

# A Bayesian Paired Comparison Approach for Relative Accident Probability Assessment with Covariate Information

P. Szwed<sup>†</sup>, J. Rene van Dorp<sup>\*</sup>, J.R.W.Merrick<sup>•</sup>, T.A. Mazzuchi<sup>°</sup> and A. Singh<sup>°</sup>.

*Abstract* – **One of the challenges managers face when trying to understand complex, technological systems (in their efforts to mitigate system risks) is the quantification of accident probability, particularly in the case of rare events. Once this risk information has been quantified, managers and decision makers can use it to develop appropriate policies, design projects, and/or allocate resources that will mitigate risk. However, rare event risk information inherently suffers from a sparseness of accident data. Therefore, expert judgment is often elicited to develop frequency data for these high-consequence rare events. When applied appropriately, expert judgment can serve as an important (and, at times, the only) source of risk information. This paper presents a Bayesian methodology for assessing relative accident probabilities and their uncertainty using paired comparison to elicit expert judgments. The approach is illustrated using expert judgment data elicited for a risk study of the largest passenger ferry system in the U.S.**

*Keywords* – **Applied Probability, Expert Judgment, Risk Analysis.**

---

<sup>†</sup> Same as Corresponding Author, As of August 2003: Department of Management, U.S. Coast Guard Academy, 31 Mohegan Avenue, New London CT 06320-8103 E-mail: PSzwed@cga.uscg.mil

<sup>\*</sup> Corresponding Author, Engineering Management and Systems Engineering Department, The George Washington University, 1776 G Street NW, Suite 110, Washington DC, 20052, E-mail: dorpir@gwu.edu.

<sup>•</sup> Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, P.O. Box 843083, VCU, Richmond, VA 23284-2014. E-mail: jmerric@vcu.edu.

<sup>°</sup> Same as corresponding author, E-mail Mazzuchi: mazzu@gwu.edu, E-mail Singh: amitas@gwu.edu

## 1. INTRODUCTION

The concepts of risk analysis and management is becoming more and more relevant in our complex technological environment. Numerous papers and books have been written in the last 20 years on this topic (see, e.g., Shrader-Frechette (1985), Paté-Cornell (1996), Kumamoto and Henley (1996), Kaplan (1997), Koller (2000) and Bedford and Cooke (2001)). Risk analysis, also known as risk assessment, is widely recognized as a systematic, science-based process for quantitatively describing risk (see, e.g. Vose (1996)). Risk, itself, is commonly defined as a quantitative measure combining the likelihood of the occurrence of an undesirable event (accident) and its consequences. Assessment of risk may be separated into the quantitative assessments of accident probabilities and consequences. Kaplan (1997) among others discusses the definition of risk in more detail. Regardless of exactly how these quantitative measures are combined into a single risk measure, separate information about accident probability and consequences are critically important to managers who are charged with risk mitigation because different risk interventions follow from accident probability reduction and consequence reduction.

The quantification of risk models for policy and decision-making often requires the elicitation of expert judgments (see, e.g. Moslesh et al. (1988), Bonano et al. (1989), Morgan and Henrion (1991) and Cooke (1991)). In fact, as long as the fundamental mechanisms that drive a system remain poorly known, the encoding of expert knowledge will be required (see, Pate-Cornell (1996)). Nevertheless, as noted by Anderson et al. (1999), expert judgment must be used with care. It is not evidence per se, but an individual's or group's inference based on available evidence. Kahneman et al. (1982) ( a Nobel Prize winner in 2002) discuss the numerous biases and heuristics that are introduced when humans process information and attempt to provide judgments.

Winkler (1996) points out that due to the general belief that "several heads are better than one", information is usually elicited from several experts. Numerous techniques exist for the aggregation of multiple experts' responses (see, e.g., Morris (1974), Winkler (1981), Genest and

Zidek (1986), Clemen (1989), Mendel and Sheridan (1989), Cooke (1991) and DeWispelare et al. 1995)). In recent reviews of the techniques, Clemen and Winkler (1990, 1999) note that often the simple aggregation techniques may work just as well as the more complex methods. The Bayesian paradigm, however, seems to supply at the present the most natural and unambiguous approach towards the aggregation problem while addressing uncertainty in the expert judgment at the same time.

While a number of different elicitation methods are available (see, e.g. Cooke (1991) for an excellent overview), the paired comparisons elicitation method seems to be quite popular. The elicitation method to be discussed in this paper belongs to this class. In the next section we reflect on the origins of the paired comparisons elicitation method.

### **1.1. Paired Comparisons Elicitation Approaches**

Origins of this class can be traced back to Thurstone's (1927a,b) pioneering work where Weber's and Fechner's law were used to quantify the intensity of psychophysical stimuli using a discriminative process. An extension of this concept found application in the field of consumer research (see, Bradley (1953)) via the Bradley-Terry (1952) paired comparisons method. An examination of the latter method is provided by Cooke (1991), among other numerous sources.

Another popular paired comparison elicitation technique is called the Analytical Hierarchy Process (AHP) developed by Saaty (1977, 1980). The AHP Process is primarily used for the construction of value functions  $V(\underline{X})$  involving multiple contributing factors  $\underline{X} = (X_1, X_2, \dots, X_p)$  (see, e.g. Foreman and Selly (2002)). The construction of a value function in this manner extends the construction of a utility function based on paired comparisons. The theoretical foundation for developing the latter has been provided by the Nobel Laureate G. Debrue (see, e.g., Debrue (1986)). The popularity of the elicitation methods above can perhaps be contributed to the observation that experts are more comfortable making paired comparisons rather than directly assessing a quantity of interest. It should however be mentioned that paired

comparisons may lead occasionally to the so-called Simpson paradox-lack of transitivity (see, Simpson (1951)).

To the best of our knowledge, Pulkkinen (1993, 1994a,b) was first to introduce a Bayesian paired comparison aggregation method for the elements of a multivariate random vector  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  by multiple experts. Experts are asked to compare the pair of random variables  $\beta_i$  to  $\beta_j$ ,  $i \neq j$ ,  $i = 1, \dots, p$  and respond in terms of an indicator function  $1_{[\beta_i \geq \beta_j]}$  (i.e. 1 when the expert judges  $\beta_i \geq \beta_j$  and 0 otherwise). The paired judgments in Pulkkinen's analysis are assumed to be consistent. Pulkkinen's (1993, 1994a,b) exposition is mainly theoretical and limited to a discussion of mathematical properties of the aggregation method, but mentions that applications of his method in the reliability engineering and system safety domain are self-evident.

We shall report herein on what appears to be a novel paired comparison elicitation method for accident probabilities. We take as an application of this approach an actual case study "The Washington State Ferry (henceforth WSF) Risk Assessment" where paired comparisons were elicited from experts. The next section discusses an overview of the WSF Risk Assessment (see also van Dorp et al. (2001) for a more detailed description).

## 1.2. Overview of the WSF Risk Assessment

The WSF system is the largest ferry system in the United States. In 1997, total ridership for the ferries serving the central Puget Sound region was nearly 23 million, a 4 percent increase over 1996 ridership, and more passengers than Amtrak, the US passenger rail carrier, handles in a year. Figure 1 shows the current ferry routes for the central Puget Sound region. This map illustrates the ferry system's role in linking together the Washington State highway system in the Puget Sound region.

In part due to the introduction of high speed ferries, the State of Washington established an independent Blue Ribbon Panel to assess the adequacy of requirements for passenger and crew safety aboard the Washington State Ferries. On July 9, 1998, the Blue Ribbon Panel engaged a

consultant team from The George Washington University and Rensselaer Polytechnic Institute/Le Moyne College to assess the adequacy of passenger and crew safety in the WSF system, to evaluate the level of risk present in the WSF system, and to develop recommendations for prioritized risk reduction measures which, once implemented, can improve the level of safety in the WSF system. The probability of ferry collisions in the WSF system was assessed using a dynamic simulation methodology that extends the scope of available data with expert judgment.

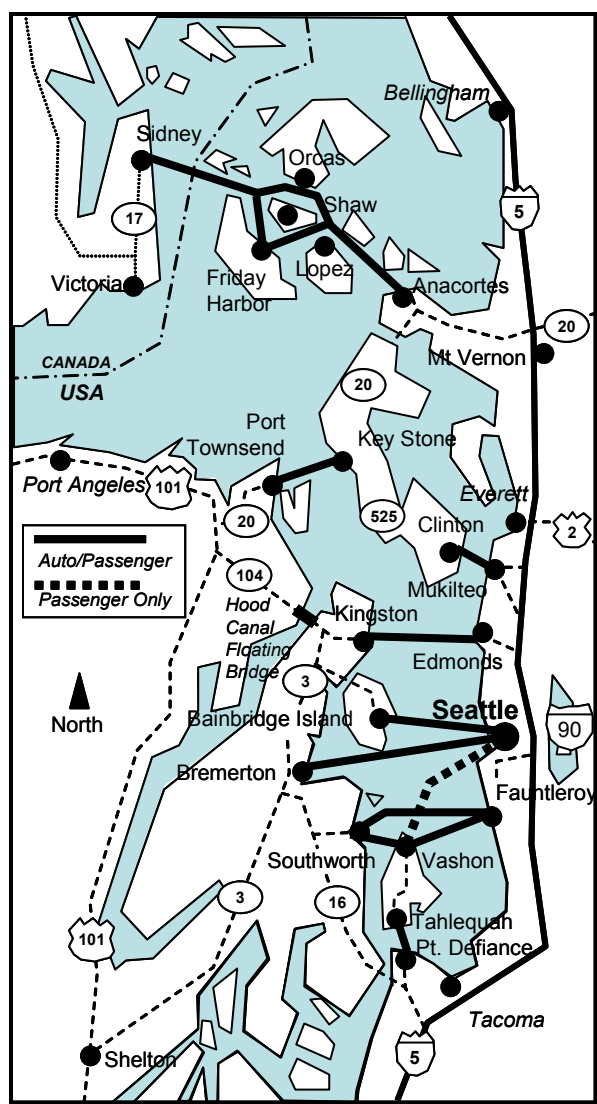


Figure 1. Washington State Ferry System Map

Experts were selected amongst WSF captains and WSF first mates who had extensive experience with all 13 different ferry routes over an extended period of time (more than 5 years). During the WSF risk assessment in 1998 expert responses to paired comparisons were aggregated by taking geometric means of their responses and using them in a classical log linear regression analysis approach to assess relative collision probabilities. The classical analysis conducted during the WSF risk assessment only resulted in point estimates of relative collision probabilities. We shall improve on the previous classical analysis by providing distributional results on these relative collision probabilities by developing a Bayesian inference engine for the paired comparison questionnaires administered during the WSF Risk Assessment. This is in compliance with the almost classical "speaking the truth in risk assessment" argument (see, e.g., Kaplan, 1997, p. 412) originating from the early 1980's when the International Society for Risk Analysis was founded: *"Since the truth is, we always have uncertainty, we say that speaking in probability curves is telling the truth"*. The paired comparison elicitation method developed herein is not limited to the maritime domain and may generally be applicable to relative accident probability estimation when limited or no data is available. The research conducted by us is part of a larger project funded by the National Science Foundation to address uncertainty in large scale maritime risk assessments in a coherent manner.

