


ORIGINAL ARTICLE OPEN ACCESS

Quantitative System Risk Assessment From Incomplete Data With Belief Networks and Pairwise Comparison Elicitation

Cristina De Persis¹ | José Luis Bosque² | Irene Huertas³ | M. Remedios Sillero-Denamiel^{4,5}  | Simon P. Wilson⁶

¹ATG-Europe for ESA, Noordwijk, The Netherlands | ²Departamento de Ingeniería Informática y Electrónica, Universidad de Cantabria, Santander, Spain | ³ESA ESTEC, Noordwijk, The Netherlands | ⁴Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Sevilla, Spain | ⁵Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Sevilla, Spain | ⁶School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

Correspondence: Simon P. Wilson (swilson@tcd.ie)

Received: 29 February 2024 | **Revised:** 20 December 2024 | **Accepted:** 17 August 2025

Keywords: Bayesian methods | fault tree analysis | risk analysis | spacecraft reentry sparse data contexts

ABSTRACT

A method for conducting Bayesian elicitation and learning in risk assessment is presented. It assumes that the risk process can be described as a fault tree. This is viewed as a belief network, for which prior distributions on primary event probabilities are elicited by means of a pairwise comparison approach. A novel and fully Bayesian updating procedure, following different observation campaigns of the events in the fault tree for the posterior probabilities assessment, is described. In particular, the goal is to handle contexts where there are limited data information (one of the challenges for elicitation), thus keeping simple the elicitation process and adequately quantifying the uncertainties in the analysis. Often, an important consideration in these contexts is the trade-off between how many of the events in the fault tree can be observed against the information that the extra data yield. How this can be addressed within this method is demonstrated. The application is illustrated through three real examples, including the motivating example of risk assessment of spacecraft explosion during controlled reentry.

1 | Introduction

The belief network, otherwise known as a Bayesian network or directed acyclic graph, is a powerful and increasingly utilized tool for the probabilistic modeling of systems of random variables. Originally proposed for use in the quantitative risk assessment of systems at least as early as Barlow (1988), its relationship to a fault tree and the reliability block diagram was recognized quite quickly (Bobbio et al. 2001; Torres-Toledano and Sucar 1998). That is, fault tree analysis studies the hierarchy structure of systems (e.g., in the field of maintenance), whereas Bayesian network can be viewed as a generalization that allows richer probabilistic relationships between events in the tree. The general methodology is well described in papers such as Langseth and Portinale (2007) and Neil et al. (2007). A large and increasing

number of applications of the Bayesian network to risk and reliability can be found; recent examples are modeling and assessing the risk of water transportation infrastructures failure (Kabir et al. 2015; Wang and Yang 2018), improving the resilience of a seaport system (John et al. 2016), or the study of safety barriers in chemical plants (see Yuan et al. 2022 and references therein), among others.

In all these applications, it is assumed that the risk assessment is made with access to expert opinion that forms an initial set of prior information on primary causal event probabilities. In fact, the use of a panel of experts for risk analysis has been studied at some length; see Cooke (1991) and Meyer and Booker (2001), or the work Cooke and Goossens (2008) where 45 expert judgments of a range of sectors, such as nuclear, chemical and gas, water

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Risk Analysis* published by Wiley Periodicals LLC on behalf of Society for Risk Analysis.

treatment, among others, are considered. Some extensions of the latter are Eggstaff et al. (2014), Colson and Cooke (2017), and Cooke et al. (2021). Also, the recent work Misuri et al. (2020) where, because of the lack of available data, a survey involving more than 40 experts was carried out to identify vulnerable safety barriers in natural-hazard triggered technological scenarios. One common approach, and the one we adopt here, is that of elicitation by pairwise comparison that has been used since the early days of elicitation methods (Gulliksen 1959; Guo and Sanner 2010). It is recognized as an indirect interrogation method and one of the simpler ways of accessing expert opinion (Por and Budescu 2017; Hassan et al. 2022). A method motivated by the Analytic Hierarchy Process (AHP) (Saaty 1980;1987) is described in this work, although equally any other pairwise comparison approach could be used. Some examples in recent literature are Hassan et al. (2022) and Dimaio et al. (2021), where the AHP with pairwise comparisons was applied to identify the main causes of pipelines leak/rupture and to include information about the condition of the barriers according to its health state in oil/gas systems, respectively. However, the inclusion of this expert information in the Bayesian networks is being studied in the recent literature (see, e.g., Barons et al. 2022 and references therein), as some problems, such as biased information on primary events from expert elicitations, are still under study (Falconer et al. 2022; Hassan et al. 2022).

The previous situation is ubiquitous when one is discussing risks associated with the emergence of a new technology or where a new risk is identified for an existing system. In fact, the development of methods to determine the probability distributions from scarce data is an important area of research in Bayesian networks (Podofillini et al. 2023; Barons et al. 2022), motivated either because of the lack of data to elicit the prior probabilities for all the primary events or because the collected data do not cover all Bayesian network relationships. Thus, in this paper, we focus on the situation where:

1. There is substantial expert opinion but that constraints on the availability of the experts, or their experience of an elicitation process, mean that the elicitation process must be kept simple;
2. There are limited data from past instances of the risky event, meaning that this information must be used to the full but that also there is uncertainty in the risk assessment that must be properly quantified. In other words, typically some of the events in the fault tree are not observed, and which are observed may change from one observation to the next.

The latter situation is highly related to one of the main challenges in the elicitation process, which is little or no data to update prior beliefs (Falconer et al. 2022). Concretely, in order to handle these sparse data contexts and adequately quantify the uncertainties in the analysis, we propose a fully Bayesian analysis, where elicitation is for the prior distribution of the probabilities of elementary events, rather than on direct specification of this probability as a number. Here, we address the above with the use of a novel elicitation by pairwise comparison, extending it to form priors on the elementary event probabilities. This is in contrast to the existing elicitation procedures for Bayesian networks in reliability, where values for those probabilities are

elicited directly (see, e.g., Hassan et al. 2022; Kabir et al. 2015; Wang and Yang 2018), rather than their prior. This extra layer of modeling allows for more flexible uncertainty quantification and accounts for the uncertainties that there are in these probabilities, particularly when the elicitation process is relatively simple, also reducing the impact of problems such as biased information from expert elicitations when they occur. The benefits of the approach described here are a light elicitation burden on the experts while keeping the proper management of uncertainties through the Bayesian paradigm.

Different methods have been proposed in the literature to deal with limited judgment in Bayesian networks with multistate nodes (Mkrtchyan et al. 2016; Podofillini et al. 2023). In particular, functional interpolation methods for approximating conditional probability distributions show the greatest modeling flexibility at the cost of growing exponentially with the number of nodes in the Bayesian network, which limits their application (Podofillini et al. 2023). Hence, this work describes a fully Bayesian assessment procedure with calibration from data which just makes use of the Bayesian paradigm. First, expert opinion is used to construct the fault tree as well as to carry out the pairwise comparison procedure, which leads to the specification of the prior distributions. As mentioned before, this elicitation process is kept simple under the previous assumptions. Second, this initial risk assessment is updated with partial data coming from observations of the risky situations/events thus obtaining the posterior distributions by using a novel methodology adapted to the case of incomplete observations which uses the logic of the fault tree for the likelihood computation.

Often the risk engineer is faced with the question of how much resource to place on observation of the system, from simply observing the final outcome to full observation of all of its components. In this paper, we also consider different observation campaigns and what extra benefit, in terms of reducing posterior uncertainty, more refined observation can bring.

The motivating example for this work comes from an application in the space industry. To limit the growth of space debris and the risk of collisions in the Low Earth Orbit Protected Region (below 2000 km altitude), it is recommended that satellites and orbital stages reenter into the Earth's atmosphere within 25 years of their mission completion, with the objective that they will largely burn up in the atmosphere. Concerning the reentry risk, the European Space Agency (ESA) policy bans space systems from uncontrolled reentry if the associated ground casualty expectancy exceeds 0.0001 per event. For such cases, a controlled reentry is mandated instead, where the reentry trajectory is designed so that any surviving components that do reach the surface will land over unpopulated areas, such as the South Pacific. A recent example, and the motivation for this work, is the ESA's Automated Transfer Vehicle (ATV), built to supply the International Space Station. In this case, a simple elicitation scheme is required, as engineers have many other demands on their time, in addition to limited data from past instances, implying that some risk events in the fault tree were not observed. Other examples of situations where the approach of this paper may be relevant are nuclear power (Verma et al. 2015), maritime safety (van Dorp and Merrick 2011; Zhang and Thai 2016), or counterterrorism (Merrick and McLay 2010), as new risks/events are constantly being added.

This paper summarizes some of the work on risk assessment from expert opinion that appeared in one of the author's PhD thesis (De Persis 2017), specifically the methodology motivated from AHP and the ATV examples. It begins with a brief description of fault trees and belief networks and their relationship and sets up notation. In Section 2.2, the use of pairwise comparison to elicit priors on the primary event probabilities is discussed. Section 3 explores how the elicited risk assessment can be updated with data that give only partial information about the events in the tree. Section 4 describes three applications of the approach to real risk assessment examples, including a study of how much additional information is gained from observing different proportions of the events in the system. Section 5 presents the main conclusions of the work.

2 | Prior Elicitation

2.1 | Fault Tree and Belief Network Terminology

Consider a fault tree describing the logical relationship between a set of events in a system, culminating in the top event of interest. The value of each node in the tree is in $\{0, 1, \text{NA}\}$, with 1 indicating that the event has been observed to occur, 0 indicating that it has been observed not to have occurred, and NA indicating that it has not been observed. Aside from the top event, and following standard terminology, a primary or base event has no causes developed further in the tree, while an intermediate event is the consequence of other events in the tree. The logic in this tree can be represented as a directed acyclic graph or belief network that brings it into the probabilistic risk modeling domain (Bobbio et al. 2001). In this representation, primary events are assumed to be independent. This logic can be enriched with probabilistic rather than deterministic relationships between events by specifying probability tables for the occurrence of each event in terms of the values of its parents in the network.

Leaving aside for the moment the issue of whether an event is observed, let there be n events in the tree, with $E_i \in \{0, 1\}$ representing whether event i occurred or not, $i = 1, \dots, n$. Let $E_{1:i} = \{E_1, \dots, E_i\}$ represent the first i of these events and let $\eta_i \subset \{E_1, \dots, E_n\}$ denote the immediate causal events of event i in the tree, also known as the parents of E_i in the language of belief networks. If event i is primary, then $\eta_i = \emptyset$. Let H_i represent all ancestors (parents, parents of parents, etc.) of E_i . Assume that there are $k < n$ primary events and that they are labeled E_1, \dots, E_k , and also assume that the top event is E_n . For the primary events, let $p_i = P(E_i = 1)$ and let $\mathbf{p} = (p_1, \dots, p_k)$ denote the set of primary event probabilities. Also note that, because of the tree structure of the network, we can label events in the tree such that the index of all ancestors of E_i are indexed before i , by first labeling the primary events, then the children of primary events, and so forth, so that $H_i \subseteq E_{1:i-1}$. It will also be useful to define the successors (children, grandchildren, etc.) of E_i as S_i .

In the belief network representation, stochastic nodes for each primary event probability p_i are added and they are assigned prior distributions, usually beta distributions for reasons of conjugacy. The network logic then implies prior distributions for the probability of intermediate events and the top event. In other words, when the belief network just models the logic of the

fault tree, prior specification of the primary event probabilities is sufficient to specify the prior of the probability of any event in the tree (Grimmett and Welsh 2014, Chapter 6). These distributions are often not in a closed form but in practice are approximated by Monte Carlo simulation from the primary event prior.

2.2 | Primary Event Probability Prior Elicitation

Following construction of the fault tree, prior distributions for the p_i are elicited from expert opinion. Many elicitation methods are now available, for example, such as Dias et al. (2018) or Garthwaite et al. (2005), and any of these can in principle be applied at this stage. In this work, focus is on constructing independent beta prior distributions for each p_i , because of their subsequent tractability. Since the amount of time that experts can devote to elicitation is often short—in our spacecraft reentry example, the experts were engineers with many other demands on their time—a simpler elicitation scheme, such as one based on pairwise comparisons, is attractive. There is an extensive literature on this approach. Bradley and Terry (1952) and Thurstone (1994) describe early approaches, in which experts are asked to compare events in pairs, assessing whether one or the other is more likely to occur. Also, Bradley and Terry (1952) are discussed and compared in some detail by Cooke (1991), where an extension called NEL model is considered, and experts indicate which event is more likely to cause an earlier failure. Szwed et al. (2006), Mazzuchi et al. (2008), Por and Budescu (2017), and Cavallo and Ishizaka (2023) are more recent uses. The idea of the AHP (Saaty 1980; 1987) has also been used for prior elicitation (Hassan et al. 2022; Dimaio et al. 2021; Cagno et al. 2000; Abastante et al. 2019), a technique that we also exploit. AHP has been widely used, which confirms that experts appreciate dealing with comparative judgments. Nevertheless, this method is not practical if the number of primary events in the tree is large, resulting in too many pairwise comparisons to be made (see Abastante et al. 2019 and references therein). Here, we describe an alternative pairwise comparison approach that exploits AHP technique and attempts to address its limitation. AHP, as an approach to mapping qualitative comparisons to a fully quantitative expression of an opinion, has well-known drawbacks, most notably that it can suffer from rank inversion or reversal (Munier and Hontoria 2021).

