

Instruments and Methods

Exploring Antarctic subglacial lakes with scientific probes: a formal probabilistic approach for operational risk management

M.P. BRITO, G. GRIFFITHS, M. MOWLEM

*National Oceanography Centre, University of Southampton, Southampton, UK
E-mail: mario.brito@noc.ac.uk*

ABSTRACT. Since their discovery, Antarctic subglacial lakes have become of great interest to the science community. It is hypothesized that they may hold unique forms of biological life and that they hold detailed sedimentary records of past climate change. According to the latest inventory, a total of 387 subglacial lakes have been identified in Antarctica (Wright and Siegert, 2011). However, exploration using scientific probes has yet to be performed. We propose a generic, formal approach to manage the operational risk of deploying probes during clean access to subglacial lake exploration. A representation of the entire probe deployment process is captured in a Markov chain. The transition from one state to the next depends on several factors, including reliability of components and processes. We use fault trees to quantify the probability of failure of the complex processes that must take place to facilitate the transition from one state to another. Therefore, the formal framework consists of integrating a Markov chain, fault trees, component and subsystem reliability data and expert judgment. To illustrate its application we describe how the approach can be used to address a series of what-if scenarios, using the intended Ellsworth Subglacial Lake probe deployment as a case study.

INTRODUCTION

The reasons why Antarctic subglacial lakes exist are now well understood. In essence, their existence stems from a combination of three factors: (1) there is a reduction of the ice melting temperature at the base, caused by the weight of the ice above; (2) the existence of an ice sheet acts as an insulator, protecting the base from the ultra-cold temperatures at the surface; and (3) the heat present under the ice can cause the ice to melt (Siegert and others, 2007). Studies of accreted ice in the lower part of the Vostok Subglacial Lake borehole have provided insights into the geomicrobiology (Prisco and others, 1999) and lake water physical, chemical and biological processes (Siegert and others, 2001) without direct access to lake waters. Several major research projects are now at the stage of planning for access to subglacial lakes with measurement and sampling instruments, namely at Ellsworth (Mowlem and others, 2011), Vostok (Lukin and Bulat, 2011) and Whillans lakes (Prisco and others, 2010; Fricker and others, 2011).

Both the Scientific Committee on Antarctic Research (http://www.scar.org/researchgroups/lifescience/SCAR_CoC_SAEs.pdf), via a sub-committee denoted as SALE (Subglacial Antarctic Lake Environments), and the US National Research Council Committee on Principles of Environmental Stewardship for the Exploration and Study of Subglacial Environments (http://www.nap.edu/catalog.php?record_id=11886) recommend that any probe deployment in a subglacial lake should not contaminate the ecosystem. Clean access is also a scientific requirement, as it will ensure the integrity of the observations and the validity of the conclusions drawn. However, clean access presents huge engineering challenges in these environments, which are likely to be met only if two conditions are fulfilled: the borehole is drilled using a clean drilling method; and the probe and all components that it interacts with are fully

cleaned prior to access to the lake. It is not surprising that accessing subglacial Antarctic lakes with science probes is a multi-risk engineering project, where technical or operational failure may be catastrophic to the campaign and project. The probe may fail to operate, it may be lost or it may contaminate the lake. In addition, delay in the deployment may incur the risk that the hole may have partially closed by the time the probe is ready to deploy.

It is therefore imperative to define a framework that can be used to quantify the probe operational risk, which includes its availability and reliability. Such a framework is needed to allow scientists and engineers to make risk-informed decisions prior to and during the deployment. While such frameworks have been used in other domains (e.g. in infection control by Nicas and Sun, 2006), there is a dearth of peer-reviewed literature on the reliability of instruments and methods in glaciology. The focus has been on the reliability of scientific measurements, for example the reliability of ice-coring analysis (Alley, 2010).

In this paper, we present a formal approach for estimating the risk of a science probe deployment into a subglacial lake. The method is particularly relevant for deployments where the borehole cannot be permanently maintained and thus the availability and reliability requirements for the probe must be high. The motivating setting for this work has been the plan to deploy a probe into Ellsworth Subglacial Lake (ESL) through a hole made by a hot-water drill, in 2013 (Siegert, 2005). Here we show how the three components of a formal approach can be applied to assess the likely success of the probe deployment. Risks with drilling the access hole are discussed elsewhere (Brito and others, in press); while these are not insignificant, hot-water drilling has been applied successfully in a number of projects (e.g. Engelhardt and others, 2000) and the issues are known through field experience. No such field experience exists for a clean probe to access a subglacial lake.

RISK ESTIMATION PROCESS FOR PROBE DEPLOYMENT

This section presents the formal approach for estimating scientific probe risk for subglacial lake exploration. The method combines three techniques: first, modelling the deployment sequence using a state diagram based on Markov chains (Feller, 1950); second, estimation of the transition probabilities between states using fault tree analysis (FTA); and finally, estimation of the probability of failure of individual failure modes, using expert judgment.

It is important to define risk in the context of this work from the several interpretations available. The most frequently referenced interpretation is the definition introduced by Kaplan and Garrick (1981), where risk is defined by three parameters, scenario S_i , likelihood L_i and consequence C_i . Here the interpretation of risk changes, depending on whether availability or reliability is being addressed. When discussing availability (the probability that the equipment will be ready in an appropriate phase when required), risk is defined by scenario S_i , which is the current state of deployment, and the likelihood L_i , which is the probability of a single transition or a set of transitions taking place. When discussing reliability, risk is defined by scenario S_i , here the failure scenario, and likelihood L_i , which is the probability of failure due to scenario S_i .

Markov chain representation of the probe deployment

The probe deployment consists of a sequence of phases, with several risks foreseen at each phase. Markov chain theory is useful for modelling a problem where there is a sequence of events and the probability of an event taking place will only depend on the outcome of the preceding trial (Feller, 1950). Given that a transition probability, $p_{j,k}$, associated with a pair of events (E_j , E_k) is known, where E_k is the event of interest and E_j is the event that took place in the previous trial, and given that the probability of being in the initial state E_j is also known, $a_{j,0}$, the probability of going from initial state E_j to the next state E_k is computed:

$$P\{(E_j, E_k)\} = a_{j,0}p_{j,k}.$$

For the general case, considering a sequence of many transitions, given that event $E_{j,0}$ precedes $E_{j,1}$ which precedes $E_{j,2}$ and so on for the remaining events, the joint probability distribution is computed using

$$P_r(E_{j,0}, E_{j,1}, \dots, E_{j,n}) = a_{j,0}p_{j,1}p_{j,2}p_{j,3} \dots p_{j,n-1}p_{j,n}.$$

Combining estimates of failure modes

Fault tree analysis is a well-established method for establishing the reliability characteristics of systems. It is a method in which the probability of an event is estimated using Boolean logic to combine probabilities of occurrence of lower-level events. The low-level events may, among others, be individual component failures, human error or design faults (O'Connor, 1995). FTA allows ranking of failure modes and identification of the minimal cut-set; this is the shortest credible way through the tree from fault to initiating event.

Estimating probability of failure

In engineering, typically, component reliability predictions are made based on data abstracted from databanks or collected from results of specific reliability testing (e.g. USDOD and Rome Laboratories, 1991; Offshore Reliability Data (OREDA) database from the offshore oil and gas

industry at <http://webshop.dnv.com/global/category.asp?c0=2631>). This is only possible if component failure data are either readily available or tests can be performed on many samples in a realistic environment. However, the combination of the polar environment, a borehole hundreds or thousands of metres deep and subglacial lakes presents an extreme environment impossible to recreate realistically in the laboratory. In addition, the instrumentation and support equipment will largely be purpose-made. Hence failure data, as frequency of occurrence from databases or tests, are not available. As a result, reliability prediction must be based on the third technique, that of eliciting experts' judgments (O'Hagan and others, 2006). The use of expert subjective judgments for reliability prediction was first introduced in the nuclear engineering industry in the 1950s; Keeney and von Winterfeldt (1991) provide a review. Since then, the approach has been adopted by many industries, and applied to space exploration systems. More recently it has been introduced for risk assessment and management of autonomous underwater vehicles (AUVs) and their missions, particularly for missions under floating ice shelves (Brito and others, 2010).

ELLSWORTH PROBE DEPLOYMENT PROCESS

The exploration of ESL is a UK-led multidisciplinary project that involves cooperation across 15 institutions (<http://www.geos.ed.ac.uk/research/ellsworth/members.html>). For the probe and associated systems, the key project partners are the UK National Oceanography Centre (NOC), the British Antarctic Survey (BAS) and the University of Edinburgh. Details of the infrastructure and logistics required to support the probe deployment have been described elsewhere (Siegert and others, 2007; Woodward and others, 2010). Reviews of the hot-water drilling system designed and deployed by BAS are given by Makinson (1993, 1994).

