them).
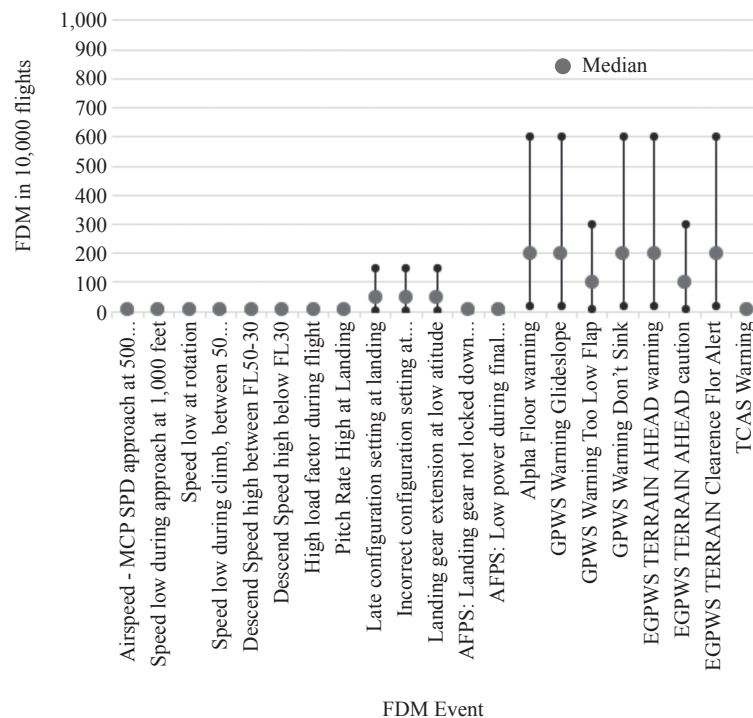
Figure 3 – Probability of FDM events contributing to the controlled flight into terrain (B777)

3.2 Contribution of FDM events to particular accident types (Boeing 777)

Each FDM event has a different contribution to different accident types. *Table 1* shows the number of accident types per FDM event for Boeing 777. The results can be presented in various ways. The table shows the name of the FDM event as well as probability distributions and accident types. One way is to take a look at the Total column, which shows the number of accident types connected to each event. According to the Boeing 777 experts, the most contributing FDM events are the ones triggered by the GPWS warnings (pull up, terrain pull up, and sink rate as depicted in *Table 1*) and the high descent rate during the approach from 500 to 50 feet. The single dangerous event appears to be the windshear warning resulting in 10 different accident types.

The second option is to look at the probability distributions. The quantile approach used in the elicitation describes the 50% median value as the one that is the closest to the expert's opinion. The 90% credible intervals (45% for quantile 5%–50% and 45% for quantile 50%–95%) are distributed around this value. The number of accidents for a particular event would fall into the 5%–95% interval with a 90% chance. The higher the numbers presented in the quantile columns, the higher the probability of the accident occurrence. To put that in an everyday

perspective, the relative frequencies of accidents in 10,000 events are higher or lower depending on the opinions. A relative frequency takes the number of the observed events divided by a number of trials. The number of trials is 10,000 occurrences of a particular event, and the number estimated by the experts in this case is the number of occurrences. It means that the higher the estimated numbers, the more likely each event occurs.

3.3 Airbus 330 results per accident type

The opinions of four experts were taken into account regarding the Airbus 330. The highest number of contributing FDM events is 23 in case of hard landing (*Figure 4*), followed by 19 events with runway excursion and in-flight damage (figures are not part of this paper). Some of the events (8 in total) were assigned a zero value above their median estimates, but they have achieved a score above the 95% quantiles. For this reason, they are included in the figure, but should be neglected as expert estimates because these events were zero.

