Application of a CREAM based framework to assess human reliability in emergency response to engine room fires on ships

Sung Il Ahn and **Rafet Emek Kurt***

Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde

ABSTRACT

For a human reliability assessment in the maritime domain, the main question is how we correctly understand the human factors in the maritime situation in a practical manner. This paper introduces a new approach based on Cognitive Reliability and Error Analysis Method (CREAM). The key to the method is to provide a framework for evaluating specific scenarios associated with maritime human errors and for conducting an assessment of the context, in which human actions take place. The output of the context assessment is, then, to be applied for the procedure assessment as model inputs for reflection of the context effect. The proposed approach can be divided into two parts: processing context assessment and modelling human error quantification. Fuzzy multiple attributive group decision-making method, Bayesian networks and evidential reasoning are employed for enhancing the reliability of human error quantification. Fuzzy conclusion of the context assessment is utilised by the model input in CREAM basic method and weighting factors in CREAM extended method respectively for considering human failure probability which varies depending on external conditions. This paper is expected to contribute to the improvement of safety by identifying frequently occurred human errors during the maritime operating for minimising of human failures.

Keywords: CREAM, Human Reliability Assessment, Maritime, Ship, Fire Fighting, Safety, Human Factors

1. Introduction

Safety is a critical issue in maritime, but it is still a challenge to predict and prevent accident occurrences because the cause of the accident consists of a variety of factors. Notably, the human factors aspects of ship operation in maritime is one of the significant contributions to the accident. The past studies show that human error is deeply related to accidents, ranging from 65 to 90 per cent. (Kristiansen (2013); Ung (2015); Akyuz et al. (2018); Kurt et al. (2016b); Antão and Soares (2019)). However, the terms of human factors and human error are often used without a clear understanding (Khan, 2008). It is due to the fact that the seafarers face many hazardous situations since they should not only carry out the navigation of ship but also have to conduct other responsibilities such as cargo loading and discharging, ballasting and de-ballasting, bunkering and maintenance work including hot and closed space work mostly independently in space away from land. Specific parts of the ship's functions have been automated, but a human still controls or interacts with most of the work. Therefore, in order to ensure safety at sea human factors, specifically Human Reliability Analysis (HRA) needs to be considered at the core of safety assessments. However, HRA has always been a concern for safety engineers and risk assessment analysts due to the fundamental limitations such as insufficient data, methodological limitations related to subjectivity of analysts and expert judgment, and uncertainty concerning the actual behaviour of people during accident conditions (Konstandinidou et al., 2006). According to Schröder-Hinrichs et al. (2011), it is more difficult to collect reliable data because human and organisational factors related to accident development and response to emergency situations are not reported enough. In this context, prospective methods for quantifying human reliability across the first generation and over the third generation HRA methods have been proposed through the nuclear and aviation sectors and recently applied to the marine sector, but the third generation methods are still in the development stage. As a representative method, cognitive reliability and error analysis method (CREAM) was first developed by Hollnagel (1998) and can be considered as one of the most popular and commonly used second-generation HRA method.

According to studies conducted by Hollnagel (1998) and later by Fujita and Hollnagel (2004), to predict human performance reliability, a context description must be provided because a discussion of what is likely to happen in a given situation must be based on a description of the specific circumstances or conditions. It is reasonable that human error probability can be determined directly from a characterisation of the context. This condition is described in terms of the degree of control presented by four characteristic control modes consist of Strategic, Tactical, Opportunistic and Scrambled mode, which identify different reliability of performance.

The CREAM can be used as both retrospective and prospective purposes and CREAM can apply to qualitative and quantitative analysis. The quantitative CREAM consists of basic and extended methods. Firstly, the CREAM basic method is a human failure probability quantification process that defines nine conditions, such as working conditions, crew collaborations, called Common Performance Conditions (CPCs) affecting human performance. In a basic predictive CREAM, it evaluates CPCs to predict human error probability concerning the contextual control modes with four different failure probability interval corresponding to a value of combined CPC scores by using mapping in the diagram of control mode. This method mainly used as screening purpose in HRA and also can be used to identify conditions that may reduce or improve the human reliability aspects of risk assessment. While subsequent and more detailed analyses of human interactions can be acquired by the CREAM extended method (He et al., 2008), the combined score of the CPCs for context assessment derived from the basic method can be an essential parameter for the extended method. The extended method will be necessary to obtain more accurate results for designated tasks of the procedures.