The Ellsworth probe (Fig. 1) is 5.17 m long and 0.20 m in diameter, and is expected to weigh ~250 kg. The outer cases are made of titanium. Power will be supplied from generators on the surface as high-voltage direct current through cables in the tether. The probe is divided into five gas-filled pressure cases. The upper pressure case contains the power and communications link to the tether. In the middle of the probe there are three pressure cases, each containing the electronics, couplings and actuator motors for a free flooded sampler system consisting of a set of eight pressure-tolerant water bottles, with a capacity of 100 mL per bottle, a filter stack for the collection of lake water filtrand, and two in situ gear pumps to drive fluid into the samplers. The pressure case at the bottom of the probe contains electronic sensor control systems, sensors and a short gravity core sediment sampler. An upward-pointing acoustic altimeter, high-definition (HD) camera and lights will be mounted on top of the upper pressure case. This system is replicated at the bottom of the probe, looking downwards.

The sediment sampler will capture at least one sample of sediment of 1 cm diameter and 20 cm length. The filtrand samplers will filter 200 L of lake water onto 0.2 μ m filters. All samplers have been purpose-designed at NOC. In situ measurements will be conducted using a series of commercially available sensors. These sensors will measure: (1) pressure (two sensors for redundancy); (2) temperature (two sensors for redundancy); (3) conductivity (two sensors for

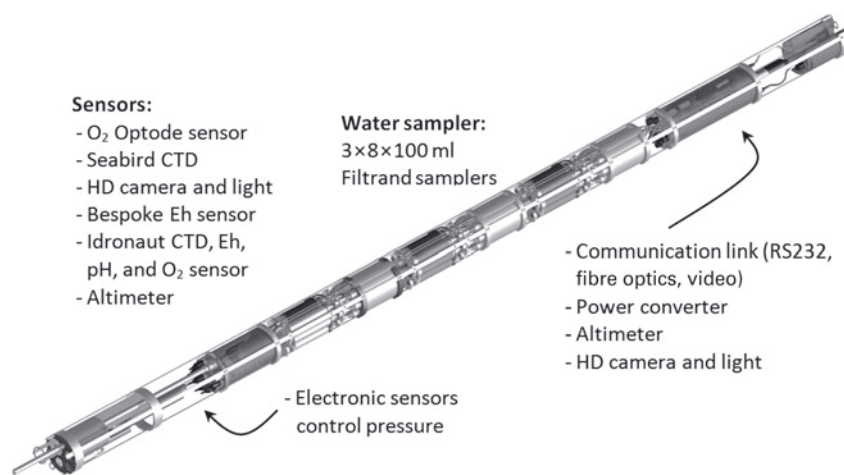


Fig. 1. Three-dimensional computer-aided design (CAD) rendering of the Ellsworth Subglacial Lake probe which consists of two gas-filled pressure cases separated by three carousels of water samplers and filters, all attached to a central core that is attached to the tether.

redundancy); (4) oxygen by Clark Electrode; (5) pH; (6) redox potential (Eh (one bespoke (doi: 10.1029/2009JC005776) and one commercial (Idronaut) sensor)); and (7) oxygen by fluorescence lifetime optode.

During deployment the probe will be connected to the surface control station via the tether. An Aramid polyester fibre tether links the top of the probe, the top sheave (of the gantry) and the winch system. Embedded in the tether are optical fibres and the electrical cables. The outer layer of the tether is made of polyurethane, for mechanical protection, to provide a cleanable surface and to prevent microbial egress from inside the tether. The probe will be lowered at a speed of $\sim 1 \text{ ms}^{-1}$.

MARKOV CHAIN MODEL OF ELLSWORTH PROBE DEPLOYMENT

Each phase of the deployment consists of a series of many activities and these are prone to faults or human error. The Markov chain captures and illustrates the desired sequence

of events for a successful deployment, and also provides paths for any necessary rework if a transition between states is not successful.

The Markov chain that represents the probe deployment is generic; it is not specific to ESL. However, the failure modes in each transition might differ for different probe designs or deployment methods. The Markov chain comprises 13 states (X_1 – X_{13}), corresponding to 13 deployment phases (Fig. 2; Table 1). State names are abbreviated to facilitate model description and subsequent analysis.

The starting point is a built and laboratory-tested probe. The deployment starts at the systems testing phase (S_t), at sea, being a more amenable environment to rectify problems than the Antarctic. All systems are run as if they were performing a real deployment; instruments, video cameras, sediment and water samplers and communications are tested. This will be, in some ways, a science mission; samples will be returned to the laboratory for analysis and the probe's instrument parameters compared with those measured by standard oceanographic instruments. If the

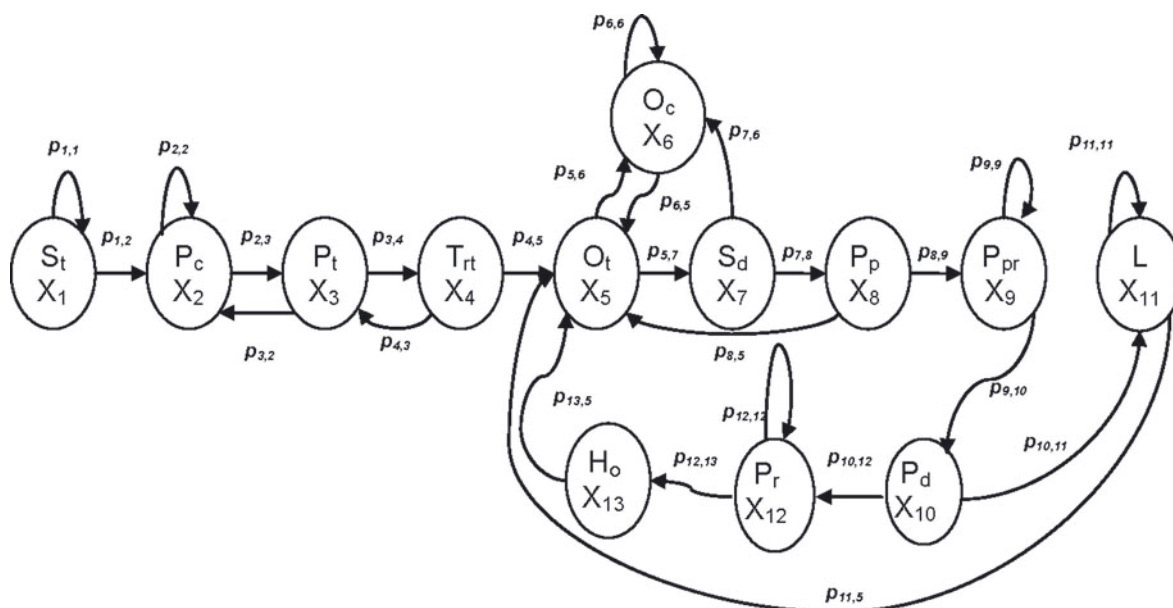


Fig. 2. Markov state space model capturing the sequence of events undertaken during the Ellsworth probe design and deployment.

Table 1. Discrete states from the state diagram presented in Figure 2

State	State description
$X_1 - S_t$	Systems test
$X_2 - P_c$	Post-test cleaning
$X_3 - P_t$	Post-cleaning testing
$X_4 - T_{rt}$	Transportation testing
$X_5 - O_t$	On-site testing
$X_6 - O_c$	On-site cleaning
$X_7 - S_d$	Sheave deployment
$X_8 - P_p$	Probe positioning
$X_9 - P_{pr}$	Probe pre-deployment
$X_{10} - P_d$	Probe deployment
$X_{11} - L$	Probe loss of functionality
$X_{12} - P_r$	Probe recaptured
$X_{13} - H_o$	Probe handed over

probe passes the systems test, it is disassembled and individual components are cleaned before reassembling. Thus, following successful systems testing (S_t) the deployment moves to the state of pre-deployment cleaning (P_c). The probe is cleaned both outside and inside. Components are individually cleaned using wet or dry methods, depending on the material. Components made of titanium (e.g. the outer case) can be autoclaved. This is a robust cleaning process where the component is steamed at 145°C for 40 min. Electronic components are cleaned using hydrogen peroxide vapour (HPV; Mowlem and others, 2011). Once the probe is cleaned and assembled, the system undergoes a pre-deployment functional test (P_t). Only the electrical systems can be tested during this phase. The probe is kept in an environment that has been cleaned to reduce microbial population. If the test is successful, the whole system of probe, tether and winch is transported from the UK to the science camp in two stages.

Neither the final cleaning after probe manufacture, nor any on-site cleaning (see below) will be verified by direct testing. Instead we will pre-qualify our cleaning procedures to ensure that the required microbial population reduction can be guaranteed. (To qualify each process we will purposely contaminate (positive control) test coupons to which the process is applied. This raises the microbiological population above the detection limit of our assessment techniques both before and after our cleaning processes. Thus an accurate measurement of the reduction achieved and the reliability of the method can be made.) This removes the risk that accessing and sampling the equipment itself introduces contamination. Only if a visible or obvious fault (e.g. a tear in a bag, a broken seal, a fault reported by the HPV machine) is encountered will a process be recognizable as having failed in these final stages.

