On the Reliability of Experts' Assessments for Autonomous Underwater Vehicle Risk of Loss Prediction: Are Optimists better than Pessimists?

Mario P. Brito[a] and Yujia Chang[a]
[a] University of Southampton, Southampton, United Kingdom

Abstract: Expert judgment elicitation is a key element of formal risk assessment. Some research in this subject has focused on identifying the best way to aggregate expert judgments. In this paper we explore this problem. Given the divergence in expert judgments, when using mathematical aggregation, it is possible to group expert judgments according to their mood, as optimists and pessimists. Using hard data, gathered after the expert judgment elicitation process, we test which group performs better. In this paper, we group the expert judgments elicited for building the risk model for the Nereid-UI hybrid autonomous underwater vehicle into optimists and pessimists. After the risk assessment the vehicle conducted 16 missions. We compared the two risk models from the risk assessment against the observed risk from actual missions. Our results showed that, for a 5 hours mission, differences between the pessimist risk model estimates and the observed risk were not statistically significant. On the other hand, differences between the predicted risk using the optimistic risk model and the observed risk were statistically significant. We conclude that for early missions in an extreme environment it is imperative to use the pessimistic risk model estimates to inform decision making.

Keywords: expert judgment aggregation, autonomous underwater vehicles, risk, risk of loss, reliability.

1. INTRODUCTION

It is irrefutable that today's risk assessment of complex systems is heavily dependent on expert judgment elicitation. This is particularly the case for problems where there is no hard data and the consequences of potential hazards can be catastrophic [1]. Expert judgment elicitation is the process of eliciting judgments on the likelihood of a hazard occurring from individuals deemed experts in a given subject matter. In engineering, research in expert judgment elicitation began in the 1970s when scientists and engineers made new efforts to improve safety risk assessment for Nuclear Power stations. Safety valve failure frequency prediction, nuclear waste safety risk prediction and many other applications were considered then by those involved in these processes [2, 3]. This research was grounded on studies in psychology, particularly on human reasoning under uncertainty. Amongst some key studies was that reported by Kahnemann and Tversky [4], which explored the mental short cuts that individuals follow when making estimates of uncertainty. When following these mental shortcuts individuals can introduce bias. Bias can also be introduced by the modeler during the process of aggregating the expert judgments or by the facilitator during the elicitation process. Formal expert judgment elicitation method is a process whereby judgments are elicited from experts with the aim of reducing the introduction of bias. Formal expert judgment elicitation provides transparency and enables reproducibility of the results, both of which are key characteristics for effective decision making. Formal expert judgment elicitation methods, such as the DELPHI method, the OTWAYs, EXCALIBUR, the SHELF-R and others, consider different ways for aggregating expert judgments. These methods apply different forms of mathematical and behavioural aggregation [5-7].

When Brito et al [12] applied the Otway expert judgment elicitation, which employs an analytical aggregation method, they concluded that two types of experts were present: the optimists and the pessimists. The first is not to be confused with optimistic bias, whereby people weight their own chances of success as better than average and their own chances of failure as lower than average [8].
When asked to estimate the probability of fault 34, loss of network leading to Autosub3 loss in a coastal environment, a number of experts assumed that the AUV would ground on a beach and be

safely recovered, while others assumed that the AUV would crash against coastal rocks [9]. It is possible that availability and representativeness mental shortcuts may have been used by the experts, after all, these types of incidents happen [10, 11].

Of course, one can make an argument that the environment should have been described in more detail. One could have specified whether the coastline to be considered is a sandy beach or a rocky shore. Given the variability in the coastline geology and the effect that the sea current and tides can have on the AUV dynamics, moving the AUV from a few hundreds of metres to a few kilometres, understandably the geology of the coastline was not specified.

Recognising that an expert can be predominantly optimistic or pessimistic, the expert judgments had to be aggregated into these two groups. Brito et al [9, 12] used cumulative distributions to identify the mood of the experts. The expert judgment aggregation was done using a weighted linear pool. The decision maker was given mission risk estimates from the optimists' and pessimists' judgment aggregation models. The decision to deploy Autosub3 in Antarctica in 2009 and 2013 was informed by the optimistic model.

In this paper we explore the validity of this approach for AUV risk model mission risk analysis. We explore the problem of experts' reliability by comparing the risk predictions made by optimist and pessimist groups with the risk model obtained from field results. We use the dataset from the Nereid-UI AUV to identify which of the two risk models produce risk estimates that are most in line with the observed risk from field results.

The manuscript is organised as follows: Section 2 presents a background to the Nereid-UI risk model, Section 3 presents the methodology used for assessing the reliability of experts' judgment predictions, Section 4 presents the case study. Section 5 presents the analysis of the results and, finally, Section 6 presents our conclusions and areas that we have identified that require further research.

