Risk analysis for Autonomous Underwater Vehicle Operations in Extreme Environments

ABSTRACT

Autonomous underwater vehicles (AUVs) are used increasingly to explore hazardous marine environments. Risk assessment for such complex systems is based on subjective judgment and expert knowledge as much as on hard statistics. Here we describe the use of a risk management process tailored to AUV operations, the implementation of which requires the elicitation of expert judgment. We conducted a formal judgment elicitation process where eight world experts in AUV design and operation were asked to assign a probability of AUV loss given the emergence of each fault or incident from the vehicle's life history of 63 faults and incidents. After discussing methods of aggregation and analysis, we show how the aggregated risk estimates obtained from the expert judgments were used to create a risk model. To estimate AUV survival with mission distance, we adopted a statistical survival function based on the nonparametric Kaplan-Meier estimator. We present theoretical formulations for the estimator, its variance and confidence limits. We also present a numerical example where the approach is applied to estimate the probability that the Autosub3 AUV would survive a set of missions under Pine Island Glacier, Antarctica in January-March 2009.

KEY WORDS: autonomous underwater vehicles; under ice operations; risk management; Kaplan-Meier.

1. INTRODUCTION

Progress in oceanography, ocean biology and climate sciences demands ever more sophisticated data collection platforms, carrying more sensors, and able to reach previously inaccessible areas. Autonomous Underwater Vehicles (AUVs) are increasingly used to support marine science campaigns [1:10]. These vehicles are complex systems, composed of mechatronic technology and software for vehicle navigation and data collection and processing. In many instances, data are transmitted by satellite to a shore laboratory, where scientists can analyse them in near real time. Losing a vehicle can be very costly: some of these vehicles cost millions to build or purchase, and the outcome of some research programmes depend on data gathered by them.

Despite this, very little has been published on AUV reliability or on the risk of operating in extreme environments. The loss of the Autosub2 AUV in Antarctica in 2005 motivated a major effort to improve the risk and reliability practices at the National Oceanography Centre, Southampton (NOCS), which operates the AUV. As a result, a novel operational risk management process model was derived to support decision making before the start of a campaign [11]. Expert judgment elicitation plays a pivotal role in this operational risk assessment. When we attempted to integrate independent expert judgment in the risk management process we concluded that existing statistical survival methods had to be modified to accommodate expert subjective probability judgment. As result, we derived an extended version of the Kaplan-Meier statistical survival model (see below).

We implemented the risk management process specifically to manage the risk in the Autosub3 AUV science campaign to the Pine Island Glacier, Antarctica where the vehicle was to operate in a more challenging environment than that where Autosub2 was lost. The campaign was successful; Autosub3 covered approximately 510km under the glacier. This was the first time

that a series of missions had been carried out by an AUV under a floating glacier. We present a brief summary of the risk management strategy carried out both before and during the campaign.

2. OPERATIONAL RISK MANAGEMENT PROCESS FOR AUTONOMOUS UNDERWATER VEHICLES

The AUV risk management process (RMP-AUV) was developed to support decision making for operations in extreme environments[11]. Many definitions of risk exist in the literature and different industrial sectors adopt different definitions. Arguably, the most widely adopted risk definition is that proposed by Kaplan and Garrick[12][13], where risk is defined by the triplet (S_i, L_i, X_i), where S_i stands for the scenario, L_i the likelihood and X_i the consequence. The RMP-AUV considers only AUV loss. The approach does not attempt to quantify financial loss, or the loss of science data. The likelihood (L) comes in first as the elicited subjective probability which refers to loss given that a fault (F) emerges in the declared environment (E). Thus our risk definition is based on the duplet subset of Kaplan-Garrick's (S, L_i), where L_i is the subjective probability $P(L|F,E)$. This risk model is static, which means that on its own it only allows us to rank faults on their risk. In a second stage, the static risk model is integrated with Kaplan-Meier statistical survival methods; the resulting risk model allows assessment of mission risk with distance. For this 'dynamic risk model', risk remains defined by the duplet (S, L_i). Although the loss scenario remains the same as for the static model, the likelihood is now the probability of loss given a specific distance (d), $P(L|d)$.

2.1 Risk Management Process

The process consists of a sequence of steps that must be performed by an individual or groups of stakeholders; Figure 1 presents a flow-diagram representation of the process. The process is generic, when applied to marine science campaigns the user is the principal investigator (PI) and the owner is the head of the laboratory. In a different industrial sector, for example commercial

or military, these roles would be performed by different individuals; however the structure of the risk management process would remain unchanged.

In phase 1, the AUV owner specifies the acceptable probability of loss (L) for a given campaign. This figure is calculated based on the vehicle's capital cost, the daily cost of operation, including vessel and staff time, the cost of spares, value of the science expedition, an element of depreciation, the campaign length, and the severity of the target operating environment.

In phase 2, the AUV user defines the mission operational requirements, presented in terms of individual mission distances in a given operating environment. The user is expected to specify a minimum required set of missions and a desired set of missions.

In phase 3, the risk assessor calculates the AUV computed probability of loss (A) for the campaign. If the computed probability of loss is lower than the acceptable probability of loss, the process jumps to phase 5, the campaign is recommended to proceed. Otherwise, if the computed probability of loss is higher then the acceptable probability of loss (A > L), a recommendation is made not to pursue the campaign but to revisit the risks. Risk mitigation actions may reduce the computed probability of loss to a value lower than the acceptable probability of loss. This feature is represented by the link from phase 6 to phase 3. The campaign acceptance decision may also be reached if the AUV owner decides to increase the acceptable probability of vehicle loss to a value greater than the computed probability of loss or the AUV user sets out requirements with lower risk.
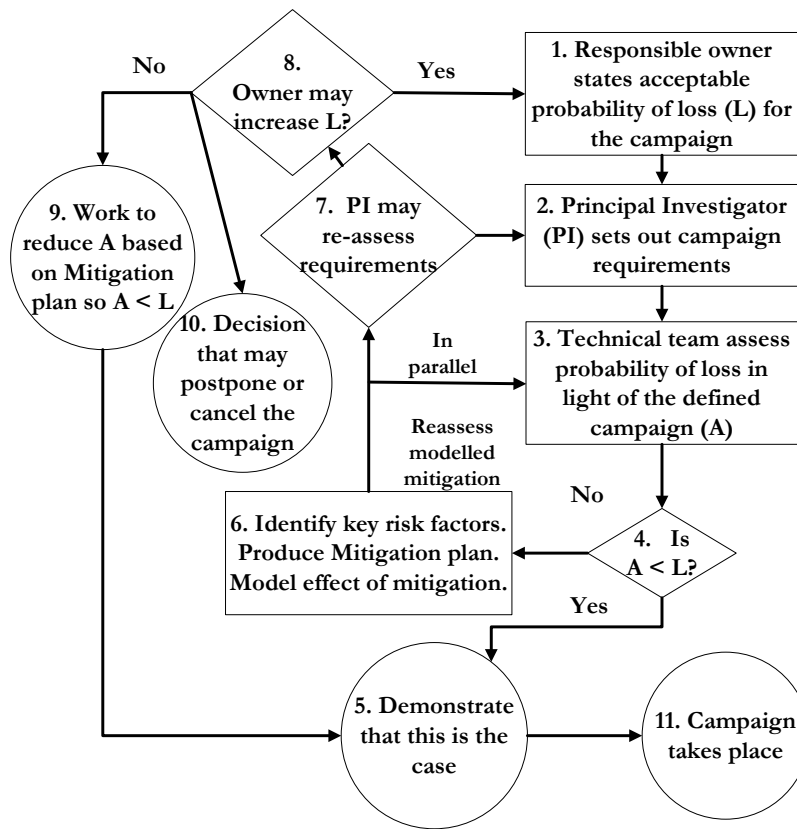
Figure 1 Operation Risk Acceptance Process proposed by Griffiths and Trembranis[11].

The operational risk model is created based on the vehicle's prior fault history. The limitations to this approach are discussed in section 7.

2.2 Basis of the risk computation – the fault incident history

The initial design of the AUV subsystems used Hazard Analysis to understand the risks and to prioritise expenditure and engineering effort. In practice, however, the revealed reliability of the vehicle during trials was more important, as many of the subsystems were new, unique, and with limited scope for testing apart from within the entire vehicle at sea. The qualitative nature of Hazard Analysis is a shortcoming in complex vehicles, as has been found, for example, in the Space Shuttle programme [14]. Despite the use of fault tree analysis, failure rate and probability data, this programme acknowledged the need for assumptions based on engineering judgment.

A total of 38 Autosub3 missions between July 2005 and April 2007 formed the basis of the dataset. Eight experts in AUV design and operations considered 63 faults on 28 missions (Appendix I). Ten missions showed no faults, included as censored data in later analyses. The missions varied in distance from 1.5 km to 302.5 km. Faults or incidents were caused by different factors such as human, operational or maintenance errors, or software or hardware design errors. We considered this history (38 missions, 1931km) representative of the likely faults to emerge over the subsequent campaign for which the risk assessment was to be performed (six missions, 495km). Results from that campaign would be added to the dataset for the next campaign, and so on.

2.3 Environments

We created the risk model for each of four contrasting AUV operating environments – open water, coastal water, sea ice and ice shelf[16]. These distinct environments have different operational challenges. A major factor in the expert judgment on risk of loss is whether recovery can be affected if the AUV encounters a problem or suffers a fault. If an incident occurs under an ice shelf, where the ice thickness may be over 100 m, the risk of loss is far higher than for the same fault in open water. Some factors are common to one or more environments, such as launch and recovery, incidents during which can lead to loss or insurance write-off[18].

