RELIABILITY OF TWO REMUS-100 AUVS BASED ON FAULT LOG ANALYSIS AND ELICITED EXPERT JUDGMENT

Gwyn Griffiths[1], Mario Brito[1], Ian Robbins[2] and Mark Moline[2]

[1] National Oceanography Centre, Southampton, UK
email (gxg, m.p.brito)@noc.soton.ac.uk

[2] Center for Coastal Marine Sciences, California Polytechnic State University
San Luis Obispo, CA, USA
email (mmoline, irobbins)@calpoly.edu

Abstract

Reliability is especially important for autonomous underwater vehicles (AUVs) that have made the transition to operational use. However, in contrast to the unmanned air vehicle community, there has been little sharing in the open literature of detailed fault histories of commercial AUVs from which an assessment of their reliability can be made. In this paper, we declare the fault history of two REMUS-100 AUVs manufactured by Hydroid Inc. and operated by the Center for Coastal Marine Sciences at California Polytechnic State University. The data set contains the faults and incidents recorded from 186 missions between 5 November 2003 and 14 January 2009. Interaction between the faults or incidents with the AUV and the operating environment complicate matters when it comes to estimating probability of loss. Through a formal process of eliciting expert judgment the paper provides optimistic and pessimistic estimates of loss of the REMUS vehicles in open water, coastal, under sea ice and under ice shelf operations. Vehicle-specific risk mitigation methods are explored, and recommendations made for reviewing the fault history data using behavioural aggregation through interaction among a group of experts rather than simple mathematical aggregation as in this paper.

1. Introduction

Reliability is especially important for autonomous underwater vehicles (AUVs) that have made the transition to operational use. A few research groups that have built their own AUVs, or modified commercial vehicles, have attempted statistical analyses of fault and incident logs (e.g. Griffiths et al., 2003; Podder et al., 2004), and a high-level analysis of the availability of the commercial HUGIN vehicles operated by C&C Technologies has been presented (Chance, 2003). However, in contrast to the unmanned air vehicle (UAV) community (e.g. OSD, 2003), there has been little sharing in the open literature of detailed fault histories of commercial AUVs from which an assessment of their reliability can be made.

With commercial UAVs used by the military it is clear that extensive record keeping, aggregation and sharing of data on faults and incidents has led to robust statistics that have been used to improve reliability over two decades, both in the US and in Israel. Their extensive experience does highlight a central issue for AUVs and for this paper: the large difference between the fault rate and what the UAV community terms the mishap rate. Mishap is defined as the loss of an aircraft, loss of human life or damage in excess of $1m. OSD (2003:27) lists the Mean Time Between Faults (MTBF) of several UAVs, notably the RQ-1A Predator at 32 hours, the RQ-2A Pioneer at 9.1 hours and the RQ-5 Hunter at 11.3 hours (post 1996). In contrast, the mishap rates for these three vehicles were 32, 334 and 55 per 100,000 hours respectively. These imply ratios of mishap-to-fault of 98, 33 and 161 respectively. With AUVs, while we may have sufficient knowledge and experience to assess MTBF, we generally have insufficient information to assess directly the mishap rate. And yet, as AUVs are being used in more challenging environments, operators and owners do need to know the risk of loss or severe damage to their vehicles.

In this paper, we set out how such an assessment can be made using a case study based on the fault history of two REMUS-100 AUVs manufactured by Hydroid Inc. operated by the Center for Coastal Marine Sciences at California Polytechnic State University (Cal Poly). The data set contains the faults and incidents recorded from 186 missions: 89 with vehicle *Boomerang I* between 18 July

2001 and 19 October 2008, and100 with *Boomerang II* between 5 November 2003 and 20 February 2009[1].

The faults and incidents with the vehicles are set out in a spreadsheet available online[2], with the key points described and discussed in section 2. Simple fault/incident history statistics are derived in section 3. The challenge of deriving the probability of loss (this is equivalent to the UAV mishap), based on expert judgment, is discussed in sections 4 and 5. Interaction between the faults or incidents with the AUVs and the operating environment complicate matters, and affect profoundly the risk of loss. We have previously argued that those interactions cannot be determined mathematically (Griffiths and Trembanis, 2007; Brito and Griffiths, 2009). However, they can be estimated through a formal process of eliciting expert judgment. The alternative, frequentist, approach to estimating the probability of loss of an AUV in different operating environments is impractical. It would require counting the number of missions and number of losses in different environments for a particular vehicle type. Obtaining such a data set is a worthy goal, difficult for a single AUV operator, but could be a community effort. This community approach has been followed with UAVs, and hence they do have frequentist estimates of mishaps. Section 6 uses an extended Kaplan Meier non-parametric analysis to provide graphical visualisations of the estimated probabilities of loss of REMUS100 with mission distance in four environments: open water, coastal, under sea-ice and under shelf ice. Such an analysis informs risk mitigation measures. Section 7 presents our conclusions.

2. The fault and incident histories of two REMUS-100 AUVs.

2.1 Background to vehicle configurations and use.

The REMUS-100 vehicles operated by Cal Poly are used for a range of missions to better characterize and improve understanding of coastal waters, Moline et al. (2005). The vehicles have been fitted with a wide range of sensors, including a non-standard Ocean Sensors OS200 CTD, a bioluminescence bathy-photometer, nets with bottles to capture animals that have passed the photometer, other optical instruments – the Wetlabs ECO-series backscatter-fluorescence sensor, a Satlantic OCR-507I irradiance sensor and a OCR-507R radiance sensor, and a 1200kHz ADCP (figure 1). Operating depths during missions may be shallower than 2m, hazards include areas of kelp and manmade structures such as piers. Different forms of navigation have been used depending on requirements. Most frequently, the vehicles operate within the coverage of a network of digital acoustic transponders, with an Ultra-Short Base Line unit on the vehicle determining range and bearing to the transponders, whose positions have been previously fixed, and their locations pre-programmed into the AUVs. Long Base Line Navigation by triangulation is also used to a maximum range of 2.5km. When acoustic navigation is not available the vehicles use dead reckoning based on compass and Doppler Velocity Log (or propeller revolutions) to update the last good acoustic or GPS fix.

Figure 1. Example of the more unusual sensors fitted to the Cal Poly REMUS-100 vehicles – a bioluminescence bathy-photometer, at the top, and two nets and collection bottles attached to the exhaust ports to collect samples, from Moline et al. (2009).

