

A BAYESIAN MODEL FOR RARE EVENT RISK ASSESSMENT USING EXPERT JUDGMENT ABOUT PAIRED SCENARIO COMPARISONS

P.S. Szwed and J.R. van Dorp

Abstract

When challenged with understanding complex, technological systems, managers often use analysis to characterize risk. Managers use this information to design projects, develop policy, and allocate resources in order to mitigate system risk. This paper presents a Bayesian risk analysis methodology for combining expert judgment with the manager's prior system knowledge to allow identification of risk mitigation opportunities. The model is demonstrated through a study of the nation's largest passenger ferry system and the results compare favorably with previous classical analyses. Hence, this methodology might be useful to engineering managers for rare event risk analysis in other applications and other disciplines as well.

1. Introduction

Risk analysis, also known as risk assessment, is widely recognized as a systematic, science-based process for quantitatively (or qualitatively) describing risk. Risk is commonly described as a combination of the likelihood of an undesirable event (accident) occurring and its consequences. Alternatively, in the context of this paper, it can be expressed as a mathematical combination of an accident's event probability of occurrence and the consequence of that event should it occur (for a detailed discussion of the definition of risk, see Kaplan (1997)). Regardless of exactly how these concepts are defined, however, information about risk is critically important to the decision making process. Engineering managers use the information gained from risk analysis to design projects, develop policy, and allocate resources in their efforts to mitigate system risk.

Often, engineering managers are interested in gaining information about rare events, such as catastrophic accidents or system failures. However, rare event risk information inherently suffers from data scarcity. While the consequences of rare events scenarios may be assessed using engineering based scenario analyses, their frequency data are usually unavailable. In such a case, engineering managers may turn to expert judgment to develop frequency data for these low frequency, high consequence events.

Expert judgment is an informed assessment or estimate (based on the expert's training and experience) about an uncertain quantity or quality of interest. An expert is a person who is recognized (by

peers, decision makers, or others) for their skills, knowledge, and expertise in a particular domain of interest. When treated properly, expert judgment is an important source of information, particularly for risk analysis (See, for example Cooke, 1991).

This paper presents a methodology for engineering managers to incorporate expert judgment as a means of obtaining accident probabilities. Managers can use this risk analysis information for identifying opportunities to mitigate system risk.

2. Mathematical Model

In order to make use of the expert judgment elicited from several experts, a mathematical inference model for aggregation and combination becomes necessary. Such a mathematical model is formulated in this paper and will provide a means for combining the decision maker's prior knowledge with the expert judgment about paired scenario comparisons.

2.1. Accident probability model. When developing probabilities to perform a risk analysis, it is generally desirable to link causal factors to the accident probability. Such models are often referred to as causal models, or accident probability models.

A causal model allows for the estimation of annual accident probability under a specific situation, or scenario. Knowing the probability of an accident per year in a specific scenario and predicting ahead the occurrence of such a scenario helps identify which precautionary measures to consider for accident prevention. For example, if it is known that the probability of an accident is unacceptably high for the scenario when three laden tows (each with twelve barges) meet in close proximity (within a half-mile of each other) during a high river stage on the Mississippi River at Algiers Point in New Orleans, then regulating to prevent such a scenario is warranted. The authorities may require vessel masters to provide one-hour advance notice for passage through that point so that traffic advisors can alert them of this potential scenario and they can slow down to avoid traffic that allows such a scenario to happen. Thus, the advantage of a causal accident probability model is that it allows decisions to be made on which measures to take for ensuring the biggest effect on reducing the probability of an accident.

One word of caution, however, is perhaps appropriate. Consider the probability of an accident during aircraft landing in low visibility at Tucson Airport in Arizona. Without a doubt, the most dangerous situation for landing air traffic is during low visibility. However, if such bad visibility situations hardly ever occurs in Tucson that airport will not get much safer by developing advanced guiding equipment for low visibility situations. Hence, both the accident probability of a particular system state and its rate of incidence are necessary to make decisions on what precautionary measures to implement.

The following accident probability model is postulated:

$$\Pr(\text{Accident} | \underline{X}) = P_0 \exp(\underline{\mathbf{b}}^T \underline{X}) \quad (1)$$

where $\underline{X}^T=(X_1, \dots, X_v)$ is a vector of v situational (causal) variables describing a scenario, $\underline{\mathbf{b}}^T=(\mathbf{b}_1, \dots, \mathbf{b}_v)$ is a parameter vector, and P_0 is a base rate probability. The model in Equation (1) has been proposed in several maritime risk assessments (see Roeleven et al. 1995, Merrick et al. 2000, van Dorp et al. 2001) and assumes that accident probability increases (or decreases) exponentially with a situational variable X_i (rather than linearly). Each situational variable X_i is assumed to be a bounded variable that may be discrete or continuous in nature. Without loss of generality, it is assumed that each situational variable X_i is normalized on a scale of [0, 1]. This normalization allows for comparison of the effects between different situational variables on the accident probability via a comparison of the elements of $\underline{\mathbf{b}}^T$. Positive values of \mathbf{b}_i indicate that the accident probability increases exponentially with X_i and vice versa. Given the model formulation in Equation (1), it follows that P_0 may be interpreted as the inherent accident probability with each situational variable X_i set to zero.

For the purposes of parameter assessment using this expert judgment elicitation, continuous variables will need to discretized in a finite number of steps. Letting s_i denote the discretization depth of situational variable X_i . The number of different scenarios that can be described by \underline{X}^T is given by the product of the discretization depths. Clearly, the number of different scenarios explodes when the number of situational variables, v , increases as well as when the discretization depth of each situational variable, s_i , increases. Since expert judgment will primarily be used to estimate the distribution of $\underline{\mathbf{b}}^T$, an application of Equation (1) would have to be parsimonious both in terms of the number of situational variables, v , and discretization depths per situational variable, s_i .