2.4 Expert judgment elicitation

The fault history and probability of loss of Autosub3's precursors were based on a one-person judgment and knowledge elicitation exercise[17]. This approach was criticised as too insular and too informal, but was taken up by some[19]. As a consequence of the criticism, the present study has involved more people and a more formal approach[16]. The purpose of the formal elicitation exercise was to model AUV risk in different operational environments, given the Autosub3

history of faults. We made the full fault history public (Appendix I) and engaged experts to estimate P(L|F,E).

2.5 Extending the Kaplan Meier survival model

Having quantified the risks as subjective probability of loss through expert judgment, the next step incorporates frequentist statistics on probability of occurrence and at what distance the faults or incidents occur. Although it is common practice to examine survival as a function of time, here we use distance, which is more applicable to the AUV example. However, the proposed method is equally applicable to survival with time. The product-limit method derived by Kaplan and Meier is a well-established nonparametric model for estimating and displaying survival functions (or equivalently P(F)) based on a small or medium data sample [20][21].

Using the survival estimator in its usual form, we can only deal with loss or survival to end of a mission (which becomes censored data), as in an earlier simplified analysis [17]. The aggregated expert judgments contain a representation of uncertainty on whether or not a given failure would lead to loss, represented as a subjective probability. To model this uncertainty, the conventional Kaplan-Meier approach must be modified, leading to the expression in equation 1. The mathematical proof is presented in appendix II.

$$\hat{S}(r) = \prod_{r_i < r} 1 - \left(\frac{1}{n_i}\right) P(e_i) \qquad (1)$$

Where n_i is the number of events at risk at range r_i, and $P(e_i)$ the probability of fault leading to loss. Thus if $P(e_i)$ is zero we have a censored case; that is, no loss is observed during the interval of interest. If $P(e_i)$ equals one, loss is observed during the interval of interest. For these extremes, the approach reduces to the original version of the Kaplan-Meier method.

8

3. FORMAL EXPERT JUDGMENT ELICITATION PROCESS

The goal of the formal judgment elicitation process is to remove biases [22][23][24]. We followed the formal elicitation process of Otway and Winterfeldt [25]. In brief, it consists of a sequence of steps that ensure that the problem is clearly presented, that experts are trained on assigning probability to events, that judgments are formally studied and misunderstandings are corrected, and that suitable methods are used for aggregating experts' judgments. The following subsections summarise how all phases of the formal judgment elicitation process were implemented.

1) **The issues.** Given the set of facts on faults and incidents with Autosub3 throughout its life to date, we sought to estimate the probability of loss of the vehicle in different operating environments. At issue is how likely was it that each fault or incident, taken in isolation, but with the expert's knowledge of the wider issues, could lead to loss in the four example environments.

2) **Selecting experts.** Here, the aim was to maximise experts' independence from the Autosub3 design, development or operation by using individuals from different backgrounds, areas of expertise and nationality. All eight experts selected were experienced in one or more areas of AUV concept design, development and operation. Here experts are referred to by their initials: TC (20 years experience in military and scientific research), BF (8 years experience in military), CJ (11 years experience in scientific and military), MM (6 experience in scientific research), RM (9 years experience in scientific research), AS (1.5 years experience in scientific research), CW (10 years experience in scientific research), and DY (15 years experience in scientific research).

3) **Clearly define the issues.** We asked experts to answer the straightforward question: "What is the probability of loss of the vehicle in the given environment X given

fault/incident Y?" The purpose is to quantify the impact of the fault or incident on the probability of loss of the vehicle, not, for instance, on the impact that the given fault might have on science delivery, or on the probability of the fault occurring or recurring.

4) **Training the experts.** We briefed six experts on the background and method of eliciting expert judgment at a presentation in August 2007 [16]. We later briefed two in person. We gave experts access to independent information. Experts were given access to independent information. First, for open water and coastal environments, Leviathan, a leading marine insurance binding authority, stated that they had not paid out on an AUV loss in the two years to June 2007. Second, out of some 150 vehicles produced by Hydroid LLC, a manufacturer of AUVs used in open and coastal waters, and under sea ice, we believe that none were lost as of August 2007. Third, through the early stages of development and operations with Seaglider (a buoyancy-driven AUV), eight out of the first ten vehicles were lost, in environments that ranged from open water to areas of sea ice, and of the next twelve built, two were lost as of September 2005.

5) **Eliciting judgments.** The experts worked independently using a pro forma that tabulated the fault or incident and required a judgment on the probability of loss and a confidence score to be given for each judgment, in each of the four environments. A text box was available for each fault for the expert to explain their judgment. The typical time taken to complete the pro forma was half to one day.

6) **Analyzing the judgments.** We conducted analyses of the expert judgments in two steps. First, we examined the distribution of the longitudinal probability judgments over all faults for all four environments. This first assessment was necessary in order to identify and understand the main differences between individual judgments of the same fault or incident. Second, we sought to understand how experts assigned their judgments. For example, an expert that used a wide range of probabilities was less

10

likely to manifest bias due to anchoring**i**. However the expert judgments' variability may vary with change in environment.

7) **Aggregating expert judgments.** Behavioural methods for combining expert probability involve eliciting probability judgments from a group of experts, where experts must all agree on each judgment. In contrast, mathematical methods make use of analytical algorithms to combine individual probability judgments; in this case we kept experts separate during the elicitation process. Previous research has shown that mathematical aggregation methods tend to perform better [29]. However, it is generally advised that one should conduct sensitivity analysis of different mathematical methods in order to study their sensitivity to individual expert judgments [30].

8) **Complete analysis and write up.** A report presented each expert's assessment in full, included detailed analyses of the experts' judgments and raised recommendations wherever misunderstandings or mistaken assumptions were indicated. The original data and the full analysis of expert judgments are presented elsewhere [26].

## 4. RESULTS OF THE ELICITATION PROCESS

With eight experts, sixty three faults and four environments, we had a possible total of 2016 expert judgments. In practice, we had 1863 because not all combinations were completed by the experts. The purpose of the analysis was to provide feedback to experts. Our intention was not to suppress differences of opinion, or to introduce our own views, or to bias the results, but to draw our experts' attention to those faults where there appeared to be resolvable differences of opinion, misunderstandings or typographical errors.

We queried opinions regarding eight faults in open water, four in coastal waters, seven under sea ice and six under shelf ice (several being common to more than one environment). This was 10% of faults; in 90% of cases there was no need to draw attention to differences between

experts' opinions. The analyses supported by these graphs are presented in the following sections, where we also draw upon the written comments of the experts, especially where there were major differences in interpretation.

4.1 Longitudinal judgments distribution

The longitudinal distribution of the probability judgments provided a visual mechanism for identifying major discrepancies in expert judgments. The method allowed us to identify major discrepancies in six faults in open water, three in coastal, five in sea ice and six under an ice shelf. Figure 2 presents the probability judgments where there were major discrepancies for the under-ice shelf environment. Failure 391_2 concerns a fault with the GPS antenna, which can significantly effect the system navigation. Some experts deemed this fault critical, but others did not consider it critical when actually under ice where it is never used. The remaining faults occurred either during recovery or when the vehicle was onboard. One might expect these faults to be deemed of low criticality. However, our aim was not to impose our views, but simply to highlight differences in opinion.
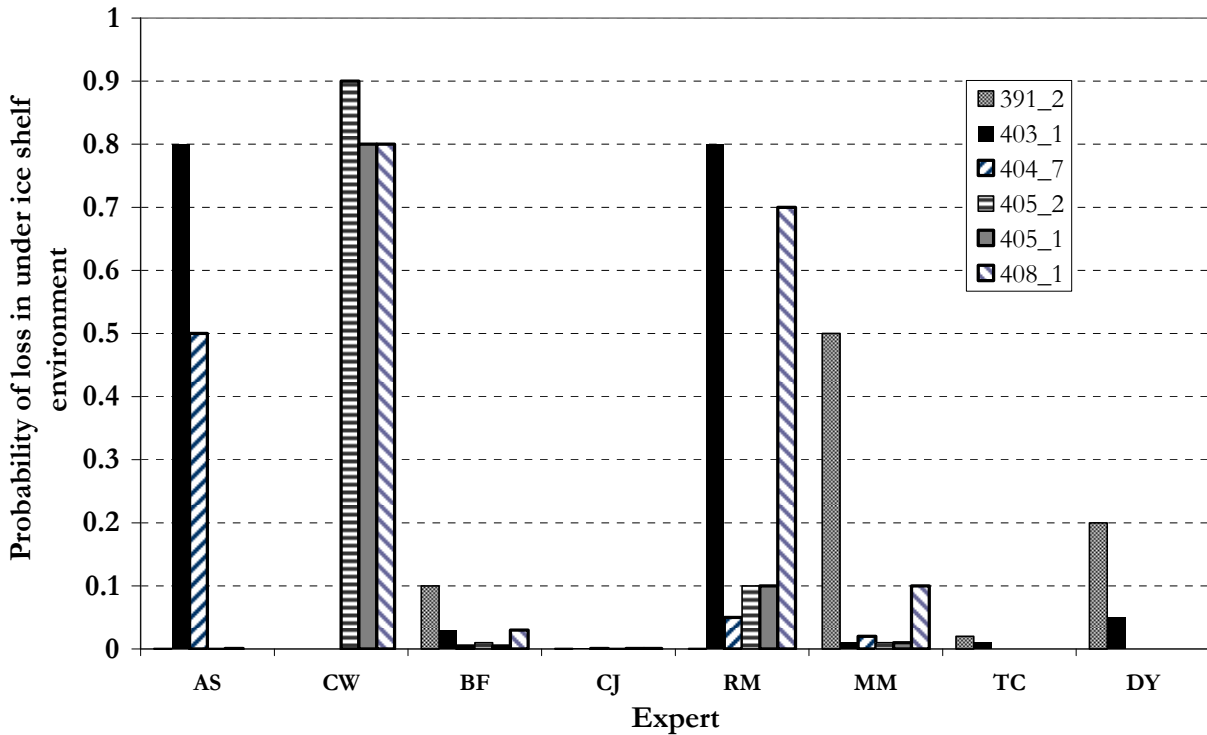
Figure 2 Comparison of expert judgments for ice shelf. Expert judgments for faults 391_2, 403_1, 404_7, 405_2, 405_1, and 408_1.