Missions have taken place in Avila Beach, Huntington Beach, Imperial Beach, La Jolla, Monterey Bay, Morro Bay, Pismo Beach, Sand City, San Diego Bay, Santa Barbara, CA; Sarasota, FL, Buzzards Bay, Hadley Harbour, Woods, Hole, MA; Castine, ME, the Great Bay area, NJ; Sequim Bay, WA and Svalbard, Norway. The median mission lengths were 14.3 and 13.9km for *Boomerang I and II* respectively, with lower quartiles of 3.8 and 5.7km and upper quartiles of 41.5 and 36.9km, and maxima of 53.8 and 59.06km. The batteries fitted – four 26V 10Ah lithium ion packs were capable of providing power for distances of up to 80km at 3 knots.

2.2 Overview of the fault and incident history

The fault and incident history for these two REMUS-100 vehicles was quite unlike that of the Autosub vehicles studied previously. Autosub missions generally exhibited different technical problems, which were almost always corrected before subsequent missions, or manifested themselves in different ways (e.g. where the root cause was an intermittent connector). Here, the REMUS

[1] Mission descriptions are available at http://www.marine.calpoly.edu/auv/REMUS/index.php

[2] Available via a link at http://www.noc.soton.ac.uk/nmf/usl/gxg/EEJ_REMUS.html

vehicles exhibited 37 unique faults or incidents[3], many recurred (some, years apart) but often in different combinations. For example, the fault 'depth sensor noisy' first encountered on mission 7 (29 July 2001) occurred alone on eight further missions and in combination with one or more other faults 33 times. This pattern of faults and incidents affected the philosophy of the subsequent analysis, as discussed in section 4.

A detailed reading of the list of unique fault and incident log will show very few instances of hardware problems, such as on mission 147, 'Batteries were low and finished mission, due to broken charging station after prior mission.' Unique faults due to software were also infrequent, such as, 'Warning: Vehicle Bioluminescence data is old, due to Software error', first encountered on mission 1 and recurring on 14 missions. More common were incidents related to the environment. For example, where the vehicle could not dive on first attempt due to sea conditions or the buoyancy trim. In the majority of these instances the vehicle tried again using a 'porpoise' mode. More complex incidents often occurred when operating close to the seabed.

A number of incidents also occurred because of mission imperatives. Missions that were run overnight, for example, from nearshore to offshore and back sought to minimise risks while offshore even at the expense of accepting greater risks during the final stages nearshore, where recovery would be easier, for example, by setting a lower-than-usual threshold for acceptable remaining battery energy.

3. Simple statistics on faults and incidents

The mission statistics in Table 1 include estimates of Mean Time Between Failures (MTBF) and its distance equivalent, Mean Distance Between Failures (MDBF). Both are defined simply as:

MTBF = Total Hours/Total Failures

MDBF = Total Distance/Total Failures.

These definitions are as used by the US Army, and more generally, for UAVs (OSD, 2003). They do not derive from records of the actual time interval between individual faults.

On first reading these statistics look unimpressive, significantly worse than the RQ-2A Pioneer UAV with a MTBF of 9.1 hours, for example. However, as noted above, a small number of recurring minor faults dominated the MTBF calculation. Seven out of 37 unique faults or incidents were responsible for 467 (92%) of the total of 507 problems logged, distributed as shown in the Pareto plot of Figure 2.

Table 1 Simple statistics on the fault or incident rates for two REMUS100 AUVs.

Vehicle	Missions	Distance (km)	No. faults	MTBF (hr)	MDBF (km)
Combined	189	4031	507	1.28	7.95
Boomerang 1	89	1811	265	1.04	6.83
Boomerang II	100	2220	242	1.54	9.17

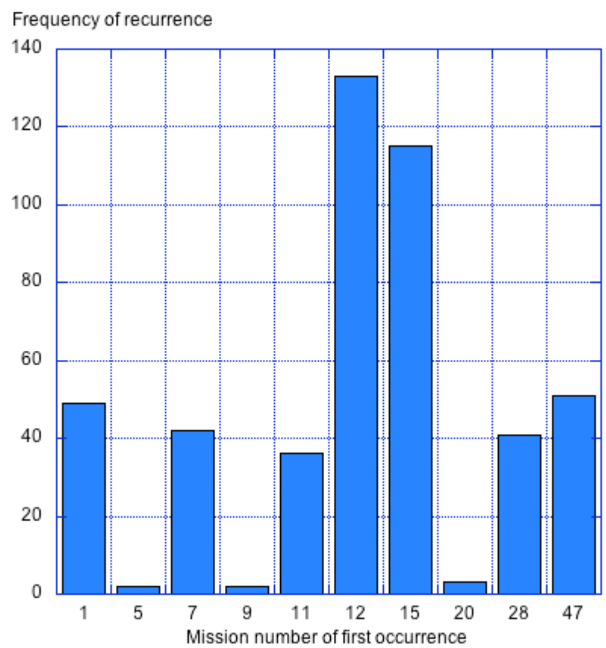


Figure 2. Pareto plot of the frequency of recurrence of those faults or incidents where there was more than one instance.

Despite these faults or incidents, 173 of the 189 missions (91.5%) were considered to have been successful. Only on 28 missions (14.8%) were there aborts not called for by the operators (the operators initiated 31 aborts for reasons such as sailboats nearby, or wanting to end missions early).

Despite the overall level of success, much would be gained by the retirement of the seven recurring items, which were described as follows (comments from a Hydroid engineer are in italic):

1 - Warning: Vehicle Bioluminescence data is old, due to software error.