According to Kurt et al. (2015) and Kurt et al. (2016a), their research conducted in the EU funded SEAHORSE Project concluded 20-30% of standard operating procedures are ineffective hence not being followed strictly during operations. This means we need to bring more attention to review procedures on board with a specific focus on human performance in order to achieve safer operations.

In this regard, this paper provides a framework for estimating human error probabilities through scenario description and procedure analysis based on the CREAM method and illustrates the practical application by proposing a way to transform human activities on board and their contextual conditions into analytical forms for HRA. With this objective, the paper is organised as follows: This section introduces HRA in the maritime and CREAM overview. The second section is a literature review, and the third section presents the proposed method based on CREAM. The case study for the procedures of the engine room fire-fighting on the ship is presented in section four. The fifth section gives the finding and discussion, followed by a conclusion in the sixth section.

2. Literature Review

Over the decades, there have been vigorous efforts to understand the mechanism of human error and to prevent maritime incidents caused by human through utilising various human reliability assessment (HRA) techniques; such as, Success Likelihood Index Method (SLIM), Human Error Assessment and Reduction Technique (HEART), Technique of Human Error Rate Prediction (THERP), Human Factors Analysis and Classification System (HFACS), Cognitive reliability and error analysis method (CREAM).

Hence, researchers put a lot of effort to condense the complex circumstances, under which ship crews are highly likely to make mistakes, into simple descriptive numbers known as Human Error Probability (HEP) by means of several uncertainty treatment methods, such as fuzzy logic, Bayesian networks, evidential reasoning, Event tree, Fault tree, and other integrated methods.

Fuzzy logic has been successfully applied in maritime context to wide range of topics concerning maritime safety and risk. For example Balmat et al. (2011) presented a fuzzy approach in order to evaluate the maritime risk assessment to pollution prevention on the open sea while Wu et al. (2019) utilised fuzzy Multiple Attribute Decision Making for ship-bridge collision alert system. Fuzzy logic has also been utilised in numerous studies related to human reliability analysis to improve the reliability and reduce uncertainty in generated results.

In following paragraphs, the details of previous maritime research studies that are conducted by using aforementioned methods (known as the first generation HRAs) are shared:

Akyuz (2016) applied the concept of the SLIM for estimating HEP when conducting the abandon-ship procedures. The fuzzy sets were used to improve the reliability of the analysis against the vagueness of expert judgments and the arbitrary measure of performance shaping factors (PSFs). Based on the SLIM, Islam et al. (2016) determined the HEPs related to marine engine maintenance tasks, where in another study Islam et al. (2017b) developed a monograph for assessing the likelihood of human error in marine operations that could be applicable for instant decision making. It was identified that with SLIM method, it is possible to estimate not only general HEPs in a given context but also HEPs in specific activities by adding particular PSFs, such as training, experience, fatigue level of a seafarer, etc. However, SLIM is overly relying on expert judgment, which makes the analysis results highly subjective and less reliable; it is because the scope of PSFs is limited to certain contexts rather than fully reflective to every aspect that affects human performance. In particular, they are weak in dealing with social and organisational aspects. To remedy the challenges posed in the SLIM, Abbassi et al. (2015) proposed the integration of SLIM with the THERP to investigate PSFs related to an offshore condensate pump maintenance task. The SLIM was used to estimate the human errors that were not covered by THERP.

On the other hand, Akyuz and Celik (2016) applied the HEART in combination of AHP to predict human errors associated with cargo operation on oil/chemical tankers. Islam et al. (2017a) developed an operational specific methodology based on the HEART in order to capture unique features of maritime environment and operation, and applied the method to the maintenance procedures of a marine engine exhaust turbocharger and also a condensate pump fitted to offshore oil and gas facilities. The HEART has a similar nature as the SLIM but it provides nominal probabilities for generic HEART tasks. Thereafter, the overall HEPs are adjusted by evaluating Error Producing Conditions (EPCs) and the proportion of effect defined by experts' judgment. As a result, like the SLIM, the multiplier values are highly relied on experts' knowledge, which leaves uncertainties in analysis results.

