Use of evidential reasoning for eliciting Bayesian subjective probabilities in human reliability analysis: a maritime case

Zaili Yang1Jin Wang1, Salman Nazir3 *, Khalifa Mohamed Abujaafar1, Zhuohua Qu2,

* Corresponding

1. Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, UK
2. Liverpool Business School, Liverpool John Moores University, UK
3. Human Factors Research Group, Department of Maritime Technology and Innovation, University College of Southeast Norway, Norway.

Use of evidential reasoning for eliciting Bayesian subjective probabilities in human reliability analysis: a maritime case

Zaili Yang[1][*], Khalifa Mohamed Abujaafar[1], Zhuohua Qu[2], Jin Wang[1], Salman Nazir[3]

1. *Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, UK*
2. *Liverpool Business School, Liverpool John Moores University, UK*
3. *Human Factors Research Group, Department of Maritime Technology and Innovation, University College of Southeast Norway, Norway.*

[*] Corresponding author: z.yang@ljmu.ac.uk

Use of evidential reasoning for eliciting Bayesian subjective probabilities in human reliability analysis: a maritime case

Abstract

Modelling the interdependencies among the factors influencing human error (e.g. the common performance conditions (CPCs) in Cognitive Reliability Error Analysis Method (CREAM)) stimulates the use of Bayesian Networks (BNs) in Human Reliability Analysis (HRA). However, subjective probability elicitation for a BN is often a daunting and complex task. To create conditional probability values for each given variable in a BN requires a high degree of knowledge and engineering effort, often from a group of domain experts. This paper presents a novel hybrid approach for incorporating the evidential reasoning (ER) approach with BNs to facilitate HRA under incomplete data. The kernel of this approach is to develop the best and the worst possible conditional subjective probabilities of the nodes representing the factors influencing HRA when using BNs in human error probability (HEP). The proposed hybrid approach is demonstrated by using CREAM to estimate HEP in the maritime area. The findings from the hybrid ER-BN model can effectively facilitate HEP analysis in specific and decision-making under uncertainty in general.

Keywords: Human reliability analysis, human error probability, evidential reasoning, Bayesian network, maritime risk.

1. Introduction

The second generation Human Reliability Analysis (HRA) methods such as the Cognitive Reliability and Error Analysis Method (CREAM) (Hollnagel, 1998) were used to proactively assess the erroneous human actions in complicated systems in a way that the context influencing human action is appropriately taken into account. While facilitating the quantitative development of HRA, these methods exposed some problems in their practical applications. For instance, the prospective assessment model of the basic approach to estimate human error probability (HEP) in CREAM (Hollnagel, 1998) cannot provide a crisp value of the consequences of human performance, and the HEP estimation mechanism is not sensitive to minor changes associated with the nine common performance conditions (CPCs) in CREAM (Yang et al., 2013; Xi et al., 2017). A fuzzy Bayesian reasoning approach was developed to deal with this problem through using Bayesian Networks (BNs) to model the parent-child relationship between the CPCs and Contextual Control Model Controlling Modes (COCOM-CMs) in CREAM (Yang et al., 2013; Abujaafar et al., 2016). However, it requires too much information about the prior conditional probabilities assigned to the node of COCOM-CMs, jeopardising the applicability of the approach. Furthermore, Bayesian inference requires probability completeness. Subjective judgements are usually used to complement the unavailability of statistical data. Even though subjective probabilities can be elicited by experts, it often raises the problems relating

to accuracy, consensus and completeness of judgements. The evidential reasoning (ER) approach (Yang and Xu, 2002) has shown its attractiveness to tackle the high uncertainty in data (e.g. incompleteness) associated with subjective judgements and has therefore been used to synthesise experts' estimates in HRA (Xi et al., 2017). However, so far the two challenging but essential features of uncertain data in HRA, interdependency among performance factors and incompleteness in subjective estimates have not been simultaneously addressed with success, in order for effective solutions to be found. Obviously, such incomplete probabilities can be effectively elicited by incorporating partial degrees of belief. In this regard, an ER algorithm has been developed on the basis of the Dempster-Shafer (D-S) theory of evidence (Dempster, 1968; Shafer, 1976), which can be well suited to modelling subjective credibility induced by partial evidence observation (Smets, 1988). The ER's synthesising capability of partial degrees of belief has enlarged the utilisation scope of the traditional probabilistic theory, particularly in describing and handling uncertain information (e.g. incompleteness and ignorance) by using the concept of degrees of belief (Yang et al., 2008; Wan et al., 2019). Therefore, it is proposed to be integrated with BNs in this research to tackle the incapability of BNs in modelling incomplete, subjective probabilities introduced by multiple experts.

This paper presents a new hybrid approach for combining an ER algorithm with BNs in a complementary way, taking into account both interdependent performance factors and incomplete subjective data simultaneously. The kernel of the proposed method is that two individual assessment scenarios involving the best and worst evaluation models for all the nodes of incomplete subjective probabilities, are created, in which the remaining probability masses (due to incompleteness) of the nodes are assigned back to their best (i.e. that contributes to the lowest HEP) and worst (i.e. that contributes to the highest HEP) grades, homogeneously and respectively. To achieve the above aim, the literature on the use of conventional CREAM and the development of the extended CREAM are reviewed to reveal the associated weaknesses and formulate the research problems in Section 2. In Section 3, a new hybrid approach by combining ER and BNs is developed to overcome the problems identified. Its applicability and feasibility are demonstrated by an illustrative example for easy understanding of the relevant mathematical algorithms in Section 4 and by studying the Deepwater Horizon accident case in Section 5. Section 6 concludes the achieved results.

2. Literature review

2.1 Traditional CREAM method

To model the causal relations, the CREAM methodology has been derived from its core, the Contextual Control Model (COCOM). COCOM focuses on the principle that human performance is the outcome of the purposive use of competence adjusted to specific working conditions rather than of the pre-determined sequence of response to given events (Yang et al., 2013). As one of the most widely known 'second generation' HRA methods, CREAM presents a consistent error classification system that

integrates individual, technological and organizational factors. The classification describes the relations between causes and effects by defining a number of sub-groups and tables, which are provided for the error modes on the one hand and the organisational causes on the other. As a kind of context-related HRA method, CREAM provides an approach to the assessment of cognitive processes during emergencies, and thus it has been widely recognised and used in the analysis of marine accidents. Among the recent examples are Zhou et al. (2017) and Ung (2018). However, CREAM (both basic and extended methods) has exposed certain practical limitations in its applications especially in the maritime industry. The failure rate intervals of Human Failure Probability (HFP) values from the basic method appear to be unacceptably wide even for the use in screening (Fujita and Hollnagel, 2004). It is also difficult to further use and interpret such failure rate intervals in practice. The extended method uses the output from the basic CREAM and appropriate data sources to calculate the probability of each cognitive function failure (Hollnagel, 1998). Lack of critical mass in statistical failure data, however, proves the tasks of adapting the extended method in the maritime area to be challenging (Xi et al., 2017). All of these limitations stimulate the development of advanced techniques in CREAM.

2.2 Extensions of CREAM with uncertainty treatment techniques

Over the past decade, advanced quantification approaches for HEP in CREAM have been proposed by using different uncertainty treatment techniques including fuzzy logic (Ung, 2015), BNs (Marseguerra et al. 2007), and ER (Xi et al., 2017). Kim et al. (2006) proposed a probabilistic method by using Bayesian networks for a better estimation of the control mode, which is able to produce mathematically correct results when levels of CPCs are given probabilistically. Konstandinidou et al. (2006) developed a fuzzy classification system for the estimation of the probability of human erroneous actions according to CREAM. The results obtained were in the form of crisp numbers, which can be used directly in other risk analysis models (e.g. fault tree model) for the quantification of specific undesired events. Although some attractiveness is observed in terms of the enhancement of CREAM in certain specific aspects though involving one (or some) of these uncertainty treatment techniques, a number of practical problems are still exposed. Examples of such problems include the loss of useful information in fuzzy Max–Min inference operations, lack of adequacy of modelling CPC dependencies and of instant human failure probability estimation, and inability of incorporating different effects/importance on human performance that CPCs may have in the practical HRA applications.

In view of the above-mentioned concerns, Yang et al. (2013) developed a generic BN-based HRA methodology, in which the prospective analysis of CREAM is modified, to facilitate the quantification of maritime human failures by effectively incorporating both fuzzy logic and Bayesian inference mechanisms. The framework has used fuzzy IF-THEN rule bases with belief structures and BNs to aggregate all the rules associated with a seafarer's task in order to estimate his/her failure probability. However, it is realised that large BNs of multi-tier nodes often exist in application domains. Their

complexity is sometimes beyond the current knowledge of domain experts. In addition, conventional mathematical methods are simply not applicable. Therefore, heuristic methods based on 'causal linkage' rather than detailed equations present a feasible way to proceed at present (McErleani et al., 1999). It is particularly important when the aforementioned BN-based HRA models fail to cope with situations where incomplete conditional probabilities are raised/assigned by multiple experts.