7 - Depth sensor noisy (example last 3 readings 5.4 5.9 6.7m). Self test failure depth sensor, pausing mission. Sensor misread, single value missed, happened before

---
[3] See tab "Unique Faults – Incidents" in the fault and incident spreadsheet available via a link at http://www.noc.soton.ac.uk/nmf/usl/gxg/EEJ_REMUS.html

mission started. *This is often seen 'on the bench' with many sub-systems running.*

11 - REMUS has ground fault: 0V reference is shorted to seawater. Pin connection poor resulting in GFI, no affect on mission. *Most often this is found to be a user problem in fitting the dummy plug. If encountered, GFI faults will halt a mission from starting, but will NOT abort a mission if they emerge when underway.*

12 - Vehicle stuck on surface; attempting to drive it down. Delay is vehicle dive due to buoyancy (10-20s), no affect on mission. *Sea conditions/buoyancy may prevent vehicle diving first time. If so, vehicle tries again using a 'porpoise' mode.*

15 - Self test failure attitude (tilt), pausing mission. Vehicle rolled to side after trying to avoid bottom, everything okay. *Vehicle uses combined ADCP beams as altimeter. When sensed as too close to bottom vehicle pitches up and propeller stops, the tilt is therefore outside normal limits. When vehicle assumes proper attitude, propeller starts and mission continues.*

28 - Vehicle at low altitude. Executing emergency climb, too close to bottom, hit bottom and bounced. *Controlled climb only good to 10° pitch, to get more than that propeller is stopped.*

47 - Ocean Sensors salinity out of range, bad conductivity reading.

4. Eliciting expert judgment on probability of loss

It is accepted that when eliciting expert judgment to build a risk model a formal elicitation process should be followed that reduces biases and the possibility for disagreements. Tversky and Kahneman (1974) studied the nature of different types of biases. Experts usually introduce biases by following mental shortcuts such as availability, representativeness, or anchoring. A formal elicitation exercise consists of the following sequential steps: 1) Specifying the purpose; 2) Selecting the experts; 3) Clearly define the issues; 4) Training the experts; 5) Analysing and aggregating; and 6) complete analysis write up. This process is generic and some phases do overlap. Whilst the sequence of the process is invariable, the activities that take place in each phase may vary. The process is typically tailored according to the type of application, expert's availability and the facilitator's[4] familiarity with advanced statistical methods.

In this research we attempt to combine features of formal elicitation process proposed by O'Hagan (2009) in the SHELF package, with features of formal elicitation presented in Otway and Winterfeldt 1992). One aim is to reduce the number of experts needed and therefore to reduce the time taken to create the risk model.

The SHELF process involves three classes of actors: the elicitation manager, the facilitator, and the experts. SHELF calls for five experts, one elicitation manager and one facilitator. Here we had responses from four experts. We used a combined elicitation manager and facilitator, an approach accepted by Otway and Winterfeldt (1992), and denoted as the decision maker.

The following sections outline the set of activities undertaken by the decision maker in each phase of the formal judgment elicitation exercise.

4.1 The purpose.

The purpose of the study is to build a risk model for Remus100 operation in four environments: open water, coastal, sea ice and ice shelf. The risk model was based on the vehicles failure and incident history and the probability of these failures resulting in AUV loss.

4.2 Selecting the experts.

The expert selection was conducted with three requirements in mind:

1) Independence between the experts and the institution responsible for the AUV operations (Cal Poly).

2) To use a number of experts who have participated in a similar risk exercise.

3) To invite experts that have not participated in this exercise but have experience relevant to this research.

We invited six experts; four were able to submit their work for this study. Table 2 summarizes the experience and subject area for the experts.

Table 2 Experts' experience in AUV design and operation; expert names are abbreviated hereafter.

Expert	Application area	Years of experience
Adam Skarke (AS)	Scientific research	4
Rob McEwen (RM)	Scientific research	10.5
Steve McPhail (SM)	Scientific research	15
Tim Boyd (TB)	Scientific research	6

4.3 Clearly define the issues.

A total of 507 occurrences of faults or incidents was noted by the AUV operations team. It would not be feasible to ask the experts to evaluate the probabilities

[4] Facilitator is the person that performs statistical analyses of the expert judgments.

(five values) for each of these, in four environments, a total for each expert of 10,140 judgments. Instead, we separated out the 37 unique faults or incidents for evaluation by the experts (a more manageable 740 judgments). An assessment showed that the faults or incidents could all be considered independent, that is, there was no interdependence of the fault clusters on any mission. In later analysis the probability of loss for each mission (where several faults or incidents may have occurred) can therefore be determined from the individual elicited values for each unique fault from:

$$P(loss) = 1-(1-P_1)*(1-P_2)\ldots(1-P_n) \qquad (1)$$

The experts were asked to provide five measures of the probability of a fault or incident resulting in loss, rather than a single probability judgment as was the case in the earlier Autosub3 exercise (Griffiths and Trembanis, 2007). By asking the expert to provide five measures, it is possible to better evaluate the skewness and tails of the distribution; the five measures being:

1. *The lower bound, L.* The minimum value that *P(loss)* can take.
2. *The upper bound, U.* The maximum possible value of *P(loss)*.
3. *The median, M.* The value for which there is 50% chance of *P(loss)* being above or below it.
4. *The lower quartile, LQ.* The value for which there is 50% chance that *P(loss)* is between *U* and *LQ* and 50% chance that it is between *LQ* and the *median*.
5. *The upper quartile, UQ.* The value for which the expert is 50% confidence that *P(loss)* is between the *median* and *UQ* and 50% confidence that it is between *UQ* and the *upper bound*.

The description of each operating environment from Griffiths and Trembanis (2007) was given to the experts.

4.4 Training experts and eliciting judgments.

Experts AS and RM received training on how to assign probabilities to events in a previous risk assessment exercise. Two Experts SM and TB were for the first time participating in this type of exercise. All experts were asked to keep in mind the fact that losing a REMUS100 is a rare event; neither of the two Calpoly vehicles had been lost in 189 missions. An example elicitation was emailed to each expert and a webpage was created to store any misconception introduced in the fault description. The experts were also asked to keep in mind the fact that some of the faults were caused because of the characteristics of the operating environment and the configuration in which the vehicle was being deployed.

4.5 Analysing and Aggregating.

It is possible to analyse in detail those faults that present huge disagreement between experts. It is the approach necessary when using behavioural aggregation with the experts present in person. It can also be done by feedback to each expert and asking for a review of their judgments. However, this process is time consuming, and, given the constraints imposed on this study, it was decided not to analyze discrepancies in judgments but to aggregate the experts mathematically into two schools of thought, the optimists and the pessimists, as we had done in the previous Autosub3 study. This provides a useful indication of the range of judgments.

A beta distribution (constrained between 0 and 1) was fitted to the five probability estimates for each unique event in each environment elicited from each expert, and subsequently aggregated after deciding whether the expert was an optimist or pessimist for each event in each environment.