The expert is asked to select the “cornerstone” primary event E_{j^*} that he or she has most confidence in giving a prior distribution of its probability. A standard elicitation approach is then used to specify a beta prior distribution of this event's probability, never done before for elicitation in Bayesian networks as far as we know. In this work, a range of values of the probability $(p^{(L)}, p^{(U)})$ is elicited and the prior is specified to be the beta distribution that has this range as its central 95% probability interval; computing this range is an easy numerical exercise since the quantiles of the beta distribution can be accurately and quickly approximated. Then the expert is asked a series of pairwise comparison questions to rate whether each other primary event is more or less likely to occur than the cornerstone event. The expert is asked to rate the probability of a primary event as being equally, moderately, strongly, very strongly, or absolutely more or less than another. At a minimum, each primary event is compared in this way with the cornerstone event. Ideally all pairwise comparisons would be elicited.

Comparison (E_1 to E_2)	Desired prior mean on p_2	Score
Absolutely less probable	0.95	0.17
Very strongly less probable	0.85	0.21
Strongly less probable	0.75	0.28
Moderately less probable	0.60	0.53
Equally probable	0.50	1.00
Moderately more probable	0.40	1.04
Strongly more probable	0.25	1.23
Very strongly more probable	0.15	1.52
Absolutely more probable	0.05	2.55

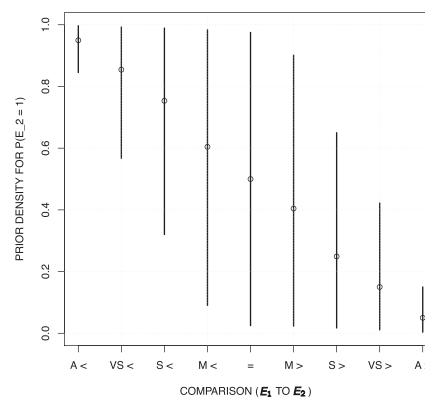


FIGURE 1 | The mapping of qualitative pairwise comparisons to a score (left) and the effect of these on the elicited prior for p_2 in the case of comparing it with a uniform prior on p_1 (right). The circle is the prior mean and the line indicates the central 95% prior probability.

Where E_i and E_j are compared, the comparison of E_j to E_i is assumed to be the reverse, for example, if E_i is strongly more probable than E_j then E_j is assumed to be strongly less probable than E_i . These nine qualitative comparisons are mapped to a numerical score, motivated by a similar approach in the AHP method for eliciting preferences (Saaty 1980). The comparison scores are placed in a matrix $Q = (q_{ij})$, with q_{ij} representing the score of the comparison between E_i and E_j . These scores are then mapped to a weight w_i for each event through the geometric mean approach of Crawford and Williams (1985), where a smaller weight indicates an event that has been elicited to have a lower probability of occurring. The derivation of subjective weights from pairwise comparison scores used here has some advantages from a statistical point of view over those for AHP. More details are given in Appendix A.

The weights are used to specify a range of values for each p_i as

$$\frac{w_i}{w_{i^*}} p^{(L)} \leq p_i \leq \min \left(1, \frac{w_i}{w_{i^*}} p^{(U)} \right), \quad (1)$$

where w_{i^*} is the weight of the cornerstone event. This interval is then mapped to a beta prior distribution that has it as its central 95% probability interval. This has the effect of shifting prior weight to lower values of p_i when E_i is rated less likely on average than E_{i^*} .

Figure 1 shows the scores that we use. These scores are derived from the following simple example of two events with E_1 as the cornerstone event. It is assumed that a uniform prior is assessed on p_1 , so with mean 0.5, giving lower and upper bounds $p^{(L)} = 0.025$ and $p^{(U)} = 0.975$. Then, the different qualitative comparisons are associated with different prior means on p_2 , for example, moderately less probable is associated with a prior mean for p_2 of 0.6, while very strongly more probable is associated with a prior mean for p_2 of 0.15, and so on. Then scores are calculated that give the different prior means for p_2 under the above procedure. For example, taking the absolutely less probable case in Figure 1, and assigning a score of 0.17 to this, leads to weights calculated by a geometric mean and a range for p_2 as in Equation (1). The beta distribution with a central 95% probability given by this range has a mean of 0.95 as required.

As a simple example, consider the system with four primary events as in Figure 2. Figure 3 illustrates the prior elicitation stage where E_1 is deemed to be the cornerstone event and the expert gives an interval (0.01,0.05) as the range for its probability, which equates to a beta distribution with parameters 2.5 and 120. The left column of the plot shows the case where a full set of pairwise comparisons is made, as given in the matrix Q , from which the method of Crawford and Williams (1985) derives weights w . The figure then shows the resulting beta prior distributions for the primary events, following Equation (1) for the other three events as well as event 1 and then finally the resulting prior distribution on the top event probability. The prior expectation of each primary event p_i and for the top event is also given. The right column shows the case where only comparisons with the cornerstone event are available.

As an extension, if the number of primary events to be compared by the expert is large and we can group them according to their meanings and relationships, as is usual in real applications such as those of Section 4, then a previous step can be done. The primary events could be grouped and the expert would be asked to select the “cornerstone” primary event for each group and perform the pairwise comparisons also per group (within each group of primary events but not between groups). This, therefore, means that the number of comparisons required is reduced, and the effort required from the expert is also significantly reduced. See Section 4 for an illustration of this extension.

The approach above assumes consultation with a single expert for the primary events or group of primary events. Using a panel of experts is also possible; the panel could agree on a set of pairwise comparisons, or experts could derive their own prior elicitation and combine them following any of the many approaches (see Keeney and Vonwinterfeldt 1991; Butler et al. 2015 for example). The only constraint is that the panel must agree on the cornerstone events in either case. The use of a panel with AHP has been studied (e.g., Petruni et al. 2019).

In this approach, which maps qualitative comparisons to a full prior distribution, prior sensitivity and robustness are important issues so that the experts and user can check that the implied top event prior matches their beliefs. These methods can show how changes in the prior distributions affect the top event prior, based

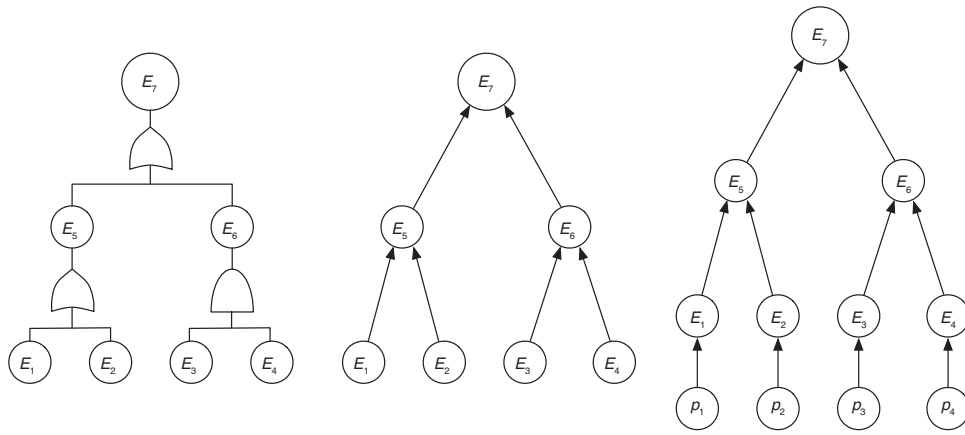


FIGURE 2 | From left to right: A fault tree for a system with four primary events (E_1 to E_4), two intermediate events (E_5 and E_6), and the top event E_7 , the equivalent representation as a belief network and the belief network extended to include the primary events probabilities $p_i = P(E_i = 1)$, $i = 1, \dots, 4$.

on which the user can modify the elicitation. Joshi et al. (2018) discuss an approach to prior robustness for event probabilities in a fault tree through the notion of a distortion function that smoothly changes the prior densities. The fault tree from the ATV example in Section 4.1 is explored in the paper and gives an idea about how a prior robustness analysis could be conducted for this model.

3 | Incorporation of Partial Data About the Event

Once the fault tree is defined and primary event probability prior distributions are elicited, the risk assessment can be updated with data about the event. In this section, it is derived for two cases: the complete case, where all n events in the tree are observed, and the incomplete case, where only a subset of the events are observed.

3.1 | Likelihood for a Complete Observation

The likelihood is $P(E_1, \dots, E_n | \mathbf{p})$ in this case. Recalling the standard result of a belief network that the joint distribution of the variables in the network is the product of the probabilities of each variable given its parents, we have

$$P(E_1, \dots, E_n | \mathbf{p}) = \prod_{i=1}^n P(E_i | \eta_i, \mathbf{p}).$$

Since E_1, \dots, E_k are the primary events that are assumed independent, with $P(E_i | p_i) = p_i^{E_i} (1 - p_i)^{1-E_i}$, one has

$$\begin{aligned} p(E_1, \dots, E_n | \mathbf{p}) &= \left(\prod_{i=1}^k P(E_i | p_i) \right) \left(\prod_{i=k+1}^n P(E_i | \eta_i, \mathbf{p}) \right) \\ &= \left(\prod_{i=1}^k p_i^{E_i} (1 - p_i)^{1-E_i} \right) \left(\prod_{i=k+1}^n P(E_i | \eta_i, \mathbf{p}) \right), \\ &= \left(\prod_{i=1}^k p_i^{E_i} (1 - p_i)^{1-E_i} \right) \left(\prod_{i=k+1}^n P(E_i | \eta_i) \right); \quad (2) \end{aligned}$$

the last line comes from the independence of nonprimary E_i from \mathbf{p} given η_i . The value of these E_i are logically derived from the fault tree. If E_i is the result of an AND gate then $E_i = 1$ if and only if $E_j = 1$ for all $E_j \in \eta_i$, while if it is the result of an OR gate then $E_i = 1$ if and only if $E_j = 1$ for some $E_j \in \eta_i$. Hence, $P(E_i | \eta_i)$ is either 1 or 0, depending on whether the values of E_i and η_i are consistent with the logic of the fault tree or not. To summarize, the likelihood for \mathbf{p} from a complete observation, given by Equation (2), is a product of the Bernoulli likelihoods of each primary event as long as the logic of the fault tree is respected for all the nonprimary event values. If it is not respected then the likelihood is 0.

3.2 | Likelihood for an Incomplete Observation

The more common scenario in situations that interest this paper is that only a subset $\mathcal{E} \subset \{E_1, \dots, E_n\}$ is observed. The likelihood is now $P(\mathcal{E} | \mathbf{p})$.

A general approach to deriving this likelihood is by marginalization of the complete likelihood over the events that are not in \mathcal{E} :

$$\begin{aligned} P(\mathcal{E} | \mathbf{p}) &= \sum_{\substack{E_i=0,1 \\ E_i \notin \mathcal{E}}} P(E_1, \dots, E_n | \mathbf{p}) \\ &= \sum_{\substack{E_i=0,1 \\ E_i \notin \mathcal{E}}} \left(\prod_{i=1}^k p_i^{E_i} (1 - p_i)^{1-E_i} \right) \left(\prod_{i=k+1}^n P(E_i | \eta_i) \right). \quad (3) \end{aligned}$$

Factoring out any observed primary events from the sum gives

$$\begin{aligned} P(\mathcal{E} | \mathbf{p}) &= \left(\prod_{\substack{i=1 \\ E_i \in \mathcal{E}}}^k p_i^{E_i} (1 - p_i)^{1-E_i} \right) \times \sum_{\substack{E_i=0,1 \\ E_i \notin \mathcal{E}}} \left(\prod_{\substack{i=1 \\ E_i \notin \mathcal{E}}}^k p_i^{E_i} (1 - p_i)^{1-E_i} \right) \\ &\quad \left(\prod_{\substack{i=k+1 \\ E_i \notin \mathcal{E}}}^n P(E_i | \eta_i) \right). \quad (4) \end{aligned}$$

$$Q = \begin{pmatrix} 1.00 & 0.21 & 1.04 & 0.53 \\ 1.52 & 1.00 & 1.04 & 1.52 \\ 0.53 & 0.53 & 1.00 & 0.17 \\ 1.04 & 0.21 & 2.55 & 1.00 \end{pmatrix}$$

$$w = (0.162, 0.444, 0.120, 0.273)$$

$$\mathbb{E}(p_{1:4}) = (0.020, 0.068, 0.015, 0.043)$$

$$\mathbb{E}(p_7) = 0.088$$

(a) Elicitation of all pairwise comparisons

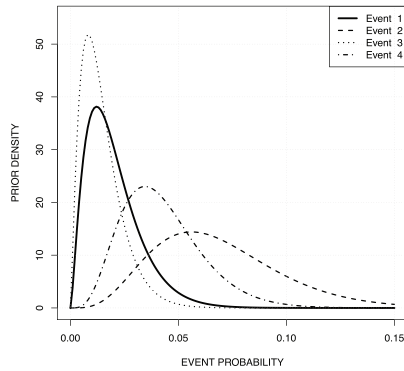
$$Q = \begin{pmatrix} 1.00 & 0.21 & 1.04 & 0.53 \\ 1.52 & 1.00 & - & - \\ 0.53 & - & 1.00 & - \\ 1.04 & - & - & 1.00 \end{pmatrix}$$

$$w = (0.136, 0.425, 0.148, 0.291)$$

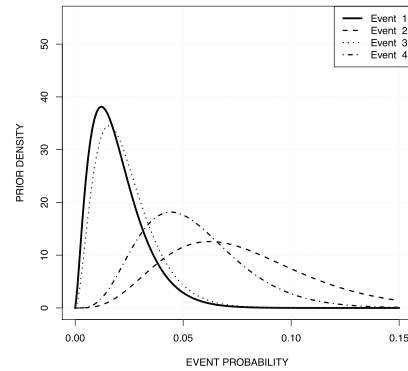
$$\mathbb{E}(p_{1:4}) = (0.020, 0.077, 0.023, 0.054)$$

$$\mathbb{E}(p_7) = 0.097$$

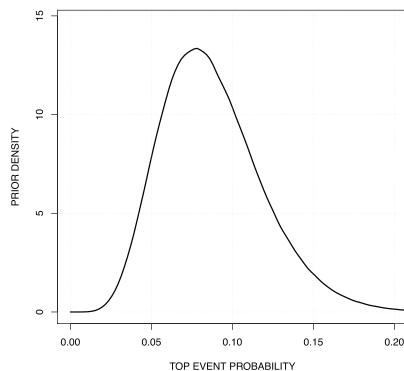
(b) Elicitation considering only pairwise comparisons with the cornerstone event