For the first stage, on board a ship to Punta Arenas, Chile, the probe will be in a single container. The container will be isolated and split into cleaned and sealed compartments. The probe will be kept inside a cleaned plastic bag, which in turn will be inside a hard case; all these components will have been cleaned prior to transportation. The case will be insulated and heated. The temperature inside the case will be monitored during transportation. The corers will also be transported in cleaned boxes. The minimum temperature in the box should be -20°C and the maximum 40°C . If the temperature in the box drops below -20°C the probe cannot

be used because this may damage the sensors or sample bottles through freezing of electrolyte in the sensors or water held as a control in one sample bottle in each sampler carousel. Once the system arrives in South America, the system will be tested; this is the post-stage 1 transportation test (T_{rt}). The aim is to ascertain if, during transport, the system suffered unacceptable temperatures or vibrations. If temperatures or vibrations exceeded set limits, the probe would be removed from the case and an electronic test of the system carried out. The second stage of the transportation takes it to the science camp. This will be via a combination of an aircraft flight to the Ellsworth Mountains plus tractor train to the lake site. Once the probe arrives at the science camp, the probe will undergo an on-site test (O_t). This involves connecting to a test lead, and then the tether through the winch and the probe switched on. The electronic status of each sensor is tested. There is a risk that the system may be contaminated during the on-site testing phase (O_t). If so, the probe or the individual component will have to be re-cleaned, and the process moves to the on-site cleaning (O_c) state. The on-site cleaning will consist of using HPV, ultraviolet (UV) radiation treatment and ethanol wipes; it will not be as complete as the pre-deployment cleaning. Some components treated in the laboratory cannot be cleaned in the field.

The next phase is the sheave deployment (S_d). At this time the 36 cm diameter borehole will have been completed, and will refreeze at $\sim 6 \text{ mm h}^{-1}$, giving ~ 26 hours for the probe deployment and recovery. This stage is therefore extremely time-critical. The sheave is kept in a protective environment to avoid contamination, and is integrated with the winch container. When the probe is ready, the sheave is positioned above the borehole by moving the winch container on rails to stops marking the correct position. During this phase the borehole is capped with a gland that is securely fixed to the structure. The gland includes a shut-off valve and a lining through the porous firm ice to enable a degree of blowout protection (Brito and others, in press). If a failure takes place during this phase, the sheave and container are retracted, cleaned and functional tests are conducted, hence the transition $p_{7,6}$ in Table 2.

During the probe positioning (P_p) phase, the probe is craned into place and the bottom of its protective case attached to the well head. The pre-deployment phase (P_{pr}) is arguably the most complex phase of the deployment. In simple terms: the outer protective hard probe case is removed, revealing a soft flexible inner case with valves and airlocks at either end; the valve that gives access to the borehole is opened via an airlock and the probe is partially lowered using the crane into the borehole; the sheave container is connected to the top of the flexible probe case (via a flexible polymer tube and another airlock) and the probe is connected to the tether (using an integral glove box). There are also cleaning processes where all valves, airlocks and the glove box are cleaned, using alcohol wipes, heating to above 0°C and treatment with HPV. UV treatment is used as a reserve method if the HPV machine fails. The probe and tether are then lowered through the gland and the probe deployed (P_d). The probe is lowered to the lake, the samplers are remotely controlled from the surface and so are all other sensors.

After hauling in, the probe is recaptured (P_r) into its protective case and handed over to the science team; this is the handover state in the Markov chain (H_o). A large sediment corer will then be deployed (Siebert and others,

2012) and if the ice hole allows, a second probe may be deployed, going through the same process as the first probe. There is an event associated with each transition probability; a brief summary is presented in Table 2. Above, the successful transitions have been described; unsuccessful transitions loop back to the same state or to a previous state.

ESTIMATING RELIABILITY OF THE PROBE AND PROCESSES

Having established the topology of the Markov chain, the next task is to estimate the values of the transition probabilities. A transition probability of 1 means that there is a 100% chance that the deployment will move from one phase to the next and the probability of failure is 0. The probability of failure $F(t)$ for a transition is a measure of the reliability of the components and processes deployed in that particular phase:

$$R(t) = 1 - F(t),$$

where $F(t)$ is the cumulative failure distribution. For example, the probe pre-deployment phase, X_9 , described earlier, involves a sequence of steps with several possible failure modes, including ‘top valve fails to open’, ‘winch failure caused by human error’ and ‘failure to clean the latex membrane using UV treatment’. The central predicament is how to calculate the probability of a successful transition from state X_9 to X_{10} based on the probabilities of multiple failure modes. Fault tree analysis (FTA), event tree analysis (ETA) and failure mode and effect analysis (FMEA) are techniques commonly used to elicit and structure failure modes that are likely to drive the probability of failure of a system (O’Connor, 1995). We chose FTA because its graphical structure provides the best insight into system functionality.

Fault tree analysis

FTA uses a graphical logical representation of the system functionality to represent fault propagation in complex systems. The graphical representation consists of one end event (rectangle; this is the event for which we are interested in computing the probability of failure), a number of base events (triangles) and a number of logical AND and OR gates. Base events are not dependent on any other events or combination of events. Once the probabilities of failures for all base events have been estimated, the mathematical model encoded by the graphical structure can be used to calculate the probability of failure for the end event, in this example $p_{9,9}$. Figure 3 shows the fault tree for this phase. The transition probability $p_{9,10}$ of the desired outcome is computed using the probability of failing the deployment, as indicated in Figure 2, i.e. $1 - p_{9,9}$.

The base events on the fault tree each capture a potential failure mode. Some failure modes are of mechanical actuation or of a structural nature; some are related to electronics and software systems. Human errors are also captured.

FORMAL EXPERT JUDGMENT ELICITATION

The probability judgments for each failure mode were obtained during eight workshops, of 3.5 hours each, spread over a 2 month period from October to December 2010. Formal judgment elicitation is a technique used to ensure that experts are properly trained in the task of assigning

Table 2. Notation and conditions for the Markov model in Figure 2. $p_{i,j}$ is the transition probability from state i to state j

Notation	Conditions for transition
$p_{1,1}$	Failed systems test
$p_{1,2}$	Passed systems test
$p_{2,2}$	Pre-deployment cleaning failed
$p_{2,3}$	Pre-deployment cleaning passed
$p_{3,2}$	Pre-deployment functional test failed
$p_{3,4}$	Pre-deployment functional test passed
$p_{4,3}$	Transportation test failure
$p_{4,4}$	Failure to transport probe safely to Ellsworth site
$p_{4,5}$	Probe safely transported to Ellsworth site
$p_{5,6}$	Probe contaminated and cleaning is required
$p_{5,7}$	Probe and systems pass on-site test
$p_{6,5}$	Probe successfully cleaned
$p_{6,6}$	On-site cleaning failed
$p_{7,6}$	On-site cleaning due to sheave deployment failure
$p_{7,7}$	Failure to deploy sheave
$p_{7,8}$	Sheave correctly deployed
$p_{8,5}$	Failure to position probe due to systems failure
$p_{8,9}$	Probe is successfully positioned above the well
$p_{9,9}$	Probe pre-deployment fails
$p_{9,10}$	Probe is deployed
$p_{10,11}$	Probe loss of functionality during deployment
$p_{10,12}$	Probe is successfully deployed
$p_{11,5}$	Probe is found and pre-deployment is restarted
$p_{11,11}$	Probe is permanently lost
$p_{12,12}$	Failure to recapture the probe
$p_{12,13}$	Probe handed over to science team
$p_{13,5}$	On-site testing of second probe
$p_{13,13}$	Failure to hand over probe to science team

probability to events. A formal elicitation is needed to ensure that the process is transparent and repeatable and that biases are removed (Keeney and von Winterfeldt, 1991). We adopted the SHEffield ELicitation Framework (SHELF) package (<http://www.tonyohagan.co.uk/shelf/index.html>) because the aim is to elicit a probability distribution rather than a single probability judgment, allowing the capture of uncertainty. In addition, the SHELF method encourages experts to agree on the final judgment, denoted as behavioural aggregation. The debate and intensive questioning seeks to remove any bias introduced by individual judgments. A summary on how the SHELF package was applied to estimate the risk posed by each failure mode is presented below.