The HFACS is firstly proposed by Shappell and Wiegmann (2000). As a qualitative method, it adopts a taxonomic nature for better understanding of human behaviour. To obtain quantified outcomes, some researchers proposed the combination of the HFACS with a Fuzzy Analytical Hierarchy Process (FAHP) or Fault Tree Analysis (FTA). Celik and Cebi (2009) generated an analytical HFACS with the concept of the FAHP, in order to identify the role of human errors in boiler explosions onboard bulk carrier. This study provides an analytical foundation and group decision-making functionality in order to achieve a quantitative assessment of shipping accidents. Zhang et al. (2019) introduced a modified model of the HFACS for collision accidents between a ship and an icebreaker. Then, the FTA model was utilized to analyse the fundamental collision risk factors according to the statistical analysis of accident reports and experts' judgment based on the HFACS-SIBCI model. Collision risk factors during icebreaker assistance were identified and classified under the initial HFACS framework. However, the past research showed HFACS would not fully address the specifics of marine incidents. For example, Salmon et al. (2012) explained the main problems to apply HFACS to the outside of aviation is that it was developed specifically for aviation, a number of the error and failure modes are aviation specific.

Furthermore, de Maya et al. (2019b) proposed MALFCM approach incorporated with BNs which is based on the concept and principles of fuzzy cognitive maps (FCMs) to represent the interrelations amongst accident contributor factors. As a weakness, although this database-driven research has led to successful results, the applicable range of the database is far limited to some specific cases rather than general ones.

Unlike the HRA studies mentioned above, Vagias (2010) investigated specific factors relating to human fatigue. BNs were utilised to predict fatigue prevalence and its importance, given the information regarding workload, environment, and ergonomic factors, prior to the occurrence of the accident. This study also provides comprehensive information about Human Factors and human error.

There have also been attempts to develop models that could directly estimate overall HEPs using BNs. Islam et al. (2018) introduced a BN model to estimate HEP by using priority probability and CPT (conditional probability table) from expert groups. In aforementioned study the impact of internal and external factors on human performance were defined in a case study for ship maintenance activities. The BN model provides flexible HEPs that could be obtained based on new information inputted to variables. As such, it is capable to predict HEPs across various maritime scenarios effectively. Despite its effectiveness on HEPs, the BN models may be subjected to produce uniform results against dissimilar activities. Hence, the direct inference logic model is hard to consider the significant differences among subtasks under the similar situations. This is because contributing factors does not fully address the characteristics of the different level of tasks.

According to the past research presented above, it can be concluded that the first generation HRA methods have relied on context assessment to estimate HEP and/or to determine performance shaping factors that may cause human errors or misbehaviours against certain features of the maritime tasks. However, those tools are less considerate for organisational factors and their interaction among PSFs.

To remedy the weakness of the first generation methods, cognitive reliability and error analysis method (CREAM) has been introduced as the second HRA generation where the individual events and their success or failures are further detailed and examined. The CREAM provides a framework of the subjective HEP estimation from expert judgement by evaluating PSFs in basic method and also provide a nominal probability for each subtask provided that subtask is converted to one of the cognitive activities. This means CREAM makes it possible to estimate overall HEP by evaluating context with PSFs. At the same time, CREAM provides nominal probabilities for cognitive activities. This makes it possible to generate more reliable data especially useful when there is unavailability of past data.

Fujita and Hollnagel (2004) introduced systematic procedures for calculating mean failure rates as a function of the CPC, without making any assumptions about individual human actions by establishing a simple mathematical manipulation. Konstandinidou et al. (2006) have developed a fuzzy modelling system for the estimation of the probability of erroneous human action in specific industrial and working contexts based on CREAM methodology. The developed fuzzy logic consists of 9 input variables similar to CPCs and if-then knowledge-based fuzzy inference system to predict a crisp value that is a failure probability of human operation. He et al. (2008) provided a simplified CREAM prospective quantification process to provide an easily practicable process to get the numeric results, and it can apply to both the basic method and extended method.

Since the introduction of the initial concept of the CREAM, numerous follow-up studies have been conducted at different disciplines to achieve highly advanced CREAM methods through which HEPs could be combined in different ways such as giving customised changes to reflect characteristics of the specific industry and its application to critical operations.