### **1.3. Bayesian Paired Comparison Approach for Relative Accident Probabilities**

Similar to the AHP process, we are interested in the functional relationship between contributing factors  $\underline{X} = (X_1, X_2, \dots, X_p)$  and an accident probability (rather than a value function). Our accident probability behaves much like a value function. That is, not only is the order amongst different sets of contributing factors (or covariates)  $\underline{X}$  important, but also the differences in their values. Whereas Pulkkinen's focus (1993, 1994a,b) is on the multivariate distribution of a random vector  $\underline{\beta}$ , our focus is more applied and based on the distribution of an accident probability  $Pr(Accident|Incident, \underline{X})$  defined by

$$Pr(Accident|Incident, \underline{X}) = P_0 Exp(\underline{\beta}^T \underline{X}) \quad (1)$$

where  $\underline{X} = (X_1, X_2, \dots, X_p)$  describes a system state during which an incident (e.g. a mechanical failure) occurred. The accident probability model (1) has been proposed in previous maritime risk assessments (see, e.g., Roeleven et al. (1995), Merrick et al. (2000) and Van Dorp et al. (2001)), resembles the well-known proportional hazards model originally proposed by Cox (1972) and builds on the assumption that accident risk behaves exponentially rather than linearly with changes in covariate values. Our goal is to establish the uncertainty distribution of the accident probability  $Pr(Accident|Incident, \underline{X})$  in entirety rather than a point estimate. Similarly to Pulkkinen (1993, 1994), our aggregation method of the expert judgment paired comparisons will follow the Bayesian paradigm. A questionnaire of paired comparisons is used to elicit the relative contribution of the elements of  $\underline{X}$  to the accident probability and update its uncertainty, initially captured by (1) and a prior multivariate distribution of the random vector  $\underline{\beta}$ .

The Bayesian analysis conducted herein exploits the structure of (1) to result in a conjugate analysis (i.e. the prior and posterior distributions belong to the same family of distributions) involving a multivariate normal prior for the parameter vector  $\underline{\beta}$  and a univariate gamma prior on an expert's precision (or, perhaps more appropriately, imprecision). In Section 2, we provide some background surrounding the use of the accident probability model (1) in large maritime risk assessments drawing primarily from the Washington State Ferry (WSF) Risk assessment (see, Van Dorp et al. (2001)). The likelihood of the expert responses to the paired comparison questionnaire is presented in Section 3. The prior distribution on the parameter vector  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  and the expert judgment's precision is discussed in Section 4. The conjugate analysis deriving the posterior distribution of  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  and the expert judgment's precision is presented in Section 5. In addition, parameter uncertainty in  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  and uncertainty in the expert judgment is propagated through the accident probability model  $Pr(Accident|Incident, \underline{X})$  to arrive at closed form expressions for prior and posterior distributions of relative accident probabilities. A calculation example is presented using expert

judgment data elicited during the WSF risk assessment (see, Van Dorp et al. (2001)) in Section 6. Some concluding remarks are provided in Section 7.

## 2. ACCIDENT PROBABILITY MODEL

An accident is not a single event, but can be considered to be the culmination of a series of cascading events (see, Garrick (1984)) starting with a triggering incident. In the maritime accident probability model in Merrick et al. (2000) and Van Dorp et al. (2001), triggering incidents have been further categorized as mechanical failures and human errors. Accidents and triggering incidents occur within the context of a system defined by ever changing combinations of contributing factors. Contributing factors may be further classified in organizational factors (OF) and situational factors (SF). In the WSF risk assessment an example of an organizational factor is a specific ferry route and ferry class combination (since operating teams are assigned by ferry class and route), whereas examples of situational factors are the changing weather conditions and traffic patterns while a ferry is underway. Figure 2 provides an example of an accident probability model, the time sequence of the accident event chain and the influence of contributing factors on this chain. The accident probability model in Figure 2 is based on the notion of conditional probability. The levels of conditional probability reflected in Figure 2 are

- $Pr(OF, SF)$ : the probability that a particular set of organizational and situational factors occur in the system
- $Pr(Incident|OF)$ : the probability that an incident occurs given the organizational factors and
- $Pr(Accident| Incident, OF, SF)$ : the probability that an accident occurs given that a triggering incident has occurred under the organizational and situational factors.



To perform an assessment of the annual accident risk and its uncertainty using this model, each term in the probability model and its uncertainty distribution needs to be estimated and propagated through the law of total probability.

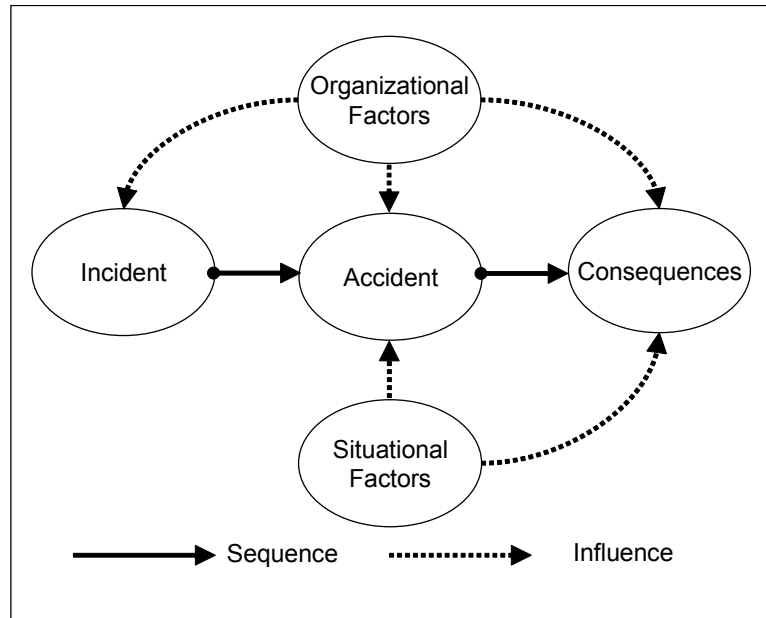


Figure 2. The accident probability model

Bayesian simulation techniques may be used to assess the exposure distribution of contributing factors, i.e. the distribution of  $Pr(OF, SF)$  (see, e.g., Merrick et al. 2003). As more data tends to be available at the triggering incident level rather than at the accident level, the distribution of  $Pr(Incident|OF)$  may be assessed utilizing the traditional Bayesian estimation techniques. For example, by updating a Poisson process for the occurrences of mechanical failures with a gamma prior distribution on the rate of occurrences, with mechanical failure data. In this paper we shall concentrate on the assessment of  $Pr(Accident| Incident, OF, SF)$  where the contributing factors  $(OF, SF)$  are described by a vector  $\underline{X} = (X_1, X_2, \dots, X_n)$  and only limited accident data is available.

For example in the WSF Risk Assessment only two collisions occurred over a period of 11 years (see, Van Dorp et al. (2001)). As an example, Table 1 provides a description of the

contributing factors used in the WSF risk assessment. The heading "discretization" in Table 1 indicates the different number of possible scenarios for a contributing factor. For example, any of the following four traffic scenarios applies to the factor TS\_1: meeting, passing, crossing astern and crossing the bow. Note that from the description in Table 1 it follows that a WSF Ferry may be interacting with more than one vessel at the same time.

Table 1. Description of 10 contributing factors to  $Pr(Accident| Incident, \underline{X})$  in WSF Risk Assessment

	<b>Designation</b>	<b>Description</b>	<b>Discretization</b>
$X_1$	FR_FC	Ferry route - class combination	26
$X_2$	TT_1	1st Interacting vessel type	13
$X_3$	TS_1	Scenario of 1st interaction	4
$X_4$	TP_1	Proximity of 1st interaction	Binary
$X_5$	TT_2	2nd Interacting vessel type	5
$X_6$	TS_2	Scenario of 2nd interaction	4
$X_7$	TP_2	Proximity of 2nd interaction	Binary
$X_8$	VIS	Visibility	Binary
$X_9$	WD	Wind direction	Binary
$X_{10}$	WS	Wind speed	Continuous

The calculation model suggested for the accident probability given contributing factors  $\underline{X}$  is given by (1), where  $\underline{X} \in [0, 1]^p$ ,  $\underline{\beta} \in \mathbb{R}^p$  and  $P_0 \in (0, 1)$ . The covariate  $X_i$ ,  $i = 1, \dots, p$  are normalized so that  $X_i = 1$  describes the "worst" case scenario and  $X_i = 0$  describes the "best" case scenario. For example, for the 10-th attribute  $X_{10}$  in Table 1,  $X_{10} = 1$  relates to the maximum wind speed typically observed in the given geographic area and  $X_{10} = 0$  relates to a wind speed of 0 knots. The calibration constant  $P_0$  equals the accident probability when  $\underline{X} = \underline{0}$ .

In the previous example (dealing wind speed) the ordering from worst to best as it relates to an accident probability is self-evident, but this may not be the case for, for example, the second covariate in Table 1 indicating *vessel class*. In that case, a scale needs to be constructed ranking

interacting vessel types according to a level of concern (from a collision perspective) when WSF captains or first mates encounter them on the water way. In the WSF risk assessment (see, Van Dorp et al. (2001)) a separate Bradley-Terry (1952) paired comparison procedure was used for that purpose, involving also WSF captains and first mates as experts. The Bradley Terry procedure assumes that each object  $i$  is associated with a true scale value. For example, the value  $X_2(i)$  is the scale value associated with the vessel type  $i$ ,  $i = 1, \dots, 13$ , of the first interacting vessel (see, Table 1). Next, experts are asked to respond whether a traffic interaction with a vessel of type  $j$  would be preferred over that of type  $i$ ,  $i, j = 1, \dots, 13$ ,  $j \neq i$ . Figure 3 presents the resulting scale values  $X_2(i)$ ,  $i = 1, \dots, 13$ , from the Bradley-Terry analysis for the second covariate in Table 1 involving 13 different vessel types.