3. Methodology

Use of the ER-BN approach in HRA is demonstrated through its application in the CREAM framework given that BNs have been widely used in CREAM-based HRA modelling due to their feature, which takes into account the interaction between the nine CPCs (which are "Adequacy of organisation (#1)", "Working conditions (#2)", "Adequacy of man-machine interface and operational support (#3)", "Availability of procedures and plans (#4)", "Number of simultaneous goals (#5)", "Available time (#6)", "Time of day (#7)", "Adequacy of training and experience (#8)" and "Crew collaboration quality (#9)"). In this research, the proposed ER-BN model consists of the following four steps:

Step 1. The rule base of modelling probabilistic causal relation between parent-child nodes in the BN-based CREAM (e.g. Figure 2) is developed (Yang et al., 2013; Abujaafar et al., 2016). It reflects the interaction among the nine CPCs originally defined in CREAM (Hollnagel, 1998). During this process, the conditional probabilities of the parent-child nodes are elicited in either a complete or incomplete format by a group of domain experts.

Step 2. The ER approach (Yang and Xu, 2002) is used to synthesise the complete/incomplete conditional probabilities and aggregate the child nodes' conditional probabilities that are symmetrically affected by the nodes associated with the nine CPCs.

Step 3. The unknown/remaining probability masses (i.e. the unassigned probabilities to the grades of the child node) due to incomplete judgements are assigned back to the best (i.e. that contributes to the lowest HEP) and worst (i.e. that contributes to the highest HEP) grades of the nodes. Two BN models, representing the best scenario in which all the remaining probability masses of the nodes having unknown probabilities are assigned to the best grades of their corresponding nodes, and the worst scenarios in which they are assigned to the worst grades of the nodes, are constructed respectively.

Step 4. The two results from the best and worst scenarios are aggregated to obtain a crisp HEP using ER. Different weights can be assigned to the results obtained from the best and worst scenarios to present the HEPs in a range from optimistic to pessimistic perspectives.

3.1 Developing the rule base of modelling parent-child nodes and eliciting the complete or incomplete degrees of belief of COCOM-CMs

To assess HEP, it requires the evaluation of the CREAM context through the effect levels of the nine CPCs. By using a BN, it is possible to graphically map the nine CPCs in a convergent connection to infer the probabilities of the COCOM-CMs (in which the four characteristic control modes are "Scrambled", "Opportunistic", "Tactical" and "Strategic"). In this context, each CPC is described by a number of discrete states including four states for three CPCs and three states for the remaining six CPCs, according to the original CREAM. Different levels of each CPC along with their individual effect on human performance are described in Table 1, and the relation between CPCs and the control modes is depicted in Figure 1.

Table 1. Description of CPCs and associated linguistic variables (Hollnagel, 1998)

| CPC		CPC Levels	Effects
#1	Adequacy of organisation	**Deficient**	**Negative**
		Inefficient	**Negative**
		Efficient	Neutral
		Very Efficient	*Positive*
#2	Working Conditions	**Incompatible**	**Negative**
		Compatible	Neutral
		Advantageous	*Positive*
#3	Adequacy of man machine interface (MMI) and operational support	**Inappropriate**	**Negative**
		Tolerable	Neutral
		Adequate	Neutral
		Supportive	*Positive*
#4	Availability of procedures and plans	**Inappropriate**	**Negative**
		Acceptable	Neutral
		Appropriate	*Positive*
#5	Number of simultaneous goals	**More than actual capacity**	**Negative**
		Matching current capacity	Neutral
		Fewer than actual capacity	Neutral
#6	Available time	**Continuously inadequate**	**Negative**
		Temporarily inadequate	Neutral
		Adequate	*Positive*
#7	Time of day	**Night** (0:00-7:00hr) (unadjusted)	**Negative**
		Night (17:00-24:00hr) (unadjusted)	**Negative**
		Day (6:00-18:00hr) (adjusted)	Neutral
#8	Adequacy of training and experience	**Inadequate**	**Negative**
		Adequate with limited experience	Neutral
		Adequate with high experience	*Positive*
#9	Crew collaboration quality	**Deficient**	**Negative**
		Inefficient	Neutral
		Efficient	Neutral
		Very efficient	*Positive*
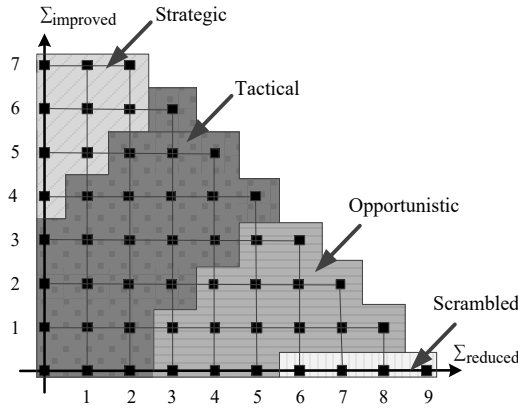
Figure 1. The relation between CPCs and control modes (Hollnagel, 1998)

Such convergent connection will result in 46,656 ($4^3 \times 3^6$) discrete conditional probabilities to be assigned. The configuration of such a large number of discrete conditional probabilities subjectively by domain experts will be of great difficulty. Therefore a divorcing method is introduced to simplify the task of assigning subjective probabilities by adding three attributes (the second tier) and two sub-attributes (the third tier) shown in Figure 2. The three attributes (nodes) are "Action load", "Working environment" and "Operator preparedness" directly influencing COCOM-CMs' probability (Marseguerra et al., 2007). Each attribute is associated with different CPCs according to the reasoning in CPCs' evaluation by Hollnagel (1998). The attribute "Working environments" is influenced by five CPCs. To further simplify its conditional probability table (CPT) assignment, two new sub-attributes, "Adequacy of working culture" and "Adequacy of perception conditions" are also introduced. The CPTs of the attributes and sub-attributes are assigned based on the uniformly defined states, "Inappropriate", "Acceptable" and "Appropriate" which present the reduced, satisfactory and improved effects to human reliability, with respect to the defined grades of the nine CPCs in CREAM. A divorcing concept has no significant effect on modelling mathematical inference if attributes and sub-attributes' CPTs are assigned properly (Kim et al., 2006). The use of a divorcing concept simplifies the assignment of CPTs of the developed BN-based CREAM model. It also makes it possible to introduce fuzzy rule bases (FRB) to facilitate the elicitation of subjective CPTs of the child nodes. For example, the interactive logical relation between the effect levels of the three attributes "Action load (A)", "Working environments (W)" and "Operator preparedness (O)", and the COCOM-CMs is described as follows while the CPT of the COCOM-CMs under the three parents is established in Table 5. The interaction among the nine CPCs (at the fifth and sixth tiers) and their relationship with the four adjusted CPCs (at the fourth tier) are modelled based on the original CREAM method (Hollnagel, 1998). Since the development of this part (Tier 4 to Tier 6) of the network has been described in Yang et al., (2013), it is not repeated in this paper.
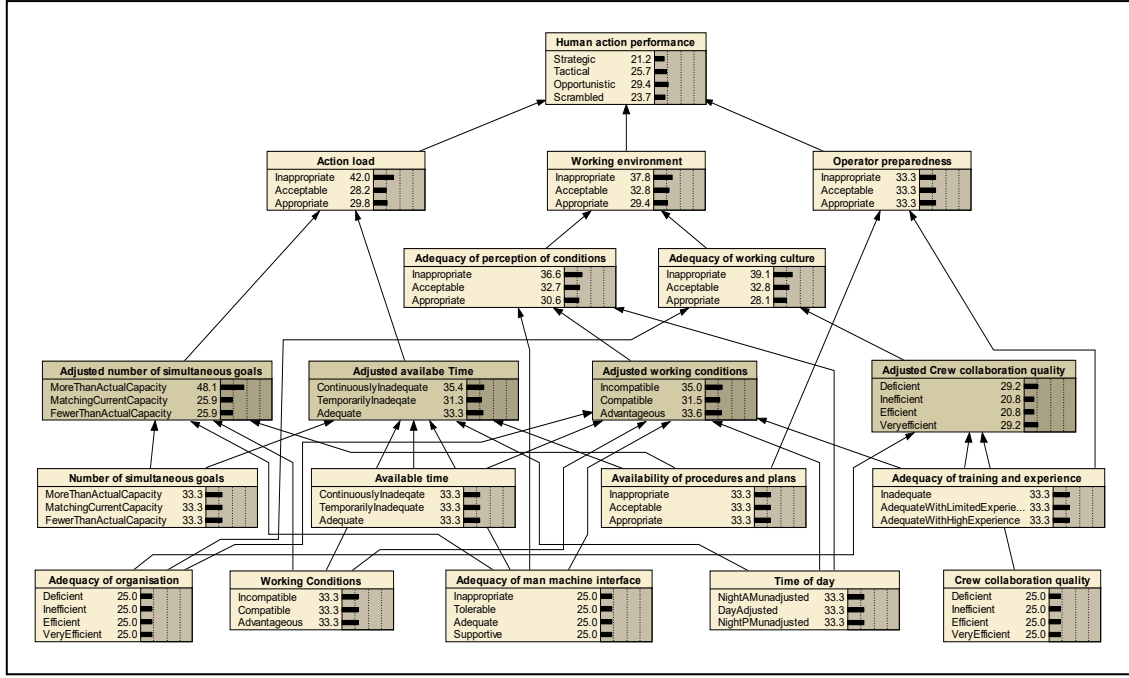
Figure 2. BN based CREAM generic model for human performance reliability assessment