4.6 Complete analysis and write up.

In a formal judgment elicitation exercise, once the expert judgments are analysed a preliminary report is prepared and sent to the experts. The aim is to allow experts to review their judgments and if possible to rectify their judgments according to any misconceptions. This article is the first publication of our preliminary assessment.

5. Remus100 Risk Model

5.1 Judgment Analysis.

The cumulative distribution of the expert judgments provides a means to assess visually whether an expert uses lower probability ranges or, on the contrary, an expert uses high probability ranges for mission failures or incidents. As an example, figure 3 presents the cumulative distribution for the judgments provided by each expert for mission 140, here the environment is coastal water.

The probability judgement for mission 140 is a result of the combination of judgments provided for single failures 12 and 15 (see section 3 for descriptions) together with a low battery state. From figure 3 it is notable that experts AS and SM use lower probability ranges than experts RM and TB. Thus for this mission and environment, experts AS and SM's judgments were aggregated to form the optimistic group and experts RM and TB's judgements were aggregated to form the pessimistic group.

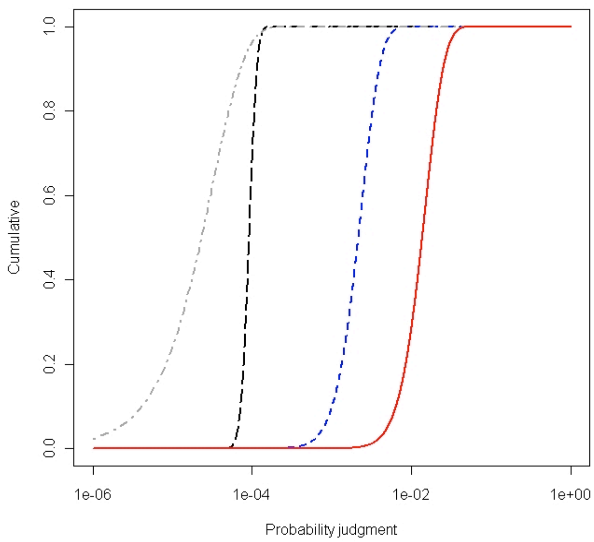


Figure 3. Cumulative distribution of the probability judgments provided for mission 140, coastal water environment, as an illustrative example. SM in dashed gray; AS in dashed black; RM in dashed blue and TB in red.

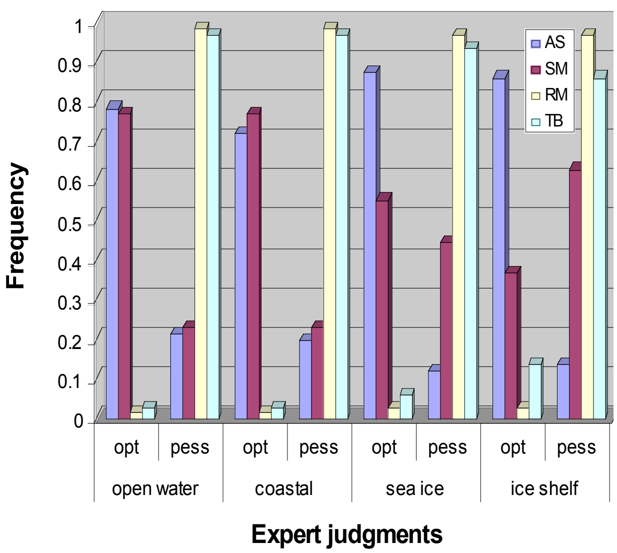


Figure 4. Frequency distribution for the cluster of optimistic (opt) and pessimistic (pess) experts.

This example is not untypical; in many cases experts' judgments spanned two orders of magnitude or more in probability of loss. Figure 4 presents the frequency distribution showing how often each expert was deemed optimistic or pessimistic. For all environments, judgments provided by RM and TB were consistently classed as pessimists. For open and coastal water, and under sea ice, judgments from SM were more often optimistic than pessimistic. For the ice shelf environment SM was classed as pessimistic in 63% of the missions and optimistic in 37%. For all environments, AS had over 70% of his judgments classed as optimistic.

It would be illuminating to revisit this data set and analysis using a behavioural aggregation approach, where the experts would be required to reach a single agreed distribution. The arguments for lower or higher probability of loss would need to be made and consensus reached. The factors discussed would undoubtedly cast light on the judgment process and the perceived severity (or not) of the faults and incidents.

Given the large number of instances of the same few faults we sought to establish whether they dominated the overall results. Table 3 shows the implications for coastal water (the environment in which the missions were run). Where the experts are declared optimists (AS and SM), with the single exception of mission 28 for AS, the contributions to the total probability of loss from these recurring faults are low compared to the higher probabilities of loss given for more serious single mission faults or incidents by these experts. This was because their estimated probabilities of loss from the single occurrence of each recurring fault were very low.

However, for TB the total probabilities from these recurring faults are of the same magnitude as for the more serious faults and incidents, while for RM two out of the seven recurring faults collectively reach a similar magnitude to the more serious faults. As these do have a profound effect on the overall results for these two experts, a means of reaching consensus is needed to bottom out the issues and perceptions with these recurring faults.

Table 3 For the coastal water environment, the elicited median probabilities from each expert ('Single', P_s) for the faults that recur frequently (shown by mission number of first occurrence and the frequency count C), together with the total contribution to the probability of loss when the single mission value is applied to all occurrences (P_t=1-(1-P_s)^C).