2. NEREID-UI RISK MODEL AND OPERATIONAL DATA

Stokey et al [13] initiated he discussions on finding the methods of increasing the reliability and to reduce the risks of AUV systems. To develop this further, Griffiths et al [14] used statistical survival methods to evaluate the probability of risk of loss for Autosub2 operations under ice. AUV loss is a real risk, and there is a body of anecdotal evidence concerning AUV loss. For example, reported in the public domain are the losses of Autosub2 under the Fimbulisen ice-shelf in 2005 [15] and the loss of Autonomous Benthic Explorer (ABE) off the Chilean coast in 2010 [16].

Methods for estimating AUV risk of loss based on hard data have been criticised for the level of subjectivity involved in the failure criticality classification. Developments to elicit this subjectivity in a formal way were suggested by Griffiths and Trembanis [12]. This informed subsequent risk models of AUV loss. Up to this point the models were based on historical data and expert judgments.

At the time of the risk analysis the Nereid-UI was a novel vehicle, with no historical operational data. The risk model was developed based on an analysis of the vehicle design and its functional and operational requirements. The following subsections describe the risk model in more detail and the operational data.

2.1. Nereid-UI risk model

There are three key elements to the risk model of the Nereid-UI. The first element is a Markov chain model of the phases that comprise the deployment of the vehicle. This model extends that proposed for the Autosub3 AUV [18]. Eighteen states of operation and their associated transitions were identified; these include pre-mission states and states capturing different modes of operation: tethered and un-tethered. The probability of failure in the transition from one state to another was calculated using fault trees. Thirty fault trees were developed. For this study we consider the fault tree for Transit in

under ice state transition to itself [26]. This fault tree has 113 independent failure modes. A subtree of this fault tree is the Core Control System Failure fault tree depicted in Figure 1, below.

The fault tree and the judgments were provided by five experts: SDM, LB, JO, MJ and DY. These experts had more than 10 years of experience individually. The experts worked for oceanography centres in the United States, France and in the United Kingdom. The experts assessed the likelihood of each failure occurring. The experts provided the median, the lower bound and the upper bound. Assessments were provided for all 113 failures. The workshop was held at Woods Hole Oceanographic Institute, Cape Cod, United States on the 20 June 2012.