To model the interactive relations between the new attributes and COCOM-CMs in a logical form, fuzzy logic can be used to construct IF-THEN rules. Each of IF-THEN rules includes two parts: an antecedent that responds to the fuzzy input of the three attributes (each of which has three grades) and a consequence associated with the COCOMs' four control modes as the fuzzy output. In this study, a collection of multiple-input multiple-output FRB (consisting of 27 rules (i.e. 3×3×3)) is defined as follows (Yang et al., 2009; 2010):

$$R_l: \text{IF } L_1^{k,l} \text{ and } L_2^{k,l} \text{ and } L_3^{k,l}, \text{THEN } (\beta_{1,l}, \beta_{2,l}, \beta_{3,l}, \beta_{4,l}) \tag{1}$$

In a fuzzy rule R_l ($l = 1, 2, …, 27$), if the input satisfies the antecedent linguistic vector(s) $L_r^{k,l}$ ($r = 1, 2, 3$; $l = 1, 2, …, 27$; $k = 1, 2, 3$), the output $\beta_{j,l}$ ($j = 1, 2, 3, 4$; $l = 1, 2, …, 27$) represents the belief degree(s) to which a control mode D_j ($j = 1, 2, 3$ or 4) is believed to be the consequence. Linguistic vector $L_r^{k,l}$ is defined with its nature of having "Appropriate" (improved), "Acceptable" (not significant) or "Inappropriate" (reduced) effects on COCOM-CMs. Obviously, if $L_1^{k,l}$ is "Action load", then $L_1^{k,l}$ can be any of the three linguistic variables used to describe "Action load", which are Inappropriate ($L_1^{1,l}$), Acceptable ($L_1^{2,l}$), and Appropriate ($L_1^{3,l}$). The following illustrative rule is developed to interpret the rules with a belief structure.

● R_2: IF the "Action load" is Inappropriate AND the "Working environments" are Appropriate AND "Operator preparedness" is Inappropriate, THEN the belief degrees of operator COCOM-CM would be 0% "Strategic", 0% "Tactical", 10% "Opportunistic", and 90% "Scrambled".

It can be further simplified and presented as:

- R_2: IF $L_1{}^1$ AND $L_2{}^1$ AND $L_3{}^2$ THEN (D_1, 0), (D_2, 0), (D_3, 0.1), (D_4, 0.9)

where each $L_r{}^{k,2}$ (r = 1, 2, 3; k = 1, 2, 3) in Rule 2 indicates the kth linguistic variable descriptor associated with the rth attribute. The set of degrees of belief $\beta_{j,2}$ (j =1, 2, 3, 4) = (0, 0, 0.1, 0.9) represents the combined subjective conditional probabilities from domain experts. The way of calculating $\beta_{j,l}$ is given in Section 3.2.2. Such a rule base represents the possible functional mappings of uncertainty between the three new attributes and the four control modes. It provides a more informative, realistic scheme than a simple IF-THEN rule base does on uncertain knowledge representation. However, the challenge lies in the incompleteness knowledge encounters by the experts when assigning degrees of belief in the rule base modelling the relation among O, A, W and COCOMs. In other words, the problem appears in a situation, where the sum of the elicited degrees of belief is less than 1. In order to incorporate them into the estimate of COCOM-CMs probabilities in a convergent connection of a BN, the synthesizing capability of the ER algorithm is investigated accordingly.

3.2 Synthesising the complete and incomplete expert judgements' degrees of belief

In order to investigate the capability of the ER approach in synthesising incomplete assessments, a hierarchy of two levels of attributes is considered, where the upper level represents the synthesised states D_j (j = 1, 2, 3, 4) of the child node (i.e. COCOM), and the lower level represents the states of the parent nodes (i.e. O, A, W) that are denoted by $L_r{}^k$ ($r = 1, 2, 3; k = 1, 2, 3$).

In this respect, the assessment of the conditional probability β_j^i of D_j (j = 1, 2, 3, 4) by the i_{th} expert E_i from a group of M ($i = 1, 2, \dots M$) conditional on $L_r{}^k$ mathematically, is represented by the following distribution:

$$P(E_i \mid L_r{}^k) = (D_j , \beta_j^i), (i = 1, 2, \dots M; j = 1, 2, 3, 4; r = 1, 2, 3; k = 1, 2, 3) \qquad (2)$$

where, $0 \leq \beta_j^i \leq 1, \sum_{j=1}^4 \beta_j^i \leq 1$ and β_j^i denotes a conditional degree of belief assigned to the j_{th} state of the COCOM-CMs node by the i_{th} expert. The above distribution reads that the conditional probability β_j of the child node has been subjectively assessed using the evaluation grade(s) D_j distinctively and conditionally on the parents' evaluation grades $L_r{}^k$ combined with a conditional degree of belief β_j^i.

An assessment by E_i is complete if $\sum_{j=1}^4 \beta_j^i = 1$ and incomplete if $\sum_{j=1}^4 \beta_j^i < 1$. Such partial or complete ignorance is not rare in many distinctive evaluation problems.

Suppose the importance or the relative weight of the expert E_i is given by the weight ω_i ($i = 1, 2, \dots, M$) with the condition that $0 \leq \omega_i \leq 1$. In this regard, the relative importance of E_i plays an important role

in a group assessment. Collectively, ω_i ($i = 1, 2, \dots, M$) has to be normalized for the consistency of the assessment.

To capture the non-linear relationship between different experts E_i ($i = 1, 2, \dots, M$), the ER approach is used to combine all β_j^i ($j = 1, 2, 3, 4$) from each E_i and generate a final conclusion. Having represented belief degree distributions β_j^i, the ER approach can be implemented as follows. First, it is required to transform the degrees of belief β_j^i for all $j = 1, 2, 3, 4$, and $i = 1, 2, \dots, M$ into basic probability masses using the following equations (Yang and Xu, 2002; Liu $et\ al.$, 2005):

$$m_j^i = w_i \beta_j^i, \tag{3}$$

$$m_D^i = 1 - \sum_{j=1}^{4} m_j^i = 1 - w_i \sum_{j=1}^{4} \beta_j^i, \tag{4}$$

$$\overline{m}_D^i = 1 - w_i, \tag{5}$$

$$\widetilde{m}_D^i = w_i \left(1 - \sum_{j=1}^{4} \beta_j^i \right), \text{ for all } j = 1, 2, 3, 4 \text{ and } i = 1, 2, \dots, M. \tag{6}$$

where m_j^i are individual degrees to which E_i supports the final synthesised conclusion D; w_i represents the relevant importance of E_i and thus $\sum_{i=1}^{M} w_i = 1$; and $m_D^i = \overline{m}_D^i + \widetilde{m}_D^i$ for all $i = 1, 2, \dots, M$. The probability mass of E_i (m_D^i) unassigned to the final synthesised conclusion D, which is unassigned to any individual output variables D_j, is split into two parts, one caused by the relative importance of E_i ($\overline{m}_D^i$), and the other due to the incompleteness of the belief degree assessment β_j^i ($\widetilde{m}_D^i$).

Then, it is possible to aggregate all the outputs from E_i ($i = 1, 2, \dots, M$) to generate the combined degree of belief (β_j) in each possible D_j of D. Suppose $m_j^{C(i)}$ is the combined belief degree in D_j by aggregating all the outputs from the M experts and $m_D^{C(i)}$ is the remaining belief degree unassigned to any D_j. Let $m_j^{C(1)} = m_j^1$ and $m_D^{C(1)} = m_D^1$. Then the overall combined belief degree in D_j is generated as follows (Liu $et\ al.$, 2005).

$$\{D_j\}: m_j^{C(i+1)} = K_{C(i+1)}[m_j^{C(i)} m_j^{i+1} + m_j^{C(i)} m_D^{i+1} + m_D^{C(i)} m_j^{i+1}] \tag{7}$$

$$m_D^{C(i)} = \widetilde{m}_D^{C(i)} + \overline{m}_D^{C(i)}, i = 1, 2, \dots, M-1 \tag{8}$$

$$\{D\}: \widetilde{m}_D^{C(i+1)} = K_{C(i+1)}[\widetilde{m}_D^{C(i)} \widetilde{m}_D^{i+1} + \widetilde{m}_D^{C(i)} \overline{m}_D^{i+1} + \overline{m}_D^{C(i)} \widetilde{m}_D^{i+1}] \tag{9}$$

9

$$\overline{m}_D^{C(i+1)} = K_{C(i+1)}[\overline{m}_D^{C(i)}\overline{m}_D^{i+1}] \tag{10}$$

$$K_{C(i+1)} = \left[1 - \sum_{j=1}^{4}\sum_{\substack{t=1 \\ t \neq j}}^{4} m_j^{C(i)} m_t^{i+1}\right]^{-1}, \quad i = 1, 2, \ldots, M-1 \tag{11}$$

$$\{D_j\}: \beta_j = \frac{m_j^{C(M)}}{1 - \overline{m}_D^{C(M)}} \quad (j = 1, 2, 3, 4) \tag{12}$$

$$\{D_j\}: \beta_D = \frac{\widetilde{m}_D^{C(M)}}{1 - \overline{m}_D^{C(M)}} \tag{13}$$

where β_j indicates the normalised belief degree assigned to D_j in the final synthesised conclusion D and β_D represents the normalised remaining belief degree unassigned to any D_j.