2.2. Likelihood function. Suppose an expert is asked to compare the accident probability in scenario $\underline{X}^T=(X_1, \dots, X_v)$ compared to scenario $\underline{Y}^T=(Y_1, \dots, Y_v)$ from the question, ‘‘How much more likely is it to obtain an accident in scenario \underline{X}^T compared to scenario \underline{Y}^T ?’’ and suppose the expert responds with the number \tilde{y} . From Equation (1), it may be concluded that:

$$\frac{\Pr(\text{Accident} | \underline{X})}{\Pr(\text{Accident} | \underline{Y})} = \frac{P_0 \exp(\underline{\mathbf{b}}^T \underline{X})}{P_0 \exp(\underline{\mathbf{b}}^T \underline{Y})} = \exp(\underline{\mathbf{b}}^T \{\underline{X} - \underline{Y}\}) = \tilde{y} \quad (2)$$

Taking natural logarithms of both sides of Equation (2) yields:

$$\text{Ln} \left(\frac{\Pr(\text{Accident} | \underline{X})}{\Pr(\text{Accident} | \underline{Y})} \right) = \text{Ln}(\tilde{y}) \quad (3)$$

Note that the base rate P_0 cancels in Equation (2). Hence, only relative probabilities can be derived using this method of paired comparison. In cases where annualized statistics of accident rates are available the base rate P_0 can be determined and relative probabilities can be converted to absolute probabilities (see Merrick et al. 2000, van Dorp et al. 2001).

Further, since human are not perfect assessors; such imperfection can be modeled by the assumption that the expert judgment is uncertain following a particular distribution. For convenience and the purposes of our model, it will be assumed that:

$$\text{Ln} \left(\frac{\Pr(\text{Accident} | \underline{X})}{\Pr(\text{Accident} | \underline{Y})} \right) \sim \text{Normal}(\underline{\mathbf{m}}(\underline{\mathbf{b}}), \underline{\mathbf{s}} \mid \underline{X}, \underline{Y}) \quad (4)$$

where $\underline{\mathbf{m}}(\underline{\mathbf{b}})=\underline{\mathbf{b}}^T \underline{Z}$, $\underline{Z}=(\underline{X}-\underline{Y})$, and $\underline{\mathbf{s}}=C^E$. The parameter C^E will be referred to as the *calibration coefficient* and represents a measure of the expert’s calibration. For example, C^E is small when an expert is well calibrated and large when the expert is not. The difference vector \underline{Z} summarizes the paired comparison of scenario \underline{X}^T and scenario \underline{Y}^T . A series of n difference vectors comprises a questionnaire of n paired comparisons that can be assembled in a matrix $\underline{\mathbf{Z}} = (\underline{Z}_1, \dots, \underline{Z}_n)$. The expert responses, $\text{Ln}(\tilde{\mathbf{y}})$ about these n paired scenario comparisons can be summarized in a vector of judgments, $\underline{\mathbf{D}} = (y_1, \dots, y_n)$, where $y_j = \text{Ln}(\tilde{y}_j)$. The

likelihood of an expert responding D may be derived using Equation (4) as follows:

$$L\{D | \underline{b}, C^E, Z\} = \prod_{j=1}^n \frac{1}{C^E \sqrt{2p}} \exp\left(-\frac{y_j^2 - 2y_j \underline{b}^T \underline{Z}^j + \underline{b}^T \underline{Z}^j (\underline{Z}^j)^T \underline{b}}{2(C^E)^2}\right) \quad (5)$$

Gathering terms in Equation (5) yields:

$$L\{D | \underline{b}, C^E, Z\} \propto \exp\left(-\frac{\sum_{j=1}^n y_j^2 - 2\underline{b}^T \left(\sum_{j=1}^n y_j \underline{Z}^j\right) + \underline{b}^T \left(\sum_{j=1}^n \underline{Z}^j (\underline{Z}^j)^T\right) \underline{b}}{2(C^E)^2}\right) \quad (6)$$

Hence,

$$L\{D | \underline{b}, C^E, Z\} \propto e^{-\left(\frac{1}{2} \underline{b}^T A \underline{b} - \underline{b}^T \underline{b} + c\right)} \quad (7)$$

where:

$$A = \frac{1}{(C^E)^2} \sum_{j=1}^n \underline{Z}^j (\underline{Z}^j)^T \quad (8)$$

$$\underline{b} = \frac{1}{(C^E)^2} \left(\sum_{j=1}^n y_j \underline{Z}^j\right)^T \quad (9)$$

$$c = \frac{1}{2(C^E)^2} \sum_{j=1}^n y_j^2 \quad (10)$$

Note that in Equation (8), $A^T = A$, hence the matrix A is symmetric. Furthermore, for $\underline{x} \neq \underline{0}$ it follows that:

$$\underline{x}^T A \underline{x} = \underline{x}^T \left(\frac{1}{(C^E)^2} \sum_{j=1}^n \underline{Z}^j (\underline{Z}^j)^T\right) \underline{x} = \frac{1}{(C^E)^2} \sum_{j=1}^n (\underline{x}^T \underline{Z}^j)^2 > 0 \quad (11)$$

Hence, with $\text{Range}(\underline{Z}^1, \dots, \underline{Z}^n) = \mathbb{R}^v$ it follows that the matrix A is a positive definite symmetric matrix and therefore invertible.