Expert selection

There are different schools of thought as to what is the ideal number of experts for a judgment elicitation exercise. At one extreme, attempting to quantify the probability of failure of a nuclear power station safety valve Keeney and von Winterfeldt (1991) elicited judgments from 40 experts. Cooke and Goossens (2000) state, ‘in general the maximum number of experts should be used, but at least four’. Experts should be chosen to obtain a spectrum of views and to cover a range of specialisms. An expert is someone that has great knowledge of the subject matter; however, expertise also involves how the person organizes and uses that knowledge (O’Hagan and others, 2006, p. 27). The SHELF package does not specify a minimum number of experts, but research has shown that a group of experts does not perform better than the best expert in the group (the difficulty being knowing who is the best expert). The group of experts for this task was

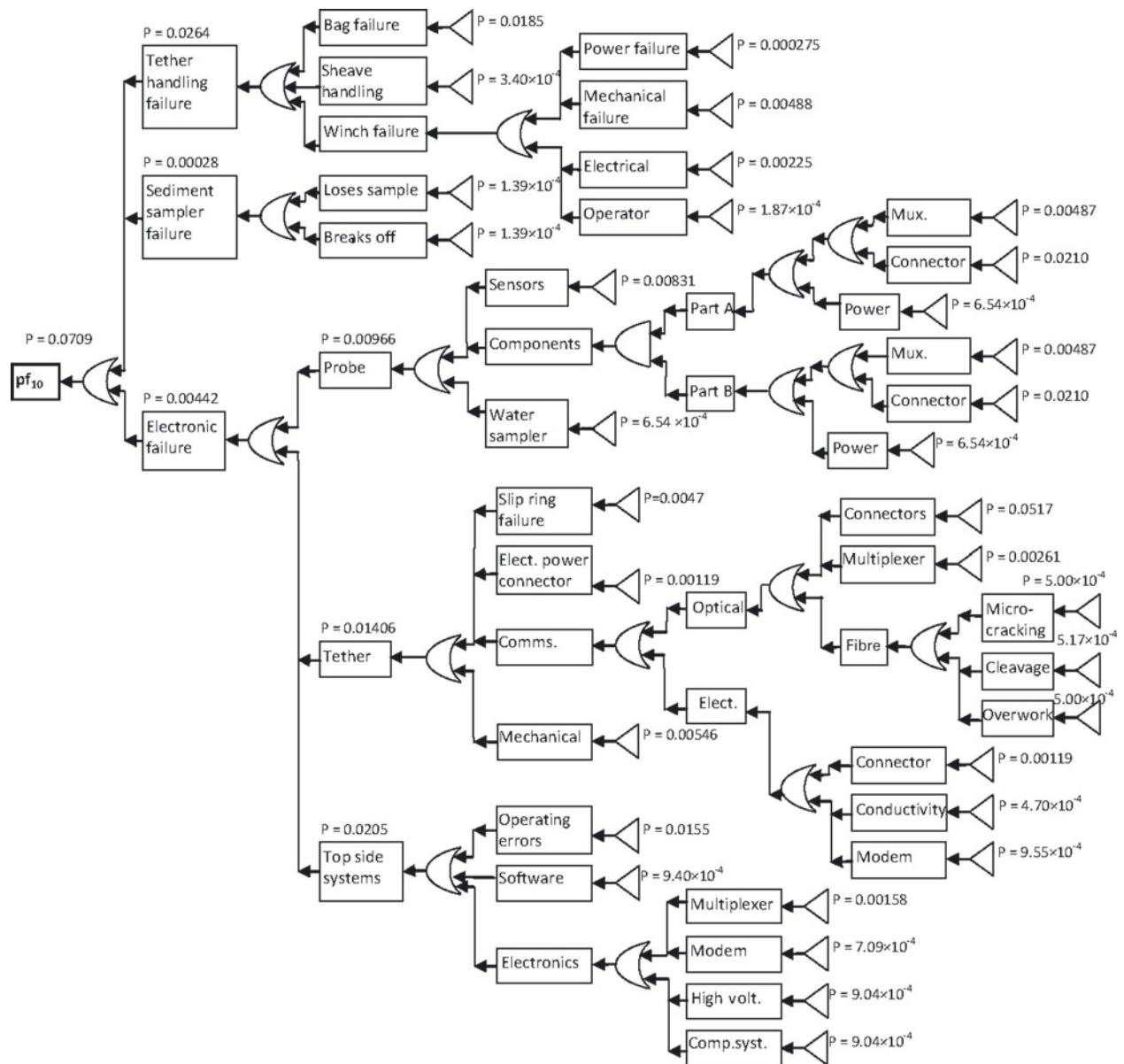


Fig. 3. Fault tree used for deriving transition probability $p_{10,11}$. Thirty-three base events are shown; the probability of failure for each base event is the 95% quantile of the agreed assessment. The probability pf_{10} stands for failure to complete phase 10.

formed by engineers and scientists working on the project. A total of ten experts took part in the judgment elicitation exercise: four mechanical engineers (with a combined 48 years' experience in oceanography); three electronic engineers (with a combined 48 years' experience); one postdoctoral scientist in chemistry (3 years' experience); the principal scientist (10 years' experience); and an external mechanical engineer from BAS (30 years' experience) provided assessments for the crane failure modes.

Background information

Experts were briefed on the purpose of the elicitation process in a 1 day seminar. All experts and project managers took part in this exercise.

Training

Experts were trained on basic probability concepts. The elicitation was conducted by the facilitator (Brito). Experts were briefed on judgment elicitation, some fallacies of assigning probabilities to events (e.g. biases) and how a

formal elicitation process aims to address fallacies (Tversky and Kahneman, 1974; Kynn, 2008). This was followed by an expert judgment elicitation exercise facilitated by one of the experts for three failure modes. Experts were asked to assess optical connector failure, optical multiplexer failure and optical fibre failure.

Eliciting probabilities

Experts were asked to provide a plausible range for the probability of failure $P(X)$. For each failure mode, they were asked to estimate the upper and lower bound of $P(X)$, so that $L < P(X) < U$. During this stage, experts discussed factors pertaining to each failure mode and how, in the worst-case scenario, these factors could increase $P(X)$, and also how, in the best-case scenario, the probability of failure could be reduced. Experts had to agree that $P(X)$ was in this range. Next, experts were asked to draw the shape of a unimodal probability distribution. Based on the shape of the probability distribution, they were asked to individually identify the median and to record evidence that supported

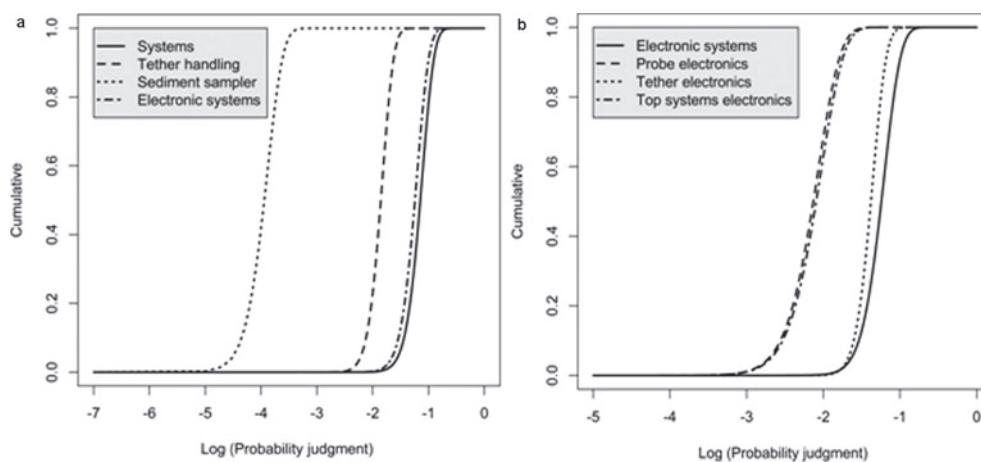


Fig. 4 Probability of failure cumulative distribution for (a) the first-level failure modes in the fault tree of Figure 3 and (b) the second-level failure modes for the probe systems.

their assessment. In the next step the experts were asked to assign the lower (LQ) and upper quartiles (UQ) so that $L < P(X) < LQ = 0.25$ and $LQ < P(X) < M = 0.25$. Similarly, for the upper quartile so that $M < P(X) < UQ = 0.25$ and $UQ < P(X) < U = 0.25$.

Behavioural aggregation

The cumulative distributions for each expert’s judgments were plotted using the median, upper and lower quartile. Experts were then asked to discuss the reasons supporting their assessments, to jointly specify *M*, *LQ* and *UQ*, and to record any disagreements.

SUBGLACIAL LAKE ELLSWORTH PROBE AND DEPLOYMENT PROCESSES RELIABILITY

This section presents and describes the results, phase by phase, of combining the elicitation process with fault trees, focusing on the final probability of failure for each transition.

The assessments for each failure mode were propagated in a fault tree, as in Figure 3, before fitting a beta distribution to the parameters of each probability distribution. As an example, Figure 4 shows the fitted cumulative probability distributions for the first and second levels for the fault tree of Figure 3. They provide a visual method of communicating and comparing the relative importance of the different subsystems to overall reliability. They also show the uncertainty in the expert judgments; the steeper the curves, the less uncertainty over the probability judgment. Fitting a beta distribution allows 95% confidence probability judgments to be estimated, and these are the measures presented here. Thus users of the probe can be 95% confident that the failure probabilities are no greater than those presented. As the fault trees and cumulative frequency plots are voluminous, only this one example is presented here; the full set is presented in an open-access report (available by request from <http://eprints.soton.ac.uk/187347/>).

Systems test (*X*₁)

The aim of this phase is to simulate the probe deployment in the Antarctic with a probe test at sea. All systems will be tested. Experience with testing of AUVs has shown that faults can be discovered at this stage, even though thorough testing

has taken place in the laboratory. The overall elicited probability of failure for this transition was 0.139.