3.3 Distributing the unassigned probability masses in the COCOM-CMs BN to obtain a HEP interval

The unassigned probability mass β_D caused by the incompleteness of judgements is assigned back to B_1 (i.e. Strategic) representing the best scenario with the lowest possible HEP and to B_4 (i.e. Scrambled) indicating the worst scenario with the highest possible HEP, respectively. Similarly, all the unassigned probability masses of the other child nodes in Figure 2 are assigned to their own CPTs with respect to the best and worst scenarios. Consequently, two individual COCOM BNs are established, from which the CPTs associated with the best and worst cases will be used to calculate the lowest and highest HEP values. The highest and lowest HEPs can be used as the two limits of an interval. It reflects the fact that the HEP analysis with incomplete input delivers its values in an interval, in which the actual HEP exists.

3.4 HEP quantification and ranking

To quantify the human failures, each D_j ($j = 1, \ldots, 4$) requires the assignment of an appropriate utility value U_{Dj}. The values can be obtained by using a Weighted Mean of Maximum (WMoM) method as 2.24×10^{-4}, 0.01, 0.0708 and 0.316, respectively (Yang et al., 2013). A new HEP index can be calculated as:

$$HEP = \sum_{j=1}^{4} \beta_j U_{D_j} \tag{14}$$

The larger the value of HEP is, the lower the reliability level of human performance. However using Eq 14, the highest and lowest HEPs with respect to the best and worst scenarios can only construct a

HEP interval. Human action is more reliable than the other if and only if its highest value is smaller than the lowest one of the other. It is worth noting that such an approach is not preferred for a ranking purpose. A new coefficient, α, is introduced to indicate evaluators' perception on the two sets of β_j (i.e. β_j^+ and β_j^-) with regards to the best and worst scenarios. More specifically, α means the extent to which the evaluators believe the HEP belongs to the best scenario and $1-\alpha$ represents the extent to which HEP belongs to the worst scenario. If the evaluators are optimistic, $\alpha = 1$ and the final HEP is the lower limit of the HEP interval. If the evaluators are pessimistic, $\alpha = 0$ and the final HEP is the upper limit for the HEP interval. If $0<\alpha<1$, the final HEP can be calculated by using Eq. 15.

$$\beta_j = \alpha\beta_j^+ \cup (1-\alpha)\beta_j^- \tag{15}$$

where $\cup$ means the combination of the two sets by the ER algorithm in Eqs. 3 - 13 and α is set as 0.5 when the evaluators are neutral. The final crisp HEP is then calculated by applying the combined β_j to Eq. 14.

4. Case study of proposed methods in the Deepwater Horizon accident

In this section, a case study of the Deepwater Horizon accident is conducted to illustrate the feasibility and applicability of the hybrid ER-BN model in facilitating the HEP analysis, and the evaluation results are compared with those obtained from traditional CREAM methods. The main reasons of using the proposed ER-BN model to investigate the ***Deepwater Horizon*** accident include that 1) there were several main governing factors symmetrically affecting the effect levels of the nine CPCs over the whole period of the drilling operations, and 2) the uncertainty associated with the available information during the final stages of drilling operations was high.

4.1 Background information of the Deepwater Horizon accident

In the evening of April 20, 2010, a well control event allowed hydrocarbons to escape from Macondo well onto Transocean's Deepwater Horizon, resulting in explosions and fire on the rig. 11 people lost their lives, and 17 others were injured. The fire, which was caused by the hydrocarbons from the well, continued for 36 hours until the rig sank. Hydrocarbons continued to flow from the reservoir through the wellbore and the Blow Out Preventer (BOP) for 87 days, causing a spill of a national significance.

Deepwater Horizon was located approximately 50 miles south of Venice, LA at Mississippi Canyon 252. The accident on April 20, 2010, involved a well integrity failure, followed by a loss of hydrostatic control of the well. This followed a failure to control the flow from the well with the BOP equipment, which allowed the release and the subsequent ignition of hydrocarbons. Ultimately, the BOP emergency functions failed to seal the well after the initial explosions (BP, 2010).

4.2 Aggregating multi attribute effects on the root cause nodes (i.e. the nine CPCs)

The evaluation of CPCs in this case study is based on the accident investigation team's analysis results specifically presented in *Appendix T of the Deepwater Horizon Accident Investigation Report* (BP, 2010). The report describes the relevant practices, procedures, and expectations, comparing them with the rig crew's actions in monitoring the Macondo well and managing the well control event on 20 April 2010. It includes the documents that governed the drilling operations on board the *Deepwater Horizon* at the time of the accident; the available real-time data; and the witness account interview. In this respect, Table 2 summarizes the specified functional assessment attributes, the identified evidence, and their evaluation.

Table 2. Identified relevant practices, procedures, and expectations of rig crew's actions in monitoring the Macondo well and managing the well control event on 20 April 2010

Functional assessment attributes			Investigation team review results	
			Identified evidence	Evaluation
1	Task responsibilities		The manager was not clearly defined	The investigation team could not verify whether anyone fitted the description of manager or had task responsibilities, and who should have made enquiries regarding the results of the negative pressure test that had been conducted to prove that the well structure integrity was intact at the time the negative pressure test results were concluded.
			The well driller's responsibility is to detect a well control situation and shut down the well quickly, and to minimize the kick size used to enhance the safety of a well control operation.	Neither the driller nor the tool-pusher realized that there were impending well control events.
2	Preparation procedures		The review of well control preparation procedures has not occurred	There is no evidence.
3	Prevention	Procedures	On April 20, 2010 between 13:28 and 17:17 hours drilling mud fluid volume monitoring equipment was not properly used; in addition, it was not known what equipment they were using.	Pressure and flow variations should have been available that would have indicated an abnormality with the oil well. In this regard, "extreme caution" could include factors such as pressure changes and flow increases. It would also include isolated individual volume monitoring to enhance well structure intact integrity.

		Witness accounts 1	On April 20, 2010 from 13:28 hours to 17:17 hours, mud was transferred to the supply vessel. Transferring mud from the pits to the supply vessel impaired the ability of mud-loggers to reliably monitor the pit levels. Mud-logger stated this concern was raised with the assistant driller. The response was that the assistant driller would notify him when the mud transfer was completed and monitoring could resume. Mud-logger indicated that this notification did not occur after mud transfer to the supply vessel stopped at 17:17 hours.	Mud-logger did not effectively monitor pit volumes for the remainder of that day.
		Witness accounts 2	There is no evidence to suggest that either the driller or assistant driller was monitoring the well mud fluid volumes and flow. Although mud-loggers' well monitoring equipment was installed and working, it was apparently not being used due to mud transfer to the supply vessel and mud pit cleaning activities.	A more timely response to well conditions may have occurred if "constant, accurate observation and recording of mud volume" was implemented as defined in high pressure high temperature drilling guide lines stated in the documents governing the drilling operation.
4	Detection	Procedures	Mud pumps were stopped at 21:31 on April 20, 2010, but the driller and the tool-pusher both apparently were trying to understand the deferential pressure just prior to the accident.	Neither the driller nor the tool-pusher realized that there was an impending well control event.
		Real-time data 1	There was an increase in return flow from the well at 20:58 hours on April 20, 2010, approximately 51 minutes before the first explosion. However, drill pipe pressure also increased and went unnoticed. The real time data indicts that a 39 bbl gain was taken in the mud pits at that time.	Interim reports and the real time data indicate that the trip tank was being emptied at that time. This may have masked the volume change caused by flow from the well.
		Real-time data 2	At 21:08 hours on April 20, 2010, pumping was stopped, and the sheen test intended to indicate the presence of free oil was performed on the spacer returning from the well. From this time forward, the fluid returning from the well was discharged overboard.	If the driller's flow metre had been operating properly, increasing return flow would have been detected at this time.
		Real-time data 3	While fluids were being discharged overboard, the mud loggers' flow meter bypassed.	The mud loggers were unable to monitor flow.
		Witness accounts 1	Mud-logger indicated that mud flow would not be seen if the flow diverter was activated or going through the dump line. The mud logging system is far more accurate.	