2.3. Prior distribution. Following the Bayesian paradigm for the parameters, \underline{b}^T , a prior distribution needs to be specified. To allow for a conjugate analysis it will be assumed that \underline{b}^T follows a multivariate normal distribution with a mean value vector $\underline{m}^T = (m^1, \dots, m^v)$ and (co)variance matrix, S .

$$\Pi(\underline{b}) = \frac{\mathbf{1}}{\sqrt{(2p)^n |\Sigma|}} e^{-\frac{1}{2} (\underline{b} - \underline{m})^T \Sigma^{-1} (\underline{b} - \underline{m})} \quad (12)$$

Specification of the initial values for \underline{m}^T and S will be discussed in more detail in the section on elicitation methodology. Typically, however, a manager or decision maker may have a notion of the relative contribution of the situational variables (causal factors) to accident probability described in Equation (1) which can be incorporated as prior information. As a starting point, a diagonal matrix may be specified for S (that is, assuming prior independence between the prior marginal distributions).

The prior mean values of \underline{m}^T and the prior (co)variance matrix S will be updated using the n paired scenario comparisons in the questionnaire and the structure of the likelihood function given in Equations (7) through (10) following a Bayesian analysis. The resulting posterior distribution of \underline{b}^T will be derived next.

2.4. Posterior distribution. Utilizing Bayes theorem it follows from Equations (7) through (10) and (12) that the posterior distribution, $\Pi(\underline{b} | D, C^E, Z)$ is proportional to:

$$\exp\left(-\left(\frac{1}{2} \underline{b}^T A \underline{b} - \underline{b}^T \underline{b}\right) - \frac{1}{2} (\underline{b} - \underline{m})^T \Sigma^{-1} (\underline{b} - \underline{m})\right) \quad (13)$$

Realizing that,

$$\frac{1}{2} (\underline{b} - \underline{m})^T \Sigma^{-1} (\underline{b} - \underline{m}) = \frac{1}{2} \underline{b}^T \Sigma^{-1} \underline{b} - \frac{1}{2} \underline{m}^T \Sigma^{-1} \underline{b} - \frac{1}{2} \underline{b}^T \Sigma^{-1} \underline{m} + \frac{1}{2} \underline{m}^T \Sigma^{-1} \underline{m} \quad (14)$$

and writing,

$$\underline{b}^T \underline{b} = \frac{1}{2} \underline{b}^T \underline{b} + \frac{1}{2} \underline{b}^T \underline{b} \quad (15)$$

It follows that the posterior distribution, $\Pi(\underline{b} | D, C^E, Z)$, is proportional to $\exp(-g(\underline{b}))$ where:

$$g(\underline{\mathbf{b}}) = \left(\frac{1}{2} \underline{\mathbf{b}}^T \underline{\mathbf{A}} \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{b}}^T \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{b}}^T \underline{\mathbf{b}} \right) + \left(\frac{1}{2} \underline{\mathbf{b}}^T \underline{\Sigma}^{-1} \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{m}}^T \underline{\Sigma}^{-1} \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{b}}^T \underline{\Sigma}^{-1} \underline{\mathbf{m}} \right) \quad (16)$$

Let the updated (co)variance matrix, $\underline{\mathbf{S}}^u$, be defined such that:

$$\left(\underline{\Sigma}^u \right)^{-1} = \underline{\mathbf{A}} + \underline{\Sigma}^{-1} \quad (17)$$

Note that $\underline{\mathbf{A}}$ is symmetric and $\underline{\Sigma}^{-1}$ is symmetric. Thus, it follows that $\left(\underline{\mathbf{S}}^u \right)^{-1}$ is symmetric. Introducing $\underline{\mathbf{m}}^u$ implicitly as follows

$$\frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^{-1} \underline{\mathbf{m}} + \underline{\mathbf{b}} \right) = \frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{m}}^u \quad (18)$$

for arbitrary $\underline{\mathbf{b}}$, it follows that:

$$\left(\underline{\mathbf{S}}^{-1} \underline{\mathbf{m}} + \underline{\mathbf{b}} \right) = \left(\underline{\mathbf{S}}^u \right)^{-1} \underline{\mathbf{m}}^u \Leftrightarrow \left(\underline{\mathbf{A}} + \underline{\mathbf{S}}^{-1} \right) \underline{\mathbf{m}}^u = \underline{\mathbf{S}}^{-1} \underline{\mathbf{m}} + \underline{\mathbf{b}} \Leftrightarrow \left(\underline{\mathbf{S}} \underline{\mathbf{A}} + \underline{\mathbf{I}} \right) \underline{\mathbf{m}}^u = \underline{\mathbf{m}} + \underline{\mathbf{S}} \underline{\mathbf{b}} \Leftrightarrow \underline{\mathbf{m}}^u \quad (19)$$

Using Equation (17) and (19) it follows that Equation (12) may be rewritten as:

$$g(\underline{\mathbf{b}}) = \frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{b}} - \frac{1}{2} \left(\underline{\mathbf{m}}^u \right)^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{m}}^u + \frac{1}{2} \underline{\mathbf{m}}^T \underline{\Sigma}^{-1} \underline{\mathbf{m}} + c \quad (20)$$