The most critical fault source was ‘probe systems’, which includes probe electronics, tether and topside electronics, at 0.11. Next was tether handling with three failure modes: ‘bag failure’, ‘winch failure’ and ‘sheave handling’. Bag failure was assigned a probability of 0.0185, winch failure 0.0075 and sheave handling 0.00034. The sediment sampler had two failure modes: ‘loses sample’ and ‘breaks off’. The probability of failure for each of these failure modes was <0.000139.

Probe cleaning (*X*₂)

Having passed the systems test, the probe is then cleaned inside and outside. Inside cleaning includes the sensors and inside of the case. Outside cleaning involves cleaning the outer case of the probe and other external components that may contaminate the lake. Two assessments were conducted for this phase of the deployment. The first was at the start of the project, on 24 November 2010, the second on 24 October 2011. The estimated risk of failure to clean the probe at the start of the project was 0.976. This is because at that time the team had not tested the effectiveness of HPV, UV and other potential treatments. There was also significant uncertainty with respect to the type of materials that would comprise different sensors and external systems. The HPV treatment was considered very inefficient while the UV treatment was considered to cause severe material degradation. With this very high probability of failure it was almost certain that the deployment would not leave this state. Consequently, a major review of cleaning was instigated, and changes implemented (Siegert and others, 2012).

The results of these improvements were considered at the second workshop on the cleaning process. The probability of failure of the revised process was assessed as 0.000353. Based on the fault tree for this phase, with 26 failure modes, the probability of failure to clean the inside of the probe was 0.000350, and 0.000108 for failure to clean the outside. Of the failure modes for the inside of the probe, failure to clean the motors and the Optode oxygen sensor were deemed most likely, with a probability of 0.000175. For the remaining failure modes, the probability of failure was in the order of 10⁻⁷. The single biggest factor for the reduction in estimated risk was successful experimental results with

testing of our cleaning procedures. While the reliability and performance for the procedures we use is well documented for ideal and laboratory conditions, our uncertainty arose from lack of knowledge as to how well these could be implemented with our specific materials and structures.

Results of laboratory HPV testing have shown that this technique is more effective than initially anticipated in cleaning the outside components of the probe. (A panel of test coupons was contaminated separately with both *Geobacillus stearothermophilus* and *Pseudomonas fluorescens* and passed through our cleaning process consisting of detergent wash, 70% ethanol wash, rinsing and drying before HPV or UV treatment. The specific sequence and extent of treatment depends on the material and structure. However, we have demonstrated reliable log₆ reduction in population. Log₆ is a standard accepted by the US Environmental Protection Agency. This was assessed using cell staining and fluorescence microscopy.) Five failure modes were identified for the outside systems; these were failure to clean: the hot-water drill hose (this has been cleaned during testing prior to shipment (flushing with hot water) and this verified (no culturable microorganisms detected). It will be further cleaned on-site with alcohol wipes, and is further cleaned during operation as it is flushed with hot water); the ethylene propylene terpolymer o-rings; the titanium case of the probe; the jacket of the probe tether; and the winch. Failure to clean the hot-water drill hose was most likely, with a probability of 6.15×10^{-5} , with failure to clean the ethylene propylene terpolymer o-rings second most likely, with a probability of 4.68×10^{-5} . HPV was considered so effective that the remaining failure modes were assigned a probability of 0.

Post-cleaning electronic test (X_3)

The fault tree for this phase is a subset of the fault tree for the deployment phase (Fig. 3), with the mechanical systems and the sediment sampler not tested. However, both probes are tested in X_3 . The probability of failure for this phase was 0.112. An electronic failure would result in the need to disassemble the probe, correct the faulty component, test and re-clean the component.

Probe transportation (X_4)

Two distinct hazards present a risk to the probe during transportation: vibration and temperature. Probe vibration will be minimized by a damping mechanism that supports the case. However, excessive vibration may break mechanical links, or human error (e.g. poor or incorrect assembly) may occur. Consequently there are three vibration failure modes: (1) excessive vibration, beyond design limits; (2) failure of the damping system; (3) procedural failure.

The probe can also fail due to exposure to excessively low or high temperatures. The temperature inside the container will be controlled by a mechanism designed in-house. Once on-site, heating systems failures could cause the probe to experience $<-20^\circ\text{C}$, causing sensor degradation, potentially reducing the quality of the observations.

The probability of failure during transport was assessed at 0.027. The top failure mode was 'on-site failure to control probe temperature' at 0.0167.

Probe on-site test (X_5)

The on-site test consists of performing a probe electronic test, identical to that carried out during phase X_3 with the

same probability of failure of 0.112. A failure on-site would mean the case(s) having to be opened, the fault found and rectified, followed by on-site cleaning (X_6). Another way of presenting this assessment is that there is an 89% probability that the probe will *not* have to be opened on-site (and subsequently cleaned).

Probe on-site cleaning (X_6)

On-site cleaning may need to take place if either the probe or sheave is dropped, or if the electronic test fails and thus the probe needs dismantling in order to assess and replace faulty components. The environmental conditions and reduced number of staff limit the type of cleaning that can be performed on-site. As a result, only HPV and UV can be applied. The HPV may fail to clean due to an equipment failure or due to a failure of technique, or it may cause material degradation. The two failure modes for UV treatment are failure to clean and failure due to material degradation. The overall probability of failure was 0.0399.

Sheave deployment (X_7)

The sheave is removed from the container and manually carried through a polythene tube into position, the tether pulled over it (in an airlock) partly consisting of a flexible polymer tube. The failure modes identified were: (1) tether handling, (2) design layout, (3) design fault, (4) handling of the sheave and (5) polymer tube failure. The tether handling failure mode was dependent on two other failure modes: (1) design mitigation and (2) procedural mitigation. The probability of failure is <0.124 . On-site, this means that the problems would need to be corrected and the sheave would need to be cleaned (X_6) before attempting a second deployment.

Probe positioning (X_8)

The probability of failure to position the probe is <0.00479 . The first-level failure modes are (1) probe freezes, (2) probe manually dropped, (3) failure to align the probe, (4) crane failure to lift the probe and (5) probe dropped by the crane. The probabilities for the crane-related failure modes were estimated by the crane designer based at BAS at <0.00478 .

Probe pre-deployment (X_9)

The first-level failure modes for this phase are: (1) valves fail to open; (2) tether failure to mate; (3) failure to move the glove box; (4) failure to align either the glove-box flange or the top valve; (5) failure to detach the hard case; (6) crane failure; (7) probe bag damaged; (8) failure to clean either the glove box or valves; (9) winch failure; and (10) failure to release the probe. Some of these failure modes were decomposed into subsidiary modes (e.g. failure to open the valves was decomposed into four, one for each valve). The probability of failure to position the probe pre-deployment is <0.0941 . The probe would be held in this position while remedial action was taken if possible (the $p_{9,9}$ transition), or if the problems were more substantial, a dummy deployment would take place and the path via X_{10} back to on-site testing X_5 would be followed.

Probe deployment (X_{10})

The probe deployment is divided into two stages. First, the probe is lowered to the lake, controlled by the winch driver. Second, operators in the control room send signals to activate all sensors and samplers and monitor the data and

the probe performance. The fault tree is similar to that defined for probe testing phase X_5 . The only difference is that here we consider that only one probe is deployed rather than two. The probability of failure is <0.13 .

Probe recapturing (X_{12})

During probe recapturing, the probe is switched off and lifted to the surface controlled by the winch driver. The probability of failure is <0.0263 .

Probe loss (X_{11})

Probe loss may come as a result of failing to recapture; thus the probability of loss, at 0.0263, is determined by the Markov condition, that the probabilities from a node must equal 1, applied to node X_{12} .

Summary of the expert judgments

Experts assessed the probability of failure for all base events in the fault trees. We had a total of 387 individual expert assessments. For some failure modes, judgments were elicited from five experts. For other failure modes the panel agreed that one expert would best represent the group view of the risk assessment; this expert would normally be the most experienced individual in the topic under consideration. This is the case for, for example, crane failures during probe positioning phase: experts agreed that Steve White (from BAS) would best represent the group's view.

Clearly, experts will not always agree, so it is often necessary to conduct an analysis of the expert judgments. Given the large number of assessments, in this paper we review the top ten most critical failure modes only. These are: (1) tether handling failure (during sheave deployment, state X_7), with a UQ of 0.143; (2) communication optical connector failure (during system testing phase), with a UQ of 0.04; (3) design failure during sheave deployment, with a UQ of 0.04; (4) handling of the sheave (during sheave deployment), with a UQ of 0.0333; (5) on-site failure to control probe temperature during on-site test (state X_5), with a UQ of 0.0133; (6) tether handling failure caused by bag damage during systems pre-deployment test (state X_1), with a UQ of 0.0133; (7) tether electronic connector failure during systems pre-deployment test (state X_1), with a UQ of 0.0133; (8) topside systems failure due to human error during systems pre-deployment test (state X_1), with a probability of 0.01; (9) electronics sensor failure, i.e. at least one sensor fails, during pre-deployment systems tests (state X_1), with a UQ of 0.005; and (10) tether mechanical failure (state X_1), with a UQ of 0.004.