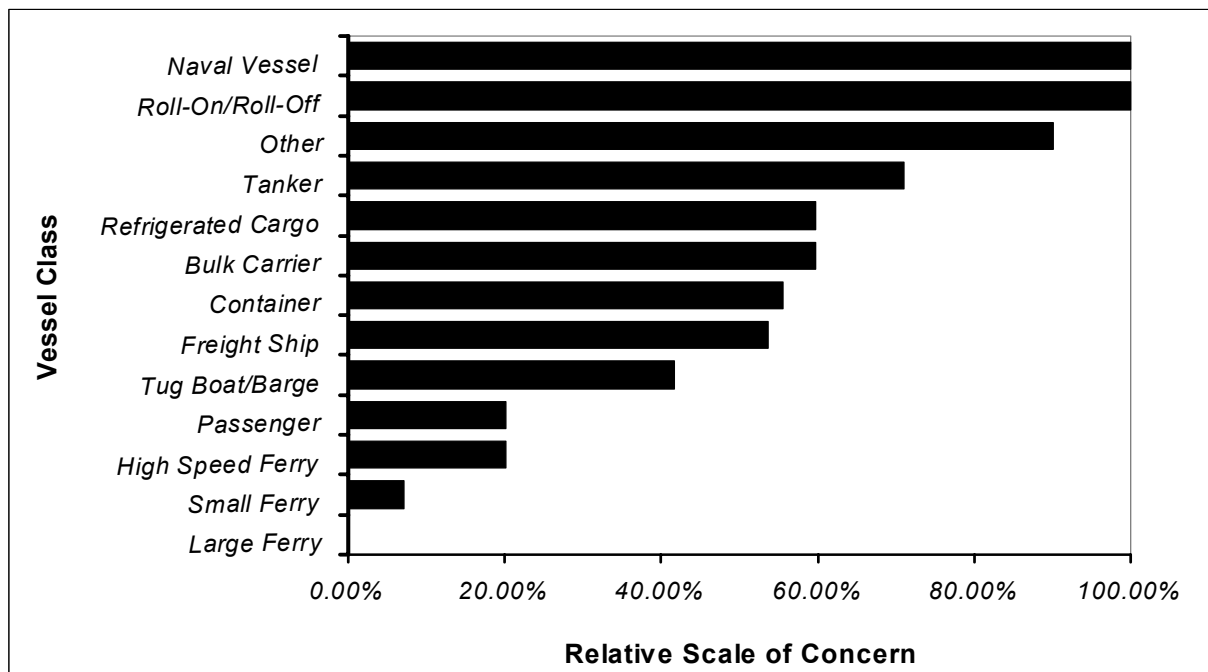


Figure 3. Constructed Covariate Scale for Interacting Vessels

It follows from Figure 3 that when encountering these vessel types, the level of concern is the largest when encountering a Naval Vessel and the smallest when encountering a large WSF Ferry. One may argue that the construction of the scale in Figure 3 introduces a motivational bias

as Washington State Ferries consistently received the lowest rankings. On the other hand, when these results were presented to the Blue Ribbon Panel on Ferry Safety (see, Van Dorp et al. (2001)) it was noted that WSF Ferries interacting with WSF Ferries is an everyday occurrence involving common actors, rather than the far less frequent Naval Vessel whose captain is unknown to the WSF Ferry operators. In a similar manner, covariate scales had to be constructed for  $X_1, X_3, \dots, X_7$  to allow for the use of (1) and their contribution to  $Pr(Accident|Incident, \underline{X})$ . Note that, some of the elements in  $\underline{X}$  may be used to describe interaction effects. For example, if  $X_1$  relates to the Ferry Route-Ferry Class combination and  $X_2$  relates to the traffic type of the first interacting vessel, one may introduce an 11-th factor  $X_{11}$  equal to  $X_1 \cdot X_2$  to model that accident probability may increase more (or less) as a result of a combined increase in both  $X_1$  and  $X_2$ . In principle more complex interactions can be included.

Having selected the contributing factors for  $Pr(Accident|Incident, \underline{X})$  and having constructed the covariate scales of the elements in  $\underline{X}$ , a paired comparison questionnaire may be designed, each question comparing two different system states  $\underline{X}^1$  and  $\underline{X}^2$ . Figure 4 provides an example question appearing in one of the questionnaires used in the WSF risk assessment (see, Van Dorp et al. (2001)). For ease of comparison  $\underline{X}^1$  and  $\underline{X}^2$  (Situations 1 and 2 in Figure 4) differ only in one contributing factor. By circling a "1" or the midpoint of the scale, the expert has indicated that he/she judges the likelihood of a particular accident type to be the same in system state  $\underline{X}^1$  as in system state  $\underline{X}^2$ . If he/she circles, e.g. the number 9 towards Situation 2 (i.e. to the right) we interpret that he/she considers the likelihood of a particular accident type to be 9 times as high in  $\underline{X}^2$  as in  $\underline{X}^1$  given a particular incident has occurred. In the WSF risk assessment (see Van Dorp et al. 2001) the focus was on collision accidents and incidents were further classified as propulsion, steering and navigation equipment failures, and human error.

Question: 32

48

Situation 1	Attribute	Situation 2
Super	Ferry Class	-
SEA-BAI	Ferry Route	-
Naval Vessel	1st Interacting Vessel	-
Crossing the bow	Traffic Scenario 1st Vessel	-
1 to 5 miles	Traffic Proximity 1st Vessel	-
Deep Draft	2nd Interacting Vessel	-
Crossing the bow	Traffic Scenario 2nd Vessel	-
1 to 5 miles	Traffic Proximity 2nd Vessel	-
more than 0.5 mile	Visibility	less than 0.5 mile
Along Ferry	Wind Direction	-
40 knots	Wind Speed	-
9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9		
Situation 1 is worse	<=====X=====>	Situation 2 is worse

Figure 4. An example question appearing in one of the questionnaires used in the WSF risk assessment

If one is interested in paired comparison of accident risk between two different systems states  $\underline{X}^1$  and  $\underline{X}^2$  given an incident occurred, it is sufficient to estimate the parameter vector  $\underline{\beta}$ , as the relative accident probability in  $\underline{X}^1$  compared to  $\underline{X}^2$  (denoted by  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$ ) follows from (1) yielding

$$P(\underline{X}^1, \underline{X}^2 | \underline{\beta}) = \text{Exp}\{\underline{\beta}^T (\underline{X}^1 - \underline{X}^2)\}. \quad (2)$$

Note that the relative accident probability is not restricted to the support  $[0, 1]$  but

$P(\underline{X}^1, \underline{X}^2 | \underline{\beta}) \in [0, \infty]$  and

$$\text{Log}\{P(\underline{X}^1, \underline{X}^2 | \underline{\beta})\} = \underline{\beta}^T (\underline{X}^1 - \underline{X}^2) \in (-\infty, \infty) \quad (3)$$

If one is interested in an absolute accident probability one is required to estimate  $P_0$  in addition to the parameter vector  $\underline{\beta}$ . The calibration constant  $P_0$  may be estimated by applying the law of total probability using all probability terms in Figure 2, the maritime system simulation and average annual accident data, for example the 2 collisions over an 11 year period as was the case in the WSF risk assessment (see, Van Dorp et al. (2001)). In the following sections, the discussion will be limited to presenting prior and posterior analysis for relative accident probabilities given by (2).

### 3. THE LIKELIHOOD OF A SINGLE EXPERT'S RESPONSE

Let  $Y_j$  be the response of an expert to a paired comparison question  $j$ , comparing two different situations  $\underline{X}_j^1$  and  $\underline{X}_j^2$  in terms of accidents proneness given an incident has occurred (e.g. a navigation equipment failure), i.e.

$$Y_j = \text{Experts response to ratio } \frac{Pr(\text{Accident}|\text{Incident}, \underline{X}_j^1)}{Pr(\text{Accident}|\text{Incident}, \underline{X}_j^2)}$$

Define

$$Z_j = \text{Log } Y_j, j = 1, \dots, n$$

to be experts' log response to question  $j$ . The response of the expert to such a question is uncertain and will assumed to be normal distributed such that

$$(Z_j | \mu_j, r) \sim N(\mu_j, r) \quad (4)$$

where  $r = 1/\sigma^2$  is the precision that does not depend on the question index  $j$  and  $\sigma$  is the standard deviation of the normal distribution in (4),  $\sigma > 0$ . This is the most common uncertainty model encountered in practice, which seems to be appropriate at least given the support indicated by (3). Utilizing the structure of the accident probability model (1) and (3) we set

$$\mu_j = q_j^T \underline{\beta}, \quad (5)$$

where  $q_j = (\underline{X}_j^1 - \underline{X}_j^2)$  is a  $p \times 1$  vector. The relevance of the paired comparison of situations  $\underline{X}_j^1$  and  $\underline{X}_j^2$  appears in the distribution (4) of  $Z_j$  only via the vector  $q_j$  (cf. (5)). The likelihood of an expert responding  $z_j$  to question  $j$ ,  $f_{Z_j}(z_j)$ , follows from (4) as

$$f_{Z_j}(z_j) \propto \sqrt{r} \exp\left\{-\frac{r}{2}(z_j - \mu_j)^2\right\}, \quad (6)$$

where the symbol  $\propto$  means "being proportional to".

Suppose the expert answers  $n$  paired comparison questions defined by the vectors  $q_j = (\underline{X}_j^1 - \underline{X}_j^2)$ ,  $j = 1, \dots, n$ , define  $Q$  to be the  $p \times n$  questionnaire matrix

$$Q = [q_1, \dots, q_n], \quad (7)$$

and let the answers of the expert be summarized in the  $n \times 1$  response vector

$$\mathcal{Z} = (z_1, \dots, z_n). \quad (8)$$

Assuming conditional independence between an individual expert's responses to different questions given the precision  $r$  and parameter vector  $\underline{\beta}$ , the likelihood  $\mathcal{L}(\mathcal{Z} | \underline{\beta}, r, Q)$  of an expert responding  $\mathcal{Z}$  to questionnaire  $Q$ , may be derived from (6) as being proportional to

$$r^{\frac{n}{2}} \exp \left\{ -\frac{r}{2} \left( \sum_{j=1}^n z_j^2 - 2 \sum_{j=1}^n \mu_j z_j + \sum_{j=1}^n \mu_j^2 \right) \right\}. \quad (9)$$

The conditional independence assumption implies that the sole source for *dependence* amongst an individual expert's responses to the different questions are the unknown precision  $r$  and the unknown parameter vector  $\underline{\beta}$  (which seems to be reasonable.) In addition, in a Bayesian analysis the standard conditional independence assumption given the unknown parameters is quite natural and is often not explicitly mentioned (see, e.g. Pulkkinen (1994a)). Substituting  $\mu_j = q_j^T \underline{\beta}$  (cf. (5)) in (9), yields

$$\begin{aligned} \mathcal{L}(\mathcal{Z} | \underline{\beta}, r, Q) &\propto r^{\frac{n}{2}} \exp \left\{ -\frac{r}{2} \left( \sum_{j=1}^n z_j^2 - 2 \left[ \sum_{j=1}^n q_j z_j \right]^T \underline{\beta} + \underline{\beta}^T \left[ \sum_{j=1}^n q_j q_j^T \right] \underline{\beta} \right) \right\} \\ &\propto r^{\frac{n}{2}} \exp \left\{ -\frac{r}{2} (c - 2 \underline{b}^T \underline{\beta} + \underline{\beta}^T A \underline{\beta}) \right\}. \end{aligned} \quad (10)$$

where

$$A = \sum_{j=1}^n q_j q_j^T; \quad \underline{b} = \sum_{j=1}^n q_j z_j; \quad c = \sum_{j=1}^n z_j^2 \quad (11)$$

The matrix  $A$  will be referred to as the design matrix of the questionnaire  $Q$ . Note that,  $A^T = A$ . Hence,  $A$  is symmetric. Furthermore, for  $\underline{x} \neq \underline{0}$  it follows that

$$\underline{x}^T A \underline{x} = \underline{x}^T \left[ \sum_{j=1}^n \underline{q}_j \underline{q}_j^T \right] \underline{x} = \sum_{j=1}^n \underline{x}^T \underline{q}_j \underline{q}_j^T \underline{x} = \sum_{j=1}^n (\underline{x}^T \underline{q}_j)^2 > 0 \quad (12)$$

as long as the columns  $\underline{q}_j$  of  $Q$  span  $\mathbb{R}^p$ . If the latter condition holds for the questionnaire matrix  $Q$ , it follows from (12) that  $A$  is positive definite and symmetric and therefore invertible.