3.4 Airbus 330 results per FDM event

'Airspeed - MCP SPD approach at 500 feet' accompanied with 'Speed low during approach at 1000 feet' are the events with most accident types

Table 1 – Number of accident types per FDM event (Boeing 777)

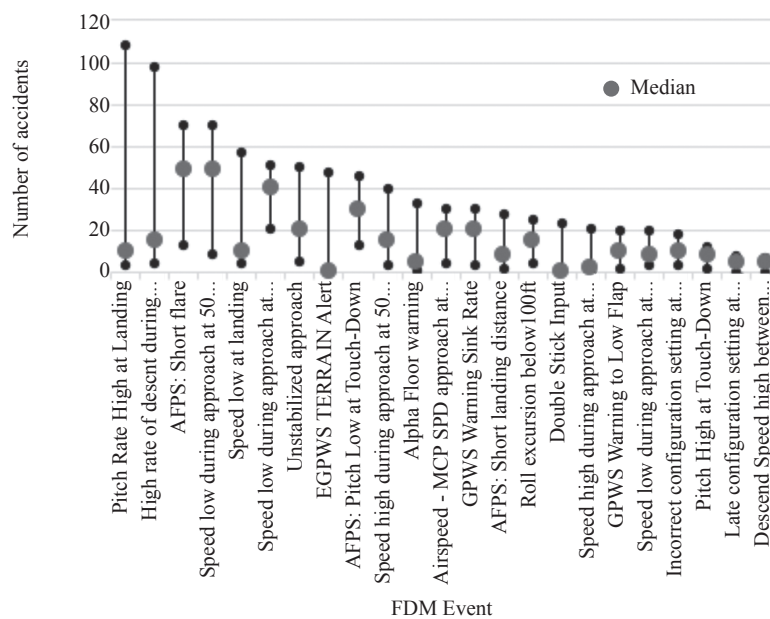| FDM event	5%	50%	95%	Accident type	Total
GPWS Warning Pull Up	50	500	1000	CFIT	9
	50	500	1000	Gear Up Landing – Gear Collapse	
	50	500	1000	Hard Landing	
	50	500	1000	Tailstrike	
	50	500	1000	Undershoot	
	20	200	600	Loss of Control	
	20	200	600	Off-Airport Landing – Ditching	
	10	100	300	In-Flight Injuries	
	1	5	15	Mid-Air Collision	
GPWS Warning Sink Rate	50	500	1000	Gear Up Landing – Gear Collapse	8
	50	500	1000	Hard Landing	
	50	500	1000	Tailstrike	
	50	500	1000	Undershoot	
	20	200	600	Loss of Control	
	20	200	600	Runway Excursion	
	20	200	150	CFIT	
	1	5	15	Mid-Air Collision	
GPWS Warning Terrain Pull Up	50	500	1000	CFIT	9
	50	500	1000	Gear Up Landing – Gear Collapse	
	50	500	1000	Hard Landing	
	50	500	1000	Tailstrike	
	50	500	1000	Undershoot	
	20	200	600	Loss of Control	
	20	200	600	Off-Airport Landing – Ditching	
	10	100	300	In-Flight Injuries	
	1	5	15	Mid-Air Collision	

Figure 4 – Hard Landing (A330)

related to them – eight per each (*Table 2*). Some of the events were assigned a zero value above their median estimates, but they have achieved a score above the 95% quantiles. For this reason, they are neglected.

3.5 Boeing 747 results per accident type

All of the accident types were quantified with probabilities. Most of the events were related to the loss of control accident type with 38 in total (*Figure 5* depicts half of them). A significant number of events contribute to tailstrike (37) and gear-up landing (32) (figures are not part of this paper). Runway incursion was found with only one contributing event (the figure is not part of this paper). Some of the events (4 in total) were assigned a zero value above their median estimates, but they have achieved a

score above the 95% quantiles. For this reason, they are included in the figure, but should be neglected as expert estimates because these events were zero.