4.2 Frequency distribution of experts' judgments

We analysed variability in judgments in terms of how often experts used different ranges of probability judgments. We considered nine intervals of probability. We examined expert judgment variability across different environments and then looked at variability in the totality of assessments between experts. A summary of our analysis is presented below.

4.2.1 Open water

The cumulative distribution (Figure 3) shows that some experts use a wider range of probability in their judgments than others. A P(L| F,E) of 0 is only assigned by five experts; a minimum P(L| F,E) of 0.001 was assigned by two experts and a minimum of 0.01 by one. Only one expert assigns P(L| F,E) values between 0.001 and 0 (0.0001 in four instances). Six experts

have at least 98% of all their P(L| F,E) judgments under 0.03; two are more pessimistic, one with 90% and one with 92% below 0.03.
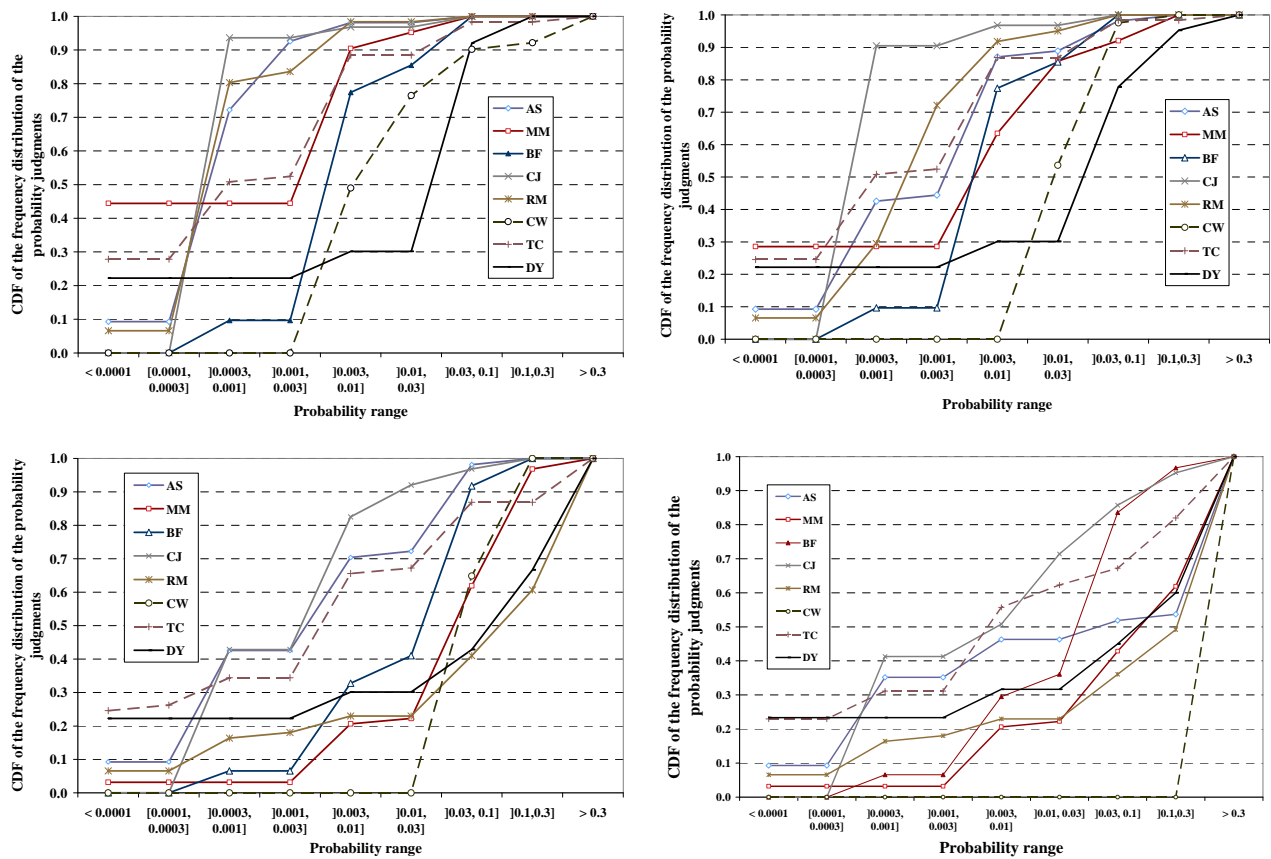
Figure 3 Comparison of the cumulative probability Cumulative frequency of the expert probability judgments. Open water (top left), Coastal water (top right), Sea ice (bottom left) and Ice shelf (bottom right).

4.2.2 *Coastal water*

In coastal water, two experts reduce the number of times that they assign a P(L| F,E) < 0.0001 compared with open water: MM assigned P(L| F,E) < 0.0001 in 29% of judgments compared with 44% for open water; TC assigned P(L| F,E) < 0.0001 in 25% of judgments compared with 29% for open water. A cluster of experts (formed by AS, MM, BF, RM and TC) quite frequently used the mid-lower ranges of [0.0003, 0.001] and [0.001, 0.003].
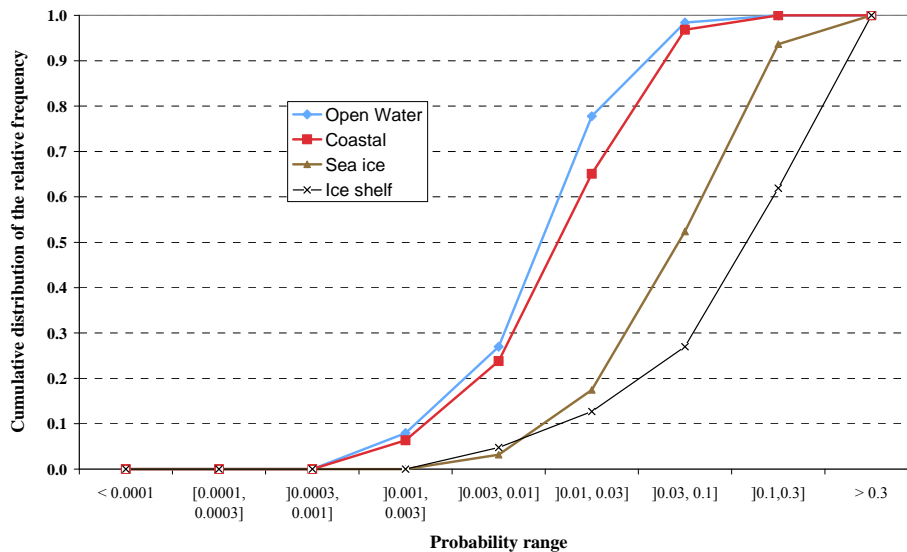
4.2.3 *Sea ice and Ice shelf*

In this environment, some experts' cumulative distribution had a narrow 'S' shape (AS, BF,

CJ, CW, MM and TC), whereas others displayed a broad 'S' shape (DY and RM). Together with the clustering in the median probability, this inspired us to classify experts as optimists (experts whose cumulative distribution follows a narrow 'S' shape) or pessimists (experts whose cumulative distribution displays a broad 'S' shape). The existence of two "schools of thought", especially for operations under sea ice and ice shelf demands caution when using an average over all experts.

The main difference between the judgments provided for sea ice and ice shelf are in the number of judgments that lie in the range of [0.3, 1.0]. Six experts assign $P(L|F,E)$ values in this range when assessing the ice-shelf scenario, whereas two experts are optimists. The other experts who were optimists for sea ice became pessimists for ice shelf.

4.3 Probability of loss across the four environments

If we compute the un-weighted average of the expert judgments for all four environments we obtain the distribution depicted in Figure 4.

Figure 4 Comparison of the cumulative probability distribution of the un-weighted average for four operating scenarios. Open water (blue), Coastal water (red), Sea ice (brown) and Ice shelf (black).

The distributions across all four environments cluster into two groups. Group 1 comprising open water and coastal water and group 2 sea ice and ice shelf. These two groups are separated

15

by a significant gap. For ice shelf, the average judgment is dominated by an increased view of risk from six of the eight experts. The distributions presented in Figure 3 do not capture the influence of each expert's confidence on his/her risk assessment. The following section considers how the experts' self assessment of confidence influences the aggregated opinion for each of the four environments.