Using the fact that $\left(\underline{\mathbf{S}}^u \right)^{-1}$ is symmetric, it follows that the posterior distribution, $\Pi(\underline{\mathbf{b}}|\underline{\mathbf{D}})$, is proportional to:

$$e^{- \left(\frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{b}} - \frac{1}{2} \left(\underline{\mathbf{m}}^u \right)^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{b}} - \frac{1}{2} \underline{\mathbf{b}}^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{m}}^u + \frac{1}{2} \left(\underline{\mathbf{m}}^u \right)^T \left(\underline{\Sigma}^u \right)^{-1} \underline{\mathbf{m}}^u \right)} = e^{- \frac{1}{2} \left(\underline{\mathbf{b}} - \underline{\mathbf{m}}^u \right)^T \left(\underline{\Sigma}^u \right)^{-1} \left(\underline{\mathbf{b}} - \underline{\mathbf{m}}^u \right)} \quad (21)$$

where the updated or posterior mean is:

$$\underline{\mathbf{m}}^u = \left(\underline{\Sigma} \underline{\mathbf{A}} + \underline{\mathbf{I}} \right)^{-1} \underline{\mathbf{m}} + \left(\underline{\Sigma} \underline{\mathbf{A}} + \underline{\mathbf{I}} \right)^{-1} \underline{\mathbf{b}} \quad (22)$$

and the updated or posterior (co)variance matrix is:

$$\underline{\Sigma}^u = \left(\underline{\mathbf{A}} + \underline{\Sigma}^{-1} \right)^{-1} \quad (23)$$

Thus, the posterior distribution, $\Pi(\underline{\mathbf{b}}|\underline{\mathbf{D}})$, can be recognized as multivariate normal. Since $\Pi(\underline{\mathbf{b}})$ and $\Pi(\underline{\mathbf{b}}|\underline{\mathbf{D}})$ belong to the same family of distributions, it follows that the above analysis is conjugate. Note that the posterior uncertainty, $\underline{\mathbf{S}}^u$, is a function of the prior uncertainty, $\underline{\mathbf{S}}^{-1}$, the calibration coefficient, $\underline{\mathbf{C}}^E$, and the particular questions to which the expert responded, $\underline{\mathbf{Z}}$. Note also that $\Pi(\underline{\mathbf{b}})$ can be updated even when an expert does not compare a full set of situational vectors.

This mathematical model formulation will provide the foundation for updating the decision maker's prior distribution with the expert judgments. Combination using multiple experts follows immediately from the single expert case above using sequential Bayesian updating.

3. Elicitation Methodology

Expert judgment is an important source of data for risk analysis. And like any source of data, expert judgment must be carefully and purposefully treated in order to yield meaningful information. For without proper treatment, all sorts of unnecessary and unwanted randomness and bias may be introduced and, as a result, the integrity of the data may become tainted. Therefore, the treatment of expert judgment must be methodical, not unlike the handling of observed or measured data during experimentation.

The general process for gathering, and ultimately, combining the expert data with the prior knowledge of the decision maker can be broken down into seven basic steps:

1. Pre-elicitation preparation
2. Elicitation of expert judgment (including calibration)
3. Extraction of decision maker prior distribution
4. Computation of expert posterior distribution
5. Diagnosis of expert judgment
6. Selection of experts
7. Combination of expert judgment

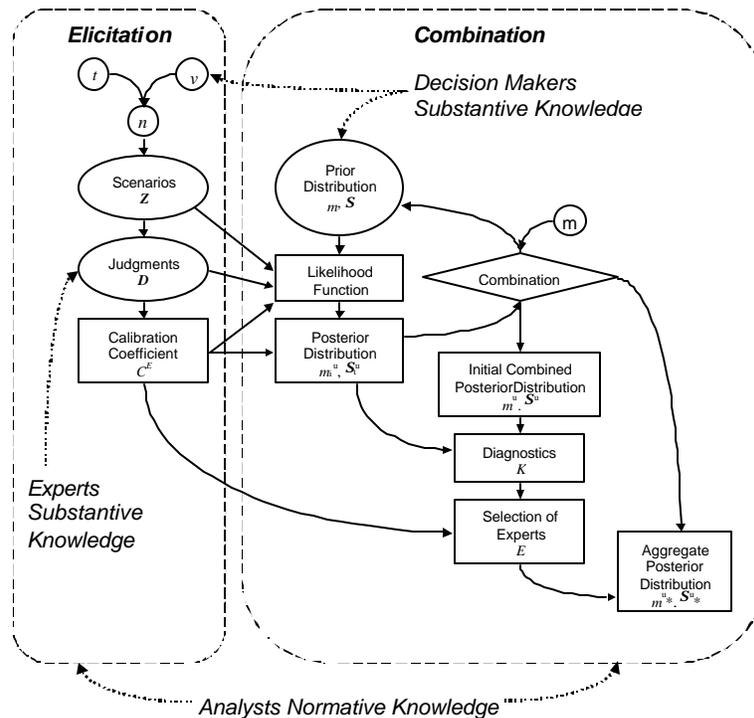
Exhibit 1 provides an overview of the approach. There are two overarching phases to the model presented – the *elicitation* phase and the *combination* phase. Generally speaking, these phases proceed from left to right and the processes, or steps, proceed top-down. Note that the first two steps (elicitation and pre-elicitation preparation) are part of the elicitation phase on the left and the combination phase, on the right, is made up of the remaining five steps (extraction of decision maker prior, computation of expert posterior, diagnosis of expert judgment, selection of experts, and combination of expert judgment).

This exhibit is intended to give the reader an understanding of how all steps in this methodology fit together. The circles represent inputs to the model, the squares outputs, and the diamond the iterative combination process. The arrows indicate relationships or flows of data between the different steps. For example, as described earlier, the likelihood function is derived from the scenarios, Z , the judgments, D , and the resulting calibration coefficient, C^E .

Once the situational variables have been identified, the manager and the analyst work together to determine how to define each of them. Usually, situational variables are defined by their states, and the number of states indicates the degree of discretization. The desire for increased discretization will result in increased model complexity and this tradeoff should be carefully considered.