3.6 Boeing 747 results per FDM event

As was the case with Boeing 777, the windshear warning is related to the most of the accident types (11) (*Table 3*). The same score was achieved by the descend speed high between FL50–FL30 (the figure is not part of this paper). There are several more events related to 10 accident types, namely airspeed – MCP SPD approach at 500 feet, descend speed high below FL30, GPWS warning pull up, terrain pull up, sink rate, terrain low and too low gear, then high load factor during the flight, unstabilized approach, and also all speed low during approach events (for 1000, 500 and 50 feet).

Table 2 – Number of accident types per FDM event (Airbus 330)

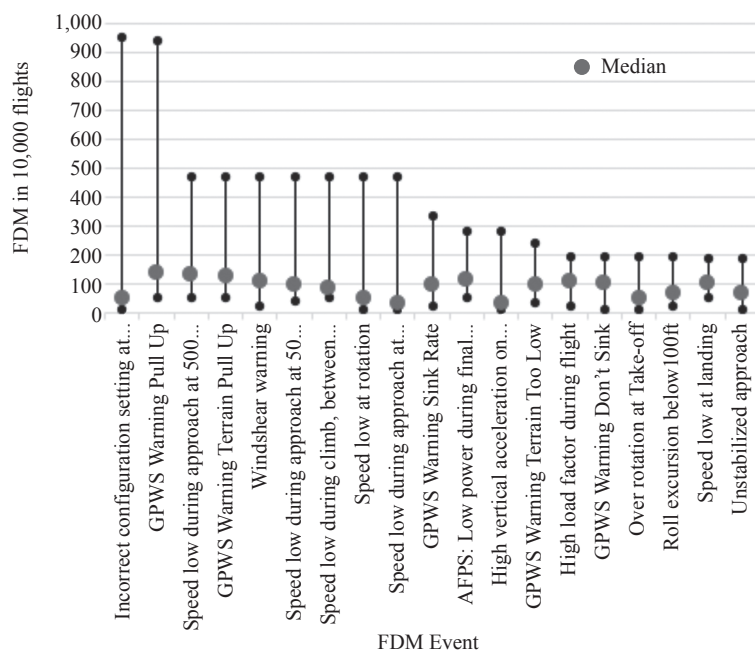| FDM event	5%	50%	95%	Accident type	Total
Airspeed - MCP SPD approach at 500 feet	20	200	600	Runway Excursion	4
	1	5	15	CFIT	
	1	5	15	Loss of Control	
	1	5	15	Runway Incursion	
Speed low during approach at 1000 feet	20	200	600	Hard Landing	6
	20	200	600	Tailstrike	
	10	100	300	Loss of Control	
	10	100	300	Undershoot	
	1	5	15	CFIT	
	1	5	15	Gear Up Landing - Gear Collapse	

Figure 5 – Loss of Control in Flight (B747)

Table 3 – Number of accident types per FDM event (Boeing 747)

FDM event	5%	50%	95%	Accident type	Total
Windshear warning	16	48	767	Hard Landing	11
	4	18	474	Controlled Flight Into Terrain	
	5	19	473	In-Flight Damage/Injuries	
	24	107	473	Loss of Control	
	21	58	473	Runway Excursion	
	6	29	473	Undershoot	
	41	90	287	Tailstrike	
	4	10	284	Gear Up Landing	
	0	6	25	Ground Damage/Injuries	
	2	6	20	Mid-Air Collision	
	0	1	4	Off-Airport Landing	

4. DISCUSSION

According to the expert judgment, different FDM events can lead to different accident types. In case of the Boeing 777 experts, the model has given 50 FDM events in total contributing to the accidents. In case of the Airbus 330, it was 48 events, and in case of the Boeing 747, the number of events was 45 (these are not part of the paper due to the page count limitation). This can be partially caused by the different number of FDM events in the elicitation sheets. The second reason is that the best experts simply do not believe that some of the events can contribute to certain accidents, whilst others do. If one wants to standardize all these events into one figure, the classical model does not allow it. *Table 4* shows the same FDM event for three aircraft types. Whereas for 777 and 330 the number of the accident types was the same (6), in case of the 747 it was higher (10).