Yang et al. (2013) proposed a modified CREAM to facilitate human reliability quantification in marine engineering by incorporating fuzzy evidential reasoning and Bayesian network based on inference logic. They extend the traditional CREAM method to a fuzzy environment to quantify human failure probabilities by incorporating Bayesian reasoning to model the dependency among CPCs. The multiple-input multiple-output rule concept, together with evidential reasoning, estimates human failure probabilities reasonable in the way of being sensitive to the minor changes of fuzzy input. It also makes it possible to realise the instant calculation of human failure probabilities in specific task analysis on-board ships. The developed method was demonstrated through an illustrative example where an oil tanker’s Cargo Oil Pumps (COPs) shutdown scenario was analysed.

Ung and Shen (2011) proposed a systematic procedure to compute probabilities of operator action failure in CREAM, then in a further study Ung (2015) developed a weighted fuzzy CREAM method. The features of aforementioned model include; the consideration of the weight of each CPC, refinement of the logicality between the CPCs and Contextual Control Modes (COCOM) and the deliberations of useful information from each input for the oil tanker’s COPs shutdown scenario same with the scenario of Yang et al. (2013). Furthermore, Zhou et al. (2017a) adopted the eight customised CPCs to better capture the essential aspects of the work situations and conditions for on-board tankers with the weighting of the CPCs by employing Fuzzy Analytical Hierarchy Process (FAHP). Lee et al. (2011) suggested a customised CPC called Cognitive Speaking Process (CSP) which focus on communication error in a nuclear plant.

Some studies illustrated a risk assessment combining the CREAM method. For example, Zhou et al. (2017b) utilised the CREAM method with a modified fault tree model for LNG spill accident during LNG carriers’ handling operations for risk assessment Ung (2019) demonstrated risk assessments of human error contribution to oil tanker collision by using the Fault Tree Analysis (FTA) structure under which a modified Fuzzy Bayesian network which is also based on Cognitive Reliability Error Analysis Method (CREAM) .

Even though newly developed CREAM methods can be considered as more reliable and sensitive quantification models, most of the advanced and modified CREAM methods focused on CREAM basic method to predict overall HEPs by evaluating contexts. Hence they would fail to utilise the extended CREAM method, which can predict individual cognitive failure probability for each task in operating procedures.

Meanwhile, a simplified CREAM method introduced by He et al. (2008) provided a different view to the CREAM basic and extended method. Akyuz (2015) and Akyuz and Celik (2015) analysed the critical maritime operating procedures by adopting both simplified CREAM basic and extended methods. Xi et al. (2017) introduced a modified CREAM methodology utilising an Evidential Reasoning (ER) approach and a Decision Making Trial and Evaluation Laboratory (DEMATEL) technique to make human error probability quantification in CREAM rational which applies to the CREAM basic and extended method. A simplified CREAM method is an easily accessible process to obtain the numeric results, but numerous assumptions were inevitably made to estimate the uncertainties posed in the over-simplification idea. For example, it is possibly misrepresented as two different scenarios, which may have an identical level of negative and positive impacts, will have the same failure probabilities.

Finally, the previous research studies on CREAM which focus on maritime sector are summarised in **Error! Reference source not found.**. The commonly used advanced CREAM methods are evaluated with 5 criteria to describe the characteristic of the proposed method in **Error! Reference source not found.**.

Table 1 Existing studies utilising CREAM method in maritime domain

CREAM studies for the maritime application	Case study	Methods
Yang et al. (2013)	Oil tanker’s Cargo Oil Pumps shutdown scenario	CREAM incorporated with fuzzy evidential reasoning and Bayesian network based on inference logic
Ung (2015)	Oil tanker’s Cargo Oil Pumps shutdown scenario	Weighted fuzzy CREAM
Akyuz and Celik (2015)	Cargo loading process of LPG tanker	Quantified CREAM utilising a context influence index
Akyuz (2015)	Gas inerting process in LPG tanker	Quantified CREAM utilising a context influence index
Wu et al. (2017)	Ship capsizing accident	CREAM incorporated with fuzzy evidential reasoning
Xi et al. (2017)	The collision avoidance of a particular scenario in Shanghai coastal waters	Modified CREAM based on an Evidential Reasoning (ER) approach and a Decision Making Trial and Evaluation Laboratory (DEMATEL) technique
Zhou et al. (2017a)	The general seafarers' human reliability when performing tasks under the operation circumstance in tanker shipping	Quantified CREAM incorporated with fuzzy analytical hierarchy process (FAHP) for the weighting of the CPCs
Zhou et al. (2017b)	LNG carrier spill accidents	Incorporating CREAM and MCS into fault tree analysis
Zhou et al. (2018)	The general seafarers' human reliability when performing tasks under the operation	A fuzzy and Bayesian network CREAM model