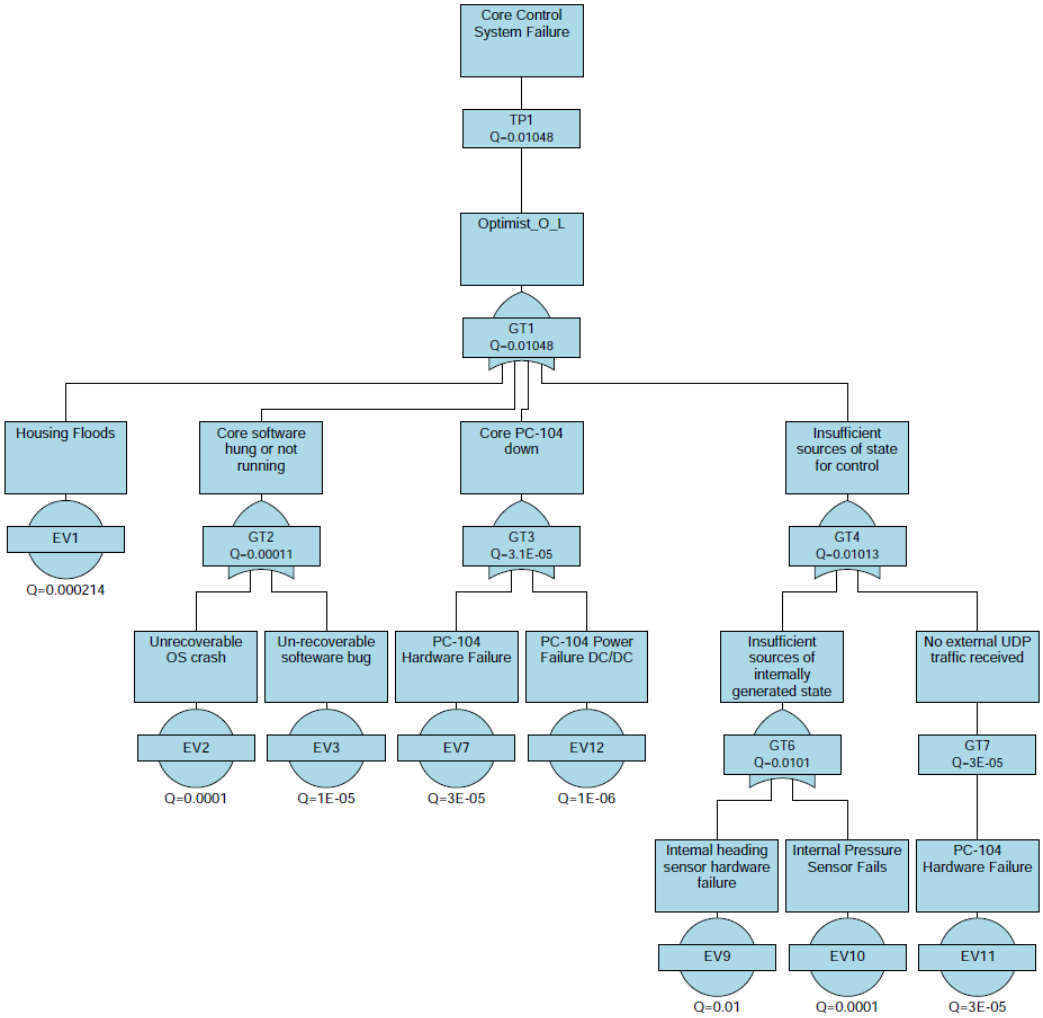

Figure 1: Core Control System Failure fault tree model

2.2. Nereid UI Operation data

The Nereid-UI is a hybrid autonomous underwater vehicle, designed to work close to the downside of glaciers/sea-ice and around the sea-floor. The vehicle is denoted a hybrid autonomous underwater vehicle because it can work connected to a ship with a fibre optical tether, or not connected to the ship – in autonomous mode. The vehicle provides a unique capability for polar research by enabling more detailed observation and sampling of the environment in cavities of the ice shelf. When operating in a tethered mode the vehicle provides live video feed of the environment [19]. The Nereid-UI AUV design was informed by the design of the SENTRY AUV, developed by the same team. Core missions of Nereid UI require a high degree of reliability; thus, assessing the risk for under ice deployments is of great importance. The missions considered in this paper are those conducted up to the submission of this paper. Some missions were technology development missions, while the remaining missions were science missions [20-23]. Nereid-UI's missions nui009-012 were test

missions for deployment and recovery. Short runs were conducted whilst underwater. Mission nui013 aborted due to software configuration error.

Table 1: Mission summary of the Nereid-UI AUV

Mission	Date	Dive Duration
nui0001	7/21/2014	05:08
nui0002	7/23/2014	05:05
nui0003	7/26/2014	05:01
nui0004	7/28/2014	05:20
nui0005	15/09/2015	02:01
nui0006	16/09/2015	01:40
nui0007	17/09/2015	01:17
nui0008	17/09/2015	01:47
nui0009	02/05/2016	00:30
nui0010	03/05/2016	00:30
nui0011	04/05/2016	00:30
nui0012	07/05/2016	00:30
nui0013	08/05/2016	00:00
nui0014	27/09/2016	08:24
nui0015	30/09/2016	11:55:00
nui0016	08/10/2016	05:07:00

3. METHODOLOGY

3.1. Un-weighted Linear Pool Aggregation Method

For the expert judgment elicitation conducted by Brito et al [9], experts provided a weight that reflected their confidence in the assessment. On the other hand, for the Nereid-UI expert judgment elicitation, experts provided a probability range. Therefore, for the Nereid-UI judgment aggregation the un-weighted linear pool method was used to aggregate the experts' judgements into two separate groups based on the different experts' moods: optimistic and pessimistic. The mean estimates were calculated for the Lower bound (L), the Median (M) and the Upper bound (U). Where the Lower bound is the minimum probability of failure possible, the Upper bound is the maximum probability of failure possible. The Median is the estimated probability of failure for each, there is a 50% chance that the true probability of failure is higher than this probability of failure and lower than the Upper bound, and a 50% chance that real probability of failure is below this value and the Lower bound.

3.2. Reliability Function

Several reliability functions are presented in the literature. For new components and systems, it is appropriate to use the exponential reliability function [24]. This reliability function is univariate and it takes the variable time, t. The other parameter in this function is the failure rate λ, or the mean time to failure (MTTF), which is 1/λ. The failure probability density function is presented in Eq. 1.

$$f(t) = \lambda \cdot \exp(-\lambda \cdot t) \qquad (1)$$

In addition, based on this, the probability of no failures occurring before time t is obtained by integrating Eq. 1 between 0 and t - obtaining the cumulative probability of failure - and subtracting from 1. The reliability, or survival function is presented below.

$$R(t) = S(t) = 1 - \int_0^t f(t)dt = \exp(-\lambda \cdot t) \quad (2)$$

The probabilities elicited from experts' judgement in the research were based on a mission length of 20h. The maximum mission length from the data was approximately 11h. In order to compare the

experts' risk estimates with the observed risk estimates, we must calculate the risk for a time shorter than 11h. We chose 5h as 5 out of 13 missions had this length of operation.

3.3. Binomial Distribution Calculation

The loss or survival of a given deployment can be seen as a trial in a Bernoulli experiment. The binomial distribution is a well-known discrete probability distribution which enable us to estimate the probability of observing a number r successful outcomes out of n experiments. Since the total number of missions carried out by Nereid UI was 16, we considered this as the total number of experiments. The binomial distribution is presented below.

$$P(x=r) = \begin{pmatrix} n \\ r \end{pmatrix} \cdot p^r \times (1-p)^{n-r} \qquad (3)$$

where, n =1,…,16. P means the probability of success after n trial, the p is the probability of success in q single trial.

The binomial distribution enabled us to estimate the number of successes and failures for each of the two models generated by the expert judgments. The X^2 (chi square) non-parametric test enabled us to estimate whether the differences between the predicted successes and observed successes were statistically significant.