		Real-time data 4	Real-time data indicates that circulation continued after flow increased and pump pressure fluctuated between 20:58 hours and 21:31 hours.	By the time the mud pumps were shut down at 21:31hours, an estimated 300bbl gain had been taken into the wellbore and the well was flowing.
		Real-time data 5	Well flow modelling indicates that between 21:36 hours and 21:38 hours a valve was opened and closed on the rig floor, presumably to bleed off pressure from the drill pipe.	Based on wetness accounts, the investigation team concluded that this occurred approximately 4 minutes before mud started flowing onto the rig floor.
		Witness accounts 2	Mud was seen shooting all the way up to the derrick for several seconds, and then it just quit and went down for several seconds after that, and then all of a sudden the degasser mud started to come out of the degasser very strongly onto the deck. Mud flow volume through the rotary table at the surface was significant.	Based on the procedure defined for equipment handling gas in the riser, the mud flow should have been routed overboard. Instead, the mud flow was routed through the mud gas separator. Based on gas dispersion and explosion analyses, the investigation team concluded that, if the rig crew had diverted mud flow to the overboard discharge line rather than to the mud gas separator, the consequences of the event would have been reduced.
5	Blowout emergency response		The emergency response procedure that should be developed jointly by the management and the operator to be used in case of well blowout was requested.	Such document was not received at the time of investigation.
6	Containment		Events stated do not support a conclusion that action was taken to shut the well in the shortest possible time, as required by the documents governing the drilling operation, following the sequence for shutting down a well when either tripping or drilling.	In the opinion of the investigating team, despite the guidance provided in the documents governing the drilling operation, wellbore monitoring did not identify the influx until after hydrocarbons were in the riser, and the subsequent action taken prior to the explosion suggests the rig crew was not sufficiently prepared to manage an escalating well control situation.

The inherent variability effects that shaped operators' actions and observations in the context of events are used in CPCs' effect level evaluations. The evaluations listed in Table 3 have been conducted in a way in which 1) if there is direct evidence from Table 2 supporting a particular effect level of CPCs, then a 100% degree of belief is assigned accordingly, 2) if there is no evidence or relevant information available to support the evaluation with respect to a particular effect level of CPCs, then average degrees of belief are assigned across all the effect levels to reflect the unknown situation, and 3) if it is irrelevant to the effect, (x) is applied. Given the functional assessment attributes are exclusive factors influencing the effect level of the CPCs, their evaluations can be considered as pieces of evidence to support the performance of the CPCs, the ER algorithm (i.e. Eqs 3-13) is used to synthesise them to obtain the effect levels of the 9 CPCs within the context of the Deepwater Horizon case. The intelligent decision system (IDS) software (Yang and Xu, 2002) is used to aggregate the evaluated degrees of belief of each

functional assessment attribute to obtain the unconditional probabilities of their associated CPCs' effect levels, as shown in Table 4.

Table 3. Evaluation of functional assessment attributes affecting CPCs' effect levels/descriptors

CPCs	Levels/descriptors	Functional assessment attributes					
		Task responsibilities	Preparation	Prevention	Detection	Emergency response	Containment
Adequacy of organisation CPC 1	Very Efficient	0	0	0	0	25	0
	Efficient	0	0	0	0	25	0
	Inefficient	0	0	100	100	25	0
	Deficient	100	100	0	0	25	100
Working conditions CPC 2	Advantageous	x	33.3	33.3	33.3	33.3	33.3
	Compatible	x	33.3	33.3	33.3	33.3	33.3
	Incompatible	x	33.4	33.4	33.4	33.4	33.4
Adequacy of MMI and operational support CPC 3	Supportive	x	x	0	0	25	0
	Adequate	x	x	0	0	25	0
	Tolerable	x	x	50	50	25	0
	Inappropriate	x	x	50	50	25	100
Availability of procedures/plans CPC 4	Appropriate	0	0	0	0	33.3	100
	Acceptable	100	0	100	100	33.3	0
	Inappropriate	0	100	0	0	33.4	0
Number of simultaneous goals CPC 5	Fewer than capacity	x	33.3	33.3	33.3	33.3	33.3
	Matching current capacity	x	33.3	33.3	33.3	33.3	33.3
	More than capacity	x	33.4	33.4	33.4	33.4	33.4
Available time CPC 6	Adequate	x	100	100	100	33.3	33.3
	Temporarily inadequate	x	0	0	0	33.3	33.3
	Continuously inadequate	x	0	0	0	33.4	33.4
Time of day (circadian rhythm) CPC 7	Day-time (6:00-18:00hr) (adjusted)	x	33.3	50	50	33.3	50
	Night(17:00-24:00hr) (unadjusted)	x	33.3	50	50	33.3	50
	Night-time(0:00-7:00hr) (unadjusted)	x	33.3	0	0	33.4	0
Adequacy of training and expertise CPC 8	Adequate, high experience	33.3	33.3	0	0	33.3	0
	Adequate, limited experience	33.3	33.3	0	0	33.3	0
	Inadequate	33.4	33.4	100	100	33.4	100
Crew collaboration quality CPC 9	Very efficient	25	25	0	0	25	0
	Efficient	25	25	0	0	25	0
	Inefficient	25	25	0	0	25	0
	Deficient	25	25	100	100	25	100

Table 4. CPCs effect levels/descriptors and the assigned degrees of belief aggregation with IDS

CPCs	Level/descriptors	Functional assessment attributes aggregated degrees of belief
Adequacy of organisation CPC 1	Very Efficient	3.51
	Efficient	3.51
	Inefficient	57.09
	Deficient	35.89
Working conditions CPC 2	Advantageous	33.33
	Compatible	33.33
	Incompatible	33.34
Adequacy of MMI and operational support CPC 3	Supportive	5.18
	Adequate	5.18
	Tolerable	29.52
	Inappropriate	60.12
Availability of procedures/plans CPC 4	Appropriate	19.95
	Acceptable	60.10

Number of simultaneous goals CPC 5	Inappropriate	19.95
	Fewer than capacity	33.33
	Matching current capacity	33.33
	More than capacity	33.34
Available time CPC 6	Adequate	78.82
	Temporarily inadequate	10.59
	Continuously inadequate	10.59
Time of day (circadian rhythm) CPC 7	Day-time (6:00-18:00hr) (adjusted)	36.25
	Night(17:00-24:00hr) (unadjusted)	49.13
	Night-time(0:00-7:00hr) (unadjusted)	14.62
Adequacy of training and expertise CPC 8	Adequate, high experience	14.01
	Adequate, limited experience	14.01
	Inadequate	71.98
Crew collaboration quality CPC 9	Very efficient	6.43
	Efficient	6.43
	Inefficient	6.43
	Deficient	80.70

4.3 Calculate the CPTs in two BNs for the best and worst scenarios

The use of the ER algorithm in the formation of CPTs in BN with incomplete information (e.g. subjective judgements) is demonstrated in this part. To obtain the information needed for constructing the CPT of the COCOM-CMs node (i.e. the node of human action performance at the top of Figure 2), three maritime experts E_i ($i = 1, 2, 3$) with significant domain knowledge were interviewed to provide their subjective elicitation on the evaluation grades of the COCOM-CMs in terms of conditional degrees of belief as defined by Eq. 1. Three offshore/marine engineers provided their input data within the context of marine engineering operations. The careful selection of the representative experts within the maritime industry is conducted to reduce the bias involved in the subjective judgements. Each of the three selected experts has over 15-year working experience on board offshore rigs or commercial ships and holds a high position in his/her companies. This contributes to the high quality of the initial data from experts. According to the collected feedback (i.e. Table 5), the initial judgements of the three experts keep a very high consistency, which proves that the judgements are at large in harmony and the data quality is assured. Their inputs are listed in Table 5.

Table 5. Elicitation of evaluation grades' conditional degrees of belief for the three attributes, O, W and A

R	E_1				E_2				E_3			
	$D_{1,1}$	$D_{2,1}$	$D_{3,1}$	$D_{4,1}$	$D_{1,2}$	$D_{2,2}$	$D_{3,2}$	$D_{4,2}$	$D_{1,3}$	$D_{2,3}$	$D_{3,3}$	$D_{4,3}$
	$B_{1,1}$	$B_{2,1}$	$B_{3,1}$	$B_{4,1}$	$B_{1,2}$	$B_{2,2}$	$B_{3,2}$	$B_{4,2}$	$B_{1,3}$	$B_{2,3}$	$B_{3,3}$	$B_{4,3}$
1	0	0	0	1	0	0	0	0.9	0	0	0	1
2	0	0	0.1	0.9	0	0	0.1	0.8	0	0	0.2	0.8
3	0	0	0.2	0.8	0	0	0.2	0.7	0	0	0.3	0.7
4	0	0	0.2	0.8	0	0	0.2	0.7	0	0	0.4	0.6
5	0	0.1	0.3	0.6	0	0	0.3	0.6	0	0	0.7	0.3
6	0	0.2	0.3	0.5	0	0	0.4	0.5	0	0.2	0.8	0
7	*0*	*0*	*0.3*	*0.7*	*0*	*0.1*	*0.3*	*0.3*	*0*	*0*	*0.4*	*0.6*
8	0	0.2	0.4	0.4	0	0.2	0.4	0.3	0	0	0.7	0.3
9	0	0.3	0.4	0.3	0	0.2	0.5	0.2	0.4	0.6	0	0
10	0	0	0.2	0.8	0	0.1	0.8	0	0	0	0.4	0.6
11	0.1	0.3	0.3	0.3	0	0.2	0.7	0	0	0.8	0.2	0