4.4 Aggregating expert Judgments

Our approach was to use a simple linear opinion pool, having discounted the logarithmic opinion pool because of the veto given to experts assigning a zero probability.

We applied a linear weighted opinion pool to combine expert judgments concerning all 63 faults [28], defined as follows: a single probability judgment is created by summing the products between an individual expert's own confidence (w_i) and their judgments ($p_i(\theta)$) for the n experts. Where θ is the uncertain event (which in our case is AUV loss). Some researchers argue that self-rating is an effective way to quantify the experts efficiency [31]; we have not altered any self-ratings. Equation 2 sets out the linear pool aggregation method.

$$p(\theta) = k \sum_{i=1}^{n} w_i p_i(\theta) \qquad with \quad k = \frac{1}{\sum_{i=1}^{n} w_i} \qquad (2)$$

The confidence varies in a range of 1 to 5; 5 indicating a high and 1 a low confidence.

Figure 5 depicts the linear weighted cumulative distribution for each of the four environments. Comparing the relative frequency of the judgments for open and coastal water, the distribution shows that whilst the shape of both distributions is similar, the coastal water distribution presents a shift of judgments towards greater risk. Similarly to open water, the majority of probability judgments for coastal water lie in the range [0.01, 0.03]. However there was 17% increase in judgments in the range [0.03, 0.1] and two failures are assigned a P(L| F,E) greater than 0.1 but lower than 0.3.
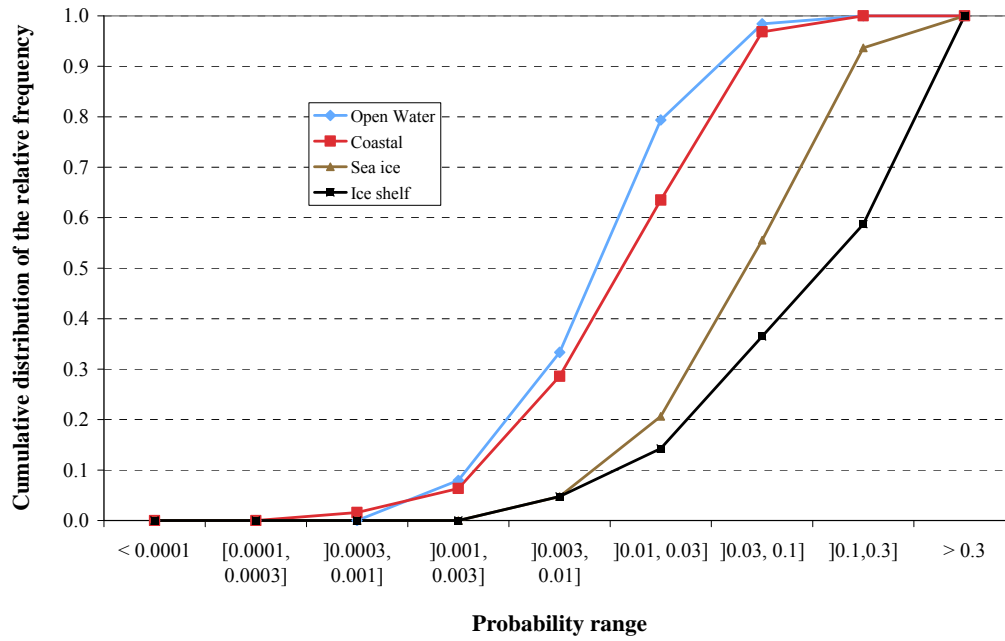
Figure 5 Comparison of the cumulative probability distribution of the linear-weighted average for four operating scenarios. Open water (blue), Coastal water (red), Sea ice (brown) and Ice shelf (black).

The shift of probability judgments towards greater risk becomes more evident in the aggregated judgments for sea ice and ice shelf. Ice shelf is the most severe environment where 41% of all failures are assigned a $P(L| F,E) > 0.3$. A summary of the statistical properties of the $P(loss)$ distributions obtained using the linear aggregated opinion pool is presented in Table I.

Table I Statistical properties of the linear aggregated opinion pool.

Statistics	Environment			
	Open water	Coastal	Sea ice	Ice shelf
Quartile 25%	0.0083	0.0083	0.045	0.072
Median	0.018	0.021	0.088	0.17
Quartile 75%	0.026	0.037	0.17	0.40
Quantile 95%	0.049	0.090	0.36	0.75

5. RISK MODEL FOR EXTREME ENVIRONMENTS

The risk model is based on two components, the subjective probability of a fault or incident leading to loss in a declared environment and, independently, the frequentist probability of the fault occurring. The first component has been established through the expert judgment

17

elicitation. For the second, we have the fault history, mission by mission, and we have those missions on which no fault occurred.

5.1 Static risk model

The expert judgment analysis identified optimists and pessimists. One might argue that a behavioral aggregation method could be used to encourage consensus between expert judgments. However from reading the explanations, we concluded that some of the disagreements were simply due to the fact that an expert had an optimistic view of the possible outcome, whereas the others would have a pessimistic view. The risk model must in this case capture both schools of thought; a linear pool averaging of all experts would remove these distinct views. We used the cumulative distribution of the frequency of the probability judgments to select which experts followed optimistic or pessimistic profiles. Table II presents the different groups for all operating environments.

Table II   Experts groups description

Model		Experts
Open water	Optimist	MM,CJ,RM,TC and
	Pessimist	BF, CW and DY
Coastal	Optimist	AS, CJ, RM and TC
	Pessimist	BF, MM, CW and
Sea ice	Optimist	TC, CJ and AS
	Pessimist	MM,BF, RM, CW
Ice shelf	Optimist	CJ and TC
	Pessimist	AS,MM,BF,RM,CW

5.2 Dynamic risk Model

It is only possible to apply statistical survival methods to a single AUV if it is considered to be a repairable system. That is, after a fault or incident, the item in question is repaired or replaced, such that the AUV is put back into the state it was in before the mission started. In this model, the serial missions of the single AUV are treated analogously to the parallel items in a parts

failure analysis. Instead of time to failure, we use the range achieved by the AUV before a fault or incident occurred. If no failure or incident occurred, we treat the mission as censored data, with its associated range.

In practice, the assumption of the AUV as a repairable system does have weaknesses. Whenever faults occur, the engineers seek a repair or replacement that increases reliability. However, the pattern of faults that have occurred shows no discernable reliability growth over two years, see Appendix I. At this stage in the vehicle's use, therefore, we consider the past history to be a good guide as to the immediate future. Despite this limitation, the risk model is transparent, traceable and easily understood.

We implemented equation 1 in a Visual Basic program running on Excel 2003. Figure 6 presents the survival distribution obtained for open water, coastal water, sea ice and ice shelf, separately for the optimistic and pessimistic subgroups.
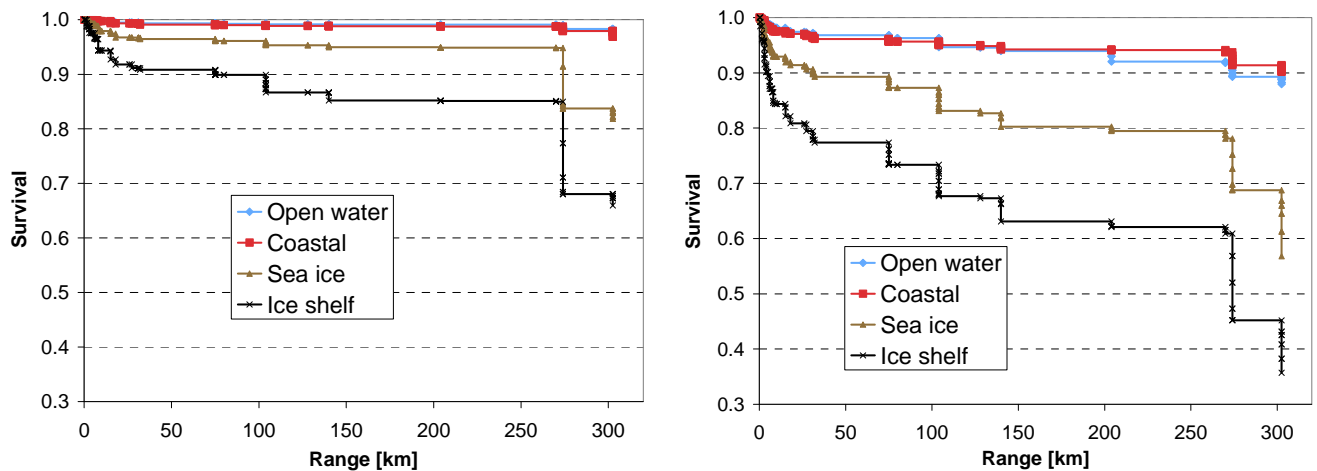
Figure 6 Kaplan Meier optimistic assessment (left) and pessimistic assessment (right) for the survival of Autosub3, for all four operational environments prior to any mitigation measures.

The optimistic Kaplan-Meier survival distribution for operations under ice shelf (Figure 6) shows a steep decline in the probability of survival at shorter distances, whereas at mid distances the survival distribution is almost flat. For managing the risk, if we can monitor the AUV at these shorter ranges and recover the vehicle to address any problems if they emerge, the risk

posed by those failures will significantly reduce, thus reducing the probability of loss when under ice. In practical terms, this means that if the AUV is about to undertake an operation under ice, it should cover some distance in open water before diving under ice. Mathematically, the mitigated risk can be estimated using conditional probability as described in section 6.1.