4. ANALYSIS

4.1. Results of the un-weighted linear pool aggregation

With 113 potential failures, five experts and three assessments per expert, we had a total of 1695 estimates. Using the un-weighted linear pool aggregation method, statistics for the judgements by two different expert groups in both optimistic and pessimistic risk attitudes were calculated. Table 2, below, presents a summary of the aggregated assessments provided by the experts for failures 3 to 11.

Table 2: Summary of the aggregated assessments provided for failures 8 to 11.

		Optimist				**Pessimist**			
No	Failure description	L	M	U	Experts	L	M	U	Experts
8	Thruster Reliability	0.0002	0.0023	0.0045	SDM,JO, DY	0.0295	0.265	0.305	MJ,LB
9	Depth Sensor Reliability	0.0002	0.0005	0.00165	MJ,SDM ,JAN,DY	0.009	0.01	0.011	LB
10	Phins Reliability	0.0002	0.002	0.0039	SDM,JA N,DY	0.0095	0.015	0.0555	MJ,LB
11	Microstrain Reliability	3.67E-05	0.0004	0.0018	SDM,JA N,DY	0.005	0.01	0.0305	MJ,LB

For each same failure, the group of pessimists always gave a higher probability to the failure compared with the optimistic group, and the Upper bound value for the optimists was even less than the Lower bound in the pessimists' group. This difference may have originated from systematic errors of subjective expert judgement due to the heuristics and biases [4], which may be eliminated by the weighted linear pool aggregation method, but this is not possible to do if we elicit the distribution density function instead.

4.2. Results of Fault Tree Analysis

Six fault tree analyses were performed; three fault tree analyses for the optimistic model, using estimates for the Lower bound, Median and Upper bound, and three fault tree analyses were estimated for the pessimistic model, one for each of the three parameters. Table 3, below, presents the probability of failure calculated for the 14 sub-trees and the combined probability of failure.

Table 3: Probability of failure calculated using fault trees and expert judgments for the 14 critical components for a 20h mission. The total probability of failure for the Nereid-UI transit under ice is presented in the last row

No.	Failure	**Optimist**			Pessimist		
		L	M	U	L	M	U
1	Core Control System Failure	0.01048	0.02328	0.121	0.2486	0.2875	0.3956
2	PDS Failure	0.0001	0.0010	0.0030	0.0046	0.0075	0.0305
3	Battery Net Working	0.0203	0.0301	0.0404	0.1058	0.1667	0.3498
4	Dry Cable Failure	0.0001	0.0050	0.0100	0.0075	0.0150	0.0325
5	Vehicle Shape Changes	0.0176	0.0376	0.0501	0.0936	0.1678	0.6639
6	Bad-Wet-Connection	0.0107	0.0113	0.0118	0.2464	0.3080	0.3605
7	Housing Floods	0.0002	0.0014	0.0055	0.2383	0.2633	0.2876
8	Failed Electrical Board	0.0004	0.0013	0.0032	0.1304	0.1667	0.2329
9	OA Failure	0.0167	0.0220	0.0455	0.1376	0.2327	0.3531
10	Vehicle Stuck on Bottom	0.0029	0.0061	0.0202	0.1388	0.1758	0.4621
11	Vehicle Stuck on Underside of Ice	0.1270	0.2401	0.3492	0.1462	0.2680	0.3888
12	ROV Commands Bad Motion	0.0004	0.0014	0.0120	0.0224	0.0395	0.0443
13	FO Failure	0.1340	0.1598	0.2738	0.4236	0.4996	0.5708
14	Pilot Error	0.0478	0.0870	0.1733	0.2049	0.3138	0.4377
15	Transit in Underwater Ice	0.2280	0.5497	0.6004	0.8058	0.9097	0.9915

Considering the median, the pessimist expert aggregated model gave a 0.9097 probability of loss of the vehicle under ice in a 20h mission, based on design information. The optimist experts gave a much smaller probability of loss than the pessimist model, 0.5497. Although the result given by the optimist experts was significant smaller than the pessimist experts, more than a 50% probability of loss still illustrates a high level of risk for the vehicle. However, this figure is comparable to that of Autosub 3 risk estimates without mitigation.

For the results of the 14 main systems' failures, Table 3, the optimist experts and the pessimists showed significantly large differences. Figure 2, below, presents a graph of the aggregated assessments for each model. The pessimists gave much higher failure probabilities in most of the events, except for PDS Failure, Dry Cable Failure and ROV Commends Bad Motions. For these events, the assessments from both groups were very similar. Significant differences between the two groups are observed for the assessments given for the Core Control System, the Battery Net Working and the FO Failure. It is interesting to notice that the value of the Upper bound of the pessimists was much higher than that of Median, such as in Vehicle shape changes and Vehicle Stuck. This may indicate that the pessimist experts had lower confidence in their judgement for these kinds of failures. Moreover, most of the probabilities of failure from the optimists were less than 0.05. However, the failure probabilities for the Vehicle Stuck on Underside of Ice and the FO Failure had been given the median of 0.2401 and 0.1498, respectively. It is also reasonable to consider that this was main reason for the high failure probability of Transit in Underwater Ice phase.