The judgments provided for the probability of tether handling failure ranged between 0.05 and 0.2, showing that the panel was very confident in this assessment. It is not unusual for expert judgments to span over three orders of magnitude (Keeney and von Winterfeldt, 1991). The fitted probability distributions for this failure mode are presented in Figure 5a. Tether handling is an operator procedure. Experts agreed that this procedure would be rehearsed prior to the deployment. Three experts (N.R., K.S. and E.W.) provided the same assessments for the lower quartile, median and upper quartile. Experts J.C. and J.W. were slightly more optimistic, but the differences with experts N.R., K.S. and E.W. are very small. There is no evidence of biases in this assessment.

Three experts (J.C., E.W. and L.F.) assessed the probability of optical connector failure; the fitted probability distributions are presented in Figure 5b. All three experts are

experienced electronics and communications engineers. L.F. is the most experienced in fibre optics, particularly for remote operating vehicle applications (ROVs), where he has experienced this type of failure mode before. J.C. and E.W. provided the most optimistic assessments: both provided assessments in the order of 10^{-3} . L.F. was more pessimistic, his assessments being in the order of 10^{-2} . We believe L.F. used the representativeness heuristics for estimating the probability of optical connector failure. Representativeness heuristics is a mental shortcut for assessing the probability of event A based on the degree of resemblance with event B (Tversky and Kahneman, 1974). Biases can be introduced using this heuristic; however, given that the probe optical connector will be provided by the same supplier and given that the probe will operate in a similar fashion to an ROV, we have no evidence to support the argument that L.F. has introduced biases.

Of the ten most critical failure modes, this is the assessment for which we have observed the largest discrepancy between expert judgments. Following a discussion, the group agreed that L.F.'s assessments would better represent the panel view.

The probability distribution for design failure during sheave deployment is presented in Figure 5c. With a probability range 0.01–0.1 we can see that the difference in expert judgment is small. For this failure mode, disagreements are more evident in the upper quartile. The agreed upper quartile was 0.04; E.W. was the most optimistic expert, assigning an upper quartile of 0.0181; N.R. was the most pessimistic expert, assigning an upper quartile of 0.0667. The agreed UQ presents a good representation of the experts' assessment.

Sheave handling is a human operator activity. The panel agreed that the risk of sheave handling failure should be smaller than the risk of tether handling failure. The fitted probability distributions for this failure mode are presented in Figure 5d. J.W. was the most optimistic expert, with LQ of 0.0125, M of 0.0167 and UQ of 0.0333. J.C. assigned the same values for LQ and M, but a UQ of 0.05. Following a discussion between experts, the panel agreed that assessments provided by J.W. would best represent the panel view, provided that the sheave handling procedure is rehearsed prior to the deployment.

Controlling the probe temperature is key to the success of the deployment, as some sensors can be permanently damaged if the ambient temperature is below -26°C . For example, the Idronaut platform uses a reference that contains a gel which can freeze if the temperature drops below -26°C . The protective system will consist of a case with foam. The panel agreed that the judgments of J.W. and N.R. best represent the group view. Following a brief discussion, both experts assigned the same values to all parameters of the probability of failure distribution. The probability distribution for this failure mode is presented in Figure 5e.

The panel agreed that the failure distribution for tether bag failure should be bounded between 0.005 and 0.02. The fitted probability distributions are presented in Figure 5f. Experts provided similar values for the median and upper quartile. For the lower quartile, N.R. was slightly more pessimistic than the rest of the group, with a LQ of 0.01. The panel agreed that the LQ assessment provided by J.W. and E.W., of 0.0667, would best represent the panel view. We have no reason to believe that there was bias in this assessment.

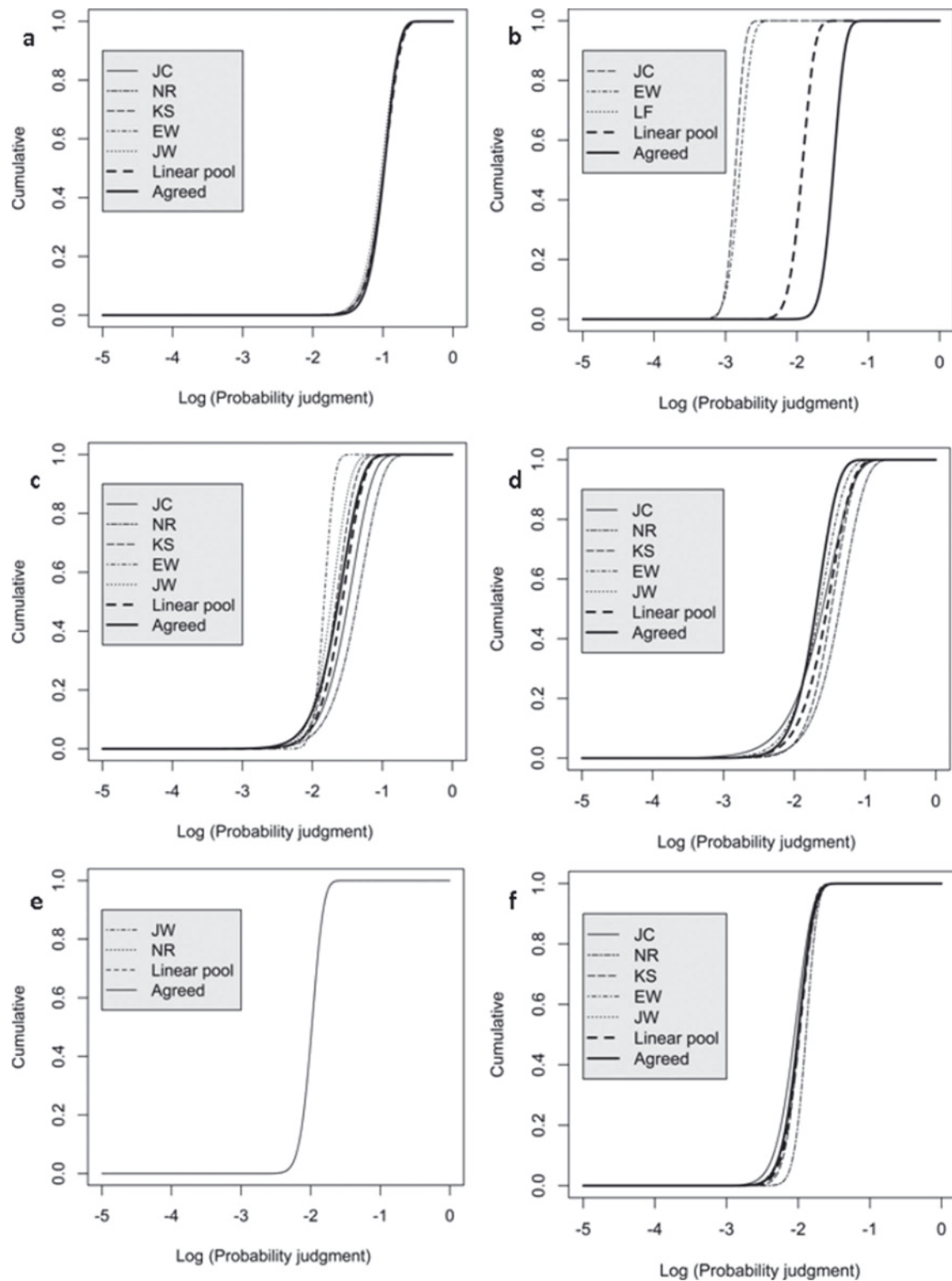


Fig. 5. Expert judgments for the top ten most critical failure modes: (a) tether handling failure; (b) communication optical connector failure; (c) design failure during sheave deployment; (d) handling of the sheave; (e) tether handling failure caused by bag damage during systems pre-deployment test; (f) tether handling failure caused by bag damage during systems pre-deployment test; (g) tether electronic connector failure during systems pre-deployment test; (h) polythene failure during sheave deployment; (i) topside systems failure due to human error during systems pre-deployment tests; (j) electronics sensor failure, i.e. at least one sensor fails, during pre-deployment systems test.

The probability distribution for the tether electronic connector failure is bounded between 0.0002 and 0.002. All experts assigned a median of 0.000667. J.C. was the most optimistic expert, with an assessment of 0.000284 for the LQ and a UQ of 0.0008; E.W. assigned a LQ of 0.0003333 and UQ of 0.001. The panel agreed on a more confident distribution, presented in Figure 5g.

For the topside system operator error failure mode, the panel agreed that the probability distribution is bounded between 0.002 and 0.02. As shown in Figure 5h, the panel agreed on the most pessimistic view presented by experts. The probability distribution for sensor failure is bounded between 0.001 and 0.01. As shown in Figure 5i, the agreed distribution is similar to the unweighted linear opinion pool.