As a general rule of thumb, each elicitation session should last no longer than one hour (see, Cooke

Exhibit 1. Overview of modeling approach including seven steps for handling expert judgment



Step 1: Pre-elicitation preparation. This step involves identifying the appropriate situational variables for modeling the system and preparing a survey instrument, or questionnaire, for eliciting the expert judgment about those variables.

Together, the engineering manager, who possesses general system knowledge, and the analyst, who has knowledge about processes for handling expert data, determine how many, v , and which situational variables to consider in order to learn about the system. This information may be drawn directly from the manager's substantive knowledge, obtained by polling a group of internal experts, or by some other means. In this case, the manager is seeking the probability of a certain system failure mode occurring and this will inevitably dictate some variables to consider. Examples of situational variables may include natural properties such as temperature or pressure, system settings, or states within the system.

(1991)). This will allow experts the opportunity to answer about 50 – 60 questions in a single session. Any more than this may lead to expert fatigue which can manifest itself in inaccuracy or indifference.

Most systems being studied through expert judgment are non-trivial and therefore judgments about all possible combinations cannot realistically be elicited. So, a representative, but non-exhaustive sampling of the set of all possible scenario pairs must be made. Referring back to Exhibit 1, the number of situational variables, v , and the contact time available with the experts (t in total), will dictate the number of questions, n . Distributing the number of questions, n , evenly across all variables, v , is one way to develop an effective, balanced questionnaire.

Each question will consist of a pair of scenarios that differ in only one dimension (i.e. only one situational variable will change states). Using a 9-point Likert scale, experts can judge which scenario will

have a greater effect on system performance. These scenario pairs are compressed into a single difference vector, \underline{Z} , by subtracting the first scenario vector from the second (\underline{Z} will contain zeros across all situational variables except the one differing between the two scenarios). All difference vectors will be assembled into a matrix, \mathbf{Z} , that forms the basis of the questionnaire.

Step 2: Expert elicitation (including calibration). The elicitation step itself is fairly straightforward. Experts compare pairs of scenarios (defined by their difference vectors, \underline{Z}) and judge the relative likelihood for a specific accident in a particular failure mode. For example, an expert may judge that the first scenario is twice as likely as the second scenario. This response will be recorded as a “-2.” The sign represents which scenario is more likely to result in an accident. A response of “1” will indicate indifference between the two scenarios.

Whenever expert judgment is relied upon for decision-making, the reliability of those experts is immediately called into question. In order for the decision to be sound, the expert judgment upon which it is made must not only be sound, it must be elicited and combined in a coherent and defensible process. This can be accomplished in many ways. One way is to examine the calibration and entropy of the experts (Cooke 1991).

In this methodology, a calibration coefficient will be used to indicate whether the sense of the expert’s response is *correct*. So for each expert and each question, the sign of the non-zero element of the difference vector, \underline{Z} , and the sign of the expert’s corresponding (transformed) judgment from the \mathbf{D} matrix will be compared and an indicator will be assigned as follows:

$$I = \frac{|D * Z|}{(D * Z)} \quad (24)$$

To get a measure of overall expert calibration, the indicators are summed across all questions to develop the expert’s calibration coefficient:

$$C^E = \frac{\left| n - \sum_{i=1}^n I_i \right|}{n} \quad (25)$$

The calibration coefficient, C^E , ranges from [0, 2]. A *well-calibrated* expert will have a low value of C^E and the calibration coefficient of an *uncalibrated* expert will be greater than “1”.

Step 3: Extraction of decision maker prior distribution. Before any combination can begin, the manager, or decision maker, must provide some notion about the situational variables as well as uncertainty about those judgments. Otherwise, the expert judgments can be combined in classical fashion (see von Dorp et al. 2001) without using prior decision maker knowledge. However, in cases where there is prior knowledge, this information should be used in the model so that all sources of information are exploited. The decision maker’s prior distribution will take the form of a mean value vector, \underline{m} , and a (co)variance matrix, \mathbf{S} . The decision maker’s knowledge can be extracted using many different methods. Information can be taken from previous historical studies of the system under examination, a compilation of findings on the individual situational variables of interest, global statistics translated for the application, or in rare cases through eliciting the expert judgment of the decision maker.

Step 4: Computation of the expert posterior distribution. Applying the mathematical model presented earlier provides the expert’s posterior distribution. Each expert’s posterior distribution is computed using the Bayesian paradigm. The posterior mean value estimates, \underline{m}^u , and (co)variance matrix, \mathbf{S}^u , are determined using Equations (22) and (23) respectively. It is helpful to develop a spreadsheet or employ some form of software coding to manage the considerable matrix manipulations involved.

Step 5: Diagnosis of expert judgment. As discussed previously, not all experts provide information that is equally useful. Some experts may have hidden or even overt bias, some may have misunderstood the instruction of the elicitation process or context of the accident being evaluated, and there’s always the possibility that those selected as experts possess no particular expertise of relevance. Tversky and Kahneman (1977) provide excellent examples of expert bias and how to treat them. Regardless of why the expert may not perform as desired (i.e. provide sufficient information about the system), some form of diagnostic is necessary to gauge their performance. This section provides three forms of diagnostics.

The calibration coefficient in Equation (25) can be used to coarsely diagnose the calibration of the expert.

Relative entropy can be developed using information theoretic approaches (see Kullback 1959 or Soofi and Retzer 2002). Cross-entropy between two distributions can be used to understand the goodness of fit, or measure of closeness, between a model distribution and a reference distribution (ideally the true distribution). In this case, the ideal distribution is

what is sought and therefore unknown. Cross-entropy can be derived from the discrimination information function (Kullback and Liebler 1951):

$$K(f : g) \equiv \int \log \frac{f(y)}{g(y)} dF(y) \quad (26)$$

where $f(y)=dF(y)$ is a probability density (mass) function $f_Y(y)$, absolutely continuous with respect to the reference distribution, $g(y)$. Cross-entropy is widely used in part due to its intuitive appeal and analytical tractability. $K(f;g) \geq 0$ is a measure of discrepancy between the two distributions. The equality holds if and only if $f(y)=g(y)$ everywhere.