6. CAMPAIGN TO THE PINE ISLAND GLACIER, ANTARCTICA

The science expedition to the Pine Island Glacier was a joint UK-US scientific programme, sponsored by the Natural Environment Research Council, UK and by the National Science Foundation, US. Its main goal was to gather data needed to understand how warm Circumpolar Deep Water (CDW) gets beneath the glacier and how it determines the rate of glacier melting. Subsidiary objectives were to map the seabed beneath the glacier as well as the glacier's underside, and to determine where and how heat is transferred from the inflowing CDW to the outflowing ice-ocean boundary layer.

We used the survival function obtained from this analysis to estimate the risk of a scientific campaign to the Antarctic that would use Autosub3[ii]. We derived the following four scenarios from the operational requirements:

- Scenario 1 – Minimum set of three 60km open-water and three 60km under-ice shelf missions.
- Scenario 2 – Minimum set as above with 540km under sea ice over the six missions, no open water.
- Scenario 3 – Desirable set of three 60km open-water, three 60km and three 120km under-ice shelf missions.
- Scenario 4 – Desirable set as above but with 720km under sea ice over the nine missions.

We used the extended survival function to compute the probability of Autosub3 loss for the four scenarios. For each scenario, we computed a probability of loss using an optimistic and pessimistic model of the expert judgments. We present the results in Table III.

Table III   Probability of losing Autosub3 based on Kaplan-Meier analyses for each scenario. The optimistic assessment is the first number in each cell, the pessimistic the second.

Analysis Model	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Kaplan-Meier	0.26 – 0.56	0.40 - 81	0.53 – 0.86	0.64 – 0.96

The computed probability of loss exceeded the acceptable limits defined by the Autosub3 responsible owner, namely: 10% for scenario 1, 17% for scenario 2, 20% for scenario 3 and 23% for scenario 4.

6.1   Modeling the effect of mitigation

The difference in risk acceptance shown above comes from a risk model that takes into account the different mission environments, and the number of missions. As a consequence, risk mitigation measures are necessary before the risk of loss will become acceptable. These will need more than rectification of the faults found, which would merely return the vehicle to its pre-fault state. Rather, we undertook a series of technical measures aimed at reducing the incidence of faults. In addition, this statistical risk analysis suggested a further mitigation strategy.

The flat shape of the nonparametric Kaplan-Meier survival distribution does not lend itself to an analysis to quantify the effect of mitigation measures such as varying the monitoring distance. This effect is better captured by a parametric survival distribution such as a Weibull distribution derived from the failure distance and the expert judgments on the probability of loss for each failure. We took a simulation-based approach to deriving the Weibull parameters for the case of loss given our experts' judgments. For each fault or incident we generated 1000 copies, with integer($1 - P(e_i)$)*1000 entries censored, the others marked as losses. We then obtained the parameters of the Weibull distribution (Figure 7) using JMP software package from SAS.
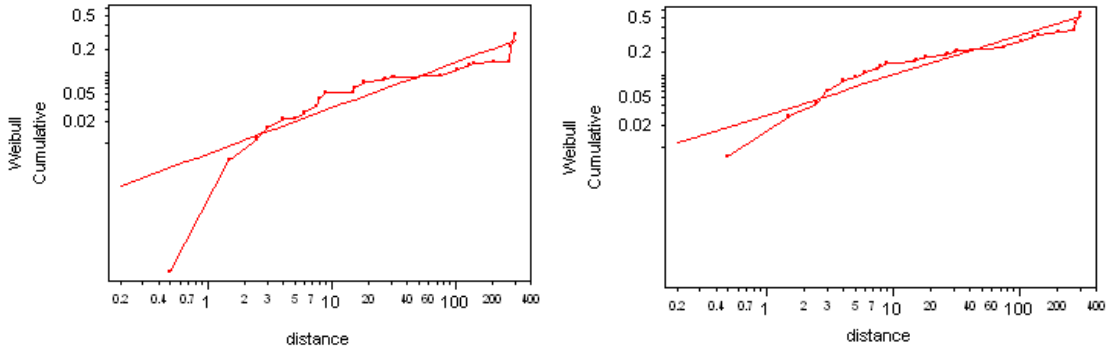
Figure 7 Weibull survival distribution (left) optimistic assessment and (right) pessimistic assessment for loss under an ice sheet before any mitigation measures. The straight red line is the best fit used to estimate the Weibull parameters alpha and beta.

To study the effect of mitigation through varying the monitoring distance we take the conditional probability of the AUV surviving distance *X* given that it has survived distance *Y*, where *Y* corresponds to the monitoring distance, as given by equation 3.

$$P(X < x \mid X > y) = \frac{F(x) - F(y)}{1 - F(y)} \quad (3)$$

Where *F(·)* is the Weibull cumulative distribution function. For the example shown in Table IV we used the experts' optimistic assessments. We excluded in a reanalysis the faults on one mission where the causes were very well understood and modifications made that put the vehicle into a state where its reliability should be higher than before the faults occurred. We included open-water trial missions before the under-ice missions of the expedition, and assumed a monitoring distance Y of 48 km. Under these conditions, the mitigated probability of loss in for the four scenarios is reduced when compared with the unmitigated case (see Table IV).

Table IV Probability of losing Autosub3 based on Weibull analyses for each scenario based on the optimistic experts' assessments without and with mitigation.

Analysis Model	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Weibull unmitigated	0.29	0.47	0.57	0.72
Weibull mitigated 48km	0.035	0.09	0.18	0.23

Although a monitoring distance of 48km is adequate for scenario 4 it is too onerous for scenarios 1, 2 and 3 because it would involve monitoring the vehicle for 8 hours. The P(loss) of 0.035 for scenario 1 is well below the acceptable P(loss) of 0.10 and would result in unnecessary use of ship time. A monitoring distance of 28km for scenario 1 would provide a P(loss) ~ 0.10. Using the same argument, a monitoring distance of 33km would give P(loss) ~0.17 for scenario 2, and a monitoring distance of 43km would give P(loss) ~ 0.20 for scenario 3. These are all within the acceptable risk margins defined by the responsible owner, and therefore formed guidance for the operations team at sea depending on the conditions they faced.

6.2 Autosub3 campaign results

On 5 January 2009 the US ice breaker, the *Nathaniel B. Palmer*, with Autosub3 onboard departed Punta Arenas, Chile, for Pine Island Glacier, Antarctica[32][33]. The vessel arrived in the working area on 16 January 2009. The environment conditions were benign for initial operations; there was no sea ice in front of the ice shelf and the sea was relatively calm. The PI decided to run scenario 1, which began with three 60km missions under the ice shelf (30km into and out of the cavity, respectively). The campaign started with an open-water test mission (mission 427) during which Autosub3 reached a depth of 836m and covered a distance of 37.5km. A fault was observed on one scientific sensor, which was later fixed. Subsequent science missions (428, 429 and 430) were successful, although a minor fault occurred during mission 429. On mission 428, which took 18.4hrs, Autosub3 travelled 101km, including 60km under the ice shelf. On mission 429, Autosub3 travelled 113km in 21hrs and on mission 430 the vehicle travelled 107km in 19hrs. The navigation plan, for all these missions, was for the vehicle to track the bottom of the seabed when travelling into the cavity, and to maintain a distance of 100m from the ice when travelling out of the cavity.

The above set of missions corresponds to scenario 1. The PI decided to move on to scenario 3 (desirable) with three additional, longer, under ice missions. During the first of these, mission 431, the vehicle collided with the ice on its way out of the cavity having entered a large, complex fissure. The collision avoidance system initially prevented impact but in attempting to avoid the ice ahead the vehicle subsequently turned in to the ice wall. It then scraped along, until a gap allowed it to turn. However, the vehicle then scraped along the opposite wall until diving out of the fissure. The vehicle's return to a safe depth, away from the ice, implied that it was unable to profile the ice on its way out of the ice shelf. Lack of information about ice topography and the behaviour of the collision avoidance system were the main causes for this incident.

The discovery of a ridge running across the ice cavity during mission 431 led to a new science requirement. The PI requested two more missions. The first of these (mission 433) involved a run on the north side of the cavity, travelling approximately 50km into the cavity and 50km out, using bottom track for navigation. Mission 434 involved a run of approximately 30km into the cavity (from point 2 to point 5 in Figure 8), then a turn south, travelling 15km (from point 5 to point 4 in Figure 8) whilst profiling the new ridge. It would then turn north and travel 30km (from point 4 to point 6 in Figure 8).