#### 4. PRIOR DISTRIBUTION

To allow for a conjugate Bayesian analysis a multivariate normal/gamma prior is proposed for the joint distribution of  $(\underline{\beta}, r)$  similar to the one described in West and Harrison (1989).

Conjugate Bayesian analysis is motivated mainly by the desire to simplify calculations of the posterior probability. Nevertheless it proved to be a reliable approach yielding invariably meaningful results.

A *Gamma* $(\frac{\alpha}{2}, \frac{\nu}{2})$  will be defined on the precision  $r$  and is given by the pdf

$$\prod(r | \alpha, \nu) = \frac{\frac{\nu}{2}^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2})} r^{\frac{\alpha}{2}-1} \exp(-\frac{r}{2}\nu). \quad (13)$$

The distribution of  $(\underline{\beta} | r)$  is assumed to be multivariate normal (*MVN*) with a prior  $p \times 1$  dimensional mean vector  $\underline{m}$  and  $p \times p$  precision matrix  $r\Delta$ , i.e.

$$\prod(\underline{\beta} | r) \propto r^{\frac{p}{2}} \exp\left\{-\frac{r}{2}(\underline{\beta} - \underline{m})^T \Delta (\underline{\beta} - \underline{m})\right\}. \quad (14)$$

Hence, from the structure of the *MVN* it follows that  $(r\Delta)^{-1}$  is the variance covariance matrix of  $(\underline{\beta} | r)$ . The joint prior distribution on  $(\underline{\beta}, r)$  follows from (13) and (14) to be

$$\prod(\underline{\beta}, r) \propto r^{\frac{\alpha}{2}-1} \exp(-\frac{r}{2}\nu) \times r^{\frac{p}{2}} \exp\left\{-\frac{r}{2}(\underline{\beta} - \underline{m})^T \Delta (\underline{\beta} - \underline{m})\right\}. \quad (15)$$

The marginal distribution of  $\underline{\beta}$  may be derived from (15), yielding

$$\prod(\underline{\beta}) \propto \left[1 + \frac{1}{\nu}(\underline{\beta} - \underline{m})^T \Delta (\underline{\beta} - \underline{m})\right]^{-\frac{\alpha+p}{2}} \quad (16)$$



and is recognized as a  $p$ -dimensional multivariate  $t$ -distribution with  $\alpha$  degrees of freedom, location vector  $\underline{m}$  and precision matrix

$$\frac{\alpha}{\nu} \Delta. \quad (17)$$

Note that,  $\alpha/\nu$  in (17) is the mean value of the precision  $r \sim \text{Gamma}(\frac{\alpha}{2}, \frac{\nu}{2})$  and hence the marginal distribution of  $\underline{\beta}$  integrates the precision given by (13) and that of  $(\underline{\beta} | r)$  (cf. (14)).

The marginal distribution of  $\beta_i, i = 1, \dots, p$ , follows from (16) as a univariate  $t$ -distribution with  $\alpha$  degrees of freedom, location parameter  $m_i$  and precision parameter  $\frac{\alpha}{\nu} \delta_{ii}$ , given by

$$\prod (\beta_i) \propto \left[ 1 + \frac{\delta_{ii}}{\nu} (\beta_i - m_i)^2 \right]^{-\frac{\alpha+1}{2}}, \quad (18)$$

where  $\delta_{ii}$  is the  $i$ -th diagonal element of the precision matrix  $\Delta$ . From (16) and (3) follows that the log-relative probability  $\text{Log}\{P(\underline{X}^1, \underline{X}^2 | \underline{\beta})\}$  has a prior  $t$ -distribution with mean

$$\underline{m}^T (\underline{X}^1 - \underline{X}^2) \quad (19)$$

and precision

$$\frac{\alpha}{\nu} (\underline{X}^1 - \underline{X}^2)^T \Delta (\underline{X}^1 - \underline{X}^2). \quad (20)$$

The prior distribution of the relative probability  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  (cf. (2)) thus follows a log- $t$  distribution (see, e.g., McDonald and Butler (1987)) with parameters specified via (19) and (20).

#### 4.1. Prior Parameter Specification

A prior chi-squared distribution with  $\alpha$  degrees of freedom (equivalent to a gamma distribution  $\text{Gamma}(\frac{\alpha}{2}, \frac{\nu}{2})$  with  $\nu = 1$ ) will be selected for the prior distribution of precision  $r$  requiring only specification of the prior parameter  $\alpha$ . From (13) it follows that  $E[r | \alpha, \nu=1] = \alpha$ . The prior parameter  $\alpha$  will be set equal to the reciprocal of the variance of an expert responding to the  $n$  paired comparison questions completely at random and depends on the scale that is used in the paired comparison questions to collect the expert responses. In the example of Figure 3, responses range from  $\frac{1}{9}, \frac{1}{8}, \dots, \frac{1}{2}, 1, 2, \dots, 9$  totaling 17 possible responses per question. With

different responses being equally like and mutually independent for an expert responding at random and noting that  $\text{Log}^2(x^{-1}) = \text{Log}^2(x)$  it follows that a priori

$$\alpha = E[r|\alpha, \nu=1] = \frac{1}{\frac{2}{17} \sum_{k=2}^9 \{\text{Log}(k)\}^2} \approx 0.380341. \quad (21)$$

Consistency within an individual expert's response can be observed when the posterior variance decreases as compared to an expert responding at random. The conjugacy of the posterior analysis will allow for straightforward sequential updating using the responses of the  $k$  individual experts. Agreement amongst the experts can be identified by further reduction (increase) in the posterior variance (precision) using sequential updating.

During the WSF risk assessment in 1998 geometric means amongst the expert responses were used in a classical log-linear regression analysis approach to assess relative accident probabilities given by (2). Using a best subset regression approach 6 interactions indicated Table 2 were selected and will also be used herein to allow for a comparison in Section 6 between the classical and Bayesian point estimates. Hence, the vector  $\underline{\beta}$  to be utilized in our example in Section 6 will be a  $1 \times 16$  vector.

Table 2. Interaction Variables associated with the contributing factors in Table 1.

	<b>Designation</b>	<b>Description</b>
$X_1$	FR_FC·TT_1	Interaction
$X_2$	FR_FC·TS_1	Interaction
$X_3$	FR_FC·VIS	Interaction
$X_4$	TT_1·TS_1	Interaction
$X_5$	TT_1·VIS	Interaction
$X_6$	TS_1·VIS	Interaction

For the distribution of  $(\underline{\beta}|r)$  we may select a priori a location vector

$$\underline{m} = (0, \dots, 0)^T \quad (22)$$

and the unit precision matrix

$$\Delta = \begin{pmatrix} 1 & & \emptyset \\ & \ddots & \\ \emptyset & & 1 \end{pmatrix}, \quad (23)$$

as long as the resulting marginal distributions of  $\beta_i$  (cf. (18)) are flat, or (perhaps more importantly) as long as the resulting prior distribution on the relative accident probabilities (2) are non-informative. The motivation for a non-informative prior is to "let the evidence speak" (i.e. the expert judgment) (see, e.g., Kaplan (1997), p. 414). Expression (22) specifies that a priori none of the attributes contribute to accident risk and expression (23) indicates a priori independence between the elements of the parameter vector  $\underline{\beta}$ .

Figure 5 below depicts the prior distribution on  $(\underline{\beta}, r)$  utilizing (21), (22) and (23). Figure 5A depicts a graph of the prior density function of the precision  $r$ . Figure 5B displays the 90% credibility intervals of  $\beta_i, i = 1, \dots, 16$  and Figure 5C provides a graph of prior distribution of the relative probability  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  associated with the paired comparison in Figure 4. The probability density in Figure 5C is one of a log- $t$  distribution (see, e.g., McDonald and Butler (1987)) with prior parameters (cf. (19) and (20))

$$\underline{m}^T (\underline{X}^1 - \underline{X}^2) = 0, \alpha = 0.380341, \nu = 1, \delta_{ii} = (\underline{X}^1 - \underline{X}^2)^T \Delta (\underline{X}^1 - \underline{X}^2) = 4.$$

The prior median of  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  equals 1 (indicating indifference in collision likelihood between system states  $\underline{X}^1$  and  $\underline{X}^2$ ). A 50% credibility interval of  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  in Figure 5A equals  $[0.181, 5.515]$ . A 75% credibility interval of  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  equals  $[2.012 \cdot 10^{-5}, 4.971 \cdot 10^4]$  (which is quite wide) and hence our prior specification utilizing (21), (22) and (23) may be viewed as sufficiently non-informative.