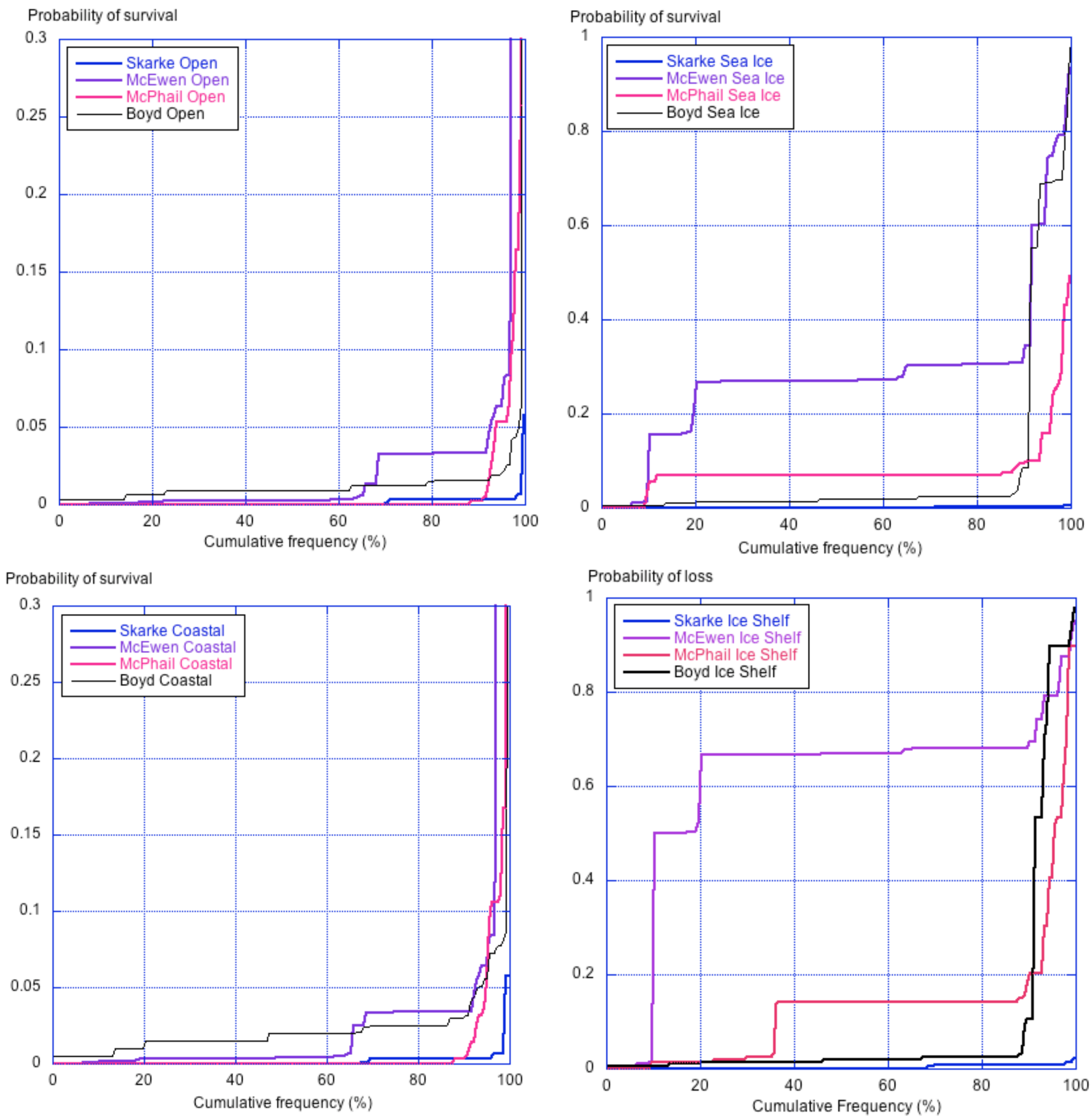
Mission	Count	TB		AS		RM		SM	
		Single	Total	Single	Total	Single	Total	Single	Total
1	49	5e-3	2.2e-1	0	0	0	0	2e-6	9.8e-5
7	42	5e-3	1.9e-1	6e-5	2.5e-3	1.5e-3	6.1e-3	1e-5	4.2e-4
11	36	5e-3	1.7e-1	0	0	2.5e-4	9e-3	2e-5	7.2e-4
12	133	5e-3	4.9e-1	8e-5	1e-2	1.5e-3	1.8e-1	1e-5	1.3e-3
15	115	5e-3	4.4e-1	6e-6	2.2e-4	5e-4	5.6e-2	1e-6	1.2e-4
28	41	5e-3	1.9e-1	3e-3	1.2e-1	3e-2	7.1e-1	4e-5	1.6e-3
47	36	5e-3	1.7e-1	6e-6	2.2e-4	3e-4	1.1e-2	2e-6	7.2e-5

Table 4 For the coastal water environment, the elicited median probabilities from each expert for their top risk(s) from the list of 37 unique faults/incidents, together with the fault or incident description (with the text in italic being the comment of a Hydroid engineer, and the text in blue has been added in response to the questions raised by TB) and the justifying text from each expert where provided.

Expert	**Mission**	**Median *P(loss)***	**Fault or incident description**	**Expert's justifying text (if provided)**
TB	18	0.5	Sank at base of pier, retrieved. Was not moving. This occurred because one of the attached sensor's O-ring failed; the AUV filled with water, changing buoyancy. The only reason it returned to the pier was because the RPMs were high enough to maintain flight. When it reached its destination it sunk to the bottom, where it was located and retrieved.	Why did the vehicle sit motionless on the bottom? Did it leak? Was it ballasted improperly? The vehicle should be slightly positively buoyant. Assuming all that is known of the fault is as described here: vehicle sank on deployment, then the issue (of loss) is relatively simple: in water depths > rated depth (100m) *P(loss)* is 1. Assuming 'coastal' means <100m, probability of successful recovery is pretty good. Probability of sinking due to a leak (assuming pressure vessel was checked) is relatively low. Most likely is ballasting error, e.g. for wrong salinity, and the vehicle might be floating at an intermediate depth, which would be harder to recover from. (Edited form)
AS	94	0.05	Transponders reversed so vehicle got confused, almost went into rocks until we rescued it. *Transponders have unique identity and preprogrammed locations in vehicle command file. Mismatch may lead to vehicle navigation errors.*	*P(loss)* is significant given the proximity of the AUV to a rocky shoreline during the navigational fault. Collision with a rock shoreline at speed would certainly result in loss. A high se state would also increase this danger.
	97	0.05	Under pier, stopped mission to prevent vehicle damage.	*P(loss)* is significant since I assume the AUV is among numerous pilings that support the pier. Collision potential is very high. This risk would be heightened by a high sea state, which could batter the AUV against the pier pilings.
RM	18	0.5	As above	
	52	0.5	Science bay leaking.	
	87	0.5	Caught in eelgrass, but freed itself. *Vehicle will try both forward and reverse thrust if it considers it is not making way.*	
	94	0.5	As above.	
SM	189	0.8	Caught under kelp raft, was not able to call or get GPS for 4 hours, had to do old fashion current and wind calculation to get vehicle general area and use distance pinger to find vehicle. Success, we found it.	I'm assuming 50% loss probability in such a scenario, based on the description.