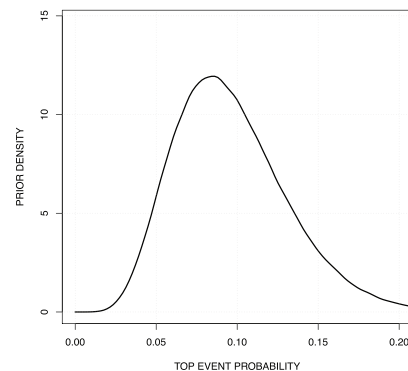
(c) Prior densities for the primary events when elicitation in Panel (a) is considered



(d) Prior densities for the primary events when elicitation in Panel (b) is considered



(e) Prior density for the top event when elicitation in Panel (a) is considered



(f) Prior density for the top event when elicitation in Panel (b) is considered

FIGURE 3 | Elicitation of four primary events. Event 1 is the cornerstone event with elicited interval (0.01,0.05). On the left is the result of an elicitation of all pairwise comparisons, while on the right is an elicitation using only pairwise comparisons with the cornerstone event.

The final term $\prod_{i=k+1}^n P(E_i | \eta_i)$ over intermediate events is either 1 or 0, as in Section 3.1, depending on whether that combination of observed and unobserved event values in E_i and η_i are consistent with the logic of the fault tree or not. Thus, the likelihood is a product of the Bernoulli likelihoods for observed *primary* events, multiplied by a sum of Bernoulli likelihoods for all combinations of unobserved primary events such that the fault tree logic is respected. Hence, one way to derive the likelihood in this case is to

go through each combination of values of the *unobserved* events, check to see if the fault tree logic is respected, and if it is then add the term $\prod_{i=1}^k p_i^{E_i} (1 - p_i)^{1-E_i}$ to a sum. Once that is done, the sum is multiplied by $\prod_{E_i \in \mathcal{E}} p_i^{E_i} (1 - p_i)^{1-E_i}$.

This approach to constructing the likelihood permits common cause events, that is, an event can be the parent of more than

one event. A disadvantage is that it does not scale to a situation where there are a large number of unobserved events, because of the large number of combinations of them that then must be summed in the marginalization.

3.3 | Construction of an Incomplete Likelihood in the Pure Tree Case

When the belief network follows a tree structure, so that each event is the parent of at most one other event, then there is a constructive approach to building the likelihood that may still be practical even when there are a large number of unobserved events. The tree structure constraint means that there cannot be common cause events; each event must be the cause of at most one other event for this approach to work.

First note that in a tree, all nodes are “converging” (e.g., two or more parents have a unique child) apart from the primary nodes, which have a unique p_i as their parent. In such a converging node, the parents are independent given that the child is not observed. Hence, one can partition \mathcal{E} into subsets that consist of an event E_i that has no observed successors, and all its observed ancestors $H_i \cap \mathcal{E}$, and these subsets will be independent. Recalling that S_i is the set of successors of E_i , we have

$$P(\mathcal{E} | \mathbf{p}) = \prod_{\substack{E_i \in \mathcal{E} \\ S_i \cap \mathcal{E} = \emptyset}} P(E_i, H_i \cap \mathcal{E} | \mathbf{p}) = \prod_{\substack{E_i \in \mathcal{E} \\ S_i \cap \mathcal{E} = \emptyset}} P(H_i \cap \mathcal{E} | \mathbf{p}) P(E_i | H_i \cap \mathcal{E}, \mathbf{p}). \quad (5)$$

If the top event E_n is observed then there is no strict partition and Equation (5) becomes $P(\mathcal{E} | \mathbf{p}) = P(H_n \cap \mathcal{E} | \mathbf{p}) P(E_n | H_n \cap \mathcal{E}, \mathbf{p})$.

As regards the two terms on the right-hand side of Equation (5):

- For $P(E_i | H_i \cap \mathcal{E}, \mathbf{p})$, either E_i is logically implied from $H_i \cap \mathcal{E}$, in which case $P(E_i | H_i \cap \mathcal{E}, \mathbf{p}) = 1$, or it is not. If it is not then we can write this probability as a function of the unobserved primary ancestor event probabilities of E_i :

$$P(E_i | H_i \cap \mathcal{E}, \mathbf{p}) = g_i(\{p_j | 1 \leq j \leq k, E_j \in H_i \cap \bar{\mathcal{E}}\}); \quad (6)$$

see Appendix B for the derivation of the function g_i .

- One can apply Equation (5) recursively to $P(H_i \cap \mathcal{E} | \mathbf{p})$, partitioning $H_i \cap \mathcal{E}$ into subsets by the events in $H_i \cap \mathcal{E}$ that have no observed successors in $H_i \cap \mathcal{E}$. The recursion ends when $H_i \cap \mathcal{E} = \emptyset$, in which case $P(E_i, H_i \cap \mathcal{E} | \mathbf{p}) = P(E_i | \mathbf{p})$ and one follows the derivation in Appendix B.

Recursive application of Equations (5) and (6) will yield the expression for the likelihood $P(\mathcal{E} | \mathbf{p})$. Note that this method favors the situation where there are few observed events, in contrast to the marginalization approach of Section 3.2.

3.4 | Posterior Computation

For ease of notation, in this section \mathcal{E} could also refer to a complete observation as well as an incomplete one. The posterior distribution of \mathbf{p} given a set of m such observations $\mathcal{E}_1, \dots, \mathcal{E}_m$ is

then

$$P(\mathbf{p} | \mathcal{E}_1, \dots, \mathcal{E}_m) \propto P(\mathbf{p}) \prod_{l=1}^m P(\mathcal{E}_l | \mathbf{p}),$$

where $P(\mathcal{E}_l | \mathbf{p})$ has been derived using of the methods of the previous parts of this section.

Unfortunately, the likelihood is not conjugate to the beta prior distributions on the p_i and so posterior calculation is implemented by Monte Carlo methods. If m is not too large then importance sampling can be used to generate samples of \mathbf{p} from the posterior distribution with $P(\mathbf{p})$ as the proposal distribution: a large sample of values $\mathbf{p}_1, \dots, \mathbf{p}_R$ is generated from the prior (an easy task as it is a product of independent beta distributions), weights $\omega_r = \prod_{l=1}^m P(\mathcal{E}_l | \mathbf{p}_r)$ are calculated and a posterior sample comes from sampling the \mathbf{p}_r with probabilities proportional to the ω_r . Alternatively, a random walk Metropolis sampling scheme can be used. This has been done with zero-mean normal proposals on the logit p_i ; given a current \mathbf{p} , propose $\lambda_i^* \sim N(\lambda_i, s_i^2)$, where $\lambda_i = \log(p_i) - \log(1 - p_i)$, from which the proposal is $p_i^* = e^{\lambda_i^*} / (1 + e^{\lambda_i^*})$. The proposed vector $\mathbf{p}^* = (p_1^*, \dots, p_k^*)$ is accepted with probability

$$\min \left\{ 1, \frac{p(\mathbf{p}^*) \prod_{l=1}^m P(\mathcal{E}_l | \mathbf{p}^*)}{p(\mathbf{p}) \prod_{l=1}^m P(\mathcal{E}_l | \mathbf{p})} \right\}.$$

This approach works better when m is sufficiently large that the prior and posterior are significantly different.

The posterior samples can be used to approximate the posterior distribution of any intermediate event or the top event by further simulation, as described in Section 2.1.

3.5 | Summary and Example

A full risk assessment procedure, with calibration from data, has now been described. The procedure starts with an expert or experts constructing a fault tree with binary events. A pairwise comparison procedure, such as that described in Section 2.2, leads to the specification of a prior distribution $P(\mathbf{p})$ and hence, by the fault tree logic, to a prior on the probability of the top event, which is usually approximated by Monte Carlo simulation. This is the output for the initial risk assessment. If a point risk estimate is required then the prior median or mean can be used.

This assessment can be updated with data following observation of an instance of the risky situation. The prior distribution of \mathbf{p} is updated to a posterior distribution that can be evaluated by Monte Carlo methods. These posterior samples can be used to generate samples from the posterior distribution of the top event probability. Similarly, the median or mean of these samples can be used as a point estimate for the probability.

Figure 3 illustrated the prior elicitation stage for the simple system of Figure 2. Figure 4 shows the result of posterior updating of the prior, as obtained from the incomplete set of comparisons in Figure 3, after 10 randomly generated observations of this system

Data

(a) Prior Elicitation

$$\begin{aligned}\mathbb{E}(\mathbf{p}) \\ &= (0.020, 0.068, 0.015, 0.043) \\ \mathbb{E}(p_7) &= 0.088\end{aligned}$$

(b) Complete data

$$\begin{aligned}\mathbb{E}(\mathbf{p} | \mathcal{E}_{1:10}) \\ &= (0.011, 0.050, 0.008, 0.072) \\ \mathbb{E}(p_7 | \mathcal{E}_{1:10}) &= 0.061\end{aligned}$$

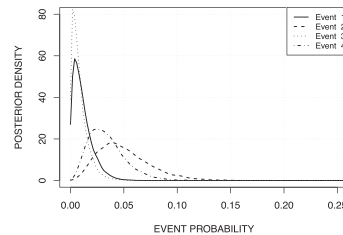
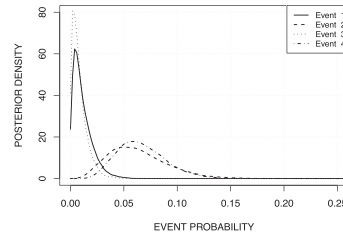
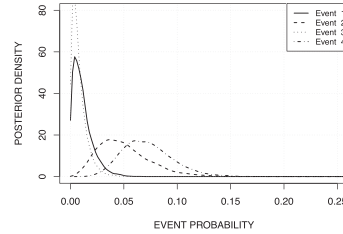
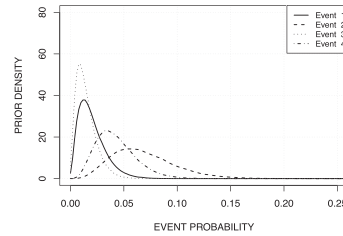
(c) Randomly incomplete (50%)

$$\begin{aligned}\mathbb{E}(\mathbf{p} | \mathcal{E}_{1:10}) \\ &= (0.011, 0.063, 0.008, 0.066) \\ \mathbb{E}(p_7 | \mathcal{E}_{1:10}) &= 0.074\end{aligned}$$

(d) Top event only

$$\begin{aligned}\mathbb{E}(\mathbf{p} | \mathcal{E}_{1:10}) \\ &= (0.011, 0.050, 0.008, 0.034) \\ \mathbb{E}(p_7 | \mathcal{E}_{1:10}) &= 0.061\end{aligned}$$

Primary Event Probabilities



Top Event Probability

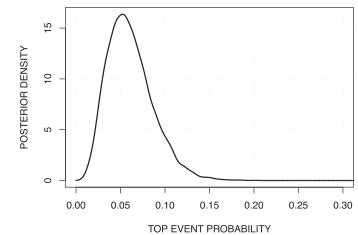
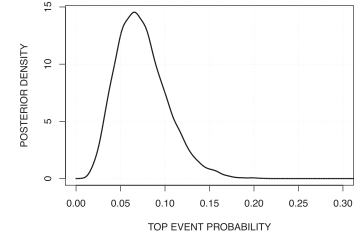
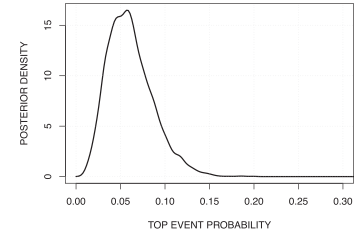
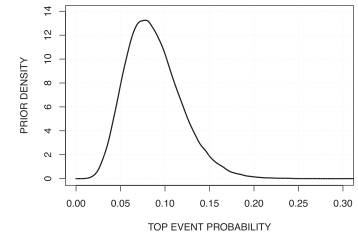


FIGURE 4 | The prior and posterior distributions for the primary event and top event probabilities. The top row is the prior and subsequent rows are the posterior after 10 simulated observations of different types with $\mathbf{p} = (0.02, 0.05, 0.05, 0.1)$ (and so a top event probability $p_7 = 0.0737$).

with primary event probabilities (0.02, 0.05, 0.05, 0.10), giving a top event probability of $p_7 = 0.074$. This is done for three cases:

- The data set is complete, with the value of all events observed;
- The same data set but incomplete, with each event randomly observed with probability 0.5;
- The same data set but with the top event observed only.