Previous credibility intervals above and those in Figure 5B were evaluated utilizing

$$A(u | \alpha, \nu, \delta_{ii}) = \frac{1}{B(\frac{1}{2}, \frac{\alpha}{2})} \sqrt{\frac{\delta_{ii}}{\nu}} \int_{m_i - u}^{m_i + u} \left[ 1 + \frac{\delta_{ii}}{\nu} (\beta_i - m_i)^2 \right]^{-\frac{\alpha+1}{2}} d\beta_i,$$

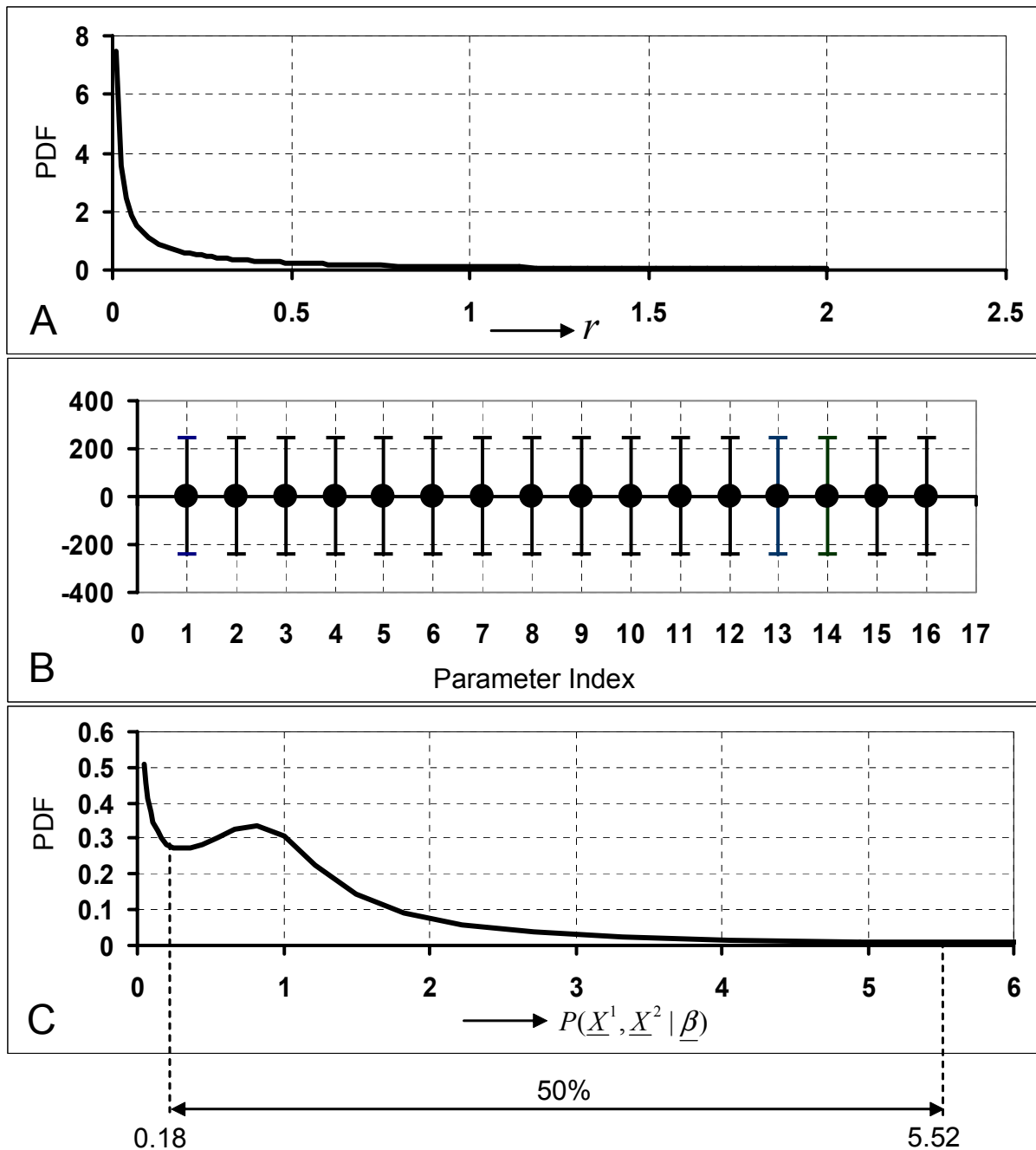


Figure 5. Prior distribution on  $(\underline{\beta}, r)$  and  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  (cf. (2))

for the two scenarios in Figure 4. A: Prior Marginal Distribution on  $r$ ; B: Prior 90% credibility intervals for the parameters  $\beta_i, i = 1, \dots, 16$ ; C: Prior distribution of relative probability  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  associated with Figure 4.

where  $A(u|\alpha, \nu, \delta_{ii})$  is the probability mass in a credibility interval  $[m_i - u, m_i + u]$  around the location parameter  $m_i$  of a  $t$ -distribution with precision  $\frac{\alpha}{\nu} \delta_{ii}$ . The latter quantity  $A(u|\alpha, \nu, \delta_{ii})$  is related to the well known incomplete beta function

$$B(x|a, b) = \frac{1}{B(a, b)} \int_0^x u^{a-1} (1-u)^{b-1} du, \quad (24)$$

where  $a, b > 0$ ,  $x \in [0, 1]$ , and  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$  via the relationship

$$A(u|\alpha, \nu, \delta_{ii}) = 1 - B\left(\frac{\nu}{\nu + \delta_{ii} u^2} \middle| \frac{\alpha}{2}, \frac{1}{2}\right),$$

(see, e.g. Press et al. (1989)). Numerical routines for evaluating the incomplete beta function (24) are widely provided in standard PC software such as Microsoft Excel. It should also be noted that due to the value of  $\alpha$  (cf. (21)), the moments of  $\beta_i$ , at least a priori, do not exist. However, since the  $t$ -distribution is symmetric around  $m_i$ , a natural point estimate for  $\beta_i$  is provided by its median value  $m_i$  indicated in Figure 5B,  $i = 1, \dots, 16$ .

## 5. POSTERIOR ANALYSIS

Applying Bayes theorem utilizing the likelihood (10), the prior distribution (15) and data specified via (7) and (8), it follows that the posterior distribution  $\prod(\underline{\beta}, r | \mathcal{Z}, Q)$  is proportional to

$$r^{\frac{n}{2}} \exp\left\{-\frac{r}{2}(c - 2\underline{b}^T \underline{\beta} + \underline{\beta}^T A \underline{\beta})\right\} \times \\ r^{\frac{\alpha}{2}-1} \exp\left(-\frac{r}{2}\right) \times r^{\frac{p}{2}} \exp\left\{-\frac{r}{2}(\underline{\beta} - \underline{m})^T \Delta(\underline{\beta} - \underline{m})\right\},$$

where  $c$ ,  $\underline{b}$ , and  $A$  are given by (11). Combining like terms we obtain

$$\prod(\underline{\beta}, r | \mathcal{Z}, Q) \propto r^{\frac{\alpha+n}{2}-1} \exp\left\{-\frac{r}{2}(1 + c + \underline{m}^T \Delta \underline{m})\right\} \times \\ r^{\frac{p}{2}} \exp\left\{-\frac{r}{2}\left(-2[\underline{b} + \Delta \underline{m}]^T \underline{\beta} + \underline{\beta}^T [A + \Delta] \underline{\beta}\right)\right\}. \quad (25)$$

Defining  $\Delta^u$  to be

$$\Delta^u = A + \Delta, \quad (26)$$

it follows from the symmetry and positive definiteness of  $A$  (cf. (12)) and  $\Delta$ , that  $\Delta^u$  is symmetric and positive definite, and hence invertible. Implicitly defining  $\underline{m}^u$  satisfying

$$[\underline{b} + \Delta \underline{m}]^T \underline{\beta} = [\Delta^u \underline{m}^u]^T \underline{\beta} \quad (27)$$

for all  $\underline{\beta}$ , it follows that

$$\underline{b} + \sum \underline{m} = \Delta^u \underline{m}^u \Leftrightarrow \underline{m}^u = (\Delta^u)^{-1} (\underline{b} + \Delta \underline{m}). \quad (28)$$

Utilizing (27) and (28) we derive from (25) that

$$\begin{aligned} \prod (\underline{\beta}, r | \mathcal{Z}, Q) &\propto r^{\frac{\alpha+n}{2}-1} \exp \left\{ -\frac{r}{2} \left( 1 + c + \underline{m}^T \Delta \underline{m} - [\underline{m}^u]^T \Delta^u \underline{m}^u \right) \right\} \times \\ &r^{\frac{p}{2}} \exp \left\{ -\frac{r}{2} [\underline{\beta} - \underline{m}^u]^T \Delta^u [\underline{\beta} - \underline{m}^u] \right\}. \end{aligned} \quad (29)$$

From (29) it follows, utilizing (11), that  $(\underline{\beta} | \mathcal{Z}, Q) \sim MVN(\underline{m}^u, r\Delta^u)$  where

$$\begin{cases} \Delta^u = \sum_{j=1}^n \underline{q}_j \underline{q}_j^T + \Delta \\ \underline{m}^u = (\Delta^u)^{-1} \left( \sum_{j=1}^n \underline{q}_j z_j + \Delta \underline{m} \right) \end{cases} \quad (30)$$

and  $(r | \mathcal{Z}, Q) \sim Gamma(\frac{\alpha^u}{2}, \frac{\nu^u}{2})$  with

$$\begin{cases} \alpha^u = \alpha + n \\ \nu^u = \nu + \sum_{j=1}^n z_j^2 + \underline{m}^T \Delta \underline{m} - [\underline{m}^u]^T \Delta^u \underline{m}^u \end{cases} \quad (31)$$

and  $\underline{m}^u$  and  $\Delta^u$  are given by (30). From (30), (31), (13) and (14) we deduce that the Bayesian updating procedure above is in fact a conjugate Bayesian analysis. In the next section we shall illustrate the inference procedure using the responses of 8 experts to a paired comparison questionnaire containing 60 questions similar to the one in Figure 4 and administered during the WSF risk assessment in 1998.

## 6. EXAMPLE WITH DATA ELICITED DURING WSF RISK ASSESSMENT

An individual questionnaire was administered to experts for each of the following possible incidents on the Washington State Ferry: propulsion failure, steering failure, navigation equipment failure, human error, as well as an individual questionnaire given an incident (either human error or mechanical failure) which occurred on the nearby vessel. As an illustrative example, we shall demonstrate our Bayesian conjugate analysis utilizing the responses of the 8 experts to the questionnaire involving the navigation equipment failure to derive the posterior distribution of the relative accident probability given by (2) associated with Figure 4. Combination of the responses of these 8 experts follows naturally by exploiting the conjugacy of the analysis in Section 3, 4 and 5 through sequential updating.

During the WSF risk assessment in 1998 expert responses were aggregated by taking geometric means of their responses and using them in a classical log linear regression analysis approach to assess relative accident probabilities given by (2). Classical point estimates for the parameters  $\beta_j, j = 1, \dots, 16$  associated with the contribution factors (the so-called main effects) in Table 1 and interaction effects in Table 2 will be compared to their Bayesian counterparts following our Bayesian aggregation method.

### 6.1. The elements $A$ , $\underline{b}$ and $c$ of the likelihood given by (11)

Expert were instructed to assume that a navigation equipment failure had occurred on the Washington State Ferry and were next asked to assess how much more likely a collision is to occur in Situation 1 (good visibility in Figure 4) as compared to Situation 2 (bad visibility in Figure 4) taking into account the value of all the contributing factors. The additional factors in Figure 4 (besides visibility) are used to assess interaction effects but also play a role in terms of designing a meaningful question. For example, a question that simply asks an expert to assess the likelihood of collision given a navigation equipment failure in bad visibility compared to good visibility is not meaningful since the expert would have to know for example whether another vessel nearby is crossing or passing and its proximity. Table 3 provides the answer of the

eight experts to the question in Figure 4. Note that Expert 8 responded (presumably inconsistently) that Situation 2 (with bad visibility) has a lower accident probability than Situation 1 (with good visibility). An expert aggregation method combines the responses in Table 3 into a single one.