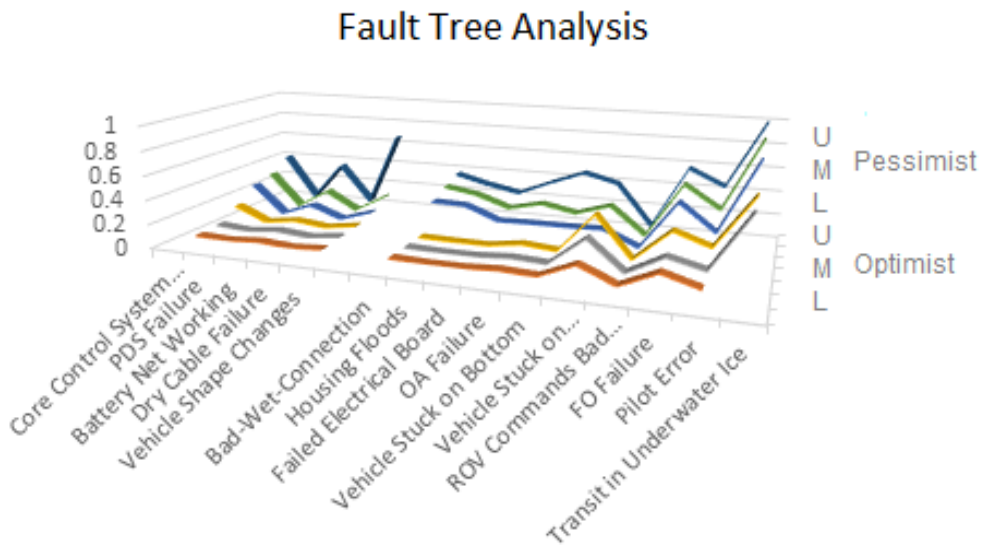

Figure 2: Results of the fault tree analysis for 14 main failures of the system based on optimistic and pessimistic experts' judgement aggregation.

4.3. Reliability Function Results

The assessments were provided for a mission of 20h. Because the average mission time of a vehicle in actual performance is around 5h and based on the given probabilities of failure of 20h mission period, the reliability function is used to calculate the mean time to failure (MTTF). The conversion results for the reliability of 20h to a 5h mission period are shown in Table 4, below.

Table 4: Reliability of experts' judgement, based on 20 h missions and the converted results to 5h

No.	Optimist			Pessimist		
	L	M	U	L	M	U
Failure(t)	0.2280	0.5497	0.6004	0.8058	0.9097	0.9915
R(t)(20h)	0.3996	0.4503	0.772	0.0085	0.0903	0.1942
MTTF	77.5194	25.0627	21.7865	12.2100	8.3195	4.1946
R(t)(5h)	0.9375	0.8191	0.7949	0.6640	0.5483	0.3036

Figure 3, below, presents a comparison of the different reliabilities, based on different mission periods.

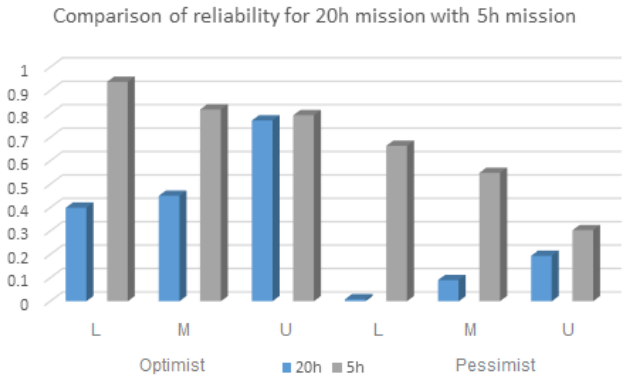

Figure 3: Reliability of 20h and 5h missions

4.4. Binomial Distribution Result of Expert Judgement

With the estimates for the reliability for a 5h mission, we could calculate the successful probability of Transit in Underwater Ice based on experts' judgements. The binomial function presented in Eq. 3 enabled us to estimate the number of successes that maximize the probability function. Table 5, below, shows the success probability for 16 missions, in both the optimistic experts' judgement and the pessimistic judgement. The number from 0 to 16 indicates the success time during the 16 missions.