Since the Bayesian event model is normal in form and the analysis is conjugate, both $f(y)$ and $g(y)$ are normal in form. The resulting cross-entropy will be:

$$K(f, g) = \frac{1}{2} \log \left(\frac{s^2}{s^{*2}} \right) + \frac{1}{2} \left(\frac{s^{*2}}{s^2} - 1 + \frac{(m^* - m)^2}{s^2} \right) \quad (27)$$

such that it will be a function of all statistics for both the reference distribution, $g(y)$, and the comparison distribution, $f(y)$. Now, in terms of the Bayesian event model, comparing the posterior distribution of an individual expert to that of the aggregated *consensus* (i.e. the combination of all-expert posterior distributions) as the reference distribution yields the following results:

$$K(\Pi_i^u, \Pi^u) = \frac{1}{2} \log \left(\frac{s_i^{u^2}}{s^{u^2}} \right) + \frac{1}{2} \left(\frac{s^{u^2}}{s_i^{u^2}} - 1 + \frac{(m^u - m_i^u)^2}{s^{u^2}} \right) \quad (28)$$

where Π_i^u is the posterior distribution for expert i and the reference distribution, Π^u is the combined distribution for all experts, or the aggregated *consensus*. Even though the reference distribution contains the compared distribution, it is important for it to remain that way because we would like a standard reference distribution when comparing experts using our inter-expert diagnostic.

Similarly, another diagnostic can be developed by comparing the posterior distribution of an individual expert, Π_i^u , to the prior distribution, Π . The cross-entropy term in that case follows as:

$$K(\Pi_i^u, \Pi) = \frac{1}{2} \log \left(\frac{s_i^{u^2}}{s^2} \right) + \frac{1}{2} \left(\frac{s^2}{s_i^{u^2}} - 1 + \frac{(m - m_i^u)^2}{s^2} \right) \quad (29)$$

Each expert's cross-entropy terms and calibration coefficient will be examined in the next step.

Step 6: Selection of experts. Using the notion of calibration and entropy as a form of scoring rule the experts may be selected. However, because the calibration coefficient measures only sense and not magnitude and, therefore, is not highly precise, the disqualification of an expert should not be considered without examining the other diagnostics. The other diagnostics (specifically the cross-entropy measures) will provide a measure of an expert's judgment in relation to either an aggregated *consensus* (i.e. the combination of all experts) or to that of the decision maker. There too, caution should be exercised so that experts with different perspectives are not disqualified merely on the basis of this differing perspective, particularly when that expert is well calibrated.

In summary, decision makers may choose to forgo selecting an expert, but only if that expert is poorly calibrated and exhibits high entropy. If the decision maker does not have sufficient rationale for disqualifying an expert, all experts should be included in the combination.

Step 7: Combination of expert judgment. The combination step is merely a compilation of other steps used throughout the process. The combination step iteratively operationalizes the mathematical model presented earlier.

This expert judgment combination step again involves the sequential combination of the selected expert posterior distributions with the decision maker's prior distribution through the likelihood function using Bayes theorem. The posterior of the first expert based on the decision maker's prior serves as the prior of the second expert and this is repeated until all selected expert's likelihood functions have been used in the updating process. The order of experts does not matter.

The final result will be a combined posterior distribution for each situational variable represented by the statistics, \underline{m}^{u*} and S^{u*} . These values will be used with the accident probability model to better understand the system and apply mitigation measures in reducing the relevant risks.

4. Application Example

Whenever a model is presented, a relevant application example often helps illustrate function and utility. Additionally, a properly selected application example often provides unique insights not otherwise obtainable. In this study, an example was selected from the maritime domain where accidents are relatively rare events and historical data are sparse. The risk analysis of the Washington state ferry system (van Dorp et al. 2001) was chosen to demonstrate the Bayesian accident

probability model presented in this paper. A brief summary of the results will be presented here.

During the summer of 1998, the Washington state legislature called upon the Washington State Transportation Commission to establish a Blue Ribbon Panel to assess the adequacy of provisions for passenger and crew safety aboard the Washington State Ferry (WSF) system. The George Washington University was selected to be a part of the consulting team for assessing the adequacy of passenger and crew safety on the WSF, evaluating the level of risk present in the WSF system, and developing recommendations for prioritized risk reduction measures to improve the level of safety in the WSF system.

To lend some perspective to the risk analysis (both the original and the model presented here), the system under investigation will be briefly described. The WSF system is the largest ferry system in the U.S. Serving the central Puget Sound, Admiralty Inlet, and the San Juan Islands; the WSF system consists of 27 vessels operating between 20 terminals on 10 routes at that time. Annual ridership has been estimated at about 26.2 million passengers annually (more passengers each year than the domestic U.S. passenger rail service, Amtrak) during that period. The Puget Sound is also the home to several major ports for domestic and Pacific Rim trade with thousands of commercial ships arriving and departing annually.

Step 1: An accident probability model was developed for the system based on the ten situational variables in Exhibit 2.