These additional missions required a new risk assessment. Data collected from missions 427, 428, 429, 430 and 431 were added to the risk model. The criticality of faults that emerged during missions 429 and 431 were assessed by the NOC's AUV experts. The computed probability of loss for the two missions was approximately 0.09, provided that a monitoring distance of 48km was set before the two missions. This was acceptable to the Responsible Owner, communicated to the ship, and subsequently Autosub3 successfully completed these missions.
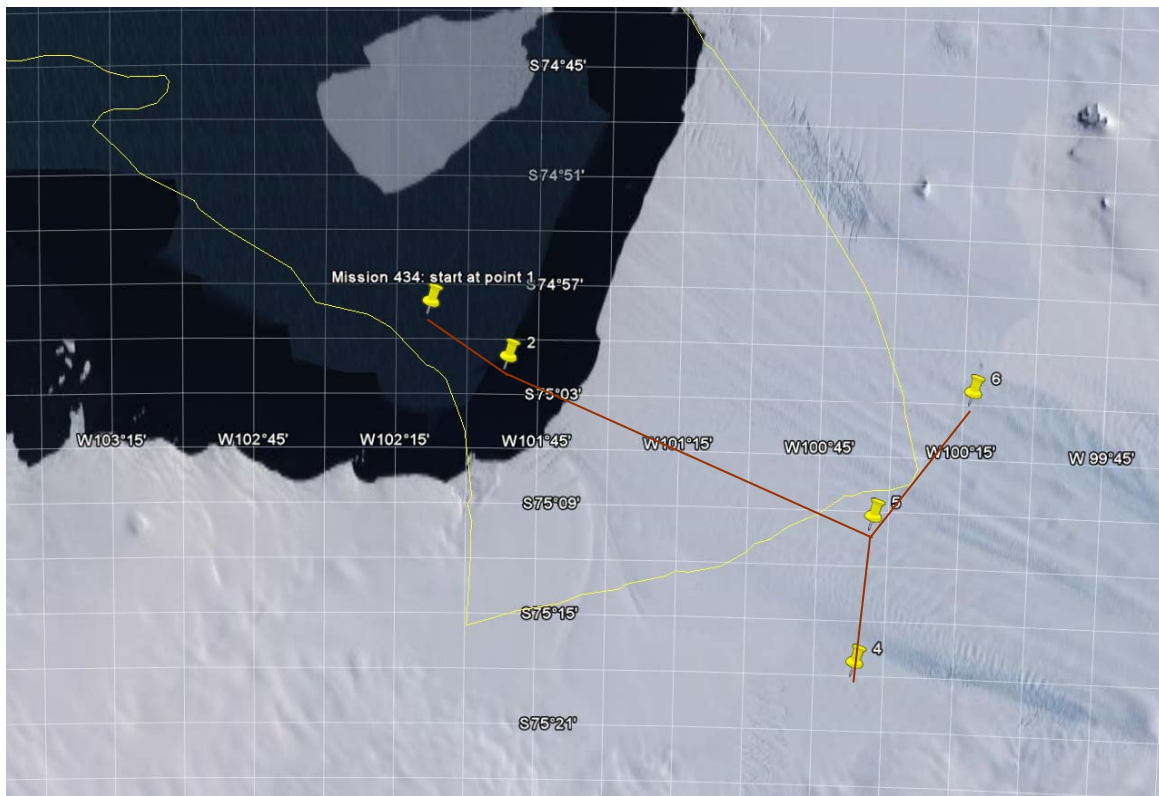
Figure 8 Plan for mission 434 (in red). Autosub3 started at point 1, -102.003° Longitude, -75.007° Latitude. The vehicle was set to perform the following path: 1→2→5→4→5→6→5→2→1. Courtesy of Google Earth. The Google Earth drawing of the coastal line, in yellow, is not exact is just an estimation.

7. DISCUSSION

All risk management models are subject to uncertainties, especially those incorporating expert judgments and other model assumptions such as descriptions of operating environments. When risk models are used to support critical 'go - no go' decisions it is imperative to understand the sources of uncertainty, how they affect the computed risk and how to reduce them. We will therefore now critically review of our methodology, focusing on uncertainties that affect risk computation and uncertainties that affect the application of the RMP-AUV itself. Uncertainties potentially affecting the decision maker's estimation of the acceptable risk are discussed elsewhere [11].

The expert judgment approach is arguably a significant source for uncertainty in the risk model. First, how expert are the experts? We followed Otway and Winterfeldt [25] in aggregating expert

judgments, combining individual probability assessments and the experts' self assessment of confidence in their own probability judgements. Cooke and Goossens[35] argue that aggregation should also take into account how well experts perform in a series of seed questions. These are questions about uncertain events relevant to the topic, but for which the facilitators know the answer. How to best to combine expert self assessment with performance in the seed questions without introducing bias is still debated [36]. In a subsequent project we used seed questions as part of expert training on assigning probability to judgements.

Second, how uncertain are the experts? For this work, experts provided a single probability judgment for each fault, so their uncertainty about individual judgments was not captured. O'Hagan and co-workers [37] require experts to provide a probability range instead of a single probability judgment. The process is based on behavioural aggregation, involving a brief discussion of individual assessments after which experts are encouraged to agree on a single cumulative distribution for each judgment.

There is also uncertainty over depiction of the operating environment. The risk models used within the RMP-AUV were developed for specific, well defined environments. Risk may alter if the real operating environment deviates from the environment description. In principle, causal relations between variables affecting risk could be defined, and the weight between, local, causal relations could be quantified, via expert judgment for instance. In this case, a Bayesian approach could be adopted for updating calculated risk according to variations in the operating environment and conditions.

Under the RMP-AUV a science campaign is authorised only if the computed risk (probability of loss) is lower than or equal to the acceptable risk. One way to cope with uncertainty is to reduce the computed risk to a level below the acceptable risk. The resulting gap between computed risk and acceptable risk is analogous to the safety factor concept used in structural engineering, and would need to be specified by the owner. Increasing the gap between the risk estimates would

increase the 'go ahead' decision robustness, not only to uncertainties, but also to future faults. The decision should be based on the number of faults that maximise the robustness for a given gap between the computed risk and the acceptable risk [37].

Minimal historic data exist for under-ice shelf AUV missions. Frequentist statistics from the three known campaigns (of which is described here is the third) are, cumulatively: 1 loss from 1 mission; 2 losses from 3 missions; 2 losses from 9 missions [34]. The latter is close to the acceptable and computed risk in this study. Experience from future under-ice missions could update this frequentist assessment.

The present dynamic risk model is limited to repairable faults, and future work should address the case of unrepairable faults. When incidents are caused by interaction between vehicle and environment, the necessary changes to behaviour algorithms generally cannot be made in the field, and hence are considered unrepairable until a proper opportunity arises.

A central predicament in risk analysis is how to predict unlikely events. The RMP-AUV is based on observed faults and successful missions. One problem is how to take into account faults that have not manifested themselves during AUV trials or subsequent science missions. One possible approach is to extend the elicitation process beyond the set of historic faults, for example with a structured 'what if' approach. The richness of the data would depend on the expertise and breadth of knowledge of experts. Nevertheless, such an approach would increase the robustness of risk estimates. Within RMP-AUV, design integrity and that the risk assessment that goes with it are no less relevant, and it is important that engineers and managers brainstorm on the operational risks before campaigns. However, we contend that the reliability of a product is given by its operational history and not by its design integrity. Historic data reveal the reliability of electronic hardware, software and mechanical components, the AUV operators, and, most importantly, how these factors interact. Nevertheless, given its clear importance, future work should integrate design integrity with the RMP-AUV.

8. CONCLUSION

A formal expert judgment exercise based on the fault history of a one-off, complex underwater vehicle in environments with different risks has led to quantitative estimates of the probability of vehicle loss. The number of experts engaged (eight) was sufficient to show that, in this very immature area of risk analysis, there were two 'schools of thought' (optimistic and pessimistic) on the risks for AUV operation under ice. This is likely to continue until more experience is gained of operating AUVs in these environments. Applying an extended version of the Kaplan-Meier model enabled estimates to be made of the probability of AUV survival as a function of range. Estimating the risk of a series of missions over a campaign allows the predicted risk of loss to be compared with the Responsible Owner's risk appetite. Although engineering best practice and the permanent removal of some of the faults that occurred provided two mitigation measures, the shape of the Kaplan-Meier survival curves suggested a mitigation strategy based on conditional probability. We derived scenario-dependent monitoring distances using a parametric analysis, now included in standard operating procedures for the AUV, thus improving the risk management decisions made at sea. However, as the shape of the Kaplan-Meier curve may be different for other AUVs, the mitigation strategy does need to be tailored to the history of each vehicle.