Table 3. Expert Response to the Paired Comparison in Figure 4

Expert Index	1	2	3	4	5	6	7	8
Response	5	5	3	9	7	9	3	0.5

The questionnaire consisted of sixty questions similar to the one displayed in Figure 4. The questions were randomized in order and were distributed evenly over the 10 contributing factors in Table 1 (i.e. 6 questions per changing contributing factor). The  $16 \times 16$  design matrix  $A$  of the questionnaire (cf. (11)) is of the following form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (32)$$

where  $A_{11}$  is a  $10 \times 10$  diagonal matrix with diagonal elements

$$(4.56, 4.33, 2.89, 6, 1.5, 2.44, 6, 6, 6, 0.375) \quad (33)$$

and associated with the contributing factors  $X_1, \dots, X_{10}$ . (The matrix  $A_{11}$  in (32) is a diagonal matrix since the paired comparison scenarios  $\underline{X}^1$  and  $\underline{X}^2$  only differed in one covariate (see Figure 4)). The matrix  $A_{22}$  in (32) is a symmetric  $6 \times 6$  matrix with elements

$$\begin{bmatrix} 3.45 & 0.33 & 0 & 1.44 & 0.76 & 0 \\ 0.33 & 3.45 & 0.44 & 0.33 & 0 & 1 \\ 0 & 0.44 & 4.11 & 0 & 1 & 2.39 \\ 1.44 & 0.33 & 0 & 1.89 & 0.36 & 0.08 \\ 0.76 & 0 & 1 & 0.36 & 3.02 & 2 \\ 0 & 1 & 2.39 & 0.08 & 2 & 6.67 \end{bmatrix} \quad (34)$$



and associated with the interaction effects  $X_{11}, \dots, X_{16}$ . Finally, the matrix  $A_{21} = A_{12}^T$  is a sparse  $10 \times 6$  matrix

$$\begin{bmatrix} 1 & 2.82 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.26 & 0 & 2.12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.13 & 0 & 0 & 0 & 0 & 0 & 0 & 3.06 & 0 & 0 \\ 0 & 2.13 & 0.52 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.02 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1.56 & 0 & 0 & 0 & 0 & 5.33 & 0 & 0 \end{bmatrix} \quad (35)$$

with only positive elements associated with the contributing factors  $X_1, X_2, X_3$  and  $X_8$  that are included in the interaction effects  $X_{11}, \dots, X_{16}$ . The questionnaire was designed in a manner such that the resulting matrix  $A$  is positive definite (and thus invertible), but equally important, involved meaningful paired comparisons consistent with realistic scenarios on the Puget Sound. The latter required maritime knowledge about the WSF Ferry system acquired by the team conducting the WSF Risk Assessment.

Figure 6 below summarizes the vector  $\underline{b}$  cf. (11) for each of the eight expert responses to 60 questions in terms of  $\sum_{j=1}^{60} q_{ij} z_j$  for each of the contributing factors  $X_i, i = 1, \dots, 10$  in Table 1 and interaction effects  $X_i, i = 11, \dots, 16$  in Table 2. Hence, Figure 6 consists of 16 histograms each one plotting the  $i$ -th element of the vector  $\underline{b}$  cf. (11) for all eight experts. From Figure 6 we may (visually) assess the consistency in the expert judgment with respect to the ordering of the covariate scale of the elements  $X_i, i = 1, \dots, 16$ . A positive (negative) value indicates agreement with the ordering of that particular scale. For example, the histogram in Figure 6 associated with the contributing factor TP1 (Traffic Proximity of first interacting vessel) shows that all experts responded (not surprisingly) that vessels further away pose less (immediate) collision risk. The histogram in Figure 6 associated with the contributing factor VIS provides a similar result to that in Table 3, i.e. that Expert 8 inconsistently rated lower visibility with lower collision risk throughout the questionnaire. The largest discrepancy with the ordering of a covariate scale amongst the 8 experts is observed in the first histogram and is associated with the variable FR-FC (Ferry Route-Ferry Class combination).

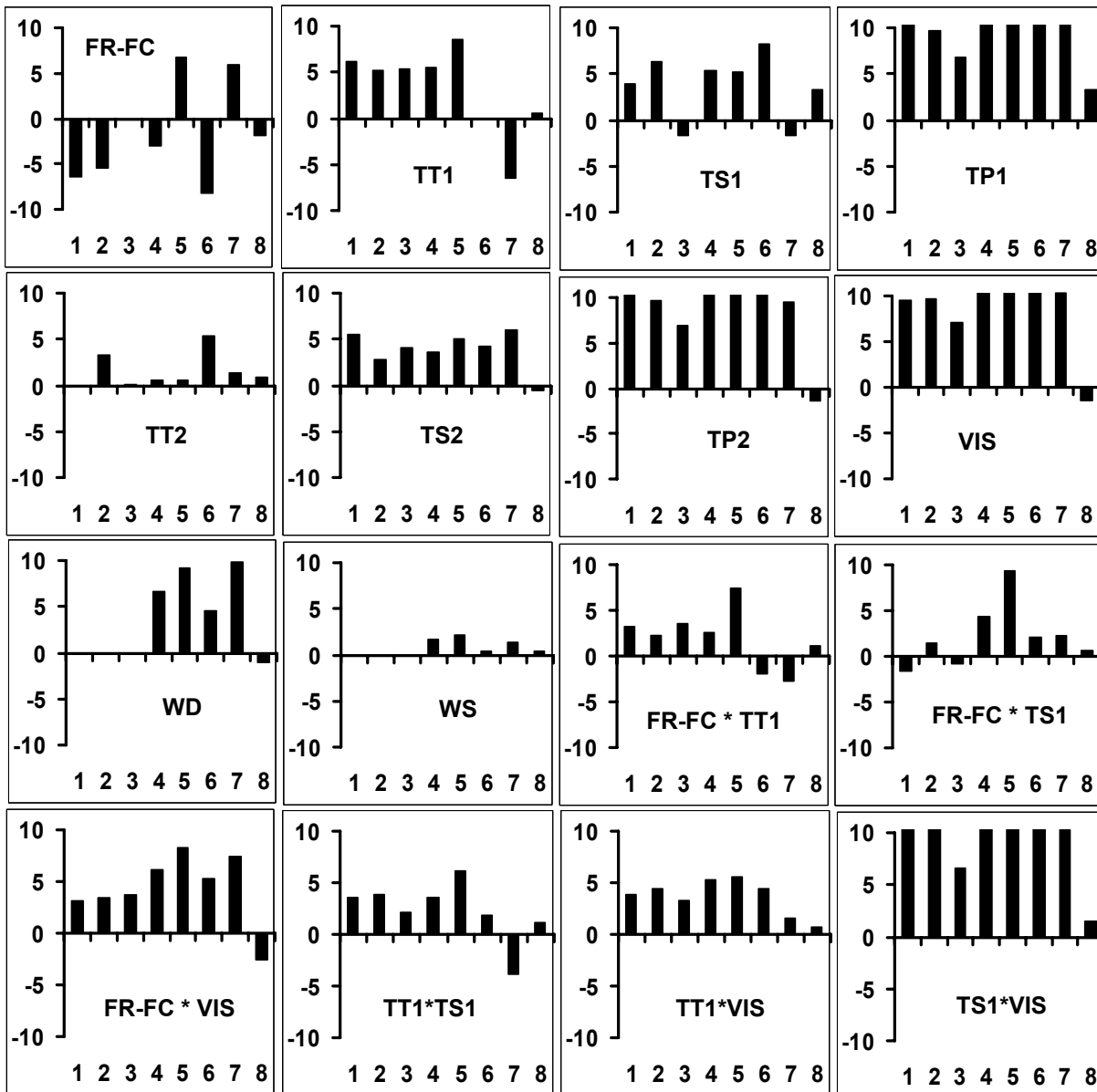


Figure 6. Summary of Individual Expert Response for 8 WSF experts in terms of  $i$ -th element of the vector  $\underline{b}$  (cf. (11)) for each of the contributing factors  $X_i, i = 1, \dots, 10$  in Table 1 and interaction effects  $X_i, i = 11, \dots, 16$  in Table 2.

The elements  $c = \sum_{j=1}^{60} z_j^2$  (cf. (11)) for each individual expert are provided in Table 4. Note

that on aggregate particularly both Expert 3 and Expert 8 assessed lower collision likelihoods in their paired comparisons questions.

Table 4. Values for  $c$  (cf. (11)) for the 8 individual experts.

Expert Index	1	2	3	4	5	6	7	8
Scalar $c$	149.07	95.28	55.74	147.93	185.71	177.30	147.12	44.94

## 6.2. Posterior Analysis

Utilizing the aggregate individual expert responses (vectors  $b$ ) in Figure 6, the matrix  $A$  specified by (32) - (35), the scalars  $c$  in Table 4, we update the prior distribution of  $(\underline{\beta}, r)$  depicted in Figure 5 in a Bayesian manner using sequential updating. The resulting posterior distribution on  $(\underline{\beta}, r)$  is displayed in Figure 7. Figure 7A contains a plot of a  $Gamma(\frac{\alpha^u}{2}, \frac{\nu^u}{2})$  density with parameters

$$\alpha^u = 480.38, \nu^u = 530.95 \quad (36)$$

Figure 7B displays 90% credibility intervals of the posterior distributions of  $\beta_i, i = 1, \dots, 16$  and the location parameters  $m_i^u$ . The posterior distribution of the parameter vector  $\underline{\beta}$  is a multivariate  $t$  distribution with location vector  $\underline{m}^u$  and precision matrix  $\frac{\alpha^u}{\nu^u} \Delta^u$ , where  $\alpha^u, \nu^u$  are given by (36) and

$$\Delta^u = \Delta + 8A$$

(cf. (26)) where the unit matrix  $\Delta$  is given by (23) and the matrix  $A$  by (32)-(35). It can be concluded from Figure 7B that traffic proximity of the first and second interacting vessel ( $X_4$  and  $X_7$ , respectively), traffic scenario of the second interacting vessel  $X_7$  and wind speed  $X_{10}$  are the largest contributing factors to accident risk. In addition, the manner in which the first interacting vessel approaches the ferry route - ferry class combination ( $X_{12}$ ), i.e. crossing, passing or overtaking, and in what visibility conditions ( $X_{16}$ ) are the largest interacting factors. The posterior location vector  $\underline{m}^u$  is displayed in Figure 8 together with their classical counterpart estimated via a log-linear regression method utilizing the geometric means of the expert responses.