The first two aircraft types differ in one FDM event – for 777 it is runway incursion, and for 330 it is hard landing. This suggests that the experts were quite close with their estimations; however, the differences in numbers still raise some questions. Therefore, it is difficult to generalize or standardize three different elicitations into one single piece, whether it is in terms of probabilities, numbers of accidents, or aircraft types. A whole new elicitation for all the experts should be conducted in this case.

The results of the calculations are expressed by three quantile values per each event. Since it is known that each of these events occurred 10,000 times, and according to the expert judgment, it is also known what the number of the resulting accidents can be, it is possible to talk about relative frequencies. Rel-

ative frequency model is a mathematical model that defines the probability of an event occurring during the monitoring of the trials of the event. That means the higher the quantile values, the higher the relative frequency and the higher the risk related to different FDM events. The differences between the aircraft types are apparent. Also, they are caused by different elicitation structures, meaning different number of events or different number of IATA accident types.

There was a different elicitation structure used for the Boeing 747 pilots with 12 IATA accident types, whereas there were 14 in case of the Boeing 777 and Airbus 330. Also, the number of FDM events used for the aircraft was different for each aircraft type. This makes it much more difficult if the intention is to generalize or standardize all the probabilities into one model. Furthermore, for this reason, all the experts are weighted separately for their aircraft types. The best rated expert is the Boeing 777 expert number 3. The suggestion for future improvement is to use this expert for creating the most accurate estimations of probabilities.

5. CONCLUSION

This paper provides answers to technical questions. It describes the implementation of the classical model into this research, then continues with the calculations and visualization of the outcomes. The paper reveals the probability distributions per each aircraft type, per IATA accident type, and per FDM event.

The classical model was found applicable for this type of data. That is why the EXCALIBUR software was used. The weighting revealed that the

Table 4 – The comparison of different results per FDM event

Aircraft type	FDM event	5%	50%	95%	Accident type	Total
Boeing 777	Descend speed high below FL30	20	200	600	In-Flight Damage	6
		20	200	600	Runway Excursion	
		5	50	150	Mid-Air Collision	
		1	5	15	Controlled Flight Into Terrain	
		1	5	15	Loss of Control	
		1	5	15	Runway Incursion	
Airbus 330	Descend speed high below FL30	0	2	35	Controlled Flight Into Terrain	6
		1	5	15	Mid-Air Collision	
		1	5	7	Hard Landing	
		2	5	7	In-Flight Damage	
		1	5	7	Runway Excursion	
		0	2	4	Loss of Control	
Boeing 747	Descend speed high below FL30	10	19	149	Hard Landing	10
		11	41	89	Loss of Control	
		2	6	79	Mid-Air Collision	
		2	8	79	Runway Excursion	
		2	9	59	Gear Up Landing	
		10	23	59	In-Flight Damage/Injuries	
		1	4	50	Controlled Flight Into Terrain	
		1	5	24	Tailstrike	
		1	4	10	Undershoot	
		0	1	4	Ground Damage/Injuries	

best expert out of the 14 is B777-3 with a weight of 0.5. The probability distributions showed the differences between the aircraft types, partially caused by the different elicitation sheets and different weighting of the experts, which is done separately per aircraft. More specifically, the probability distributions per IATA accident type in case of the A330 are different than for B777 and B747 in terms of the number of the IATA accident types, in terms of the number of events (different number for each of the aircraft type), and in terms of event probabilities. The same is applicable to the distributions per FDM event. The distributions differ in terms of the aircraft, in terms of the event, and in terms of the IATA accident type.