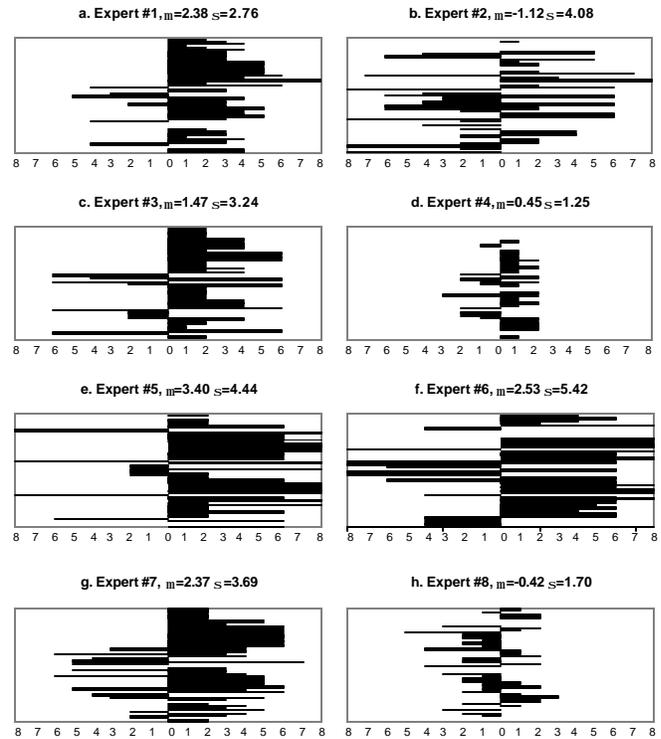
Exhibit 2. Description of the situational variables for the 10 waterway attributes examined in the WSF study.

Name	Description	Discretization
FR_FC	Ferry route-class combo	13 combos
TT_1	1 st interacting vessel type	13 types
TS_1	Scenario of 1 st interaction	4 meetings
TP_1	Proximity of 1 st vessel	Binary
TT_2	2 nd interacting vessel type	13 types
TS_2	Scenario of 2 nd interaction	4 meetings
TP_2	Proximity of 2 nd vessel	Binary
VIS	Visibility	Binary
WS	Wind speed	Binary
WD	Wind direction	Binary

Sixty difference vectors, \mathbf{Z} , were developed such that only one situational variable differed for each question. These situational variables were assembled into a useable questionnaire format.

Step 2: The questionnaires were administered to eight experts consisting of WSF captains, Puget Sound pilots, and Coast Guard officers. The following exhibit provides a graphical representation of their responses.

Exhibit 3. Individual expert judgment responses to 60 questions about the relative probability of collision between scenario pairs in the WSF system given a propulsion failure



Calibration coefficients, C^E , were calculated based on the responses above, \mathbf{D} , and the difference vector matrix, \mathbf{Z} , as described in Equations (24) and (25). The following exhibit shows each expert's calibration coefficient.

Exhibit 4. Calibration coefficients for eight experts participating in WSF study

Expert	C^E	Relative Rank
1	0.83	5
2	1.05	7
3	0.87	6
4	0.60	2
5	0.52	1
6	0.72	4
7	0.67	3
8	1.25	8

Note that experts #1, 4, 5, 6, and 7 were the most highly calibrated. This seems to intuitively correspond to Exhibit 3 since the other experts had *balanced* responses even though the questionnaire favored the second scenarios (on the right).

Step 3: Two cases were examined using the methodology presented in this paper. The first involved an *uninformed* decision maker (using flat priors) and the second an *informed* one (using the results of the WSF study as prior information; see Exhibit 5).

Exhibit 5. *Informed* decision maker prior estimates of situational variables

Name	Mean	Variance
FR_FC	0.000	5.000
TT_1	1.503	0.209
TS_1	0.642	0.299
TP_1	3.330	0.311
TT_2	0.606	0.467
TS_2	1.177	0.325
TP_2	2.736	0.310
VIS	3.343	0.310
WS	1.775	0.310
WD	3.737	0.621

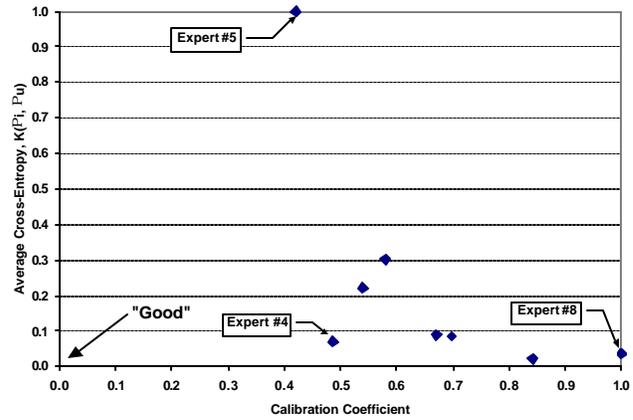
The flat priors for the *uninformed* decision maker’s situational variable have zero mean estimates and large variance much like the FR_FC waterway attribute above.

Step 4: Next, the experts’ posterior distributions were computed for each of the variables. This was accomplished using the above priors and the model developed previously. Generally, the expert posterior distributions varied depending upon the responses. The results of these responses are too extensive for presentation in this paper, but the combined responses will be provided in Step 7. (Readers interested obtaining complete results or a spreadsheet version of the model should contact the author.)

Step 5: The calibration and entropy were examined for each expert. Refer directly to Exhibit 4 for expert calibration. Cross-entropy was calculated for each expert and each variable according to all reference distributions (two cases of decision maker priors and aggregated *consensus* posterior distributions). Exhibit 6 provides one typical representation of expert performance by plotting normalized calibration in terms of normalized average cross-entropy.