Implementation was with the random walk Metropolis algorithm over 100,000 iterations with $s_\lambda^2 = 0.25$, the latter selected after a fine-tuning process. The figure shows very little difference in the inference between complete and 50% incomplete data. This is not surprising in this case as, in many cases, it is possible to logically infer most if not all of the missing values. Posterior uncertainty using the top event observation only is somewhat larger than in the other two cases.

To further explore this, each of these posterior updates was repeated 100 times, each time with data regenerated with the same true primary event probabilities. Table 1 summarizes the posterior distributions obtained over these 100 runs where a uniform prior has been assumed for the primary event probabilities, while Table 2 repeats this for the prior obtained from the incomplete comparisons in Figure 3. Each table shows the average of the posterior means, standard deviations and central 95% probability intervals, as well as the root mean square error between the posterior mean and the true value:

$$\text{RMSE} = \left(\frac{1}{100} \sum_{r=1}^{100} (p_{\text{posterior},r} - p_{\text{true}})^2 \right)^{1/2},$$

where $p_{\text{posterior},r}$ is the posterior mean of p from the r th run and p_{true} is the true value.

TABLE 1 | Summary of results for the simulation study of the system in Figure 2, using a uniform prior on each primary event probability.

	P_1	P_2	P_3	P_4	P_7
True value	0.020	0.050	0.050	0.100	0.074
Prior mean	0.500	0.500	0.500	0.500	0.812
<i>Top event observation only</i>					
Average of:					
post. means	0.019	0.138	0.008	0.035	0.155
post. std. devs	0.015	0.041	0.007	0.017	0.039
post. 95% PI widths	0.056	0.159	0.027	0.066	0.154
RMSE	0.001	0.089	0.042	0.065	0.082
<i>Randomly incomplete (50%)</i>					
Average of:					
post. means	0.038	0.101	0.027	0.058	0.137
post. std. devs	0.017	0.034	0.013	0.021	0.036
post. 95% PI widths	0.066	0.134	0.048	0.082	0.140
RMSE	0.021	0.055	0.025	0.043	0.065
<i>Complete observation</i>					
Average of:					
post. means	0.049	0.111	0.038	0.072	0.153
post. std. devs	0.019	0.035	0.015	0.023	0.037
post. 95% PI widths	0.072	0.135	0.057	0.091	0.144
RMSE	0.031	0.062	0.016	0.031	0.083

In general, the tables show that the data have the effect of moving prior distributions in the direction of the true values. For example, in Table 1, the prior mean probability for the primary event E_1 under a uniform prior is fixed at $p_1 = 0.500$, whereas its true value is 0.020. After observing the data, it turns 0.019 (with the top event observation only), 0.038 (with incomplete data), and 0.049 (with the complete observation scene). The prior coming from the elicitation is quite strong and has a large effect on the posterior, relative to the uniform prior case, even after 10 observations. It can be seen from the smaller root mean square errors in Table 2, which implies a more stable set of results under the incomplete pairwise comparison prior with smaller standard deviations. However, note that inference for the top event probability, which is the risk event of interest in most of the real-life examples, is very similar across all three data types in both cases, uniform prior and prior coming from elicitation, but the final estimate is much better for the latter.

Finally, it should be highlighted that the observed probability for the incomplete case could be changed, as in real-life applications that probability may be smaller than 0.5. After running different experiments in which the probability of being observed changes when the Randomly Incomplete campaign is taking into account, we have observed that the novel fully Bayesian methodology described in Sections 3.3 and 3.4 for the posterior distributions computation is able to deal adequately with incomplete observations thanks to use the logic of the fault tree for the likelihood computation, while being stable in estimating the probabilities for the child nodes, which are usually the important ones.

TABLE 2 | Summary of results for the simulation study of the system in Figure 2, using the incomplete pairwise comparison prior of Figure 3.

	P_1	P_2	P_3	P_4	P_7
True value	0.020	0.050	0.050	0.100	0.074
Prior mean	0.020	0.077	0.023	0.054	0.097
<i>Top event observation only</i>					
Average of:					
post. means	0.012	0.058	0.008	0.034	0.070
post. std. devs	0.010	0.026	0.007	0.017	0.027
post. 95% PI widths	0.037	0.102	0.027	0.065	0.106
RMSE	0.008	0.012	0.042	0.066	0.010
<i>Randomly incomplete (50%)</i>					
Average of:					
post. means	0.012	0.061	0.008	0.058	0.073
post. std. devs	0.009	0.027	0.007	0.021	0.028
post. 95% PI widths	0.035	0.103	0.026	0.083	0.108
RMSE	0.008	0.013	0.041	0.044	0.010
<i>Complete observation</i>					
Average of:					
post. means	0.011	0.060	0.008	0.073	0.072
post. std. devs	0.009	0.026	0.007	0.023	0.027
post. 95% PI widths	0.034	0.100	0.025	0.091	0.104
RMSE	0.008	0.010	0.041	0.027	0.010

4 | Examples

In this section, we look at three real-world examples. The first is the motivating example for this work and involved a real elicitation with experts. The other two examples are more complex and illustrate the scale of the required elicitation even when the fault tree is complicated.

In all of the examples, we look at the question of how much information about the events in the fault tree affects the posterior inference. In particular, we look at how posterior inference on the top event probability changes as we move from observation of only the top event, to observation of some of the events, to observation of all the events in the tree.

4.1 | Automated Transfer Vehicle Reentry

The Automated Transfer Vehicle (ATV) was developed by the ESA to resupply the International Space Station. Its mission consisted of a launch to the station with supplies (e.g., experimental equipments, propellants, and goods for the permanent crew); once unloaded at the station, waste was placed into it. The vehicle undocked from the station and was designed to have a controlled burn-up in the atmosphere, with any surviving fragments landing in the remote South Pacific. Five ATVs were launched between 2008 and 2014. All successfully resupplied the Space Station and were then successfully deorbited.

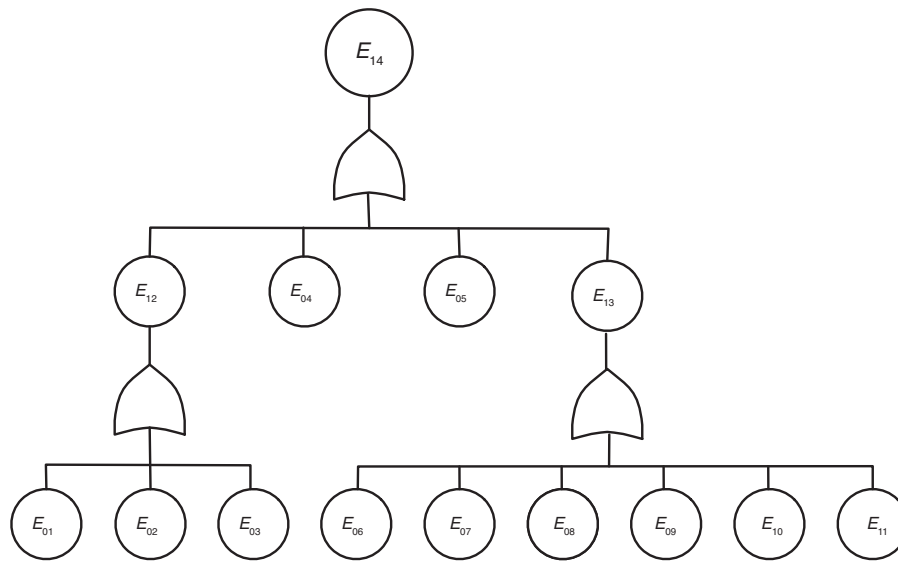


FIGURE 5 | Fault tree for unexpected explosion during reentry of the Autonomous Transfer Vehicle.

One risk associated with the ATV, and most spacecraft that are deorbited, is that the vehicle will unexpectedly explode during reentry, due to causes such as the heating of leftover fuel, which could result in scattering debris over land (Koppenwallner et al. 2005). Due to the difficulties in observing such a reentry, there is unlikely to be complete information on the causes of an explosion (De Pasquale et al. 2009). The question then arises as to what can be learned from the observation. De Persis (2017) conducted a fault tree analysis as part of an assessment of the risk of this event, from which we base this analysis.

Figure 5 shows the fault tree that was elicited from expert engineers at ESA. Appendix C gives a description of each node in the tree. Note that the tree contains only OR nodes, so that it only takes one of the primary events to occur in order for the top event to occur. Intermediate nodes are included in the diagram because they may be what is observable during the reentry.

The primary nodes of the tree come in three groups:

- Nodes 1 to 3 are events concerning propellant or the propellant tanks;
- Nodes 6 to 11 are events concerning the batteries;
- Nodes 4 and 5 are other causes of an explosion on reentry.

Separate experts were consulted about each of these three groups. Within each group, a complete pairwise comparison was made and the prior distributions were constructed separately from these complete comparisons within each group. No comparisons between the three groups were made. The details of the elicitation are in Appendix D. Figure 6 shows all the distinct prior distributions of primary events that were elicited. The resulting prior distribution for the top event probability is given in Figure 7. The mean for the top event probability is 0.17 with a central 95% probability interval (0.11,0.24).

Figure 7 shows a kernel density estimate of the posterior distribution of the top event probability, given observation that all five

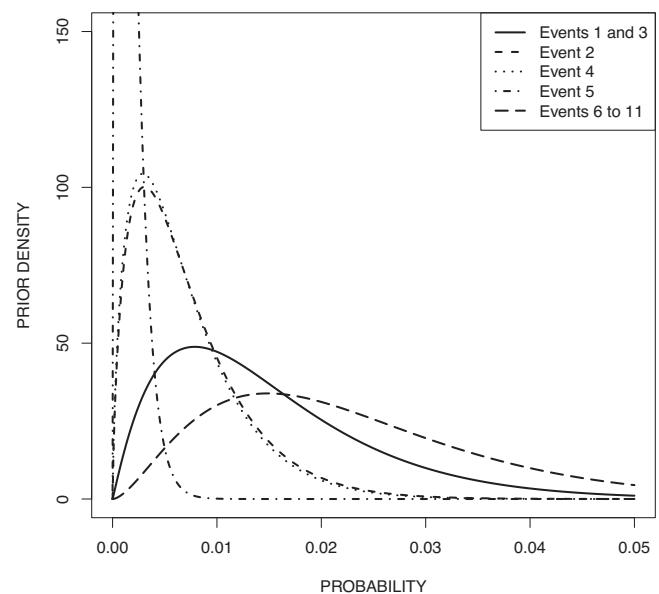


FIGURE 6 | Prior distributions on primary event probabilities.

reentries occurred without explosion. These data are just observation of the top event; however, because all nodes in the fault tree are OR gates, observation that the top event did not occur is equivalent to a complete observation that all primary events did not occur. As expected, the posterior distribution is shifted toward smaller probabilities. The posterior mean probability is now 0.10, with a central 95% probability interval of (0.02,0.25). Compared to the prior, the mean has decreased but uncertainty in the value of the top event probability has actually increased because the data are somewhat in contradiction with the prior opinion.

A detailed observation campaign of the first ATV reentry was attempted at considerable cost, using aircraft-based cameras that tracked the reentry trajectory (Lips et al. 2010). No similar observation campaign was attempted subsequently. To conclude

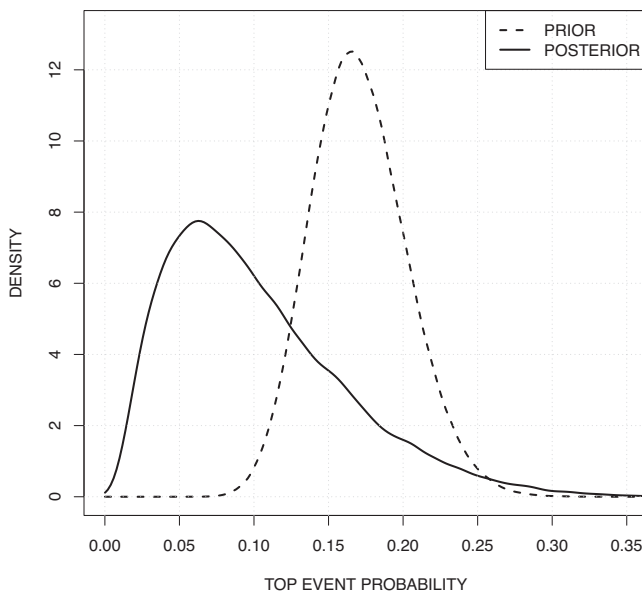


FIGURE 7 | The prior distribution of the top event probability implied by the priors in Figure 6 (dashed) and the posterior distribution given data that all five ATV spacecraft were observed to reenter without explosion (solid).

TABLE 3 | Average posterior uncertainty in the top event probability of the ATV example, for different combinations of data size and observation campaign, under the elicited prior.