Table 4 lists the top risk for each expert in the coastal water environment, with a description of the fault or incident as provided the expert and the notes from the expert when available. AS has the highest risk of loss at a median of 0.05; the other experts have medians of 0.5 (TB and RM) and 0.8 (SM) for their highest risks. RM considers four faults or incidents to be equal top risks, one of which (18) is the top risk for TB and another (94) to be one of the two top risks for AS. Only SM considers the incident on mission 189 to be a top risk.

Figure 5 Cumulative frequency plots for each environment with each expert's assessment computed from the unique faults to encompass the faults or incidents on all 189 missions. As an example of how to read these plots, the point on the Y axis corresponding to 50% on the cumulative frequency scale on the X axis is the median probability of loss from that expert, similarly the upper and lower quartiles can be read at cumulative frequencies of 75 and 25%.



Having considered the treatment by the experts of the frequently occurring faults or incidents, and their top risks, overall visual impressions of the risk profiles by expert in each environment can be gleaned from the cumulative frequency plots in Figure 5. Note that the probability (Y axis) scales for the open water and coastal environments are from 0 to 0.3 compared with 0 to 1 for the under ice environments. In open and coastal waters AS and SM's assessments show why we considered them optimists, their median and upper quartiles are at a low probability of loss. For SM the sharp rise occurs at the 90th percentile; AS also shows a sharp rise, but it is not until the ~98th percentile and it is of lower magnitude. In these two environments TB is the most pessimistic below the ~65th percentile, after which RM becomes the most pessimistic expert, a school of thought he exhibits clearly for the two under ice environments.

5.2 Aggregation.

Having considered the responses from each expert individually, the next stage is aggregation. The judgments were aggregated using a linear pool for the mean of each probability judgment. A beta distribution was fitted to each probability range elicited from the experts; an algorithm was implemented in the R statistical software package to perform this fitting automatically using a simple optimisation algorithm (R, 2009; Nelder and Mead, 1965).

The shape of the beta distribution is controlled by hyper-parameters α and β. The mean and variance for each mission were calculated using these hyper-parameters (equations 2 and 3).

$$\mu_j = \frac{\alpha}{\alpha+\beta} \tag{2}$$

$$\sigma_j^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\cdot(\alpha+\beta+1)} \tag{3}$$

The aggregated cumulative frequency plots of the probability of loss in each environment are shown in Figure 6. The school of thought - optimist or pessimist – has a greater influence that the environment within each pair – open/coastal and sea/shelf ice. There is little distinction between the probability of loss of the optimists and the pessimists in the open water and coastal environments from their judgments. For sea ice and shelf ice RM was considered to be the only overall pessimist based on the cumulative frequency plots (figure 5). The gap between his assessments and those of the other three experts does need to be reconciled.

6. Survival model

6.1 Model methodology

A model that can be used to predict the probability of survival for one of these vehicles in a declared operating environment, based on the data and judgments collected here, is useful to the responsible owner. Such a model can be used to answer questions such as, "What risk of loss is there for one of these vehicles when undertaking a 25km mission under the sea ice conditions envisaged in the study?" The simple model derived here uses the extended Kaplan Meier approach (Brito, Griffiths and Trembanis, 2008) to estimate and present visually the probability of survival with mission distance. The extension to the standard Kaplan Meier non-parametric lifetime estimator is to use the probability of survival instead of the usual censored flag, to give the probability of survival:

$$\hat{S}(r_i) = \prod_{r_i<r} \frac{n_i - p_i}{n_i} \tag{4}$$

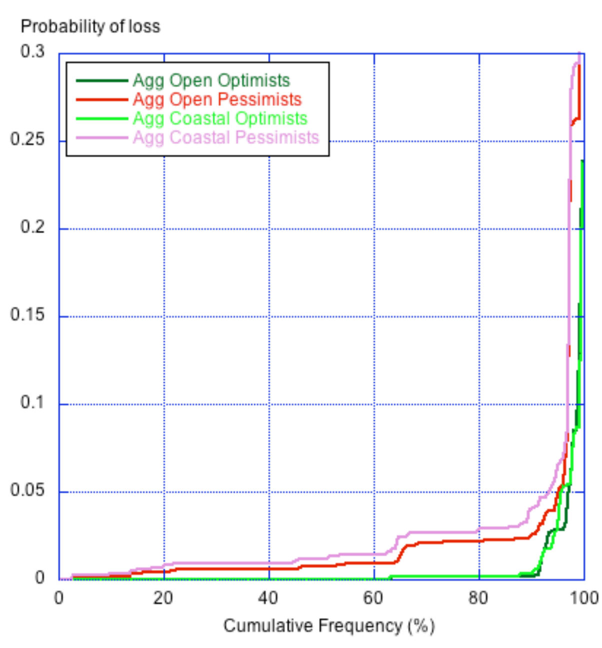
where n_i is the number (of missions) at risk immediately prior to range r_i and where there is a probability p_i that the i^{th} fault is fatal.

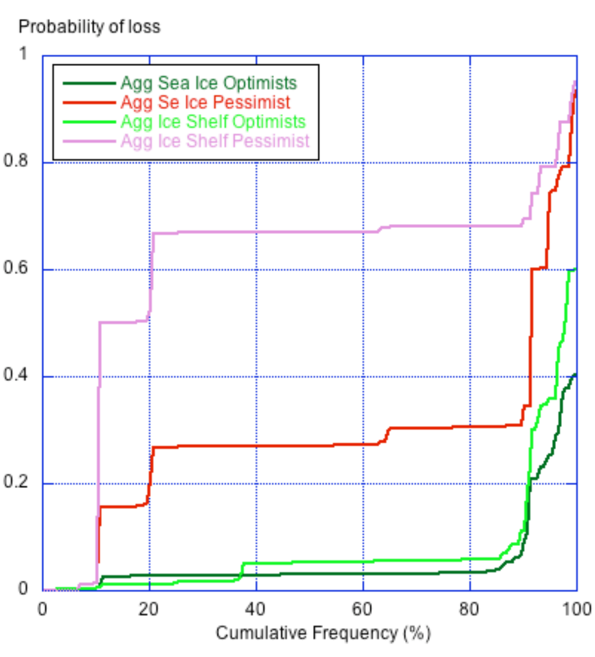
6.2 Model results

Figure 7 presents the results for all four environments. Each sub-figure shows the optimistic and pessimistic expert assessments. To give some examples of how these curves may be used:

a) What is the probability of losing the AUV in coastal waters for a single mission of median length (14km)?