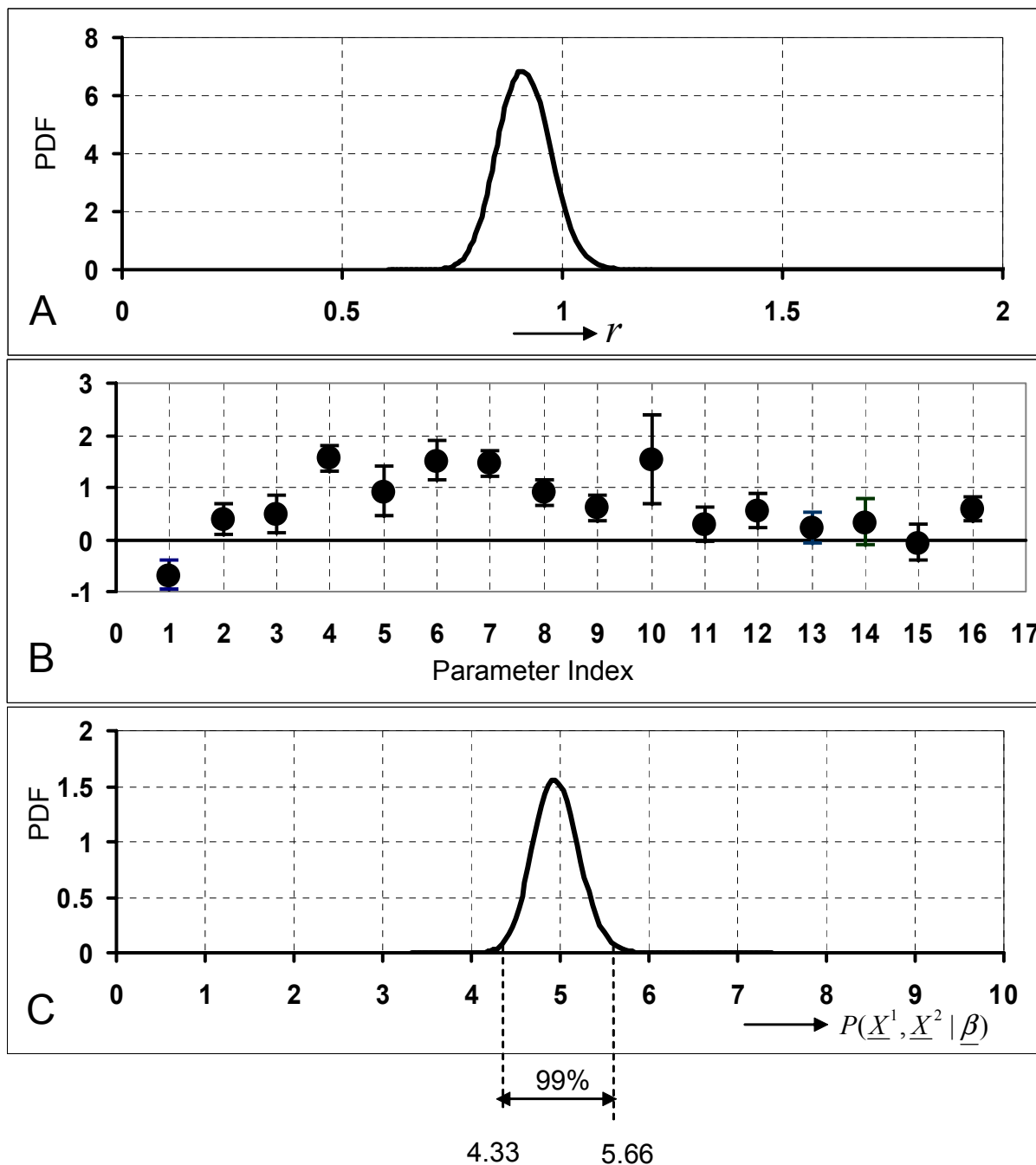


Figure 7. Posterior distribution on  $(\underline{\beta}, r)$  and  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  (cf. (2))

for the two scenarios in Figure 4. A: Posterior Marginal Distribution on  $r$ ; B: Posterior 90% credibility intervals for the parameters  $\beta_i, i = 1, \dots, 16$ ; C: Posterior distribution of relative probability  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  associated with Figure 4.

A remarkable agreement should be noted between the Bayesian and classical point estimates provided in Figure 8, except for a discrepancy associated with the contributing factor WS (Wind Speed). From Figure 7, however, it follows that the classical point estimate associated with WS in Figure 8 is well within the 90% credibility bounds of  $\beta_{10}$  depicted in Figure 7. Finally, Figure 7C displays the posterior distribution of the relative probability  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  associated with Figure 4.

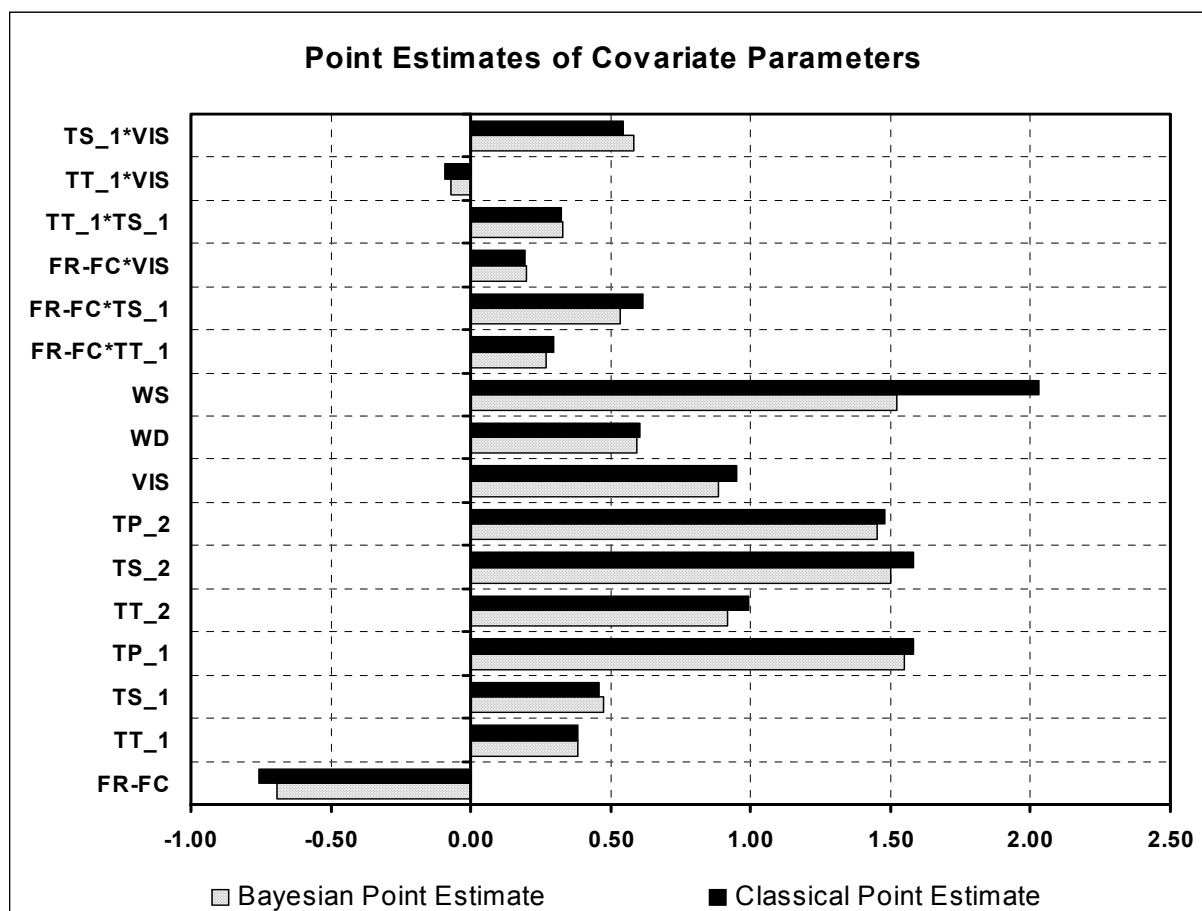


Figure 8. Comparison of Bayesian and Classical Point Estimates of the parameters  $\beta_i, i = 1, \dots, 16$ .

Compare the 50% posterior credibility interval of  $P(\underline{X}^1, \underline{X}^2 | \underline{\beta})$  of [4.78, 5.13] to the 50% prior one of [0.18, 5.52] in Figure 5C. In addition, the 99% posterior credibility interval of [4.33, 5.66]

is indicated in Figure 7C (which is remarkably narrow compared to the prior 75% credibility interval of  $[2.012 \cdot 10^{-5}, 4.971 \cdot 10^4]$ ) containing its median point estimate 4.94. Hence, Situation 2 in Figure 4 is approximately 5 times more likely to result in a collision than Situation 1 given that a navigation equipment failure occurred on the ferry.

Figure 9 below provides a posterior analysis of point estimates  $\alpha^u/\nu^u$  of the precision  $r$ , where  $\alpha^u$  and  $\nu^u$  are given by (31).

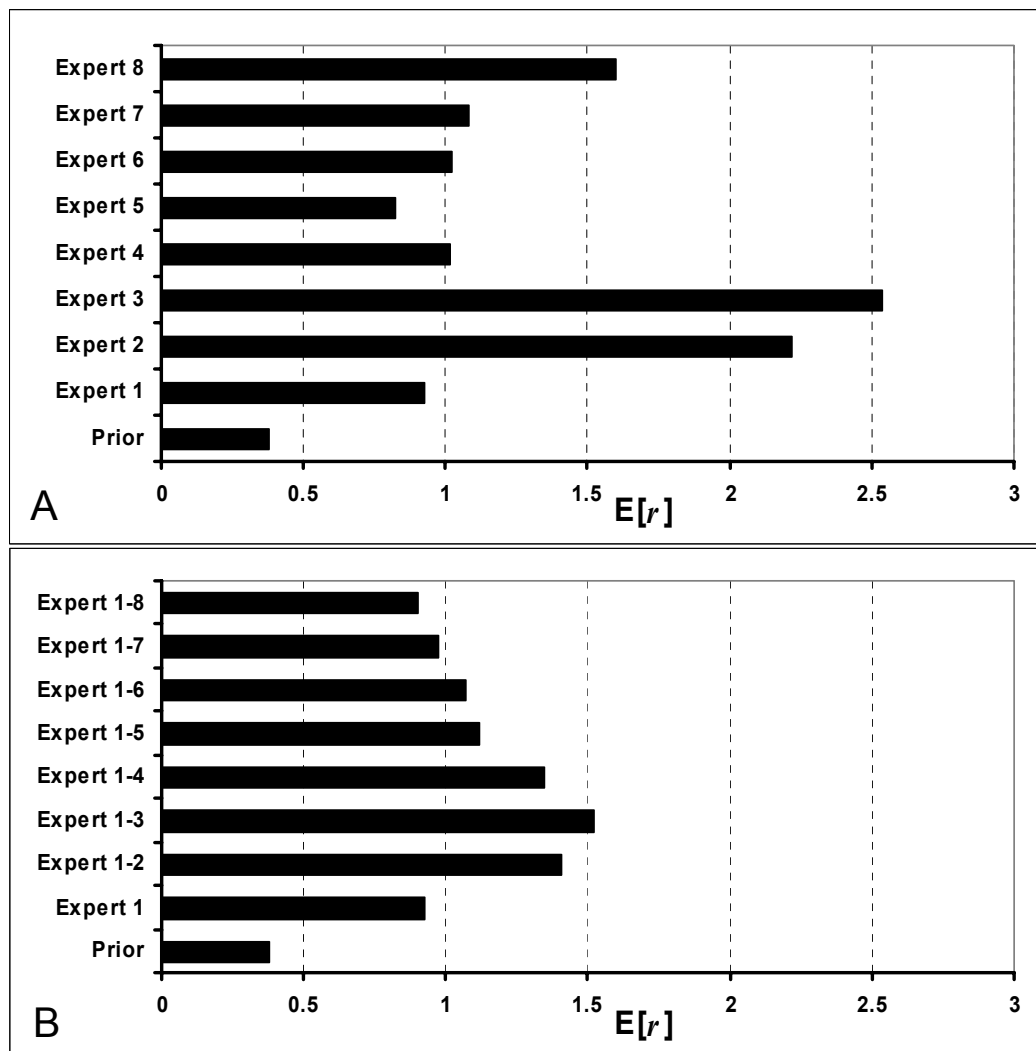


Figure 9. Prior and Posterior points estimates of the precision  $r$  (cf. (4) and (13))

A: Individual posterior estimates for Experts  $i$ ,  $i = 1, \dots, 8$ ;

B: Sequential Posterior estimates involving Experts 1 through  $i$ ,  $i = 1, \dots, 8$ .