Data size	1	3	5	10
True value	0.170	0.170	0.170	0.170
Prior mean	0.170	0.170	0.170	0.170
<i>Top event observation only</i>				
Average of:				
post. means	0.107	0.109	0.109	0.112
post. std. devs	0.027	0.027	0.027	0.027
post. 95% PI widths	0.106	0.103	0.105	0.104
<i>Randomly incomplete (50%)</i>				
Average of:				
post. means	0.107	0.108	0.109	0.109
post. std. devs	0.027	0.026	0.027	0.026
post. 95% PI widths	0.105	0.102	0.105	0.100
<i>Complete observation</i>				
Average of:				
post. means	0.108	0.109	0.109	0.112
post. std. devs	0.026	0.027	0.027	0.026
post. 95% PI widths	0.100	0.102	0.105	0.102

this example, an exploration of the value of an observation campaign, that was able to observe some or all of the intermediate and primary events in the tree, is explored through simulation. The events in the tree are simulated using the elicited prior means as the primary event probabilities. Table 3 summarizes the posterior uncertainty in the top event probability under different

TABLE 4 | Average posterior uncertainty in the top event probability of the ATV example, for different combinations of data size and observation campaign, under a uniform prior.

Data size	1	3	5	10
True value	0.170	0.170	0.170	0.170
Prior mean	0.9995	0.9995	0.9995	0.9995
<i>Top event observation only</i>				
Average of:				
post. means	0.114	0.127	0.140	0.172
post. std. devs	0.027	0.028	0.030	0.032
post. 95% PI widths	0.106	0.111	0.116	0.125
<i>Randomly incomplete (50%)</i>				
Average of:				
post. means	0.123	0.162	0.192	0.266
post. std. devs	0.028	0.031	0.032	0.035
post. 95% PI widths	0.110	0.121	0.126	0.135
<i>Complete observation</i>				
Average of:				
post. means	0.389	0.201	0.257	0.376
post. std. devs	0.034	0.032	0.034	0.036
post. 95% PI widths	0.126	0.126	0.132	0.134

combinations of data type and size. For data size, we look at one observation (being the number of observation campaigns actually undertaken), 3, 5 (being the total number of ATVs that were deorbited), and then 10, to see what is the impact of a campaign on a satellite series with more deorbits. For data type, we consider three different observation campaigns as in Section 3.5: observation of the top event only, each event is observed with probability 0.5, and observation of all events in the tree. The table is a summary of the inference of each combination over 100 simulated data sets.

The principal feature of Table 3 is that there appears to be very little difference in posterior uncertainty between the three types of observation (top event observation only; randomly incomplete with probability 0.5; and complete observation), which, as commented before, return a posterior mean probability of about 0.10 after observing the data. The fact that there are no major differences among the three observation campaigns makes sense in this case since (a) the prior expectation is that about 83% of top event observations will be that no explosion occurred and (b) such an observation is logically equivalent to the complete observation that all primary events did not occur. In other words, in most cases there is no real difference between the different data types when prior elicitation is performed. However, Table 4 repeats the analysis with uniform priors on all of the primary event probabilities, and although the data have the effect of moving the prior distributions across the three observation campaigns, it can be observed that the more complete the data, the more the probability is overestimated. In addition, it can again be seen how the standard deviations are larger than in Table 3 where the elicited prior is considered.

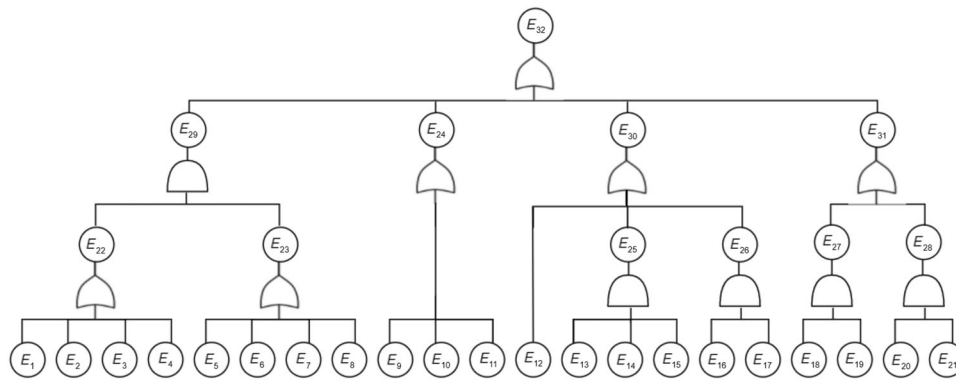


FIGURE 8 | Fault tree for the release prevention barrier of the Tesoro Anacortes refinery.

The conclusion is that, under this fault tree and under data scarcity, there is little benefit to a detailed observation campaign, at least in terms of quantifying the risk of the top event if elicited prior is considered. However, we note that there are other reasons to conduct an observation campaign outside the scope of this paper, such as characterization of the fragment size during spacecraft break-up, the altitude and velocity of those fragments, and so forth.

4.2 | Release Prevention Barrier (RPB)

The Tesoro Anacortes refinery accident occurred in April 2010. The cause of the accident was determined to be a rupture of the E-6600E heat exchanger by high-temperature hydrogen. After investigating the incident, the US Chemical Safety Board highlighted the contributory factors and arranged them into seven prevention barriers regarding release, dispersion, ignition, escalation, emergency management failure, human factors, and organizational failure.

In Adedigba et al. (2016), an analysis was made of this accident and a fault tree for the failure of the RPB, as in Figure 8, was constructed. Assigned probabilities for the primary events of the tree were also given. The primary events are split into four groups: operational, design, inspection, and maintenance errors. A description of the events in the tree is given in Appendix E.

A complete pairwise comparison can be done within each of the four groups. The group sizes are 3, 4, 6, and 8 for a total of $3 + 6 + 15 + 28 = 52$ pairwise comparisons in the complete case. For this example, we specify the pairwise comparisons so that the prior distributions roughly match the probabilities given in Adedigba et al. (2016) (see Appendix F). Figure 9 displays the priors distributions of primary events as well as the resulting prior of the top event probability which mean of 0.098 with a central 95% probability interval (0.060, 0.145). This can be compared to the top event probability of 0.0842 that is given in Adedigba et al. (2016).

As in the ATV example, Tables 5, 6, and 7 explore the effect on the posterior distribution of the three different types of data (top event observation only, randomly incomplete with probability 0.5, and complete) when only one, three, and five instances of the risky events have been observed (what would be expected

TABLE 5 | Average posterior uncertainty in the top event probability for different combinations of data size and observation campaign, under the elicited prior, for the release prevention barrier of the Tesoro Anacortes refinery.

Data size	1	3	5
True value	0.098	0.098	0.098
Prior mean	0.098	0.098	0.098
<i>Top event observation only</i>			
Average of:			
post. means	0.067	0.068	0.068
post. std. devs	0.019	0.018	0.018
post. 95% PI widths	0.072	0.071	0.071
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.068	0.067	0.067
post. std. devs	0.019	0.018	0.018
post. 95% PI widths	0.072	0.071	0.070
<i>Complete observation</i>			
Average of:			
post. means	0.080	0.068	0.068
post. std. devs	0.018	0.018	0.018
post. 95% PI widths	0.072	0.071	0.070

in the event of a critical failure). Table 5 shows this for the inference on the top event probability, Table 6 shows it for primary event E_4 (hydrogen induced cold cracking), and Table 7 shows it for intermediate event E_{22} (process upset). One hundred sets of data of different sizes are simulated using the prior means as the primary event probabilities, and summaries of the resulting posterior distributions are shown.

As can be seen from the results, for inference on primary, intermediate, and top event probabilities, observation of the top event alone or observation of incomplete data seem to be sufficient; the posterior distribution for this event changes little as more complete information is observed. The latter also shows how our methodology for fully Bayesian elicitation and risk assessment procedure proposed in this work is able to properly estimate the

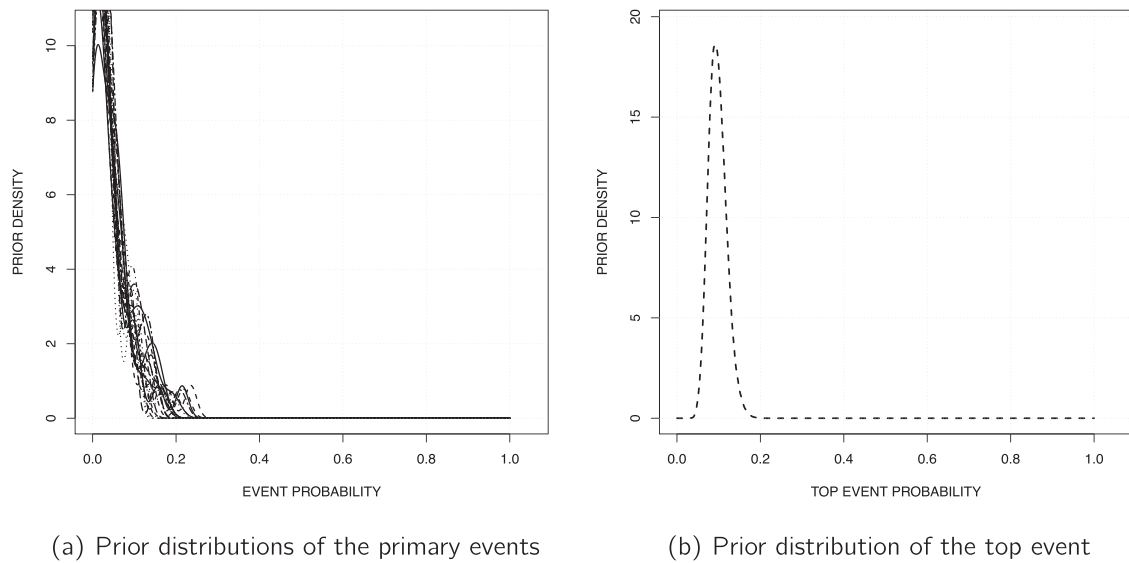


FIGURE 9 | Prior distributions of the primary events (a) and the implied prior on the top event (b) for the release prevention barrier of the Tesoro Anacortes refinery.

TABLE 6 | Average posterior uncertainty in the primary event E_4 probability for different combinations of data size and observation campaign, under the elicited prior, for the release prevention barrier of the Tesoro Anacortes refinery.

Data size	1	3	5
True value	0.068	0.068	0.068
Prior mean	0.068	0.068	0.068
<i>Top event observation only</i>			
Average of:			
post. means	0.053	0.053	0.053
post. std. devs	0.030	0.030	0.030
post. 95% PI widths	0.113	0.113	0.113
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.053	0.053	0.054
post. std. devs	0.030	0.029	0.029
post. 95% PI widths	0.111	0.111	0.110
<i>Complete observation</i>			
Average of:			
post. means	0.052	0.053	0.052
post. std. devs	0.029	0.029	0.028
post. 95% PI widths	0.094	0.110	0.108

posterior distribution of the risky top event in contexts where scarce data are available.

4.3 | Hydrogen Station

The last example considered in this work is related to the accident modeling of a hydrogen station. While hydrogen is

TABLE 7 | Average posterior uncertainty in the intermediate event E_{22} probability for different combinations of data size and observation campaign, under the elicited prior, for the release prevention barrier of the Tesoro Anacortes refinery.

Data size	1	3	5
True value	0.136	0.136	0.136
Prior mean	0.136	0.136	0.136
<i>Top event observation only</i>			
Average of:			
post. means	0.103	0.103	0.102
post. std. devs	0.033	0.033	0.034
post. 95% PI widths	0.129	0.128	0.130
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.103	0.104	0.105
post. std. devs	0.033	0.033	0.033
post. 95% PI widths	0.126	0.128	0.127
<i>Complete observation</i>			
Average of:			
post. means	0.109	0.104	0.107
post. std. devs	0.029	0.033	0.032
post. 95% PI widths	0.112	0.127	0.124

a source of energy with some excellent properties—nontoxic, renewable, sustainable, and abundant (Mazloomi and Gomes 2012)—there are safety issues concerning its well-known flammability that have to be studied for production, storage, and transportation processes. Risk assessment for this application has been studied (see Al-Shanini et al. 2014 and references therein). Like the previous example, the risk assessment framework is defined by incorporating different prevention barriers

TABLE 8 | Average posterior uncertainty in the top event (release prevention barrier failure) probability for different combinations of data size and observation campaign, under the elicited prior.

Data size	1	3	5
True value	0.376	0.376	0.376
Prior mean	0.376	0.376	0.376
<i>Top event observation only</i>			
Average of:			
post. means	0.283	0.285	0.285
post. std. devs	0.047	0.046	0.046
post. 95% PI widths	0.182	0.180	0.179
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.282	0.281	0.290
post. std. devs	0.047	0.046	0.046
post. 95% PI widths	0.182	0.177	0.179
<i>Complete observation</i>			
Average of:			
post. means	0.283	0.285	0.287
post. std. devs	0.047	0.046	0.046
post. 95% PI widths	0.181	0.178	0.179

whose failure is modeled by fault trees (see Figures 2–6 in Al-Shanini et al. 2014). In this example, the RPB in Figure 10 has been selected to illustrate our approach (see Figure 2 of Al-Shanini et al. 2014).

This tree is formed from 43 primary events that come in four groups: operational error prevention barrier failure, H₂ containment equipment/component failure, earthquake/lighting prevention barrier failure, and maintenance prevention barrier failure. Complete pairwise comparison as well as prior distributions construction within each group were performed, without comparison between groups as before. The group sizes are 18, 4, 10, and 11, implying the need for 136 + 6 + 45 + 55 = 242 pairwise comparisons in the complete case and 17 + 3 + 9 + 10 = 39 in the incomplete case. The details of a complete elicitation can be found in Appendix G, and Figure 11 displays the priors distributions of primary events as well as the resulting prior of the top event probability associated with this elicitation.