12	0.1	0.4	0.4	0.1	0	0.2	0.7	0	0	0.7	0.3	0
13	0.1	0.3	0.4	0.2	0	0.3	0.6	0	0	0.6	0.4	0
14	0.2	0.4	0.3	0.1	0	0.5	0.4	0	0.3	0.7	0	0
15	0.3	0.4	0.2	0.1	0	0.5	0.4	0	0.2	0.8	0.4	0
16	0	0.3	0.4	0.3	0	0.6	0.3	0	0	0.5	0.5	0
17	0.2	0.4	0.4	0	0	0.7	0.2	0	0.5	0.5	0	0
18	0.3	0.4	0.3	0	0	0.8	0.1	0	0.6	0.4	0	0
18	0	0	0.3	0.7	0.3	0.4	0.2	0	0	0	0.3	0.7
20	0	0.2	0.4	0.4	0.3	0.4	0.2	0	0	0.6	0.4	0
21	0	0.4	0.4	0.2	0.4	0.3	0.2	0	0	0.4	0.6	0
22	0	0.2	0.3	0.5	0.6	0.4	0	0	0	0.5	0.5	0
23	0.2	0.4	0.4	0	0.6	0.3	0	0	0.7	0.3	0	0
24	0.3	0.5	0.2	0	0.7	0.20	0	0	0.8	0.2	0	0
25	0.1	0.4	0.4	0.1	0.7	0.2	0	0	0.5	0.5	0	0
26	0.3	0.5	0.2	0	0.8	0.1	0	0	0.8	0.2	0	0
27	0.5	0.5	0	0	0.9	0	0	0	0.9	0.1	0	0

Due to the similar seniority of the three experts, equal weight was assigned to each expert when synthesising their judgements using the ER algorithm. Taking Rule No. 7 in Table 5 as an example, the first two assessments by E_1 and E_2 are synthesised, as presented in Appendix A. In a similar way, the result of combining three experts' judgements can be obtained by synthesising the combination of the first two assessments (as one set) with the third assessment (expert E_3) using the same algorithm. It is worth noting that judgements from other experts can be also combined when more feedback from a wider range of interview is collected in future research. Consequently, the synthesised human action control modes' degrees of belief B_j for the 7th rule are *Strategic* $(B_1) = 0$, *Tactical* $(B_2) = 0.0252$, *Opportunistic* $(B_3) = 0.3271$ and *Scrambled* $(B_4) = 0.5713$. The results reveal that an unknown mass of 0.0744 is involved in the 7th rule due to the expert judgements. Windows based IDS software was developed to simplify the above calculation process by Yang (2001). It is used to synthesise the basic attributes E_i of Rule 7 with the same result obtained. IDS is also used in synthesising the other combined degrees of belief (or probabilities) listed in Table 5.

Although the ER algorithm is used to synthesise experts' combined degrees of belief mass β_j, a remaining unknown mass β_D, which is not assigned to any evaluation grades, is also developed. Consequently, the remaining unassigned degrees of belief are assigned back to the best evaluation grade "*Strategic*" and the worst evaluation grade "*Scrambled*" on all rules. Accordingly, two sets of evaluation grades are generated in Table 6 and used as prior probabilities in the generic COCOM BN model to calculate HEP estimates. Consequently, there are two BNs models presenting the best and worst scenarios.

Table 6. Synthesised and combined degrees of beliefs of COCOM-CMs evaluation grades

Rule	Strategic (D_1)	Strategic (D_1) C Unknown	Tactical (D_2)	Opportunistic (D_3)	Scrambled (D_4)	Scrambled (D_4) C Unknown	Unknown
1	0	0.0211	0	0	0.9789	1	0.0211
2	0	0.0231	0	0.1019	0.8750	0.8981	0.0231
3	0	0.0244	0	0.1973	0.7783	0.8027	0.0244

4	0	0.0248	0	0.2321	0.7431	0.7679	0.0248
5	0	0.0265	0.0279	0.4316	0.5140	0.5405	0.0265
6	0	0.0277	0.1220	0.5323	0.3180	0.3457	0.0277
7	0	0.0744	0.0252	0.3271	0.5713	0.6457	0.0744
8	0	0.0271	0.1164	0.5282	0.3283	0.3554	0.0271
9	0.1227	0.1519	0.3910	0.2979	0.1592	0.1884	0.0292
10	0	0.0274	0.0274	0.4734	0.4718	0.4992	0.0274
11	0.0295	0.0576	0.4495	0.4045	0.0884	0.1165	0.0281
12	0.0284	0.0554	0.4394	0.4768	0.0284	0.0554	0.0270
13	0.0284	0.0555	0.4031	0.4845	0.0569	0.0840	0.0271
14	0.1525	0.1797	0.5789	0.2126	0.0288	0.0562	0.0274
15	0.1512	0.1784	0.6164	0.1767	0.0285	0.0557	0.0272
16	0	0.0271	0.4819	0.4058	0.0852	0.1123	0.0271
17	0.2159	0.2433	0.5757	0.1809	0	0.0274	0.0274
18	0.2851	0.3125	0.5695	0.1180	0	0.0274	0.0274
19	0.0865	0.1153	0.1153	0.2715	0.4979	0.5267	0.0288
20	0.0855	0.114	0.4205	0.3457	0.1198	0.1483	0.0285
21	0.1125	0.1406	0.3815	0.4188	0.0591	0.0872	0.0281
22	0.1758	0.2343	0.3223	0.2822	0.1612	0.2198	0.0586
23	0.5278	0.5817	0.2997	0.1186	0	0.0539	0.0539
24	0.6370	0.6635	0.2810	0.0555	0	0.0265	0.0265
25	0.4437	0.4719	0.3803	0.1182	0.0296	0.0578	0.0282
26	0.6753	0.7016	0.2432	0.0552	0	0.0263	0.0263
27	0.8153	0.8397	0.1603	0	0	0.0244	0.0244

The CPT of the COCOM-CMs node in the BN presenting the best scenario can be obtained by taking into account the values of (D1)C, (D2), (D3) and (D4), while the one for the worst scenarios is associated with the set of {(D1), (D2), (D3) and (D4)C} (See Table 6). For instance, in Rule 1 in Table 6, if *O* is inappropriate, *W* is inappropriate, and *A* is inappropriate, then COCOM-CMs are {0.0211 Strategic, 0 Tactical, 0 Opportunistic, 0.9789 Scrambled} in the best scenario; and {0 Strategic, 0 Tactical, 0 Opportunistic, 1 Scrambled} in the worst scenario. As a result, the two BNs for the best and worst scenarios are constructed and presented in Figures 3 and 4, respectively.
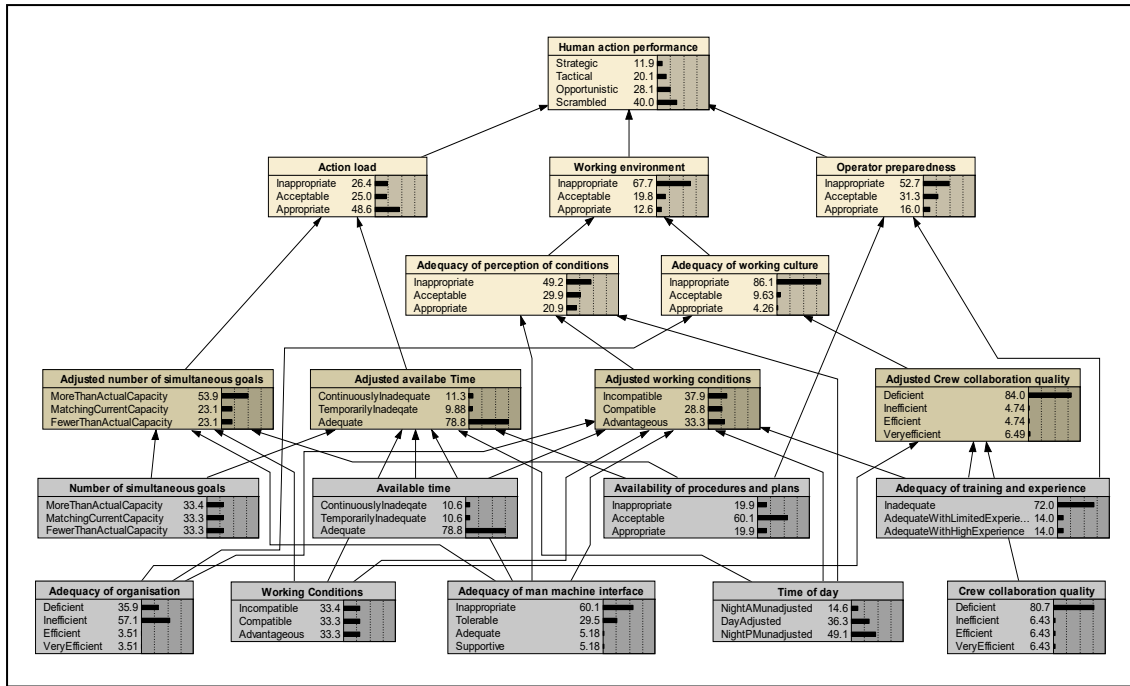
Figure 3: BN model displaying human COCOM-CMs' posterior probabilities based on the best possible set of evaluation grades
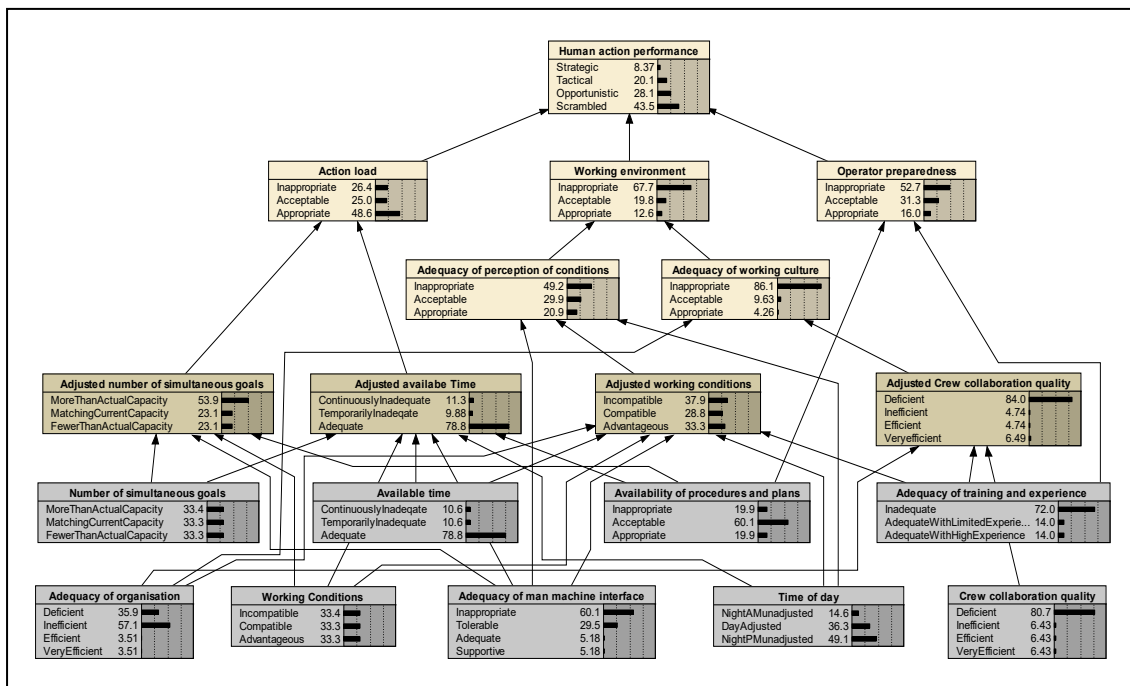