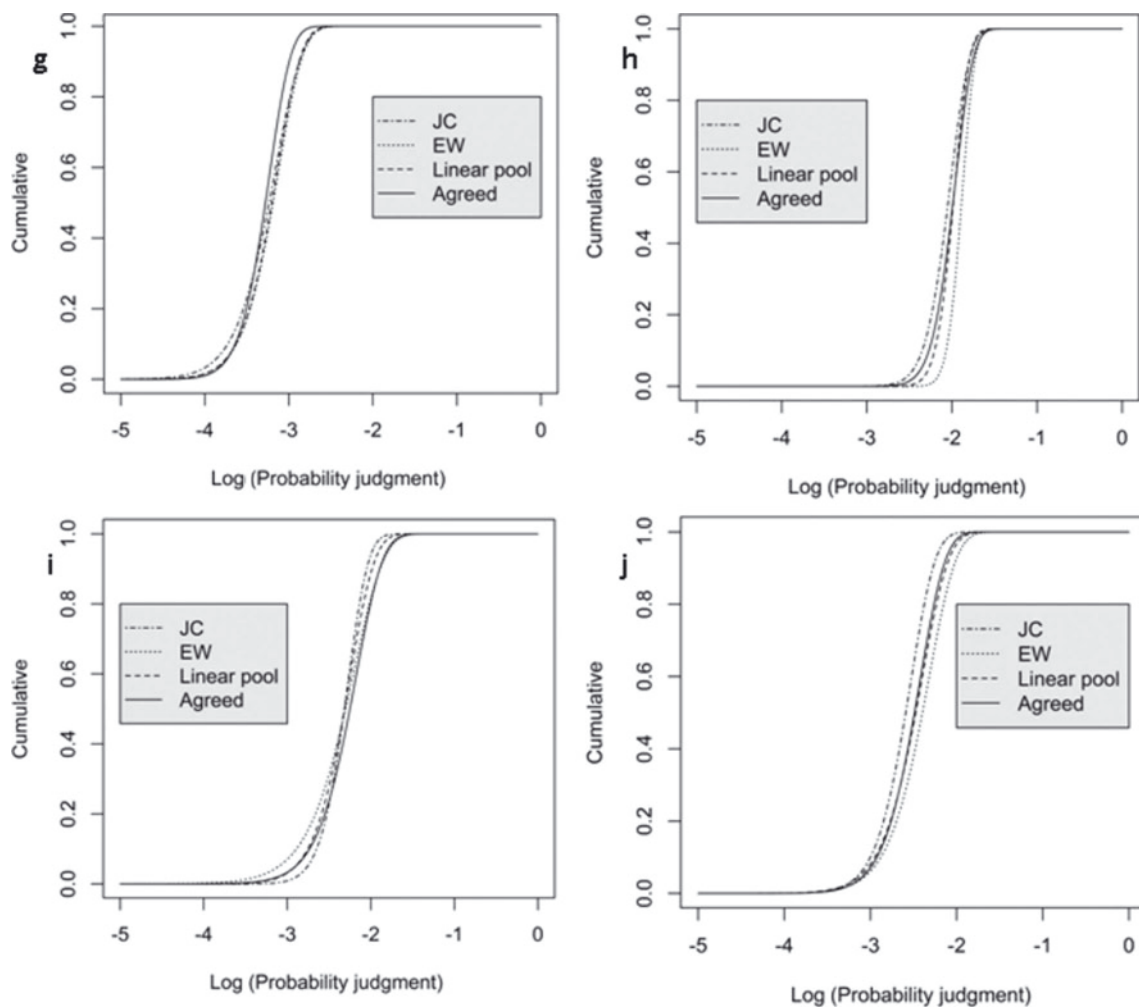


Fig. 5. Continued.

The panel agreed that J.W. would best represent the panel on tether mechanical failure mode. The probability distribution for this failure mode is bounded between 0.001 and 0.005, which represents a very confident assessment.

Expert judgment is naturally a subjective process, but this does not imply that the process is biased, as we have shown in this review of the expert assessments.

ESTIMATING PROBE AVAILABILITY

Having explained how FTA can be used to populate the transitions in the Markov chain of the probe deployment, in this section we describe how the model can be used to perform analyses of availability. First the transition matrix for the Markov chain is specified, based on the transition probabilities as elicited and set out as 95% quantiles in Figure 6. However, in order to reflect the uncertainty in the risk assessment of different phases of the deployment, all parameters of the probability distributions (*L*, *LQ*, *M*, *UQ* and *U*) have been propagated using the Markov formalism and a beta distribution fitted to these parameters to ease data analysis.

Probe availability, without any rework (i.e. all of the steps between the first and the last in the sequences below all go according to plan without any problems), was calculated for the following scenarios:

Probability that the probe will go from being ready for its system test at sea to being cleaned and ready for

transportation to Antarctica: $p_{1,2} \times p_{2,3} \times p_{3,4} = 0.76$. All the remedial work needed if the probe fails during any of these stages would be done before the probe leaves the UK.

Probability that the probe will go from passing its tests in the UK to passing its tests on-site: $p_{4,5} \times p_{5,7} = 0.86$. Remedial work would need to be done on-site, and would be done before the hot-water drilling started.

Probability that the probe will go from having passed its on-site tests to completing the pre-deployment phase: $p_{7,8} \times p_{8,9} = 0.87$. These two phases can be practised, and faults corrected, before the hot-water drilling starts, as the wellhead will be in place above an auger-drilled hole of 15 m length.

Probability that the deployment will go from successful pre-deployment to recovery: $p_{9,10} \times p_{10,12} = 0.79$. This is the time-critical period, as the hole has been drilled and will only be open for ~26 hours.

Probability of successfully deploying and recovering *either* of the two probes, given both passed their on-site testing $(1 - (1 - (p_{9,10} \times p_{10,12}))^2) = 0.96$.

It is nearly impossible to validate these predictions of availability. This would require significant field testing and process rehearsal. However, they can be compared against availability estimates for other systems of comparable

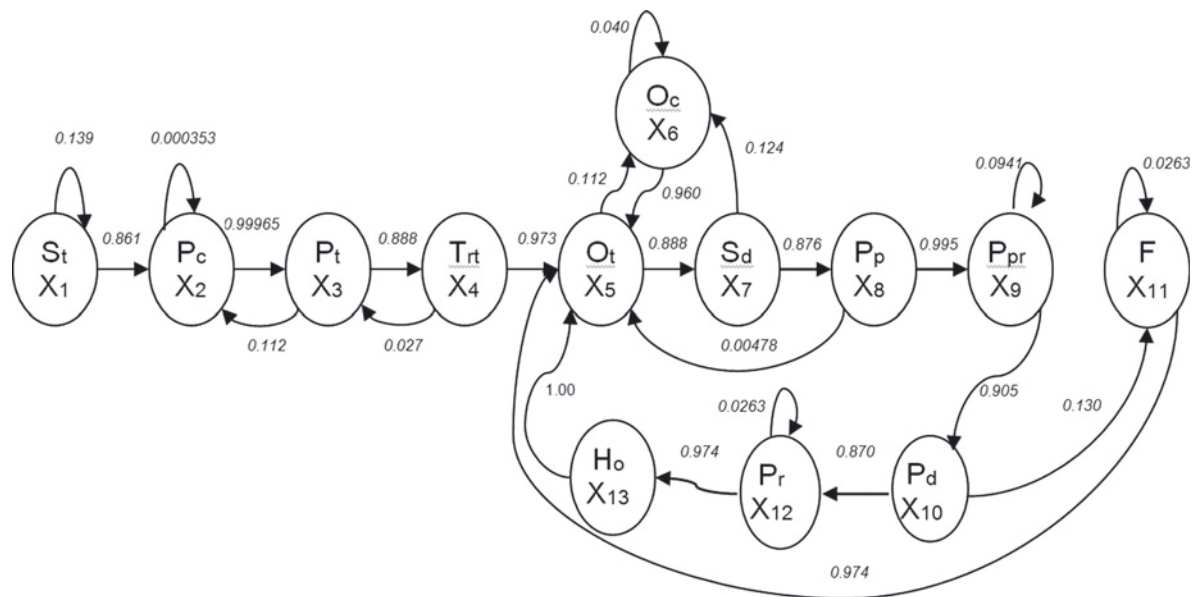


Fig. 6. Markov state space model capturing the sequence of events undertaken during the Ellsworth probe design and deployment.

complexity. Availability figures for five types of unmanned aerial vehicles (UAVs) – Predator RQ-1A (concept demonstrator), Predator RQ-1B (early production), Pioneer RQ-2A (1990–91), Pioneer RQ-2B, Hunter RQ-5 (reliability enhanced 1996–2001) – were published by the US Office of the Secretary of Defense (www.acq.osd.mil/uas/docs/reliabilitystudy.pdf, but the document is no longer available). The average availability for these vehicles was 0.77. For the Autosub3 AUV, the estimated availability is 0.75 (Brito and Griffiths, 2011). For one Ellsworth probe, the availability for deployment during the critical period after the hole has been drilled is 0.79. Given that the probe is a totally new development, the availability figure is encouraging compared with the UAV and AUVs. However, it is sufficiently low that the need for a second probe to be available is clear, both from good practice and from the elicitation and statistical propagation arguments presented here.