Appendix I – Autosub3 historical fault description

Fault code	Distance	Description
384_1	1.5	Mission aborted (to surface) due to network failure. (Much) later tests showed general problem with the cable
384_2	1.5	Loop of recovery line came out from storage slot, long enough to tangle propeller.
385_1	15.2	Autosub headed off in an uncontrolled way, due to a side effect of the removal of the upwards-looking Dopler Velocity Log (DVL).
386_1	26	GPS antenna failed at end of mission.
387_1	27.2	Homing failed, and the vehicle headed off in an uncontrolled direction. Mission was stopped by acoustic command. Problem was due to (a) the uncalibrated receiver array, and (b) a network message ("homing lost") being lost on the network.
388_1	0.5	Aborted after 4 minutes post dive, due to network failure. Logger data showed long gaps, up to 60s, across all data from all nodes, suggesting logger problem.
388_1	0.5	Depth control showed instability. +/- 1m oscillation due to incorrect configuration gain setting.
389_1	3	Vehicle went into homing mode, just before dive and headed north. Vehicle mission stopped by acoustic command. It was fortunate that the ship-side acoustics configuration allowed the ship to steam at 9kt (faster rather than 6kt with the towfish) and catch the AUV.
389_2	3	Separately, homing mode not exited after 2 minutes, as expected. It will continue on last-determined heading indefinitely – a Mission Control configuration error.
389_3	3	Problem with deck side of acoustic telemetry receiver front end, unrelated to vehicle systems.
391_1	31	ADCP down range limited to 360m, reduced accuracy of navigation.
391_2	31	GPS antenna flooded. No fix at end point of mission.
391_3	31	EM2000 swath sonar stopped logging during mission.
392_1	32	As consequence of GPS failure on M391, AUV ended up 700m N and 250m E of expected end position.
393_1	5	Acoustic telemetry giving poor ranges and no acoustic telemetry.
394_1	3	Jack-in-the-box recovery float came out, wrapping its line around the propeller, jamming it, and stopping the mission. Caused severe problems in recovery, some damage to upper rudder frame, sub-frame and GPS antenna. Required boat to be launched.
395_1	8	Jack-in-the-box line came out, wrapped around the propulsion motor and jammed.
396_1	4	Current estimation did not work, because minimum time between fixes for current to be estimated had been set to 15min; leg time was only 10min. Mission stopped and restarted with configurable time set to 5min.
397_1	4	Main lifting lines became loose, could have jammed motor.
398_1	8	Operators ended mission prematurely, they believed the AUV was missing waypoints. In fact, a couple of waypoints had been positioned incorrectly.
401_1	7.5	Configuration mistake; DVL up configured as down- looking DVL causing navigation problems through tracking sea surface as reference. This data was very noisy and put vehicle navigation out by a factor of 1.5.
401_2	7.5	Damaged on recovery, "moderately serious" to sternplane, shaft bent.
402_1	274	Stern Plane stuck up during attempt to dive, 2d 20h into mission. Stern plane actuator had flooded.
402_2	274	Abort due to network failure. Abort release could not communicate with depth control node for 403s. Possibly side-effect of actuator or motor problems.
402_3	274	Motor windings had resistance of 330 ohm to case. Propeller speed dropping off gradually during a dive.
402_4	274	Only one position fix from tail mounted ARGOS satellite telemetry transmitter.
402_5	274	GPS antenna damaged on recovery.
403_1	140	Recovery light line was wrapped around the propeller on surface. Flaps covering the main recovery lines (and where the light line was towed) were open.
403_2	140	Took over 1 hour to get GPS fix at final waypoint.
403_3	140	Propeller speed showed same problem as m402. Subsequent testing of motor with Megger showed resistance of a few kohm between windings.
404_1	75	Pre-launch, abort weight could not be loaded successfully due to distorted keeper. "If not spotted, could have dropped out during mission", considered low probability of distortion and not checked.
404_2	75	Pre-launch, potential short circuit in motor controller that could stop motor.
404_3	75	Propeller speed showed same problem as on m402 and 403.
404_4	75	CTD (scientific instrument) drop-out of 1 hour (shorter drop-outs noted in previous missions).
404__5	75	M404 recovery was complicated when lifting lines and streaming line became trapped on the rudder (probably stuck on the Bowline knot where the two were attached). Recovery from the situation required the trapped lifting lines grappled astern of the ship, attached to the gantry lines, and the caught end cut.
404_6	75	The forward sternplane was lost due to lifting line trapping between the fin and its flap on recovery.
404_7	75	The acoustic telemetry nose transducer was damaged due to collision with the ship.

Cont.

Fault code	Distance	Description
405_1	2.5	Fault found pre-launch, LXT acoustic tracking transducer had leaked water – replaced.
405_2	2.5	Fault found pre-launch, starboard lower rudder and sternplane loose.
406_1	104	AUV ran slower than expected and speed dropped off during mission, due to motor problem.
406_2	104	Current spikes of 3A and voltage drops in first part of mission.
405_1	2.5	Fault found pre-launch, LXT acoustic tracking transducer had leaked water – replaced.
406_3	104	Propulsion motor failed 500V insulation test on recovery on windings to case.
406_4	104	One battery pack out of four showed intermittent connection
406_5	104	Acoustic telemetry unit gave no replies.
406_6	104	On surfacing first GPS fix was 1.2km out.
406_7	104	Spikes in indicated motor rpm
407_1	204	Acoustic telemetry unit gave no replies at all – no tracking or telemetry.
407_2	204	Noise spikes on both channels of turbulence probe data.
408_1	302.5	Propulsion motor felt rough when turned by hand – bearings replaced before deployment.
408_2	302.5	Aborted at 50m due to overdepth as no depth mode commanded. Unless compounded by another problem, this would show itself immediately on first dive.
408_3	302.5	No telemetry from Acoustic telemetry unit.
408_4	302.5	Difficulty stopping Autosub on surface via radio command. Separate problems with the two WiFi access points.
408_5	302.5	Still spikes on motor rpm that need investigating.
409_1	1.5	No acoustic telemetry or transponding. LXT acoustic ship side ultra short baseline navigation receiver had leaked during mission giving poor bearings to sub, replaced with spare.
410_1	9	No acoustic telemetry or transponding.
411_1	128	No GPS fix at the end of the mission. GPS antenna bulkhead had water inside and had flooded.
412_1	270	No GPS fix at end of mission. After next mission, GPS fixes started coming in after vehicle power up/power down; perhaps problem was due to initialisation with receiver – and not this time the antenna.
412_2	270	Problem at start for holding pattern. Holding pattern timed out due to programming mistake.
415_1	6	Prior to dive, checks showed reduced torque on rudder actuator. Actuator replaced with new one - first use for this new design of actuator motor and gearbox. However, AUV spent most of mission "stuck" going around in circles at depth due to rudder actuator fault. The new actuator overheated, melting wires internally, the motor seized, and internal to the main pressure case, the power filter overheated. Some of the damage may have been caused by an excessive current limit (3A); correct setting was 0.3A. But this does not explain high motor current. Possible damage during testing when motor stalled on end stop? Compounded by wiring to motor held tightly to case with cable ties, and worse, covered with tape (acting as an insulator). Wires were not high temperature rated.
415_2	6	Three harness connectors failed due to leakage, affecting payload systems: EM2000 swath echo sounder tube, acoustic Doppler current profiler (ADCP) down, and Seabird CTD. Despite connector problems the system worked without glitches and failed only when the power pins had burned completely through on the connector feeding power to the electronics.
415_3	6	Although it worked properly at the start of the mission at a range of 1200m, the acoustic telemetry stopped working at the end of mission. Hence could not stop the mission acoustically when needed.
416_1	18	Not possible to communicate with vehicle at 1180m depth; holding pattern caused a timeout, and AUV surfaced. Acoustic telemetry max range was 500m for digital data.
418_1	15	When homing was stopped deliberately after 10 min, the AUV did not go into a "stay here" mode. Rather it continued on the same heading; stopped by acoustic command 500m from shore. Cause was incorrect configuration of mission exception for homing. Default in campaign configuration script was not set due to inexperience with new configuration tools.

APPENDIX II – Extended version of the Kaplan Meier Estimator

This section provides the rationale for the extended version of the Kaplan-Meier survival estimator and its associated variance.

Assumption 1. *The system is repairable. After the repair of a fault, the system is left in the state that it was before the fault.*

In its usual form, the KM nonparametric estimator of *S(r)* – the survivor function with range *r* is defined as:

$$\hat{S}(r) = \prod_{r_i < r} \frac{n_i - d_i}{n_i} \tag{1}$$

where n_i is the number (of missions) at risk immediately prior to range r_i and d_i is the number of deaths at range r_i. Instead of the instance *i* signifying a death, let it signify a failure. Consider the ith fault. With probability p_i this is fatal and

$$\hat{S}(r_i) = \frac{n_i - 1}{n_i} \hat{S}(r_{i-1}) \tag{2}$$

and with probability $(1-p_i)$ the failure is not fatal and

$$\hat{S}(r_i) = \hat{S}(r_{i-1}) \tag{3}$$

Taking expectations, the survivor function becomes

$$\hat{S}(r_i) = p_i \frac{n_i - 1}{n_i} \hat{S}(r_{i-1}) + (1 - p_i)\hat{S}(r_{i-1}) \tag{4}$$

Rearranging terms and simplifying, this becomes

$$\hat{S}(r_i) = (\frac{n_i - p_i}{n_i})\hat{S}(r_{i-1}) \tag{5}$$

Applying this relationship recursively, gives

$$\hat{S}(r_i) = \prod_{r_i < r} \frac{n_i - p_i}{n_i} \tag{6}$$

31

It has been assumed that no two failures can occur at exactly the same range. If this is not the case, then the combined p_i for (6) from the multiple failures can be obtained from

$$p_i = 1 - \prod_{n=1}^{m}(1 - p_n)$$

(7)

where there are m failures at range i each with a probability of p_n of leading to death.