Figure 4: BN model displaying human COCOM-CMs' posterior probabilities based on the worst possible set of evaluation grades

4.4 Quantification of human error in the Deepwater Horizon case

The evaluations of the CPCs' effect levels in Table 4 are used as input observations of root cause nodes in the two established BN models in Figures 3 and 4. The posterior probabilities of the two assessment models' evaluation grades and their transformed respective HEP results are shown in Table 7. During this process, Eq. 14 is used to calculate the lower and upper limits of the HEP interval from the best and worst scenarios. The result shows that the HEP estimates in the case are from 14.83% to 15.94%[1].

Finally, HEPs are presented in a utility interval rather than a crisp value. Such an interval could be used effectively to specify the uncertainty involved in the assessment. However, for a ranking purpose, a crisp value of the HEP interval can be calculated as 0.1547 using Eq. 15, as presented in Table 7. The result indicates a control mode of 'opportunistic' according to the values of probability of action failure defined in the CREAM methodology (See Table 8).

Table 7. Final HEPs of both assessed scenarios

Conditions	HEP/time
1. The worst scenario BN aggregated COCOM-CMs probability and their transformed highest HEP	$0.0838 \times 0.000224 + 0.201 \times 0.01 + 0.281 \times 0.0708 + 0.435 \times 0.316$ $= 0.1594$
2. The best scenario BN aggregated COCOM-CMs probability and their transformed lowest HEP	$0.119 \times 0.000224 + 0.201 \times 0.01 + 0.281 \times 0.0708 + 0.40 \times 0.316$ $= 0.1483$
The highest and lowest values of HEPs interval	$0.1483 \leq HEP \leq 0.1594$
The HEP crisp value (e.g. $\alpha = 0.5$)	0.1547

Table 8. The control modes and probability intervals (Hollnagel, 1998)

Control mode	Probability interval
Strategic	$0.5 \times 10^{-5} < p < 1 \times 10^{-2}$
Tactical	$1 \times 10^{-3} < p < 1 \times 10^{-1}$
Opportunistic	$1 \times 10^{-2} < p < 0.5 \times 10^{0}$
Scrambled	$1 \times 10^{-1} < p < 1 \times 10^{0}$

4.5 Comparative analysis and discussion of the results

---

[1] Such HEPs are subject to the assignment of the utility values of the four control modes. Although the utility values are cited from the work from a leading journal (i.e. Ocean Engineering), further verification of such utility values is still required to fully validate the significance of the HEP values in terms of precise risk analysis. Having said that, it is believed that the obtained HEP values can be effectively used for risk prioritisation.

Based on the identified relevant practices, procedures, and expectations of rig crew's actions from the Deepwater Horizon Accident Investigation Report (as presented in Table 5), the traditional CREAM and fuzzy CREAM methods are applied in the same case study for the quantification of human error in HRA. It allows for a more practical and effective investigation of the applicability of the proposed model through a comparative analysis. In this subsection, we mainly focus on the comparison of their pros and cons in the real-life applications, as well as the similarity and difference of results obtained from different methods. Thus, detailed information on how to conduct these CREAM-based methods is not omitted here, and more information can be found in Hollnagel (1998) and Konstandinidou et al. (2006) for further reference. The evaluation results from different methods are presented and compared in Table 9.

Table 9. Comparison of the results obtained from different methods

	Methods		
	CREAM	Fuzzy CREAM	ER-BN CREAM
Control mode	Opportunistic	Opportunistic	Opportunistic/ Scrambled
Probability interval	(0.01 , 0.5)	(0.01 , 0.5)	(0.1483, 0.1594)
HEP crisp value	-	0.1	0.1547

At this stage of validation, it is worthwhile to differentiate the results obtained from different CREAM related methods. In this context, the traditional CREAM result is calculated based on CPCs' evaluation scores improved and reduced (1, 7); the plotting of these scores on the graph shown in Figure 1 reveals the result of opportunistic control mode and its related generic probability interval. It only provides a wide failure rate interval. Without a crisp value of HEP, it is not even suitable for screening in practical applications. The result generated by the fuzzy model, which can be expressed in the form of a crisp number, can be used directly in fault tree and event tree calculations for the quantification of specific undesired events. However, it also suffers from the problem of failing to incorporate the uncertainties in data (e.g. incomplete information) involved during the assessment. Besides, altogether 46656 fuzzy rules are included in the proposed fuzzy model (Konstandinidou et al., 2006), which, due to its complexity, inevitably hinders its industrial applications. In the proposed ER-BN CREAM model, the final calculation of HEP is 0.1547 failure/time; this HEP inclusively lies within the range of original opportunistic mode in CREAM ($0.1 < HEP < 0.5$), revealing the accuracy of the result. However it improves the accuracy of the HEP interval from [0.01, 0.5] to [0.1483, 0.1594] with a HEP crispy value of 0.1547. Comparing the results with those obtained from other approaches, obviously, the specified HEP would provide a more accurate result with a better resolution that will enable an assessor to develop a rational preventative plan.

By changing the BN mode, it has been found that CPC8 (adequate training and expertise to 100% with respect to all functional assessment attributes) has the most significant impact on

the reduction of the HEP. Theoretically, such an analysis can provide scientific support on which CPC(s) should be better controls in the recommendations for avoidance of similar accidents in future. From a perspective of capability and competency, it is suggested to deepen the capabilities of personnel in key operational and leadership positions and augment existing knowledge and proficiency in management deep-water drilling and wells operations. Also, advanced deep-water well control programs that supplement current industry and regulatory training need to be developed (BP, 2010). With respect to the training and exercises of emergency response, the organization is suggested to motivate personnel to discuss safety-related concerns of the emergency drills and exercises to increase personnel's skills in the emergency. It is also important for organizations and personnel to regularly check the emergency equipment and procedures associated with their capacities, arrangement, and performance standards during the emergency drills and exercises (Norazahar et al., 2014).

5. Contribution and implication

The major contributions of the generated methods and models in this paper are explained from both practical and theoretical aspects. Practically, the uses of the BN inference and FRB structure can effectively help to forecaste the high HEPs of hazardous situations and send an early warning signal to prevent maritime accidents. Accordingly, the used techniques would provide the potential for identifying the most influencing CPC(s) and the associated functional assessment attributes (e.g. initiating events or root causes), to develop the risk control options effectively. In addition, using the ER algorithm for synthesising expert's judgments for Bayesian subjective probability elicitation is able to enhance CREAM based HRA methodology, which will facilitate the application of relevant methods in human performance reliability analysis in maritime and offshore domains where expert judgements are usually involved due to the lack of reliable data in maritime safety assessment. Theoretically, both BN probabilistic inference and ER synthesising capabilities are used to represent and process context knowledge and uncertainty. The combination provides a feasible solution to subjective elicitation of CPT in BN applications.

The developed methods possess enormous potential as valuable aids and effective alternatives to retain and improve human performance in marine engineering operations. It has the potential and flexibility to be tailored to handle the incompleteness of subjective data when using BNs to aid decision-making in other sectors.