Tables 8, 9, and 10 show the summary of the posterior distribution for the top event (RPB failure), a primary (rupture disk failure), and an intermediate (earthquake/lighting structural prevention failure) event, respectively, over 100 simulated runs from data generated with primary event probabilities given by their prior. We see in this case that, once again, inference can be accomplished well by just observing the top event or sparse data using the fully Bayesian updating procedure described above.

TABLE 9 | Average posterior uncertainty in the primary event E_4 (rupture disk failure) probability for different combinations of data size and observation campaign, under the elicited prior.

Data size	1	3	5
True value	0.056	0.056	0.056
Prior mean	0.056	0.056	0.056
<i>Top event observation only</i>			
Average of:			
post. means	0.041	0.041	0.041
post. std. devs	0.025	0.026	0.025
post. 95% PI widths	0.096	0.097	0.096
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.042	0.040	0.043
post. std. devs	0.025	0.025	0.025
post. 95% PI widths	0.096	0.094	0.094
<i>Complete observation</i>			
Average of:			
post. means	0.042	0.042	0.043
post. std. devs	0.026	0.025	0.024
post. 95% PI widths	0.097	0.094	0.093

TABLE 10 | Average posterior uncertainty in the intermediate event E_{67} (earthquake/lighting structural prevention failure) probability for different combinations of data size and observation campaign, under the elicited prior.

Data size	1	3	5
True value	0.116	0.116	0.116
Prior mean	0.116	0.116	0.116
<i>Top event observation only</i>			
Average of:			
post. means	0.050	0.050	0.049
post. std. devs	0.024	0.025	0.024
post. 95% PI widths	0.091	0.093	0.092
<i>Randomly incomplete (50%)</i>			
Average of:			
post. means	0.050	0.052	0.052
post. std. devs	0.024	0.023	0.024
post. 95% PI widths	0.092	0.088	0.092
<i>Complete observation</i>			
Average of:			
post. means	0.051	0.051	0.055
post. std. devs	0.025	0.025	0.025
post. 95% PI widths	0.094	0.094	0.095

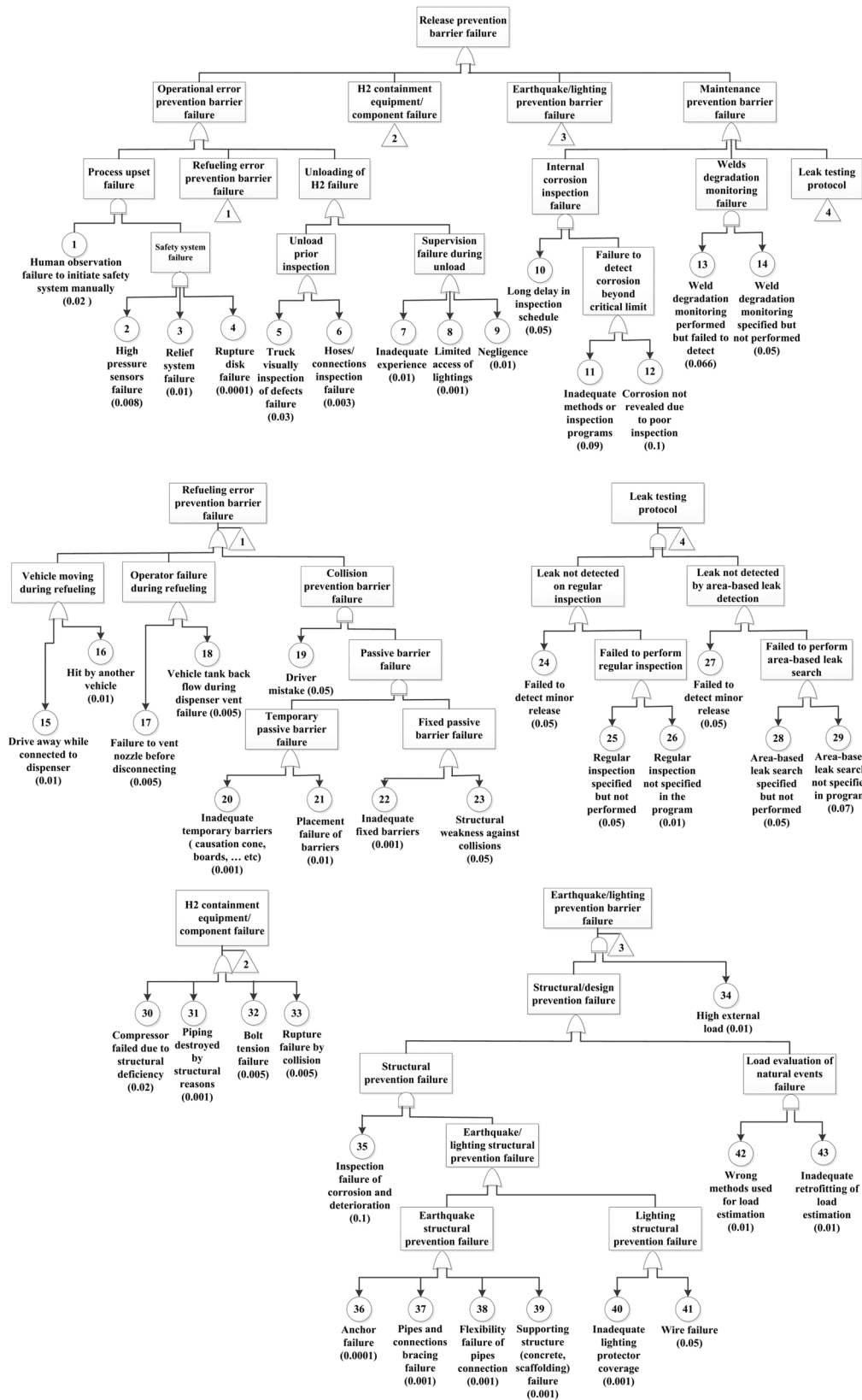


FIGURE 10 | Release prevention barrier failure fault tree model for the Hydrogen data set, from Al-Shanini et al. (2014).

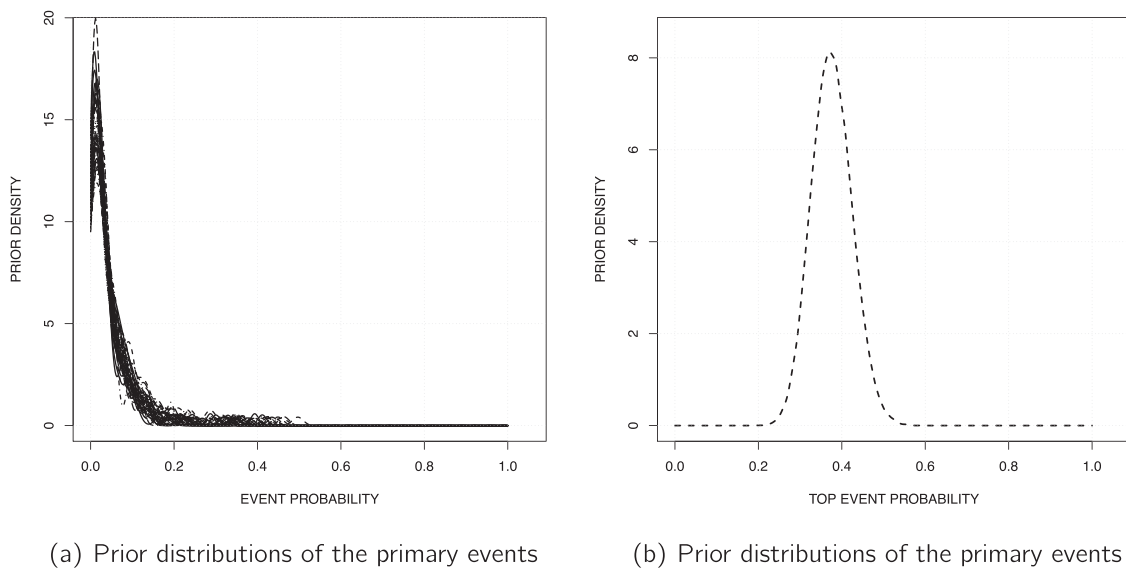


FIGURE 11 | Prior distributions on primary event probabilities (a) and the resulting prior on the top event probability (b) for the hydrogen station example, based on the elicitation described in the Appendix.

5 | Conclusion

An approach to probabilistic risk assessment for a system, using a combination of fault tree analysis, prior elicitation, and Bayesian updating, has been described. Through its use of pairwise comparisons for prior elicitation, it has particular use in circumstances where the access to experts is limited or they have little experience of elicitation methods. The method is illustrated with an application to spacecraft reentry risk, as well as two other more complicated examples from the recent literature. One application of this work is in exploring the effects of observing different refinements of data on the posterior inference. We find that, in general, inference for the top event probability needs little more than observation of the top event, while inference for other event probabilities can have higher posterior precision if more complete data are used.

Here we have used the method to explore how observation of different amounts of data in the system affects the posterior uncertainty in the probabilities of the top and other events. More formally, this question can be embedded into a decision problem where the trade-off between the cost of extra observation and the value of the additional information are quantified and the optimal decision found. Such a framework also allows factors like the timeliness of the analysis into account; in some of the applications discussed in this paper, there is an additional trade-off between the need for a timely risk assessment versus waiting longer to collect more data that reduces uncertainty.

The principal benefits of the approach are the relatively light burden on the expert for elicitation, and the proper management of uncertainties through a Bayesian network. Any other elicitation approach for the primary event probabilities can be “plugged in” and used if needed, including more comprehensive methods that would give richer information.

The use of an AHP-like method to map quantitative comparisons to a fully quantified prior densities means that it may suffer from

some of the drawbacks of that approach. We emphasize the need for prior sensitivity and robustness to partner the elicitation, to confirm that the process has led to a prior that is consistent with the expert’s belief.

Regarding future work, Falconer et al. (2022) discuss the need to address the question of reassessing the risk of a system that has undergone some modification in light of testing; a common scenario in system development. An inefficient solution is to repeat the risk assessment from the start at each redesign, and only use data from the redesigned system to update the posterior. Doing anything else raises issues such as:

- If only a part of the system is changed, one could repeat the prior assessment only for events related to those altered parts, leaving everything else unchanged.
- How to use data from previous iterations of the design. Can parts of it be used, such as coming from the observation of events related to parts of the system that have not changed?

Another future development is incorporation of prior sensitivity and robustness into the method, following the approach in Joshi et al. (2018), as discussed at the end of Section 2.2. Finally, it is worth looking further into the possible difficulties that the size of the fault tree could bring as future work, and examining more in depth how the curse of dimensionality in terms of the size of the fault trees could affect the proposed methodology for practical and complex real-life applications.

Software

There is open-source software, in the form of R code, to implement the methods described in this paper. At the time of writing, the code derives the likelihood using Equation (3) and implements inference by MCMC, as in Section 3.4. The code includes sample scripts to run the analyses that generate Figures 3, 4, and 7. The code is open source under GPL v3 license and can be downloaded

Acknowledgments

This research is partially supported by the Insight Centre for Data Analytics, funded by Science Foundation Ireland through grant 12/RC/2289-P2, and by the Spanish Ministry of Science, Innovation and Universities through grant PID2022-137818OB-I00. It is also partially supported by the NPI program of the European Space Agency (ESA) and by the Spanish Ministry of Science, Innovation and Universities, under the program "Salvador de Madariaga," grant PRX18/00128. We are immensely grateful to Dr. Guillermo Ortega of ESA, who provided insight and expertise that greatly assisted the research. We would like to thank those at the ESA who gave their time and expertise to the elicitation process for the ATV application.