Figure 6 Cumulative frequency plots for the aggregated optimists and pessimist(s) in each of the four environments.





Optimistic: 0.003
Pessimistic: 0.019

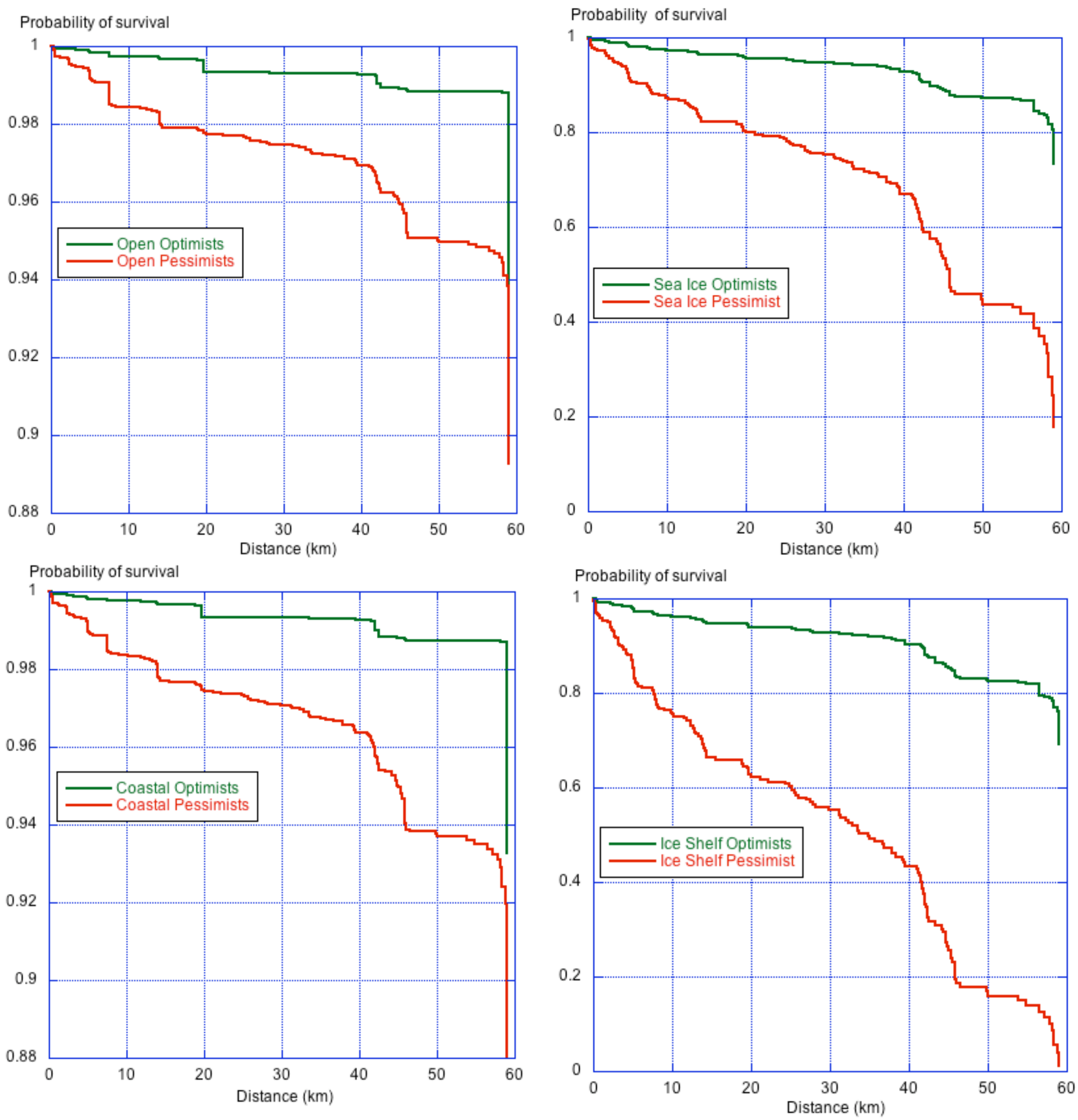
b) How many missions of this median length would there be before the probability of loss was 50%?
Optimistic: 232
Pessimistic: 36
Given the vehicles had completed 189 missions, although with testing incidents, it does suggest that the pessimistic estimates are too pessimistic.

c) What is this in mean distance to loss (mishap), and how does this compare with UAVs?
Optimistic: 3248km, ratio of 408 between faults and mishaps, compared with ~100 for military UAVs (OSD, 2003).
Pessimistic: 504km, ratio of 63.

d) What is the probability of losing the AUV for an under sea ice mission of 25km?
Optimistic: 0.046
Pessimistic: 0.22

e) How many missions of 25km under sea ice would there be before the probability of loss was 50%?
Optimistic: 15
Pessimistic: 3

Figure 7. Kaplan Meier non-parametric survival graphs for the four environments for judgments from the optimists and pessimist(s).



6.3 Risk mitigation strategy

Brito, Griffiths and Trembanis (2008) showed that the shape of the Kaplan Meier curve for Autosub3 suggested an effective risk mitigation strategy. The probability of survival decreased markedly for missions of less than 25km, followed by a plateau as range increased further. Consequently, if the AUV was monitored over an initial distance, in waters where recovery could be effected if a fault developed, the probability of a fault emerging during the unmonitored phase would be reduced. This was expressed as a conditional probability of losing the vehicle before a distance x, given that the vehicle had survived a lesser distance y, as equation 5:

$$P(X < x \mid X > y) = \frac{F(x) - F(y)}{1 - F(y)} \tag{5}$$

For the REMUS100 vehicles studied here, there is much less of a steep decline followed by a plateau in risk as a function of distance in the Kaplan Meier plots of Figure 7. The gain is marginal. In the case of a sea ice environment for the optimistic experts, the probability of losing the vehicle over 25km is 0.046 as set out above. If the vehicle is monitored for the first 15km (arrow on figure 7), $F(y)$=1-0.964=0.036, the probability that the vehicle survives to 40km (F(x)=1-0.927)=0.073), an additional run of 25km is 0.038. This is a small reduction over the 0.046 for a 25km run without an up-front monitoring distance.