6. Conclusion

ER's synthesising and aggregation capability has enlarged the scope of a BN mechanism inference viability in describing and handling uncertain information in an engineering operation context. By using the concept of degrees of belief, the ER-BN combination can model context knowledge incompleteness and ignorance explicitly at any BN assessment level. Combining degrees of ignorance with the best and worst evaluation grades can generate two BNs to describe the best and worst scenarios of COCOM-

CMs' probabilities. Subsequently, their results are transformed and presented in HEP intervals, where each could be further converted into a crisp HEP value for a ranking purpose, as demonstrated in the above case study. Consequently, a new hybrid ER-BN method is developed capable of handling the problems to which the traditional methods lack the ability to provide appropriate solutions. Applying it in HRA facilities the assessment of HEPs through the established CREAM BN generic model in a situation where incomplete subjective probability elicitation is necessary.

Acknowledgements

REFERENCES

1. Abujaafar K., Yang Z., Wang J., Nazir S. and Øvergård K.I., "Modelling Adequacy of Organisation in Human Reliability Analysis – A Case of Marine Engineering Operations", European Safety and Reliability Conference 2015 (ESREL15), September 6-10, 2015, Zurich, Switzerland.

2. Abujaafar K.M., Qu Z., Yang Z., Wang J., Nazir S. and Øvergård K.I., "Use of evidential reasoning for eliciting Bayesian subjective probabilities in human reliability analysis", 11th International Conference on System of Systems Engineering (SoSE), 12-16 June, 2016, Kongsberg, Norway.

3. BP, "Deepwater Horizon Accident investigation report", Investigation Team Internal Report, September 8, 2010.

4. Dempster A.P., "A generalization of Bayesian inference", Journal of the Royal Statistical Society (Series B), Vol. 30, pp. 205-247, 1968.

5. Fujita Y, and Hollnagel E., Failures without errors: quantification of context in HRA. Reliability Engineering & System Safety, Vol. 83, pp.145-151, 2004.

6. Hollnagel E., Cognitive Reliability and Error Analysis Method - CREAM, Oxford, UK: Elsevier Science. 1998.

7. Kim M.C., Seong P.H. and Hollnagel E., "A probabilistic approach for determining the control mode in CREAM", Reliability Engineering and System Safety, Vol. 91, pp.191-199, 2006.

8. Konstandinidou, M., Nivolianitou, Z., Kiranoudis, C., & Markatos, N. (2006). A fuzzy modeling application of cream methodology for human reliability analysis. Reliability Engineering & System Safety, 91(6), 706-716.

9. Liu J., Yang J.B., Wang J. and Sii H., "Engineering system safety analysis and synthesis using the fuzzy rule-based evidential reasoning approach," Quality and Reliability Engineering International, Vol. 21, No. 4, pp. 387-411, 2005.

10. Marseguerra M., Zio E. and Librizzi M., "Human reliability analysis by fuzzy CREAM", Risk Analysis, Vol. 27, pp137-154, 2007.

11. McErleani F.J., Bell D.A. and Guan J.W., "Modification of belief in evidential causal networks", Information and Software Technology, Vol. 41, pp. 597-603, 1999.

12. Norazahar, N., Khan, F., Veitch, B., & Mackinnon, S. (2014). Human and organizational factors assessment of the evacuation operation of BP Deepwater Horizon accident. *Safety Science, 70*, 41-49.

13. Shafer G., A mathematical theory of evidence, Princeton University Press, New Jersey, USA. 1976.

14. Smets P., "Belief function", In: Smet, P., Mamdani, E.H., Dubois, D. and Prade, H., eds, Non-Standard Logics for Automated Reasoning (Academic Press, London), pp 253-277, 1988.

15. Ung, S. T. (2015). A weighted cream model for maritime human reliability analysis. *Safety Science*, 72, 144-152.

16. Ung, & Shuen-Tai. (2018). Human error assessment of oil tanker grounding. *Safety Science, 104*, 16-28.

17. Wan, C., Yan, X., Zhang, D., & Yang, Z. (2019). A novel policy making aid model for the development of LNG fuelled ships. *Transportation Research Part A: Policy and Practice, 119*, 29-44.

18. Xi Y.T., Yang Z., Fang Q.G., Chen W.J. and Wang J., "A modified CREAM for human error probability quantification", Ocean Engineering. Vol. 138, pp. 45-54, 2017.

19. Yang J.B. and Xu D.L., "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty", IEEE Transactions on Systems, Man, and Cybernetics Part A, Vol. 32 No. 3, pp. 289-304, 2002.

20. Yang, Z. L., Bonsall, S, Wall, A., Wang, J. & Usman, M. (2013). A modified cream to human reliability quantification in marine engineering. *Ocean Engineering, 58*, 293-303.

21. Yang Z., Bonsall S. and Wang J., "Fuzzy rule-based Bayesian reasoning approach for prioritization of failures in FMEA", IEEE Transactions on Reliability, Vol. 57, No. 3, pp. 517-528, 2008.

22. Yang Z., Bonsall S. and Wang J., "Use of fuzzy evidential reasoning in maritime security assessment", Risk Analysis, Vol. 29, pp. 95-120, 2009.

23. Yang Z., Bonsall S. and Wang J., "Facilitating uncertainty treatment in the risk assessment of container supply chains", Journal of Marine Engineering and Technology, No. A17, pp. 23-36, 2010.

24. Zhou, Q., Wong, Y. D., Xu, H., Thai, V. V., Loh, H. S., & Yuen, K. F.. (2017). An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. *Safety Science*, 95, 140-147.

Appendix A

To calculate the basic conditional probability masses m_j^i as defined by Eq. 2.

$m_1^1 = 0.333 \times 0 = 0; m_2^1 = 0.333 \times 0 = 0; m_3^1 = 0.333 \times 0.3 = 0.0999; m_4^1 = 0.333 \times 0.7 = 0.2331.$

$m_1^2 = 0 \times 0.333 = 0; \ m_2^2 = 0.1 \times 0.333 = 0.0333; \ m_3^2 = 0.3 \times 0.333 = 0.0999; m_4^2 = 0.3 \times 0.333 = 0.0999.$

$m_1^3 = 0 \times 0.333 = 0; \ m_2^3 = 0 \times 0.333 = 0; \ m_3^3 = 0.4 \times 0.333 = 0.1332; m_4^3 = 0.6 \times 0.333 = 0.1998.$

Next the remaining relative importance $\overline{m}_D^i$ for all $i = (1, 2, 3)$ is obtained as follows using Eq. 5

$\overline{m}_D^1 = 1 - 0.333 = 0.667$

$\overline{m}_D^2 = 1 - 0.333 = 0.667$

$\overline{m}_D^3 = 1 - 0.333 = 0.667.$

The remaining probability mass $\widetilde{m}_D^i$ due to the possible incompleteness of any individual grade β_j^i is defined by Eq. 6.

$\widetilde{m}_D^1 = 0.333[1 - (0 + 0 + 0.3 + 0.7)] = 0;$

$\widetilde{m}_D^2 = 0.333[1 - (0 + 0.1 + 0.3 + 0.3)] = 0.0999;$

$\widetilde{m}_D^3 = 0.333[1 - (0 + 0 + 0.4 + 0.6)] = 0$

The normalizing factor $K_{C(i+1)}$ for combining the two assessments from E_1 and E_2 is calculated using Eq. 11.

$K_{C(i+1)} = [1 - (0 \times 0.0333 + 0 \times 0.0999 + 0 \times 0.0999) + (0 \times 0 + 0 \times 0.0999 + 0 \times 0.0999) + (0.0999 \times 0 + 0.0999 \times 0.0333 + 0.0999 \times 0.0999) + (0.2331 \times 0 + 0.2331 \times 0.0333 + 0.2331 \times 0.0999)]^{-1} = 1.0464.$

The remaining combined probability mass $\widetilde{m}_D^{C(i+1)}$ due to the possible incomplete assessment of β_j^i by E_1 and E_2 is defined by Eq. 8.

$\widetilde{m}_D^{C(i+1)} = 1.0464[(0 \times 0.0999) + (0.667 \times 0.0999) + (0 \times 0.667)] = 0.0697.$

The combined remaining relative importance $\overline{m}_D^{C(i+1)}$ from the two assessments conducted by E_1 and E_2 are obtained using Eq. 9.

$$\overline{m}_D^{C(i+1)} = 1.0464(0.667 \times 0.667) = 0.4655.$$

To calculate the combined probability mass β_j, Eq. 12 is employed as follows.

$\beta_1 = 1.0464[(0 \times 0) + (0 \times 0.7669) + (0.667 \times 0)] = 0;$

$\beta_2 = 1.0464[(0 \times 0.0333) + (0 \times 0.7669) + (0.667 \times 0.0333)] = 0.0232;$

$\beta_3 = 1.0464[(0.0999 \times 0.0999) + (0.0999 \times 0.7669) + (0.667 \times 0.0999)] = 0.1537;$

$\beta_4 = 1.0464[(0.2331 \times 0.0999) + (0.2331 \times 0.7669) + (0.667 \times 0.0999)] = 0.2698.$

Finally, the remaining combined probability mass β_D due to the possible incomplete assessment of E_1 and E_2 is calculated by Eq. 13.

$$\beta_D = \frac{0.0697}{1-0.4655} = 0.1304.$$