Table 5: Success probability in optimist and pessimist expert

No.	**Optimist**			Pessimist		
	L	M	U	L	M	U
1	1.003E-11	1.3153E-12	5.4210E-20	3.0601E-03	3.0033E-06	2.6389E-08
2	6.21E-10	9.5287E-11	1.3010E-17	2.1345E-02	5.8330E-05	8.3441E-07
3	1.8025E-08	3.2359E-09	1.4637E-15	6.9792E-02	5.3103E-04	1.2367E-05
4	3.2542E-07	6.8375E-08	1.0246E-13	1.4199E-01	3.0081E-03	1.1405E-04
5	4.0914E-06	1.0062E-06	4.9948E-12	2.0118E-01	1.1867E-02	7.3251E-04
6	3.7987E-05	1.0934E-05	1.7981E-10	2.1049E-01	3.4572E-02	3.4742E-03
7	2.6941E-04	9.0768E-05	4.9448E-09	1.6824E-01	7.6937E-02	1.2587E-02
8	1.4889E-03	5.8713E-04	1.0596E-07	1.0478E-01	1.3341E-01	3.5535E-02
9	6.4799E-03	2.9908E-03	1.7881E-06	5.1388E-02	1.8219E-01	7.9002E-02
10	2.2283E-02	1.2037E-02	2.3841E-05	1.9914E-02	1.9658E-01	1.3878E-01
11	6.0341E-02	3.8153E-02	2.5033E-04	6.0771E-03	1.6703E-01	1.9197E-01
12	1.2733E-01	9.4228E-02	2.0482E-03	1.4451E-03	1.1059E-01	2.0693E-01
13	2.0524E-01	1.7777E-01	1.2801E-02	2.6250E-04	5.5936E-02	1.7039E-01
14	2.4430E-01	2.4768E-01	5.9082E-02	3.5212E-05	2.0892E-02	1.0361E-01
15	2.0252E-01	2.4031E-01	1.8991E-01	3.2895E-06	5.4342E-03	4.3875E-02
16	1.0446E-01	1.4508E-01	3.7981E-01	1.9121E-07	8.7952E-04	1.1561E-02

The differences between the optimists' and the pessimists' judgements are shown in Figure 4, below. This shows that the turning point appears when both the optimistic and pessimistic experts assume that there are 11 successes in the missions. Before this point, the success probabilities given by the pessimist expert aggregated model are larger than the probabilities provided by the optimist expert aggregated model. The pessimistic experts' aggregated model estimates most of missions as not being successful. In addition, the peak value for the optimistic experts appears when the experts consider the number of successes to be 13, while the peak value for the pessimists is when the number of successes is 9. The comparison of the two results reveals that the optimistic experts had more confidence than the pessimistic experts regarding the success probability for the vehicle transits in under ice conditions.

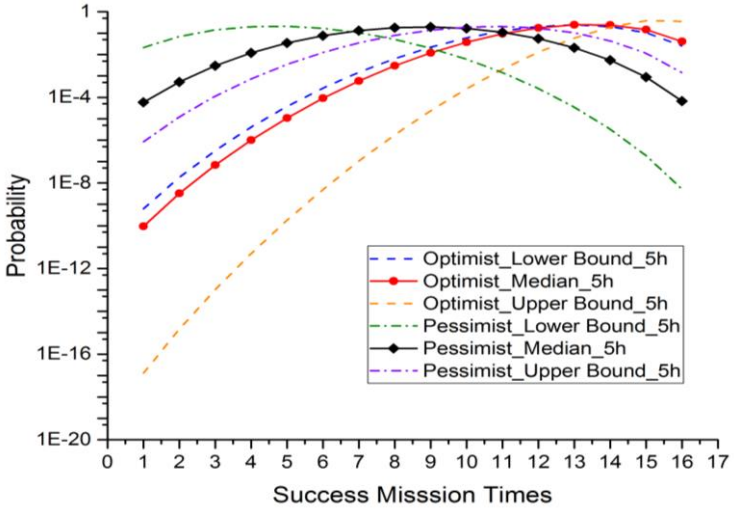

Figure 4: Success probability of different times in 16 missions of 5h

4.5. Binomial Distribution of Observed Data

We turn now to the experimental evidence on the observed data, which is the vehicle's actual performance. It is also based on the rationale of the binomial distribution. The historical reports record that, for all of the 16 missions, the vehicle was successfully launched and recovered 15 times. Therefore, we could assume the probability of success for a single mission to be 15/16. Using Eq. 3, we could calculate the probability of success for 15 out of 16 missions.

$$P(x=r=15)=\begin{pmatrix} n \\ r \end{pmatrix} \cdot p^r \times (1-p)^{n-r} = \begin{pmatrix} 16 \\ 15 \end{pmatrix} \cdot \left(\frac{15}{16}\right)^{15} \times \left(\frac{1}{16}\right) = 0.3798$$

Using this model, the expected frequency is 0.3798 * 16 = 6, and the expected un-success frequency can be 16 – 6 = 10.

4.6. Chi-square Test Results

The X^2 test is a hypothesis test method that allow us to estimate whether differences in frequency distribution are statistically significant or not. The Null hypothesis H_0 is: There is no difference between expert judgement and actual performance. The Alternative hypothesis H_1 is: There is a difference between expert judgement and actual performance. In this section, the theoretical frequency for optimist and pessimist were compared with the actual frequency from observed data. Table 6, below, shows the chi-square result for the optimistic experts.

Table 6: Optimist expert aggregated judgements, X^2 test results.