References

- Abastante, F., S. Corrente, S. Greco, A. Ishizaka, and I. M. Lami. 2019. "A New Parsimonious AHP Methodology: Assigning Priorities to Many Objects by Comparing Pairwise Few Reference Objects." *Expert Systems With Applications* 127: 109–120.
- Adedigba, S. A., F. Khan, and M. Yang. 2016. "Dynamic Safety Analysis of Process Systems Using Nonlinear and Non-Sequential Accident Model." *Chemical Engineering Research and Design* 111: 169–183.
- Al-Shanini, A., A. Ahmad, and F. Khan. 2014. "Accident Modelling and Safety Measure Design of a Hydrogen Station." *International Journal of Hydrogen Energy* 39, no. 35: 20362–20370.
- Barlow, R. E. 1988. "Using Influence Diagrams." In *Accelerated Life Testing and Experts' Opinions in Reliability*, edited by C. Clarotti and D. Lindley, 145–157. North Holland.
- Barons, M. J., S. Mascaro, and A. M. Hanea. 2022. "Balancing the Elicitation Burden and the Richness of Expert Input When Quantifying Discrete Bayesian Networks." *Risk Analysis* 42, no. 6: 1196–1234.
- Bobbio, A., L. Portinale, M. Minichino, and E. Ciancamerla. 2001. "Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks." *Reliability Engineering & System Safety* 71, no. 3: 249–260.
- Bradley, R. A., and M. E. Terry. 1952. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika* 39: 324–345.
- Butler, A. J., M. K. Thomas, and K. D. M. Pintar. 2015. "Systematic Review of Expert Elicitation Methods as a Tool for Source Attribution of Enteric Illness." *Foodborne Pathogens and Disease* 12, no. 5: 367–382.
- Cagno, E., F. Caron, M. Mancini, and F. Ruggeri. 2000. "Using AHP in Determining the Prior Distributions on Gas Pipeline Failures in a Robust Bayesian Approach." *Reliability Engineering & System Safety* 67, no. 3: 275–284.
- Cavallo, B., and A. Ishizaka. 2023. "Evaluating Scales for Pairwise Comparisons." *Annals of Operations Research* 325: 951–965.
- Colson, A. R., and R. M. Cooke. 2017. "Cross Validation for the Classical Model of Structured Expert Judgment." *Reliability Engineering & System Safety* 163: 109–120.
- Cooke, R. M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.
- Cooke, R. M., and L. L. Goossens. 2008. "TU Delft Expert Judgment Data Base." *Reliability Engineering & System Safety* 93, no. 5: 657–674.
- Cooke, R. M., D. Marti, and T. Mazzuchi. 2021. "Expert Forecasting With and Without Uncertainty Quantification and Weighting: What do the Data Say?" *International Journal of Forecasting* 37, no. 1: 378–387.
- Crawford, G., and C. Williams. 1985. "A Note on the Analysis of Subjective Judgment Matrices." *Journal of Mathematical Psychology* 29, no. 4: 387–405.
- De Pasquale, E., L. Francillout, J.-J. Wasbauer, J. Hatton, J. Albers, and D. Steele. 2009. "ATV Jules Verne Reentry Observation: Mission Design and Trajectory Analysis." In *IEEE Aerospace Conference*, 1–16. IEEE.
- De Persis, C. 2017. "A Risk Assessment Tool for Highly Energetic Break-Up Events During the Atmospheric Re-Entry." PhD diss., Trinity College Dublin.
- Dias, L. C., A. Morton, and J. Quigley. 2018. *Elicitation: The Science and Art of Structuring Judgement*. Springer.
- Dimaio, F., O. Scapinello, E. Zio, et al. 2021. "Accounting for Safety Barriers Degradation in the Risk Assessment of Oil and Gas Systems by Multistate Bayesian Networks." *Reliability Engineering & System Safety* 216: 107943.
- Eggstaff, J. W., T. A. Mazzuchi, and S. Sarkani. 2014. "The Effect of the Number of Seed Variables on the Performance of Cooke's Classical Model." *Reliability Engineering & System Safety* 121: 72–82.
- Falconer, J. R., E. Frank, D. L. L. Polaschek, and C. Joshi. 2022. "Methods for Eliciting Informative Prior Distributions: A Critical Review." *Decision Analysis* 19, no. 3: 189–204.
- Garthwaite, P. H., J. B. Kadane, and A. O'Hagan. 2005. "Statistical Methods for Eliciting Probability Distributions." *Journal of the American Statistical Association* 100, no. 470: 680–701.
- Grimmett, G., and D. J. A. Welsh. 2014. *Probability: An Introduction*. Oxford University Press.
- Gulliksen, H. 1959. "Mathematical Solutions for Psychological Problems." *American Scientist* 47, no. 2: 178–201.
- Guo, S., and S. Sanner. 2010. "Real-Time Multiattribute Bayesian Preference Elicitation With Pairwise Comparison Queries." In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 289–296. JMLR Workshop and Conference Proceedings.
- Hassan, S., J. Wang, C. Kontovas, and M. Bashir. 2022. "An Assessment of Causes and Failure Likelihood of Cross-Country Pipelines Under Uncertainty Using Bayesian Networks." *Reliability Engineering & System Safety* 218: 108171.
- John, A., Z. Yang, R. Riahi, and J. Wang. 2016. "A Risk Assessment Approach to Improve the Resilience of a Seaport System Using Bayesian Networks." *Ocean Engineering* 111: 136–147.
- Joshi, C., F. Ruggeri, and S. P. Wilson. 2018. "Prior Robustness for Bayesian Implementation of the Fault Tree Analysis." *IEEE Transactions on Reliability* 67, no. 1: 170–183.
- Kabir, G., S. Tesfamariam, A. Francisque, and R. Sadiq. 2015. "Evaluating Risk of Water Mains Failure Using a Bayesian Belief Network Model." *European Journal of Operational Research* 240, no. 1: 220–234.
- Keeney, R. L., and D. Vonwinterfeldt. 1991. "Eliciting Probabilities From Experts in Complex Technical Problems." *IEEE Transactions on Engineering Management* 38, no. 3: 191–201.
- Koppenwallner, G., B. Fritsche, T. Lips, T. Martin, L. Francillout, and E. De Pasquale. 2005. "Analysis of ATV Destructive Re-Entry Including Explosion Events." In *4th European Conference on Space Debris*, Vol. 587, 545.
- Langseth, H., and L. Portinale. 2007. "Bayesian Networks in Reliability." *Reliability Engineering & System Safety* 92, no. 1: 92–108.
- Lips, T., S. Lohle, T. Marynowsky, et al. 2010. "Assessment of the ATV-1 Re-Entry Observation Campaign for Future Re-Entry Missions." In *Making Safety Matter*, Proceedings of the fourth IAASS Conference, Vol. 680. id.47.
- Mazloomi, K., and C. Gomes. 2012. "Hydrogen as an Energy Carrier: Prospects and Challenges." *Renewable and Sustainable Energy Reviews* 16, no. 5: 3024–3033.

- Mazzuchi, T. A., W. G. Linzey, and A. Bruning. 2008. "A Paired Comparison Experiment for Gathering Expert Judgment for an Aircraft Wiring Risk Assessment." *Reliability Engineering & System Safety* 93, no. 5: 722–731.
- Merrick, J. R. W., and L. A. McLay. 2010. "Is Screening Cargo Containers for Smuggled Nuclear Threats Worthwhile?" *Decision Analysis* 7, no. 2: 155–171.
- Meyer, M. A., and J. M. Booker. 2001. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. SIAM.
- Misuri, A., G. Landucci, and V. Cozzani. 2020. "Assessment of Safety Barrier Performance in Natech Scenarios." *Reliability Engineering & System Safety* 193: 106597.
- Mkrtchyan, L., L. Podofillini, and V. Dang. 2016. "Methods for Building Conditional Probability Tables of Bayesian Belief Networks From Limited Judgment: An Evaluation for Human Reliability Application." *Reliability Engineering & System Safety* 151: 93–112.
- Munier, N., and E. Hontoria. 2021. *Uses and Limitations of the AHP Method: A Non-Mathematical and Rational Analysis*. Springer.
- Neil, M., M. Taylor, and D. Marquez. 2007. "Inference in Hybrid Bayesian Networks Using Dynamic Discretization." *Statistics and Computing* 17: 219–233.
- Petruni, A., E. Giagloglou, E. Douglas, J. Geng, M. C. Leva, and M. Demichela. 2019. "Applying Analytic Hierarchy Process (AHP) to Choose a Human Factors Technique: Choosing the Suitable Human Reliability Analysis Technique for the Automotive Industry." *Safety Science* 119: 229–239.
- Podofillini, L., B. Reer, and V. N. Dang. 2023. "A Traceable Process to Develop Bayesian Networks From Scarce Data and Expert Judgment: A Human Reliability Analysis Application." *Reliability Engineering & System Safety* 230: 108903.
- Por, H.-H., and D. V. Budescu. 2017. "Eliciting Subjective Probabilities Through Pair-Wise Comparisons." *Journal of Behavioral Decision Making* 30, no. 2: 181–196.
- Saaty, T. L. 1980. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill.
- Saaty, T. L. 1987. "The Analytic Hierarchy Process — What Is It and How It Is Used." *Mathematical Modelling* 9: 161–176.
- Szwed, P., J. R. van Dorp, J. R. Merrick, T. A. Mazzuchi, and A. Singh. 2006. "A Bayesian Paired Comparison Approach for Relative Accident Probability Assessment With Covariate Information." *European Journal of Operational Research* 169, no. 1: 157–177.
- Thurstone, L. L. 1994. "A Law of Comparative Judgment." *Psychological Review* 101, no. 2: 266–270.
- Torres-Toledano, J. G., and L. E. Sucar. 1998. "Bayesian Networks for Reliability Analysis of Complex Systems." In *Progress in Artificial Intelligence—IBERAMIA 98: 6th Ibero-American Conference on AI Lisbon, Portugal, October 5–9, 1998 Proceedings* 6, 195–206. Springer.
- van Dorp, J. R., and J. R. W. Merrick. 2011. "On a Risk Management Analysis of Oil Spill Risk Using Maritime Transportation System Simulation." *Annals of Operations Research* 187: 249–277.
- Verma, A. K., S. Ajit, and H. P. Muruva. 2015. *Risk Management of Non-Renewable Energy Systems*, Vol. 12. Springer.
- Wang, L., and Z. Yang. 2018. "Bayesian Network Modelling and Analysis of Accident Severity in Waterborne Transportation: A Case Study in China." *Reliability Engineering & System Safety* 180: 277–289.
- Yuan, S., M. Yang, G. Reniers, C. Chen, and J. Wu. 2022. "Safety Barriers in the Chemical Process Industries: A State-of-the-Art Review on Their Classification, Assessment, and Management." *Safety Science* 148: 105647.
- Zhang, G., and V. V. Thai. 2016. "Expert Elicitation and Bayesian Network Modeling for Shipping Accidents: A Literature Review." *Safety Science* 87: 53–62.

Appendix A: Derivation of the Prior Weights w_i from Pairwise Comparisons

Our approach follows that of Crawford and Williams (1985), whose derivation of subjective weights from pairwise comparison scores has some advantages from a statistical point of view over the original eigenvector approach due to Gulliksen (1959) and refined for the Analytic Hierarchy Process (AHP) in Saaty (1980) in that the weights can be interpreted as the maximum likelihood estimates from a lognormal model for the ratio of weights. Let q_{ij} be the comparison score (on the scale given in Figure 1 in our case) between events E_i and E_j .

The weights are defined as normalized geometric means of the pairwise comparison scores. In the case of elicitation of all pairwise comparisons, the unnormalized weights are

$$W_i = \left(\prod_{\substack{j=1 \\ j \neq i}}^k q_{ij} \right)^{1/(k-1)},$$

with k the number of primary events, while if only some of the pairwise comparisons have been made then the geometric mean is

$$W_i = \left(\prod_{\substack{j=1 \\ j \neq i \\ q_{ij} \text{ defined}}}^{k_i} q_{ij} \right)^{1/k_i},$$

where k_i is the number of comparisons made between E_i and other events. The weight is then normalized so that they sum to 1:

$$w_i = \frac{W_i}{\sum_j W_j}.$$

Note that this normalization is not strictly necessary as the ratio of two weights is used for the prior specification, nevertheless it does help one to compare between different groups of events that are being elicited.

Appendix B: Derivation of $P(E_i | H_i \cap \mathcal{E}, \mathbf{p})$

Recall that H_i are the ancestor events of E_i and \mathcal{E} are the observed events, so that $H_i \cap \mathcal{E}$ is the set of observed ancestors of E_i . In this section, this probability is derived in the case when $H_i \cap \mathcal{E}$ does not logically imply E_i , for which $P(E_i | H_i \cap \mathcal{E}, \mathbf{p}) = 1$. The possible logical implications are: $E_i = 0 \Leftrightarrow \exists E_j \in \eta_i, E_j = 0$ when E_i is the result of an AND gate, or $E_i = 1 \Leftrightarrow \exists E_j \in \eta_i, E_j = 1$ when E_i is the result of an OR gate. By working up the network from the observed primary events, any such implications are easy to determine.

If E_i is a primary event then $H_i = \emptyset$ and $P(E_i | H_i \cap \mathcal{E}, \mathbf{p}) = p_i$. If E_i is not a primary event then we can write it in terms of its parents η_i as

$$E_i = \begin{cases} \prod_{E_j \in \eta_i} E_j, & \text{if AND gate,} \\ 1 - \prod_{E_j \in \eta_i} (1 - E_j), & \text{if OR gate.} \end{cases}$$

If any events in η_i are in $H_i \cap \mathcal{E}$ then their value is known and the above expressions for E_i conditional on $H_i \cap \mathcal{E}$ are

$$E_i = \begin{cases} \prod_{E_j \in \eta_i, E_j \notin \mathcal{E}} E_j, & \text{if AND gate,} \\ 1 - \prod_{E_j \in \eta_i, E_j \notin \mathcal{E}} (1 - E_j), & \text{if OR gate.} \end{cases} \quad (\text{B.1})$$

Recursive application of Equation (B.1) gives E_i as a function of the unobserved primary events that are ancestors of E_i , which we denote g_i :

$$P(E_i = e | H_i \cap \mathcal{E}, \mathbf{p}) = P(g_i(\{E_j | j = 1, \dots, k; E_j \in H_i; E_j \notin \mathcal{E}\}) = e | \mathbf{p});$$

this function is a sum of products of E_j and $(1 - E_j)$ terms. Since each E_j is binary and independent given \mathbf{p} , the probability of this function is g_i with each E_j replaced by p_j :

$$P(E_i = e | H_i \cap \mathcal{E}, \mathbf{p}) = g_i(\{p_j | j = 1, \dots, k; E_j \in H_i; E_j \notin \mathcal{E}\}).$$

Appendix C: Definition of Node Events for the ATV Fault Tree

Tables C.1 and C.2 give a description of the primary and nonprimary events of the ATV fault tree in Figure 5.

TABLE C.1 | Description of the primary events of the ATV fault tree in Figure 5, partitioned into the three groups for elicitation.