Step 6 All eight experts were examined in terms of their calibration and cross-entropy. The experts who exhibited relatively high average cross-entropy (expert #5) demonstrated good calibration. Conversely, the experts who exhibited relatively high, or poor, calibration (expert #8) yielded low cross-entropy. Most experts had relatively good calibration and entropy (expert #4 for example). Therefore, none of the expert had poor calibration coupled with high cross-entropy and all experts were selected for inclusion in the final aggregation.

Exhibit 6. Normalized expert calibration vs. average cross-entropy (*informed* decision maker case)



Step 7: Since no experts were eliminated in either case (*uninformed* or *informed*), the final combination of expert judgment is exactly the same as the combination used in determining the reference distribution for the third diagnostic ($\Pi(\mathbf{b}^T \mathbf{D})$). Exhibit 7 presents the results for the case where the judgments of all eight experts were used to update an *informed* decision maker, or manager, using the Bayesian accident model presented in this paper.

Exhibit 7. Parameter estimates and uncertainty for the probability of a collision given a propulsion failure on the WSF vessel found using the Bayesian accident probability model for an *informed* decision maker

Name	Mean	Variance
FR_FC	2.032	0.081
TT_1	1.869	0.060
TS_1	2.609	0.109
TP_1	1.737	0.117
TT_2	1.863	0.221
TS_2	2.306	0.125
TP_2	1.510	0.116
VIS	1.207	0.116
WS	2.419	0.116
WD	2.422	0.339

The *uninformed* case yields similar results to those of the *informed* case, but the mean values are shifted slightly and the variances are considerably larger. When compared with the results of the classical study (von Dorp et al., 2001), the results are similar. The classical study provides a model that emphasizes natural waterway attributes and proximity attributes. This study developed a model that emphasizes natural waterway attributes and scenario attributes.

Conclusions

The model and methodology presented in this paper provides a means for gathering expert judgment through the well-established pairwise comparison technique and combining those judgments with the decision maker's prior system knowledge using the Bayesian paradigm to provide an accident probability model for estimating accident probability. By taking advantage of previous system knowledge, engineering managers have the ability to obtain more complete risk information to account for rare event scenarios. This is useful when safety, accident prevention, and risk mitigation are part of system design and policy development.

References

- Clemen, R. and Winkler, R. "Combining Probability Distributions from Experts in Risk Analysis," *Risk Analysis*, Vol. 19, (1999), pp. 187-203.
- Cooke, R. M. *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford Univ. Press (1991).
- Genest, C. and Zidek, J.V. "Combining Probability Distributions: A Critique and an Annotated Bibliography," *Statistical Science*, Vol. 1, (1986), pp. 114-148.
- Kaplan, S. "The Words of Risk Analysis," *Risk Analysis*, Vol. 17, (1997), pp. 407-417.
- Kullback, S. *Information Theory and Statistics*, Wiley and Sons (1959).
- Kullback, S., and Leibler, R.A. "On Information and Sufficiency," *Annals of Mathematical Statistics*, Vol. 22, (1951), pp. 79-86.
- Merrick, J.; van Dorp, J.R.; Harrald, J.; Mazzuchi, T.; Spahn, J.; and Grabowski, M. "A Systems Approach to Managing Oil Transportation Risk in Prince William Sound," *Systems Engineering*, Vol. 3, (2000), pp. 1:128-142.
- Meyer, M.A. and Booker, J.M. *Eliciting and Analyzing Expert Judgment: A Practical Guide*, second edition, Society for Industrial and Applied Math (2001).
- Pulkkinen, U. "Bayesian Analysis of Consistent Paired Comparisons," *Reliability Engineering and System Safety*, Vol. 43 (1994), pp. 1-16.
- Roeleven, D.; Kok, M.; Stipdonk, H.L.; and de Vries, W.A. "Inland Waterway Transport: Modeling the Probabilities of an Accident," *Safety Science*, Vol. 19 (1994), pp. 203-215.
- Saaty, T. *The Analytic Hierarchy Process*. McGraw-Hill (1980).
- Soofi, E.S. and Retzer, J.J. "Information Indices: Unification and Application," *Journal of Econometrics*, Vol. 107 (2002), pp. 17-40.
- van Dorp, J.R., Merrick, J.R.W., Harrald, J.R., Mazzuchi, T.A., and Grabowski, M. "A Risk Management Procedure for the Washington State Ferries," *Risk Analysis*, Vol. 21, (2001), pp. 127-142.
- Vargas, L.G. "An Overview of the Analytic Hierarchy Process and its Applications," *European J. of Operational Research*, Vol. 48, (1990), pp. 2-8.
- Vargas, L.G. and Whitaker, R.W. "Decision Making by the Analytic Hierarchy Process: Theory and Applications," *European J. of Operational Research*, Vol. 48, (1990), entire volume.

Commander **Paul S. Szwed** has worked for the U.S. Coast Guard in a variety of capacities for 16 years. He currently serves as the manager of an engineering design review team at the Coast Guard's Guard Marine Safety Center in Washington, D.C. He received his D.Sc. in Engineering Management from the George Washington University. He also holds an M.Eng. degree in Naval Architecture and Offshore Engineering and an M.S. degree in Industrial Engineering and Operations Research from the University of California at Berkeley, an M.S. degree in Environmental Management from the University of San Francisco, and a B.S. degree in Ocean Engineering from the U.S. Coast Guard Academy.

J. Rene van Dorp received his D.Sc. in Operations Research from the George Washington University. He holds an M.S. degree in mathematics and computer science (cum laude) from the Delft university of Technology in the Netherlands. He is currently an Assistant Professor in Engineering Management and Systems Engineering at the George Washington University. His research interests include risk management analysis, probabilistic risk analysis, reliability analysis, accelerated life testing, and engineering/expert judgment. He has also worked as a risk analyst in the Netherlands.