Figure 9A depicts  $E[r|Expert\ i]$  obtained by updating the prior precision with the individual responses of Expert  $i$ ,  $i = 1, \dots, 8$ . Figure 9B displays  $E[r|Expert\ 1-i]$  derived using sequential updating involving Expert 1 through Expert  $i$ ,  $i = 1, \dots, 8$ . From Figure 9A it may be concluded that each expert responded consistently in the sense that posterior precision increased when compared to the precision of an expert responding at random (the prior precision in Figure 9A). In addition, from Figure 9B we conclude that at first agreement is present amongst Experts 1-3 due to a continued increase in posterior precision utilizing sequential updating. From Expert 4 onward and including Expert 8, however, a continued disagreement is observed in Figure 9B due to a continued decline in posterior precision. Note the increasing pattern in Figure 9A from Expert 5 on compared to the continued decreasing pattern in Figure 9B from Expert 4 and up. The latter indicates that consistency of an individual expert response does not necessarily result in an increase in agreement amongst a group of experts.

## 7. CONCLUDING REMARKS

A Bayesian aggregation method has been developed using responses from multiple experts to a paired comparison questionnaire to assess the distribution of relative accident probabilities. The classical analysis conducted during the WSF risk assessment only resulted in point estimates of relative accident probabilities, not full posterior distributional results as indicated in Figure 7C. In addition, utilizing posterior distributional results for the parameter vector  $\underline{\beta}$  credibility statements can be made for any arbitrary paired comparison. For example setting Situation 1 in (2) to the best possible scenario ( $\underline{X}^1 = \underline{0}$ ) and Situation 2 to the worst possible scenario ( $\underline{X}^2 = \underline{1}$ ) a 99% credibility interval of  $P(\underline{X}^1, \underline{X}^2|\underline{\beta})$  equals [31142, 36749]. Therefore, informally, collision risk in the worst possible scenario differs at least by 4 orders of magnitude to that of the best possible scenario while taking uncertainty of the expert judgments into account.

Worst case scenario's however may have a very low incidence of occurrence, which is why all conditional probabilities in Figure 1 and their uncertainties need to be estimated to assess the

distribution of collision risk on for example a per year basis. This paper only provided distributional results for the relative probability given by (2). Merrick et al. (2003) assesses the distribution of  $Pr(OF, SF)$  using Bayesian Simulation techniques. A subsequent paper will integrate the approach herein with that of Merrick et al. (2003) to assess collision risk and its uncertainty in a Bayesian (and therefore coherent) manner.

#### ACKNOWLEDGEMENT

This material is based in part upon work supported by the National Science Foundation under Grant Nos. SES 0213627 and SES 0213700. Any findings, opinions, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Special thanks to our colleague Sam Kotz for his helpful comments that improved both the content and presentation of this paper. We are also indebted to the referees and the editor of *EJOR*, whose comments improved the presentation of the first version.

#### REFERENCES

- Anderson, E., Hattis, D., Matalas, N., Bier, V., Kaplan, S., Burmaster, D., Conrad, S. and Ferson, S., 1999. Foundations. *Risk Analysis*, 19 (1) 47-68.
- Bedford, T. and Cooke, R., 2001. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, Cambridge, U.K.
- Bonano, E., Hora, S., Keeney, R. and von Winterfeldt, D., 1989. Elicitation and use of expert judgment in performance assessment for high-level radioactive waste repositories. SAND89-1821, NUREG/CR-5411, Sandia National Laboratories, Albuquerque, NM.
- Bradley, R., 1953. Some statistical methods in taste testing and quality evaluation. *Biometrics*, 9 (1) 22-38.
- Bradley, R.A. and Terry M. 1952. Rank Analysis of Incomplete Block Designs I. The method of paired comparisons. *Biometrika*, 39, 324-345.



- Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5 (4) 559-583.
- Clemen, R.T. and Winkler, R., 1990. Unanimity and Compromise Among Probability Forecasters. *Management Science*, 36 (7) 767-779.
- Clemen, R.T. and Winkler, R., 1999. Combining Probability Distributions from Experts in Risk Analysis. *Risk Analysis*, 19 (2) 187-203.
- Cooke, R.M., 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, Oxford, U.K..
- Cox, D.R., 1972. Regression models and life tables (with discussion), *Journal of Royal Statistical Society, Series B*, 34 (2) 187-220.
- Debrue, G., 1986. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. Yale University Press, New Haven, CT.
- DeWispelare, A., Herren, L. and Clemen, R., 1995. The Use of Probability Elicitation in the High-Level Nuclear Waste Recognition Program. *International Journal of Forecasting*, 11 (1) 5-24.
- Foreman, E.H. and Selly, M.A., 2002. *Decision by Objectives: How to Convince Others That You Are Right*. World Scientific Pub Co, 1st Edition., River-Edge, NJ.
- Garrick, B.J., 1984. Recent Case Studies and Advancements in Probabilistic Risk Assessment. *Risk Analysis*, 4 (4) 267-279.
- Genest C. and Zidek, J., 1986. Combining probability distributions: a critique and annotated bibliography. *Statistical Science*, 1 (1) 114-148.
- Kahneman, D., Slovic, P. and Tversky, A., 1982. *Judgments Under Uncertainty*. Cambridge University Press, Cambridge, U.K.
- Kaplan, S., 1997. The Words of Risk Analysis. *Risk Analysis*, 17 (4) 407-417.
- Koller, G. R., 2000. *Risk Modeling for Determining Value and Decision Making*. Chapman & Hall, New York.

- Kumamoto, H. and Henley, E. 1996. Probabilistic Risk Assessment and Management for Engineers and Scientists. IEEE Press, Piscataway, NJ.
- McDonald, J.B., and Butler, R.J. 1987. Some generalized mixture distributions with an application to unemployment duration. *Review of Economics and Statistics*, 69 (2) 232-240.
- Mendel, M. and Sheridan, T., 1989. Filtering Information from Human Experts. *IEEE Transactions on Systems, Man & Cybernetics*, 36 (1) 6-16.
- Merrick, J.R.W., van Dorp, J.R., Harrald, J.R., Mazzuchi, T.A., Grabowski, M. and Spahn, J.E., 2000. A Systems approach to managing oil transportation risk in Prince William Sound. *Systems Engineering*, 3 (3) 128-142.
- Merrick, J.R.W., Dinesh, V., Singh, A., van Dorp, J. R. and Mazzuchi, T., 2003. Propagation of Uncertainty in a Simulation-Based Maritime Risk Assessment Model Utilizing Bayesian Simulation Techniques. *Proceeding of the 2003 Winter Simulation Conference* (eds. Chick, S. , Sanchez, P.J, D. Ferrin and D.J. Morrice), 449-455.
- Morgan, M. and Henrion, M., 1991. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, U.K.
- Morris, P., 1974. Decision analysis expert use. *Management Science*, 20 (9) 1233-1241.
- Moslesh, A., Bier, V. and Apostolakis, G., 1988. A Critique of Current Practice for the Use of Expert Opinions in Probabilistic Risk Assessment. *Reliability Engineering and System Safety*, 20 (1) 63-85.
- Paté-Cornell, M., 1996. Uncertainties in Risk Analysis: Six Levels of Treatment. *Reliability Engineering and System Safety*, 54 (2-3) 95-111.
- Press, W.H., Flannery, B.P., Teukolsky S.A. and Vettering W.T, 1989. *Numerical Recipes in Pascal*. Cambridge University Press, Cambridge.
- Pulkinen, U., 1993. Methods for combination of expert judgments. *Reliability Engineering and Systems Safety*, 40 (2), 111-118.
- Pulkinen, U., 1994a. Bayesian Analysis of Consistent Paired Comparisons. *Reliability Engineering and System Safety*, 43 (1) 1-16.

- Pulkinen, U., 1994b. Gaussian Paired Comparison Models. *Reliability Engineering and System Safety*, 44 (2) 207-217.
- Roeleven, D., Kok, M., Stipdonk, H.L. and de Vries, W.A., 1995. Inland Waterway Transport: Modeling the Probabilities of an Accident. *Safety Science*, 19 (2-3) 191-202.
- Saaty, T., 1977. A Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology*, 15, 234-281. New York.
- Saaty, T., 1980. *The Analytic Hierarchy Process*. McGraw-Hill .
- Shrader-Frechette K. 1985. *Risk Analysis and Scientific Method*. Reidel, Dordrecht, The Netherlands.
- Simpson, E.H., 1951. The Interpretation of Interaction in Contingency Tables. *Journal of The Royal Statistical Society, Series B*, 13 (2) 238-241.
- Thurstone, L.L., 1927a. A law of comparative judgment. *Psychology Review*, 34, 273-286.
- Thurstone, L.L., 1927b. Psychophysical Analysis. *American Journal of Psychology*, 38, 368-389.
- Van Dorp, J., Merrick, J., Harrald, J., Mazzuchi, T. and Grabowski, M., 2001. A Risk Management Procedure for the Washington State Ferries. *Risk Analysis*, 21 (1) 127-142.
- Vose, D. (1996). *Risk Analysis: A Quantitative Guide to Monte Carlo Simulation Modeling*. Wiley, New York, NY.
- West, M and Harrison, J., 1989. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.
- Winkler R., 1981. Combining probability distributions from dependent information sources. *Management Science*, 27 (5) 479-488.
- Winkler, R., 1996. Uncertainty in probabilistic risk assessment. *Reliability Engineering and System Safety*, 54 (2-3) 127-132.