The variance for the original Kaplan-Meier estimator is typically computed using the "exponential" Greenwood formula[21]:

$$\hat{V} = \frac{1}{(\log\hat{S}(r))^2} \sum_{ri<r} \frac{d_i}{n_i(n_i - d_i)}$$

(8)

For the case where no two failures occur at exactly the same range, noting that (6) is (1) with p_i substituted for d_i, for the probabilistic KM, the variance may be written:

$$\hat{V}_{KM} = \frac{1}{(\log\hat{S}(r))^2} \sum_{ri<r} \frac{p_i}{n_i(n_i - p_i)}$$

(9)

The variance of the extended Kaplan-Meier estimator must also take into account the variance in the expert judgments. The combined variance is:

$$\hat{V}(KM + EEJ) = \hat{V}_{KM} + \hat{V}_{EEJ} + 2 \cdot Cov(\hat{V}_{KM}, \hat{V}_{EEJ}) \quad (10)$$

Assuming that the expert judgments are aggregated using the linear opinion pool:

$$\hat{V}_{EEJ_k} = \frac{1}{n} \sum_{j=1}^{j=n}(p_j - \mu_k) \quad (11)$$

Where k stands for the fault index, n is the number of experts that provided judgments for fault k and μ is the un-weighted average of the probability judgments assigned to fault k. The combined variance must take into account that multiple faults may occur at the same range. Thus for a given range and assuming independence between faults that have occurred at the same range:

$$\hat{V}_{EEJ_r} = \sum_{k=1}^{m} \hat{V}_{EEJ_k} \quad (11)$$

The variance for the extended Kaplan-Meier formulation is:

32

$$\hat{V} = \frac{1}{(\log\hat{S}(r))^2} \sum_{r_i < r} \frac{p_i}{n_i(n_i - p_i)} + \hat{V}_{EEJr} + 2 \cdot \frac{1}{(\log\hat{S}(r))^2} \sum_{r_i < r} \frac{p_i}{n_i(n_i - p_i)} \cdot \hat{V}_{EEJr} \quad (12)$$

with asymmetric confidence intervals bound between 0 and 1:

$$\exp(-\exp(c_+(r))) < \hat{S}(r) < \exp(-\exp(c_-(r))) \quad (13)$$

where:

$$c_{\pm}(r) = \log(-\log\hat{S}(r)) \pm z_{\alpha/2}\sqrt{\hat{V}} \quad (14)$$

REFERENCES

1. Adams, S., Lewis, R.S., Bose, N., Opderbeke, J. and Bach-mayer, R., 2008. Preparing the MUN explorer for sea ice missions. In, Proceedings of IEEE AUV2008 Workshop on Polar AUVs [CDROM]. Richardson TX, USA, IEEE.

2. Vaugham, D.G., 2007. Ice/Ocean interactions: Urgent questions for AUVs. In Collins, K.J. and Griffiths, G. (eds.) 2008. Proceedings of the International Workshop on Autonomous Underwater Vehicle Science in Extreme Environments held at the Scott Research Institute, Cambridge, 11-13 April 2007. London: Society for Underwater Technology, 2002pp.

3. Dowdeswell, J.A.; Evans, J.; et al., 2008. Autonomous under-water vehicles (AUVs) and investigations of the ice-ocean interface in Antarctic and Arctic waters. Journal of Glaciology, 54, 187, 661-672.

4. Griffiths G (ed). The technology and applications of autonomous underwater vehicles. London: Taylor and Francis, 2003.

5. Banks, J.C., Brandon, M., Garthwaite, P.H., 2006. Measurement of sea-ice draft using upward-looking ADCP on an autonomous underwater vehicle. Annals of Glaciology, 44, pp.211-216.

6. Bellingham, J.G., Rajan, K., 2007. Robotics in Remote and Hostile Environments. Science (318), pp.1098-1102.

7. Nicholls, K.W., et al, 2006. Measurements beneath an Antarctic ice shelf using an autonomous underwater vehicle. Geophysical Research Letters (33), L08612.

8. McPhail S. Autosub Operations in the Arctic and the Antarctic. Chapter 3 in Griffiths G, Collins KJ (eds). Masterclass in AUV technology for polar science. London: Society for Underwater Technology, 2007.

9. Levine ER, Lueck RG. Turbulence measurement from an autonomous underwater vehicle. Journal of Atmospheric and Oceanic Technology, 1999; 16:1533–1544.

10. Eriksen CC, Osse, TJ, Light, RD, Wen, T, Lehman, TW, Sabin, PL, Ballard, JW Chiodi, AM. Seaglider: a long-range autonomous underwater vehicle for oceanographic research. Journal of Oceanic Engineering, 2003; 26: 424–436.

11. Griffiths, G. and Trembanis, A., 2007. Towards a Risk Management Process for Autonomous Underwater Vehicles. In Griffiths, G. and K.J. Collins (eds.), 2007. Proceedings of the Masterclass in AUV Technology for Polar Science at the National Oceanography Centre, Southampton, 28-30 March 2006. London: Society for Underwater Technology, 146 pp.

12. Kaplan, S., and Garrick, B.J., 1981. On the Quantitative Definition of Risk. Risk Analysis, 1(1), pp. 11-27.

13. Kaplan, S., 1997. The words of risk analysis. Risk analysis, 17 (4), pp.407-417.

14. Hamlin, T.L., Canga, M. A., Boyer, R.L., and Thigpen, E.B., 2008. 2009 Space Shuttle Probabilistic Risk Assessment Overview. PSAM 8: Proceedings of the Eighth International Conference on Probabilistic Safety Assessment and Management New Orleans, Louisiana 14-19 May 2006. [Online] available at:

http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20100005659_2010007106.pdf [Assessed 24 March 2010]

15. Ferguson, J., 2008. Adapting AUVs for Use in Under-Ice Scientific Missions. In MTS/IEEE Proceedings of OCEANS 2008, "Oceans, Poles and Climates: Technological Challenges", Quebec City, Canada, 15-18 September 2008.

16. Griffiths G, Trembanis A. Eliciting expert judgment for the probability of AUV loss in contrasting operational environments. Proceedings 15th International Symposium on Unmanned Untethered Submersible Technology. Lee, NH: AUSI.

17. Griffiths G, Millard NW, McPhail SD, Stevenson P, Challenor PG. On the reliability of the Autosub autonomous underwater vehicle. Underwater Technology, 2003; 25:175-184.

18. Griffiths G, Bose N, Ferguson J, Blidberg R. Insurance for autonomous underwater vehicles. Underwater Technology, 2007; 27:43–48.

19. Podder K, Sibenac M, Thomas H, Kirkwood W, Bellingham J. Reliability growth of autonomous underwater vehicle – Dorado. Pp. 856-862 in Proceedings Oceans 2004, Kobe, Japan. Piscataway: MTS/IEEE.

20. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 1958; 53(282):457-481.

21. Prentice, R.L., Kalbfleisch, J. D. The Statistical Analysis of Failure Time Data. Wiley: New York, 2002.

22. Finucane ML, Alhakami A, Slovic P, Johnson SM. The effect of heuristic in judgments of risk and benefits. Journal of Behavioral Decision Making, 2000; 13:1-17.

23. Kahnemann D, Tversky A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972; 3:430-454.

24. Tversky A, Kahnemann D. Judgment under uncertainty: Heuristics and biases. Science, 1974; 185:1124-1131.

25. Otway H, Winterfeldt D. Expert judgment in risk analysis and management: process, context, and pitfalls. Risk Analysis, 1992; 12:83–93.

26. Brito MP, Griffiths G, Trembranis A. Eliciting expert judgment on the probability of loss of an AUV operating in four environments. Southampton, UK: National Oceanography Centre Southampton Research and Consultancy Report 48, 2008. [Online] Available at:http://eprints.soton.ac.uk/54881/. [Accessed on 30 October 2008].

27. O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T. Uncertain judgments: Eliciting experts' probabilities. Chichester, UK: John Wiley, 2003.

28. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. Risk Analysis, 1999; 18:463-469.

29. Mosleh, A., Bier, V.M., and Apostolakis, G., 1988. A critique of current practice for the use of expert opinions in probabilistic risk assessments. Reliability Engineering and System Safety, 20, pp.63-85.

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

31. Dalkey N, Brown B, Cochran S, 1969. The Delphi method III: Use of self ratings to improve group estimates. Santa Monica, CA: United States Air Force Project RAND, RM-6115-PR .

32. McPhail, S.D., Furlong, M.E., Pebody, M., Perrett, J.R., Stevenson, P., Webb, A., White, D., 2009. Exploring beneath the PIG Ice Shelf with the Autosub3 AUV. In Proceedings of Oceans09. August 2009, Bremen Germany.

33. NBP09-01, 2009. Pine Island Cruise report. Unpublished.

34. Strutt, J.E., 2006. Report of the inquiry into the loss of Autosub2 under the Fimbulisen. National Oceanography Centre Southampton Research and Consultancy Report 12, pp.38. [Online]. Available at: http://eprints.soton.ac.uk/41098/ [Assessed on 15 March 2010]

35. Cooke, R.M., and Goossens, L.H.J., 2004. Expert judgement elicitation for risk Assessments on Critical Infrastructures. Journal of Risk Research, 7(6), pp. 643-656.

36. Aspinall, W., 2008. Expert Judgment Elicitation using Classical Model and Exaclibur. Briefing notes for the Seventh Session of the Statistics and Risk Assessment Section's International Expert Advisory Group on Risk Modelling: Iteractive risk Assessment Processes for Policy Development Under Conditions of Uncertainty/Emerging Infectious Diseases: Round

IV. [Online] Available at:

http://ssor.twi.tudelft.nl/~risk/extrafiles/EJcourse/Sheets/Aspinall%20Briefing%20Notes.pdf.

[Assessed on 26 March 2008]

37. Brito, M., and Griffiths, G., 2010. Analysis of Robustness of Statistical Survival Estimates via Multiple Objective Optimization. In proceedings of 10th International Probabilistic Safety Assessment & Management Conference (PSAM10), Seattle, 7-11 of June 2010.

---

[i] It is a heuristic in which people start with an initial estimate, also denoted as anchor, and then adjust up or down. It is a mental shortcut that can ease the assessment of repetitive situations; it can reduce mental processing time. However in some occasions people tend to stick too closely to their initial judgment or not to adjust their judgment sufficiently.

[ii] As specified by the principal investigator, Dr Adrian Jenkins (British Antarctic Survey).