CONCLUSION

Clean access to subglacial lakes is now a requirement for Antarctic science expeditions. Given the engineering and environment challenges, it is imperative that scientists adopt a formal approach for estimating the likelihood of success. The approach described for estimating the availability of a scientific probe, based on, among other factors, the effectiveness of the cleaning processes, components' and operator reliability, provides both a means of assessment and a means of communication of the risks involved.

While the three separate techniques – Markov chain, fault analysis supported by fault trees, and expert judgment elicitation – are well known, their integration in the manner described here is novel. The major limitation is the lack of verification for the quantitative results of the process. Much rests on the elicited expert judgments for a whole series of activities that have yet to take place. However, after deployment, drawing upon the experience gained, exactly the same structure of the Markov chains and fault trees as set out here can be used with newly elicited judgments to revisit the predictions. Experience may also show that other paths

in the Markov chain need to be added or existing paths removed; new states may be needed. Consequently this first model would be refined, and would gain in realism, while keeping the underlying formalism.

Perhaps more important than the initial predictions of availability are the discussions at the workshops held to determine and then populate the fault trees with failure probabilities. The first round of application of the fault tree analyses helped engineers to identify design vulnerabilities. This led to major changes with respect to the deployment and cleaning processes. For example, the sheave deployment was deemed a high-risk operation; following this assessment the design team changed the sheave deployment process, so that the sheave is deployed together with the probe.

Throughout this process, the main actors were the project engineers and scientists. The acceptable risk was not specified by the project owner. If this value had been specified it would have allowed us to follow a targeted risk approach for the probe deployment. Instead, we followed a revealed risk approach; the estimated risk was compared to that of other vehicles and platforms. The decision is left to the engineers and managers to accept or reject the revealed risk.

This paper has not discussed the different shapes of the agreed probability-of-failure distributions or the amount of agreement or disagreement between experts for all assessments. We have presented an analysis of the expert judgments for the ten most critical failures only. From this analysis we concluded that disagreements between expert judgments are not significant. We believe that this is a result of the judgment elicitation process. A key phase of this project is for experts to agree on the upper and lower bound of the probability distribution.

Future work will seek to compare the predicted and actual transition probabilities following the deployment of the probe at Lake Ellsworth in 2013.

ACKNOWLEDGEMENTS

M.B. and G.G. were supported by the UK Natural Environment Research Council (NERC) under the Oceans2025

research programme, while M.M. was supported under NERC grant NE/G004242/1. We thank the Ellsworth team at the National Oceanography Centre, Southampton, for freely taking part in the judgment elicitation exercise. We thank our colleagues at BAS, particularly Keith Makinson, for providing insightful details on the hot-water drilling process, and the experts that took part in the elicitation exercise.

REFERENCES

- Alley RB (2010) Reliability of ice-core science: historical insights. *J. Glaciol.*, **56**(200), 1095–1103 (doi: 10.3189/002214311796406130)
- Brito MP and Griffiths G (2011) A Markov chain state transition approach to establishing critical phases for AUV reliability. *IEEE J. Ocean. Eng.*, **36**(1), 139–149 (doi: 10.1109/JOE.2010.2083070)
- Brito MP, Griffiths G and Challenor P (2010) Risk analysis for autonomous underwater vehicle operations in extreme environments. *Risk Anal.*, **30**(12), 1771–1788 (doi: 10.1111/j.1539-6924.2010.01476.x)
- Brito MP, Griffiths G, Mowlem M and Makinson K (in press) Estimating and managing blowout risk during access to subglacial Antarctic lakes. *Antarct. Sci.* (doi: 10.1017/S0954102000338)
- Cooke RM and Goossens LHJ (2000) Procedures guide for structural expert judgement in accident consequence modelling. *Radiat. Protect. Dosim.*, **90**(3), 303–309
- Engelhardt H, Kamb B and Bolsey R (2000) A hot-water ice-coring drill. *J. Glaciol.*, **46**(153), 341–345 (doi: 10.3189/172756500781832873)
- Feller W (1950) *An introduction to probability theory and its applications*. Wiley, New York
- Fricker HA and 13 others (2011) Siple Coast subglacial aquatic environments: the Whillans Ice Stream Subglacial Access Research Drilling Project. In Siegert MJ, Kennicutt MCI and Bindschadler RA eds. *Antarctic subglacial aquatic environments*. American Geophysical Union, Washington, DC, 199–219 (Geophysical Monograph Series 192)
- Kaplan S and Garrick BJ (1981) On the quantitative definition of risk. *Risk Anal.*, **1**(1), 11–27 (doi: 10.1111/j.1539-6924.1981.tb01350.x)
- Keeney RL and von Winterfeldt D (1991) Eliciting probabilities from experts in complex technical problems. *IEEE Trans. Eng. Manage.*, **38**(3), 191–201 (doi: 10.1109/17.83752)
- Kynn M (2008) The 'heuristics and biases' bias in expert elicitation. *J. R. Stat. Soc.*, **171**(1), 239–264 (doi: 10.1111/j.1467-985X.2007.00499.x)
- Lukin V and Bulat S (2011) Vostok subglacial lake: details of Russian plans/activities for drilling and sampling. In Siegert MJ, Kennicutt MCI and Bindschadler RA eds. *Antarctic subglacial aquatic environments*. American Geophysical Union, Washington, DC, 187–197 (Geophysical Monograph Series 192)
- Makinson K (1993) The BAS hot water drill: development and current design. *Cold Reg. Sci. Technol.*, **22**(1), 121–132 (doi: 10.1016/0165-232X(93)90051-9)
- Makinson K (1994) BAS hot water drilling on Ronne Ice Shelf, Antarctica. *Mem. Natl Inst. Polar Res.*, Special Issue 49, 192–202
- Mowlem MC and 36 others (2011) Probe technology for the direct measurement and sampling of Ellsworth Subglacial Lake. In Siegert MJ, Kennicutt MCI and Bindschadler RA eds. *Antarctic subglacial aquatic environments*. American Geophysical Union, Washington, DC, 159–186 (Geophysical Monograph Series 192)
- Nicas M and Sun G (2006) An integrated model of infection risk in a health-care environment. *Risk Anal.*, **26**(4), 1085–1096 (doi: 10.1111/j.1539-6924.2006.00802.x)
- O'Connor PDT (1995) *Practical reliability engineering*, 3rd edn. Wiley, Chichester
- O'Hagan A and 7 others. (2006) *Uncertain judgements: eliciting experts' probabilities*. Wiley, Chichester
- Priscu JC and 11 others (1999) Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, **286**(5447), 2141–2144 (doi: 10.1126/science.286.5447.2141)
- Priscu J, Powell RD and Tulaczyk S (2010) Probing subglacial environments under the Whillans Ice Stream. *Eos*, **91**(29), 253–254 (doi: 10.1029/2010EO290002)
- Siegert MJ (2005) Lakes beneath the ice sheet: the occurrence, analysis, and future exploration of Lake Vostok and other Antarctic subglacial lakes. *Annu. Rev. Earth Planet. Sci.*, **33**, 215–245 (doi: 10.1146/annurev.earth.33.092203.122725)
- Siegert MJ and 6 others (2001) Physical, chemical and biological processes in Lake Vostok and other Antarctic subglacial lakes. *Nature*, **414**(6864), 603–609 (doi: 10.1038/414603a)
- Siegert MJ and 30 others (2007) Exploration of Ellsworth Subglacial Lake: a concept paper on the development, organisation and execution of an experiment to explore, measure and sample the environment of a West Antarctic subglacial lake. *Rev. Environ. Sci. Biotech.*, **6**(1–3), 161–179 (doi: 10.1007/s11157-006-9109-9)
- Siegert MJ and 16 others (2012) Clean access, measurement, and sampling of Ellsworth Subglacial Lake: a method for exploring deep Antarctic subglacial lake environments. *Rev. Geophys.*, **50**(RG1), RG1003 (doi: 10.1029/2011RG000361)
- Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science*, **185**(4157), 1124–1131 (doi: 10.1126/science.185.4157.1124)
- United States Department of Defense (USDOD) and Rome Laboratories (1991) *MIL-HDBK-217F Military handbook: reliability prediction of electronic equipment*. National Technical Information Service, Springfield, VA
- Woodward J and 9 others (2010) Location for direct access to subglacial Lake Ellsworth: an assessment of geophysical data and modeling. *Geophys. Res. Lett.*, **37**(11), L11501 (doi: 10.1029/2010GL042884)
- Wright A and Siegert MJ (2011) The identification and physiographical setting of Antarctic subglacial lakes: an update based on recent discoveries. In Siegert MJ, Kennicutt MCI and Bindschadler RA eds. *Antarctic subglacial aquatic environments*. American Geophysical Union, Washington, DC, 9–26 (Geophysical Monograph Series 192)

MS received 16 January 2012 and accepted in revised form 25 June 2012