The authors believe that the following questions are worth further research efforts. To what extent are expert weights sensitive to the removal of one seed question at a time? Are the variables of interest sensitive to the removal of one expert and one seed question from the set? That kind of research could provide answers to statistical questions about sensitivities and influences of single experts and seed questions on weights and on target variables. The sensitivity test could reveal whether the classical model is more sensitive to the removal of one seed question at a time than to the removal of one expert at a time. We assume that the removals would have an influence on the target questions.

As for the elicitation, standardization is recommended in terms of terminology and numbers for the same FDM event, as well as for the same IATA accident types for all the aircraft types to minimize the differences amongst them. The next step would be to create relations and correlations amongst the single FDM events. For example, there are three FDM events related to the approach speed. They only differ in altitudes. The next step can be done by connecting them together, first elicit and study occurrences at 1000 feet, then create a Bayesian Network ranked correlations with the same event at altitudes of 500 and 50 feet.

The next approach would be oriented towards the creation of a Bayesian Belief Network. The distributions are known, now they need to be used for the quantification of the nodes. The network should be designed beforehand also with the help of the experts. After the nodes are quantified, rank correlations should be computed from the data. The important step to make is to connect the FDM events with each other where possible, also with the help of the experts. The use of the classical model, however, ends with the nodes quantification. The simulation and sampling of the network can be done through the Monte Carlo method with Beta distributions of the best fit.

doc. Ing. **BENEDIKT BADÁNIK**, PhD.[1]
E-mail: benedikt.badanik@fpedas.uniza.sk
Ing. **MARTIN JÁNOSSY**[2]
E-mail: martin.janossy@gmail.com
ARTHUR DIJKSTRA, MSc.[3]
E-mail: Arthur@admc.pro
[1] Katedra leteckej dopravy, Fakulta prevádzky
  a ekonomiky dopravy a spojov
  Žilinská univerzita v Žiline
  Univerzitná 1, 010 26 Žilina, Slovakia
[2] Řízení letového provozu ČR
  Navigační 787, 252 61 Jeneč u Prahy, Česká republika
[3] Kooikerboog 7, 1394ND Nederhorst den Berg,
  Netherlands

VYUŽÍVANIE METÓDY EXPERTNÉHO ODHADU NA STANOVOVANIE PRAVDEPODOBNOSTI LETECKÝCH NEHÔD

ABSTRAKT

Zvyšovanie úrovne bezpečnosti je jedna z hlavných priorít v komerčnej leteckej doprave. V súčasnosti máme k dispozícii analýzy bezpečnosti a rizík v leteckej doprave, ktoré poskytujú detailnejší obraz o priebehu leteckých nehôd. V minulosti mali letecké spoločnosti podstatne menej podrobných informácií o leteckých nehodách a neboli preto schopné spoľahlivo odpovedať na otázky, týkajúce sa bezpečnosti v leteckej doprave. Cieľom tohto článku je zistiť a popísať možnosti využívania metódy expertného odhadu pri stanovovaní pravdepodobnosti výskytu leteckých nehôd a napomôcť tak včasnému odhaľovaniu príčin vzniku leteckých nehôd a zvyšovaniu bezpečnosti v leteckej doprave.

KĽÚČOVÉ SLOVÁ

expertný odhad; klasický model; softvér Excalibur; pravdepodobnosť nehody.

REFERENCES

[1] Ažaltovič V, Škvareková I, Pecho P, Kandera B. Calculation of the Ground Casualty Risk during Aerial Work of Unmanned Aerial Vehicles in the Urban Environment. In: Stopkova M, Bartuška L, Stopka O. (eds.) *Transportation Research Procedia - Proceedings of the LOGI 2019 conference - Horizons of autonomous mobility in Europe, 14 – 15 November 2019, České Budějovice, Czech Republic.* Amsterdam, NL: Elsevier; 2020. p. 271-275.

[2] Unal R, Keating C, Conway B, Chytka T. *Development of an expert judgment elicitation and calibration methodology for risk analysis in conceptual vehicle design.* Old Dominion University. Report number: 130012, 2004.