Event	Description
E01*	Propellant valve leakage
E02	Propellant tank destruction
E03	Propellant pipe rupture
E04*	Pressure vessel burst, with sudden release of propellant
E05	Chemical reaction between hypergolic propellants
E06*	Battery overpressure
E07	Battery short circuit
E08	Battery corrosion
E09	Battery overdischarge
E10	Battery overtemperature
E11	Battery cell degradation

* indicates the cornerstone event.

TABLE C.2 | Description of the intermediate and top events of the ATV fault tree in Figure 5.

Event	Description
E12	Chemical reaction of propellant and air
E13	Burst of a battery cell
E14	Top event. Explosion of spacecraft

Appendix D: Prior Elicitation for the ATV Example

The pairwise comparison matrix for events E01, E02, and E03 was elicited as

$$Q_1 = \begin{pmatrix} 1.00 & 1.04 & 1.00 \\ 0.53 & 1.00 & 0.53 \\ 1.00 & 1.04 & 1.00 \end{pmatrix}.$$

The cornerstone event for this set of events is E01 and its interval was elicited to be (0.01,0.04)

The pairwise comparison matrix for events E04 and E05 was elicited as

$$Q_2 = \begin{pmatrix} 1.00 & 1.23 \\ 0.28 & 1.00 \end{pmatrix}.$$

The cornerstone event for this set of events is E04 and its interval was elicited to be (0.005,0.02).

The pairwise comparison for events E06 to E11 was elicited as

$$Q_3 = \begin{pmatrix} 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \end{pmatrix}.$$

In this case, the expert evaluated that each of the primary events that concern a battery were equally likely. The cornerstone event for this set of events is E06 and its interval was elicited to be (0.014,0.055).

The resulting beta parameter values for each event probability prior are given in Table D.1.

Tables E.1 and E.2 give a description of the primary and intermediate events of the fault tree in Figure 8.

TABLE D.1 | Prior elicitation results for the primary events of the ATV fault tree in Figure 5.

Event	Prior weights w_i	Beta prior parameters	Mean	Prior Mode
E01*, E03	0.400	2.06, 134.4	0.015	0.008
E02	0.200	1.73, 237.7	0.007	0.003
E04*	0.815	1.72, 246.2	0.007	0.003
E05	0.185	1.65, 1078.5	0.0015	0.0006
E06* – E11	0.167	2.70, 113.7	0.023	0.015

Appendix E: Definition of Elementary Events for the Tesoro Anacortes Refinery Fault Tree**TABLE E.1** | Description of the primary events of the Tesoro Anacortes refinery fault tree in Figure 8, partitioned into the four groups for elicitation.

Event	Description
E01*	High-temperature hydrogen attack
E02	Difficulty with valve operation during startup
E03	Leaks from heat exchanger during startup not reported
E04	Hydrogen induced cold cracking
E05	Inexperience
E06	Job carried out without permit to work
E07	External supervision failure
E08	Wrong procedure
E09*	Poor construction material for NHT heat exchanger
E10	High mechanical stress
E11	Insufficient instrumentation to measure process conditions
E12*	Long delay in inspection schedule
E13	Inadequate methods for detecting HTHA
E14	Inadequate training of the inspectors to detect HTHA easily
E15	Failure of HTHA inspection on heat exchanger
E16	Failure to detect leaks from heat exchanger flanges
E17	Failed to detect minor release
E18*	Wrong maintenance procedure (Nelson curve methodology)
E19	Delay maintenance operations
E20	HTHA degradation monitoring performed but failed to detect
E21	HTHA degradation monitoring specified but not performed

* indicates the cornerstone event.

TABLE E.2 | Description of the intermediate events of the Tesero Anacortes refinery fault tree in Figure 8.

Event	Description
E22	Process upset
E23	Operator error
E24	Design error factor
E25	High-temperature hydrogen attack effect
E26	Leak
E27	Wrong maintenance procedure
E28	High-temperature hydrogen attack effect
E29	Operational error factor
E30	Inspection error factor
E31	Maintenance error factor
E32	Top event. Failure of the release prevention barrier

Appendix F: Prior Elicitation for the Tesoro Anacortes Refinery Example

TABLE F.1 | Prior elicitation results for the primary events of the Tesoro Anacortes refinery fault tree in Figure 8.

Event	Prior weights w_i	Beta prior parameters	Mean	Prior Mode
E01*	0.097	2.68, 107.0	0.024	0.016
E02	0.111	5.73, 144.3	0.035	0.029
E03	0.065	2.54, 162.0	0.015	0.009
E04	0.214	3.78, 51.8	0.068	0.052
E05	0.127	4.04, 99.6	0.039	0.030
E06	0.127	4.04, 99.6	0.039	0.030
E07	0.061	1.97, 150.1	0.013	0.006
E08	0.197	4.06, 60.3	0.063	0.049
E09*	0.266	1.72, 246.2	0.007	0.003
E10	0.266	1.72, 246.2	0.007	0.003
E11	0.467	1.99, 113.3	0.017	0.009
E12*	0.180	6.46, 140.4	0.044	0.038
E13	0.071	1.59, 182.3	0.009	0.003
E14	0.254	7.41, 102.1	0.068	0.060
E15	0.135	3.05, 119.1	0.025	0.017
E16	0.180	6.46, 140.4	0.044	0.038
E17	0.180	6.46, 140.4	0.044	0.038
E18*	0.369	36.8, 368.0	0.091	0.089
E19	0.232	6.60, 151.6	0.042	0.036
E20	0.167	2.40, 131.6	0.018	0.010
E21	0.232	6.60, 151.6	0.042	0.036

The primary events are split into four groups as defined in the main text, and separate elicitations are done on each, leading to the weights and priors as listed in Table F.1.

Appendix G: Prior Elicitation for the Hydrogen Station Example

Four pairwise comparison matrices were elicited for this example and are given below:

1. The matrix for events related to operational error prevention barrier failure is Q_1 , which had E01 as its cornerstone event with elicited probability interval (0.005, 0.04).
2. The matrix for events related to H_2 containment equipment/component failure is Q_2 . The cornerstone event for this first set of events is E30 and its interval was elicited to be (0.005, 0.04).
3. The matrix for events related to the earthquake/lightning prevention barrier failure is Q_3 . The cornerstone event for this first set of events is E35 and its interval was elicited to be (0.001, 0.03).
4. The matrix for events related to the maintenance prevention barrier failure is Q_4 . The cornerstone event for this first set of events is E10 and its interval was elicited to be (0.02, 0.07).

The resulting beta parameter values for each event probability prior are given in Table G.1.

TABLE G.1 | Prior elicitation results for the primary events of the Hydrogen fault tree.

Event	Beta prior parameters
E01/E30	2.407, 152.026
E02	3.701, 101.993
E03	2.911, 133.019
E04	3.424, 57.285
E05	2.310, 154.854
E06	3.386, 83.478
E07/E15/E16/E21	3.037, 139.249
E08/E20/E22	3.871, 82.296
E09	3.037, 139.249
E10/E14/E24/E25	4.870, 129.244
E11	1.942, 126.383
E12	1.832, 239.873
E13	2.052, 111.928
E17/E18	3.515, 90.768
E19/E23	2.310, 157.901
E26	6.335, 148.442
E27/E28	4.870, 129.244
E29	1.916, 114.633
E31	2.903, 40.698
E32/E33	3.414, 58.547
E34	1.744, 167.730
E35	1.764, 358.316
E36	1.710, 53.556
E37/E38/E39/E40	1.663, 78.687
E41	1.752, 209.443
E42/E43	1.744, 167.730

$$Q_1 = \begin{pmatrix} 1 & 0.28 & 0.53 & 0.21 & 1.04 & 0.28 & 0.53 & 0.28 & 0.53 & 0.53 & 0.53 & 0.28 & 0.28 & 1.04 & 0.28 & 0.53 & 0.28 & 1.04 \\ 1.23 & 1 & 1.23 & 0.28 & 1.23 & 0.53 & 1.23 & 0.53 & 1.23 & 1.23 & 1.23 & 0.53 & 0.53 & 1.23 & 0.53 & 1.23 & 0.53 & 1.23 \\ 1.04 & 0.28 & 1 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1 & 1 & 1 & 0.28 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1.04 \\ 1.52 & 1.23 & 1.23 & 1 & 1.52 & 1.23 & 1.52 & 1.23 & 1.52 & 1.52 & 1.52 & 1.23 & 1.23 & 1.52 & 1.23 & 1.52 & 1.23 & 1.52 \\ 0.53 & 0.28 & 0.53 & 0.21 & 1 & 0.28 & 0.53 & 0.28 & 0.53 & 0.53 & 0.53 & 0.28 & 1.04 & 0.28 & 0.28 & 0.53 & 0.28 & 1.04 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 1 & 1.23 & 0.53 & 1.23 & 1.23 & 1.23 & 1.04 & 1.04 & 1.23 & 0.53 & 1.23 & 0.53 & 1.23 \\ 1.04 & 0.28 & 1 & 0.21 & 1.04 & 0.28 & 1 & 0.28 & 1 & 1 & 1 & 0.28 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1.04 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 1.04 & 1.23 & 1 & 1.23 & 1.23 & 1.23 & 1.04 & 1.04 & 1.23 & 1 & 1.23 & 1 & 1.23 \\ 1.04 & 0.28 & 1 & 0.21 & 1.04 & 0.28 & 1 & 0.28 & 1 & 1 & 1 & 0.28 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1.04 \\ 1.04 & 0.28 & 1 & 0.21 & 1.04 & 0.28 & 1 & 0.28 & 1 & 1 & 1 & 0.28 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1.04 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 0.53 & 1.23 & 0.53 & 1.23 & 1.23 & 1.23 & 1 & 1 & 1.04 & 0.53 & 1.23 & 0.53 & 1.23 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 0.53 & 1.23 & 0.53 & 1.23 & 1.23 & 1.23 & 1 & 1 & 1.04 & 0.53 & 1.23 & 0.53 & 1.23 \\ 0.53 & 0.28 & 0.53 & 0.21 & 0.53 & 0.28 & 0.53 & 0.28 & 0.53 & 0.53 & 0.53 & 0.28 & 0.28 & 1 & 0.28 & 0.53 & 0.28 & 1 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 1.04 & 1.23 & 1 & 1.23 & 1.23 & 1.23 & 1.04 & 1.04 & 1.23 & 1 & 1.23 & 1 & 1.23 \\ 1.04 & 0.28 & 1 & 0.21 & 1.04 & 0.28 & 1 & 0.28 & 1 & 1 & 1 & 0.28 & 0.28 & 1.04 & 0.28 & 1 & 0.28 & 1.04 \\ 1.23 & 1.04 & 1.23 & 0.28 & 1.23 & 1.04 & 1.23 & 1 & 1.23 & 1.23 & 1.23 & 1.04 & 1.04 & 1.23 & 1 & 1.23 & 1 & 1.23 \\ 0.53 & 0.28 & 0.53 & 0.21 & 0.53 & 0.28 & 0.53 & 0.28 & 0.53 & 0.53 & 0.53 & 0.28 & 0.28 & 1 & 0.28 & 0.53 & 0.28 & 1 \end{pmatrix}.$$

$$Q_2 = \begin{pmatrix} 1 & 0.28 & 0.28 & 0.28 \\ 1.23 & 1 & 1.04 & 1.04 \\ 1.23 & 0.53 & 1 & 1 \\ 1.23 & 0.53 & 1 & 1 \end{pmatrix}.$$

$$Q_3 = \begin{pmatrix} 1 & 1.23 & 0.21 & 0.28 & 0.28 & 0.28 & 0.28 & 1.04 & 1 & 1 & 1 \\ 0.28 & 1 & 0.17 & 0.21 & 0.21 & 0.21 & 0.21 & 0.28 & 0.28 & 0.28 & 0.28 \\ 1.52 & 2.55 & 1 & 1.23 & 1.23 & 1.23 & 1.23 & 1.52 & 1.52 & 1.52 & 1.52 \\ 1.23 & 1.52 & 0.28 & 1 & 1 & 1 & 1 & 1.23 & 1.23 & 1.23 & 1.23 \\ 1.23 & 1.52 & 0.28 & 1 & 1 & 1 & 1 & 1.23 & 1.23 & 1.23 & 1.23 \\ 1.23 & 1.52 & 0.28 & 1 & 1 & 1 & 1 & 1.23 & 1.23 & 1.23 & 1.23 \\ 0.53 & 1.23 & 0.21 & 0.28 & 0.28 & 0.28 & 0.28 & 1 & 0.53 & 0.53 & 0.53 \\ 1 & 1.23 & 0.21 & 0.28 & 0.28 & 0.28 & 0.28 & 1.04 & 1 & 1 & 1 \\ 1 & 1.23 & 0.21 & 0.28 & 0.28 & 0.28 & 0.28 & 1.04 & 1 & 1 & 1 \end{pmatrix}.$$

$$Q_4 = \begin{pmatrix} 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 0.53 & 1 & 1.04 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 \\ 0.28 & 0.53 & 1 & 0.28 & 0.28 & 0.28 & 0.28 & 0.28 & 0.28 & 0.28 & 0.28 \\ 0.53 & 1.04 & 1.23 & 1 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 1.04 \\ 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 1.04 & 1.04 & 1.23 & 1.04 & 1.04 & 1.04 & 1.04 & 1 & 1.04 & 1.04 & 1.04 \\ 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 1 & 1.04 & 1.23 & 1.04 & 1 & 1 & 1 & 0.53 & 1 & 1 & 1.04 \\ 0.53 & 1.04 & 1.23 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 0.53 & 1 \end{pmatrix}.$$