Value	fo (Actual performance)	fe (Expert judgement)	fo-fe	⟦ ((fo-fe))/fe⟧ ^2
Success	6	13	-7	3.7692
Un-success	10	3	7	16.3333
Total	16	16		X^2 = 20.1025

In order to examine whether the null hypothesis is true or not, it is necessary to compare the value of X^2 with the $X^2_{critical}$. If the value of X^2 is greater than the $X^2_{critical}$ then the differences between the predicted success estimates and the observed success are statistically significant. This means that the null hypothesis must be rejected. In this case, the degree of freedom is equal to: (2-1) * (2-1) = 1. The $X^2_{critical}$ is 3.8410 at 95% confident level.

Since X^2 is greater than $X^2_{critical}$ (20.1025 > 3.8410), the null hypothesis can be rejected with a p value < 0.005. Thus, the alternative hypothesis is accepted. As a result, a difference between the optimistic expert judgement and actual performance can be identified. The X^2 result for the pessimist expert aggregated model is shown in Table 7, below. The X^2 of 2.2875 is less than the $X^2_{critical}$ (3.8410). Thus, the null hypothesis fails to be rejected. As a result, no difference between the pessimistic expert judgement and actual performance can be identified.

Table 7: Pessimist expert aggregated judgements, X^2 test results.

Value	fo (Actual performance)	fe (Expert judgement)	fo-fe	⟦ ((fo-fe))/fe⟧ ^2
Success	6	9	-3	1
Un-success	10	7	3	1.2857
Total	16	16		X^2 = 2.2857

5. CONCLUSIONS

The un-weighted linear pool aggregation method was used to aggregate the experts' judgements mathematically and fault tree analysis (FTA) technique was employed to calculate the probability of loss/reliability of the Nereid-UI AUV. The probability of loss for the Nereid-UI AUV was based on its design assessment only, with the use of no operational data. For Autosub 3 deployment under the Pine Island Glacier, the estimated probability of loss for scenario 1 was 0.33 (for the optimistic model) and 0.48 (for the pessimistic model). For the Nereid-UI AUV, the probability of loss for a 5h mission was estimated at 0.181 (for the optimistic model) and 0.452 (for the pessimistic model). There are of course differences between the two vehicles in terms of operation profile and there are differences between the mission lengths. If we were to update the risk estimates for the Nereid-UI AUV with operational data, the risk profiles between the two platforms would be much closer.

The aim of the study was to test the validity of grouping experts into optimists and pessimists. Using the Nereid-UI AUV operational data and risk assessments, we tested which risk model would better reflect the field results.

X^2 test results have shown that the judgements by optimistic experts had a significant difference with actual performance. The pessimist expert aggregated judgements model has no significance difference with the actual performance. Our results show that the pessimistic model was more reliable for risk estimates. This was the case at least for the first few missions of the vehicle, 16 in our case. If we consider both optimists and pessimists as an integral group and calculate their judgement only via mathematical average method, it will lead to significant biases and make the experts' judgements unreliable when compared to the real cases. This difference has been noticed while dealing with the data in this work and mainly comes from the systematic errors due to heuristics and bias, which may be eliminated by the weighted linear pool aggregation method.

We plan to extend this study as follows: we would like to assess the impact of the seed questions on the experts' performance. In this study we elicited the experts' judgements for 10 seed questions. We intend to use this data to provide further insight into the expert judgment elicitation process.

Acknowledgements

The authors would like to acknowledge the work of colleagues from Woods Hole Oceanographic Institute in organizing the risk workshop. The authors would like to thank, in particular, Dr. Carl Kaiser, Dr. Michael Jakuba and Mr. Andy Bowen.

References

[1] Cooke R.M.: 'Experts in Uncertainty: Opinion and Subjective Probability in Science' (Oxford University Press, 1991. 1991)
[2] Apostolakis G.: 'The Concept of Probability in Safety Assessments of Technological Systems', Science, 1990, 250, (4986), pp. 1359-1364
[3] Keeney R.L., and Winterfeldt D.v.: 'Eliciting Probabilities from Experts in Complex Technical Problems', IEEE Transactions on Engineering Management, 1991, 38, (3), pp. 191-201
[4] Kahneman D., and Tversky A.: 'Subjective probability: A judgment of representativeness', Cognitive Psychology, 1972, 3, (3), pp. 25
[5] O'Hagan A., Buck C.E., Daneshkhah A., Eiser J.R., Garthwaite P.H., Jenkinson D.J., Oakley J.E., and Rakow T.: 'Uncertain judgements: Eliciting experts' probabilities' (Wiley, 2006. 2006)
[6] Otway H., and von Winterfeldt D.: 'Expert judgment in risk analysis and management: process, context, and pitfalls', Risk Analysis, 1992, 12, (1), pp. 11
[7] Cooke R., and Goossens L.H.J.: 'Expert judgment elicitation for risk assessments of critical infrastructures', Journal of Risk Research 2004, 7, (6), pp. 643-656