[3] Tyshenko MG, El Saadany S, Oraby T, Darshan S, Catford A, Aspinall W, Krewski D. Expert judgment and re-elicitation for prion disease risk uncertainties. *International Journal of Risk Assessment and Management (IJRAM).* 2012;6(1-3): 234-248. Available from: http://www.inderscience.com/offer.php?id=47552 [Accessed 17th May 2014].

[4] Galieriková A, Materna M, Sosedová J. Analysis of risks in aviation. In: Ostaševičius V. (ed.) *Transport Means – Proceeding of the 8th International Scientific Conference, 30 September – 2 October 2008, Kaunas, Lithuania.* Kaunas, LT University of Technology; 2008. p. 1427-1431.

[5] Berlowitz I. Transport aircraft human errors quantification. In: *Proceedings of the 59th Israel Annual Conference on Aerospace Sciences 2019. 6-7 March 2019, Tel-Aviv and Haifa, Israel.* Technion Israel Institute of Technology; 2019. p. 693-786.

[6] Sedláčková AN, Kurdel P, Mrekaj B. Synthesis criterion of ergatic base complex with focus on its reliability. In: Novitzká V, Korečko Š, Szakál A. (eds.) *Proceedings of the 2017 IEEE 14th International Scientific Conference on Informatics – INFORMATICS´ 2017, 14-16 November 2017, Poprad, Slovakia.* Košice, Slovakia: TU Košice; 2017. p. 318-321.

[7] Bugaj M. Aircraft maintenance – New trends in general aviation. *Promet – Traffic&Transportation.* 2005;17(4): 231-234.

[8] Guo Y, Sun Y. Flight safety assessment based on an integrated human reliability quantification approach. *PLoS ONE.* 2020;15(4). DOI: 10.1371/journal.pone.0231391 [Accessed 16th Sept 2020].

[9] Novak A, Skultety F, Bugaj M, Jun F. Safety studies on GNSS instrument approach at Zilina airport. In: Andoga R. (ed.) *Proceedings of the MOSATT 2019 International Scientific Conference, 28-29 November 2019, Košice, Slovakia.* Piscataway, New Jersey; 2019. p. 122-125.

[10] O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Rakow T. *Uncertain Judgements: Eliciting Experts' Probabilities.* Sheffield, UK: John Wiley & Sons, Ltd.; 2006.

[11] Meyer M, Booker J. *Eliciting and analyzing expert judgment: A practical guide.* Office of Nuclear Regulatory Research, Division of Systems Research. Report number: NUREG/CR-5424, 1990.

[12] Burgman M, Fidler F, McBride M, Walshe T, Wintle B. *Eliciting Expert Judgments: Literature Review.* Australian Centre for Excellence in Risk Analysis. Report number: 1, 2006.

[13] Cooke RM. *Experts in uncertainty.* New York, NY: Oxford University Press; 1991.

[14] Aspinall W. *Expert judgment elicitation using the*

classical model and excalibur, Briefing notes for Seventh Session of the Statistics and Risk Assessment Section´s International Expert Advisory Group on Risk Modeling. Willy Aspinall; 2008. Available from: http://dutiosc.twi. tudelft.nl/~risk/extrafiles/EJcourse/Sheets/Aspinall%20 Briefing%20Notes.pdf [Accessed 17th March 2020].

[15] Aspinall WP, Bailey L, Hincks TK, Poole M, Sparks RS. *Development of a Bayesian Belief Network to assess variant Climate Change Scenarios.* Bristol, UK: University of Bristol, 2008.

[16] Kazda A, Caves R. *Airport design and operation*. 3rd ed. Bingley, UK: Emerald Group Publishing Limited; 2015.

[17] Janossy M. Email sent to Cooke RM. 5th May 2015.

[18] Cooke RM, Goossens LJ. *Procedures guide for structured expert judgment*. Delft University of Technology Delft. Report number: ETNU-CT93-0104-NL, 1999.