What appears more important from this fault and incident history and the experts' assessment of impact is to avoid missions over 40km where the probability of survival apparently decreases markedly.

7. Conclusions

With 189 missions completed spanning nearly eight years a wealth of experience has been gained with the two REMUS100 vehicles operating in mainly coastal waters. It is important to reiterate that 91.5% of the missions were considered to have been successful and that aborts not called for by the operators only occurred on 14.8% of the missions. On a number of the missions considered high risk, the operating team were prepared for the scenarios based on previous experience with the vehicles and the environmental setting. Three of the experts that took part in this study considered that loss on one or more missions was as likely or more likely, than recovery; that is, they assigned a probability of loss of 0.5 or greater.

It is important to note that REMUS was initially intended for primarily locating mines on the seabed. Operations by Cal Poly have strayed from this for oceanographic and near shore applications, which might influence performance metrics.

The existence of a small number of recurring faults, encountered in different combinations, has complicated this analysis. It meant that it was not feasible to present the full data set for expert judgment, and that a mathematical approach based on the independence of the faults had to be used. Experts assessed the likelihood of these faults leading to loss by considering one instance, not by their collective effect (e.g. the fact that the incident first encountered on mission 12 happened 133 times). Despite this, overall, three of the four experts considered (through their judgments) these faults to be of very minor consequence.

The categorisation of experts as either pessimistic or optimist, first seen in an earlier Autosub3 analysis, recurs here. However, in this study, the gap between the assessments of the optimists and pessimists is wider. This results in a large difference between the cumulative frequency distributions and the probability of loss at the longer ranges in the Kaplan Meier analyses. The difference is sufficiently large as to render this form of expert judgment analysis, reached through mathematical aggregation, of limited value.

It is only to be expected that the existence of pessimists and optimists within our cohort of experts reflect the existence of pessimists and optimists over the risks in AUV missions throughout the developer and user community. Thus, there is likely to be a tendency for readers who are optimists to see the results here from the optimists as supporting their outlook. Conversely, readers who are pessimistic over AUV risks will alight on results from the pessimists in this study. This tendency towards what is termed confirmatory bias by psychologists exists when people misinterpret new evidence as supporting their previously held beliefs (e.g. Kosnik, 2008). In an ideal Bayesian world people would use the new evidence to update their beliefs, not the opposite, of using their beliefs to interpret the new evidence.

One method to remove the tendency for readers to display confirmatory bias is to only present one set of results, and not those from optimists and pessimists. We were reluctant to use mathematical aggregation to produce this one result, given the wide range of judgments. Consequently, we plan to revisit this data set using behavioural aggregation. By gathering the experts together and facilitating a consensus on the probability of loss distribution for each unique fault we envisage a more robust outcome.

References

Brito, M.P., Griffiths, G. and Trembranis, A. (2008) Eliciting expert judgment on the probability of Loss of an AUV operating in four environments. Southampton, UK, National Oceanography Centre Southampton, 198pp. (National Oceanography Centre Southampton Research and Consultancy Report, 48) http://eprints.soton.ac.uk/54881/

Brito, M.P. and Griffiths, G., 2009. Using Expert Judgments on Autonomous Underwater Vehicle Probability of Loss to Estimate Vehicle Survivability in Four Operating Environments. Submitted to *Risk Analysis*.

Chance, T. S., 2003. AUV surveys – extending our reach, 24000 km later. *Proc 11th International Symposium on Unmanned Untethered Submersible Technology*, AUSI, New Hampshire.

Griffiths, G., Millard, N. W, McPhail, S. D., Stevenson, P. and Challenor, P. G., 2003a. On the Reliability of the Autosub Autonomous Underwater Vehicle. *Underwater Technology,* 25(4): 175-184.

Griffiths, G. and Trembanis, A., 2007. Eliciting expert judgment for the probability of AUV loss in contrasting operational environments. *Proc. 15th International Symposium on Unmanned Untethered Submersible Technology*, New Hampshire, August 2007.

Keeney, R.L. and Winterfeldt, D. 1991. Eliciting Probabilities from Experts in Complex Technical Problems. *IEEE Transactions on Engineering Management* 38 (3): 191-201.

Kosnik, L-R. D., 2008. Refusing to budge: a confirmatory bias in decision making? *Mind and Society*, 7(2): 193-214.

Moline, M.A., Blackwell, S.M., Von Alt, C., Allen, B., Austin, T., Case, J.F., Forrester, N., Goldsborough, R., Purcell, M. and Stokey, R., 2005. Remote environmental monitoring units: An autonomous vehicle for characterizing coastal environments. *Journal of Atmospheric and Oceanic Technology*, 22: 1797-1808.

Moline, M.A., Blackwell, S.B., Case, J.F., Haddock, S.H.D., Herren, C.M., Orrico, C.M. and Terrill, E., 2009. Bioluminescence to reveal structure and interaction of coastal planktonic communities. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56 (3-5): 232-245.

Nelder, J.A. and Mead, R., 1965. A simplex algorithm for function minimisation. *Computer Journal*, 7: 308-313.

O'Hagan, T., 2009. Sheffield expert knowledge elicitation software package. [online]: http://www.tonyohagan.co.uk/shelf/index.html [Accessed on 29 June 2009].

OSD, 2003. Unmanned Aerial Vehicle Reliability Study, Office of the Secretary of Defense, available from www.acq.osd.mil/uas/docs/reliabilitystudy.pdf

Otway, H. and Winterfeldt, D. 1992. Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls. *Risk Analysis* 12 (1): 83-93.

Podder, K., Sibenac, M., Thomas, H., Kirkwood, W. and Bellingham, J., 2004. Reliability growth of Autonomous Underwater Vehicle – Dorado. *Proc. Oceans 2004*, Kobe, Japan. MTS/IEEE. Pp. 856-862.

R, 2009. The R Project for Statistical Computing. [online]. Available on: http://www.r-project.org/ [Accessed on 1 July 2009].

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and Biases. *Science*, 185(4157): 1124-1131.