[8] Alicke M.D., Klotz M.L., Breitenbecher D.L., Yurak T.J., and Vredenburg D.S.: 'Personal contact, individuation, and the above-average effect', Journal of Personality and Social Psychology, 1995, 68, pp. 804-825
[9] Brito M., Griffiths G., and Trembanis A.: 'Eliciting expert judgment on the probability of loss of an AUV operating in four environments', in Editor (Ed.)^(Eds.): 'Book Eliciting expert judgment on the probability of loss of an AUV operating in four environments' (National Oceanography Centre, Southampton 2008, edn.), pp.
[10] Brito M.P., Smeed D.A., and Griffiths G.: 'Analysis of causation of loss of communication with marine autonomous systems: A probability tree approach', Methods in Oceanography, 2014, (0)
[11] Brito M.P.: 'Reliability Case Notes No. 10. Board of Inquiry: Circumstances surrounding the stranding of the AutoNaut 'Gordon' on the Plymouth coast on 7th November 2014', in Editor (Ed.)^(Eds.): 'Book Reliability Case Notes No. 10. Board of Inquiry: Circumstances surrounding the stranding of the AutoNaut 'Gordon' on the Plymouth coast on 7th November 2014' (National Oceanography Centre, 2015, edn.), pp. 1-91
[12] Brito M.P., Griffiths G., and Challenor P.: 'Risk Analysis for Autonomous Underwater Vehicle Operations in Extreme Environments', Risk Analysis, 2010, 30, (12), pp. 1771-1788
[13] Stokey R., Austin T., von Alt C., Purcell M., Goldsborough R., Forrester N., and Allen B.: 'AUV bloopers or why Murphy must have been an optimist'. Proc. Proc. 11th International Symposium on Unmanned Untethered Submersible Technology, New Hampshire1999 pp. Pages
[14] Griffiths G., Millard N.W., McPhail S.D., Stevenson P., and Challenor P.G.: 'On the Reliability of the Autosub Autonomous Underwater Vehicle', Underwater Technology, 2003, 25, pp. 175-184
[15] Strut J.: 'Report of the inquiry into the loss of Autosub2 under the Fimbulisen.', in Editor (Ed.)^(Eds.): 'Book Report of the inquiry into the loss of Autosub2 under the Fimbulisen.' (National Oceanography Centre, Southampton 2006, edn.), pp.
[16] http://oceanexplorer.noaa.gov/technology/subs/abe/abe.html, accessed 15/05/2018
[17] Griffiths G., and Trembanis A.: 'Eliciting expert judgment for the probability of AUV loss in contrasting operational environments', New Hampshire, US2007 pp. Pages
[18] Brito M., and Griffiths G.: 'A Markov Chain state transition approach to establishing critical phases for AUV reliability', IEEE Journal of Oceanic Engineering, 2011, 36, (1), pp. 139-149
[19] Bowen A.D., Jakuba M.V., Yoerger D.R., Whitcomb L.L., Kinsey J.C., Mayer L., and German C.R.: 'Nereid UI: A Light-Tethered Remotely Operated Vehicle for Under-Ice Telepresence'. Proc. OTC Arctic Technology Conference, Houston, Texas, USA, 2012/12/3/ 2012 pp. Pages
[20] Jakuba M.V., Bailey J., Kelley S., Suman S., Machado C., and Kaiser C.: 'SVC4: AUV Operations', in Editor (Ed.)^(Eds.): 'Book SVC4: AUV Operations' (Woods Hole Oceanographic Institution, 2016, edn.), pp. 1-11
[21] Jakuba M., and Laney S.: 'Shallow-water sea trails of the WHOI NEREID Under-Ice ROV to demonstrate capability in ice-covered shelf in polar oceans', in Editor (Ed.)^(Eds.): 'Book Shallow-water sea trails of the WHOI NEREID Under-Ice ROV to demonstrate capability in ice-covered shelf in polar oceans' (Woods Hole Oceanography Centre, 2017, edn.), pp. 1-4
[22] German C.R., Jakuba M.V., Bailey J., Elliott S., Judge C., McFarland C., Suman S., Whitcomb L.L., and Laney S.: 'TECHNOLOGY DEVELOPMENT: USE OF HROV NUI FOR UNDER ICE RESEARCH', in Editor (Ed.)^(Eds.): 'Book TECHNOLOGY DEVELOPMENT: USE OF HROV NUI FOR UNDER ICE RESEARCH' (Woods Hole Oceanographic Institution, 2014, edn.), pp. 1-20
[23] German C., Jakuba M., Bailey J., Branch A., Machado C., Suman S., Whitcomb L., Boetius A., Hand K., McDermott J., Purser A., Bowen A., Chien S., Kinsey J., Schaffer S., Bach W., Nakamura K.-i., and Pizarro O.: 'Use of HROV NUI for Under Ice Extreme Environment Exploration (PSTAR&ROBEX)', in Editor (Ed.) (Woods Hole Oceanographic Institute, 2016, edn.), pp. 1-12
[24] O'Connor P.D.T.: 'Practical Reliability Engineering' (Wiley, 1995, Fourth ed. edn. 1995)
[26] Brito, M.P., Jakuba, M. and Kaiser, K. 2018. A Performance-based Risk Model for Autonomous Underwater Systems Deployment. Risk Analysis (in preparation).