	circumstance in tanker shipping	
Yang et al. (2019)	Drilling rig crew's actions in monitoring the Macondo well and managing the well control event on 20 April 2010	CREAM based on an Evidential reasoning for eliciting Bayesian subjective probabilities
Shirali et al. (2019)	Petrochemical plant control room	CREAM-Bayesian network
Ung (2019)	Oil tanker collision	Fault tree analysis and modified fuzzy Bayesian Network based CREAM

Table 2 Proposed approach versus existing CREAM based approaches

	Model Input	Model output	Characteristic	Applicability	Weighting for CPCs
Fuzzy CREAM	Multi inputs for CPCs level	Single output (overall HEP based on single control mode)	If-then knowledge-based fuzzy inference system , instant calculation model	Basic CREAM	Not applicable
Simplified CREAM	Multi inputs for CPCs level	Single output (overall HEP in basic and HEP per cognitive activities in extended method based on mathematical equation)	Simple mathematical manipulation, easy for user but many assumptions applied	Both basic and extended CREAM	Not applicable
Evidential Reasoning CREAM	Multi inputs for CPCs level	Multiple outputs (overall HEP based on distribution of control modes in basic and HEP per cognitive activities in extended method)	Evidential reasoning to establish fuzzy IF–THEN rule based Bayesian network, instant calculation model	Basic CREAM	Not applicable
Proposed CREAM	Multi inputs for CPCs level and relative importance	Multiple outputs (overall HEP based on distribution of control modes in basic and HEP per cognitive activities in extended method)	Fuzzy logic, Bayesian network and Evidential reasoning applied for uncertainty treatment to distribute fuzzy conclusion for context assessment, not for HEP estimation. Then fuzzy conclusion is utilised to basic method for overall HEP and extended method for weighting respectively.	Both basic and extended CREAM	Applicable

In this respect, this research aims to develop a framework combining the CREAM applicable to entire system process in practice during maritime on-board procedures in various scenarios. In order to achieve the research objective, independent CPCs assessment process is designed from quantification models. Then results of context, fuzzy CPCs score, can be fed into quantification models for CREAM basic and CREAM extended methods, respectively.

Furthermore, this proposed method employed fuzzy theory with multiple experts with the fuzzy opinion aggregation method, Bayesian network, evidential reasoning to realise the detailed analysis close to realistic HRA outcomes. With those combined methods, the procedures of engine room fire-fighting on a general cargo ship in a specific context defined by a scenario could be evaluated to present cognitive failure probabilities per duty under the multiple contextual control modes.

3. Methodology

This section proposes a hybrid approach combining fuzzy theory, Bayesian network and evidential reasoning to CREAM in order to predict human error probability in maritime on-board procedures. Also, a fuzzy multiple attributive group decision making methodology by Ölçer and Odabaşi (2005) is employed and customised for the opinion aggregation to minimise the subjectivity of experts' judgment. According to Marseguerra et al. (2007), human performance in accidents has shown that the influence of the contextual conditions to the task is actually greater than the characteristics of the task itself. The context of a critical maritime scenario which may include factors such as time management, the external environment, proper procedures and training level of crews, is more important and safety-critical in an emergency when compared to typical operating situations. Therefore, the effect of the context should be taken into account when predicting human error. In this respect, the CREAM method is selected as an appropriate framework for the evaluation of maritime emergency procedures on ships. The reasons are that firstly, CREAM can be used to evaluate the context assessment and also apply to an analysis of cognitive activities required for individual tasks, respectively. Secondly, CREAM is a convenient structure to employ other techniques for developing an advanced approach. The flow chart of the proposed approach is shown in Figure *1*.

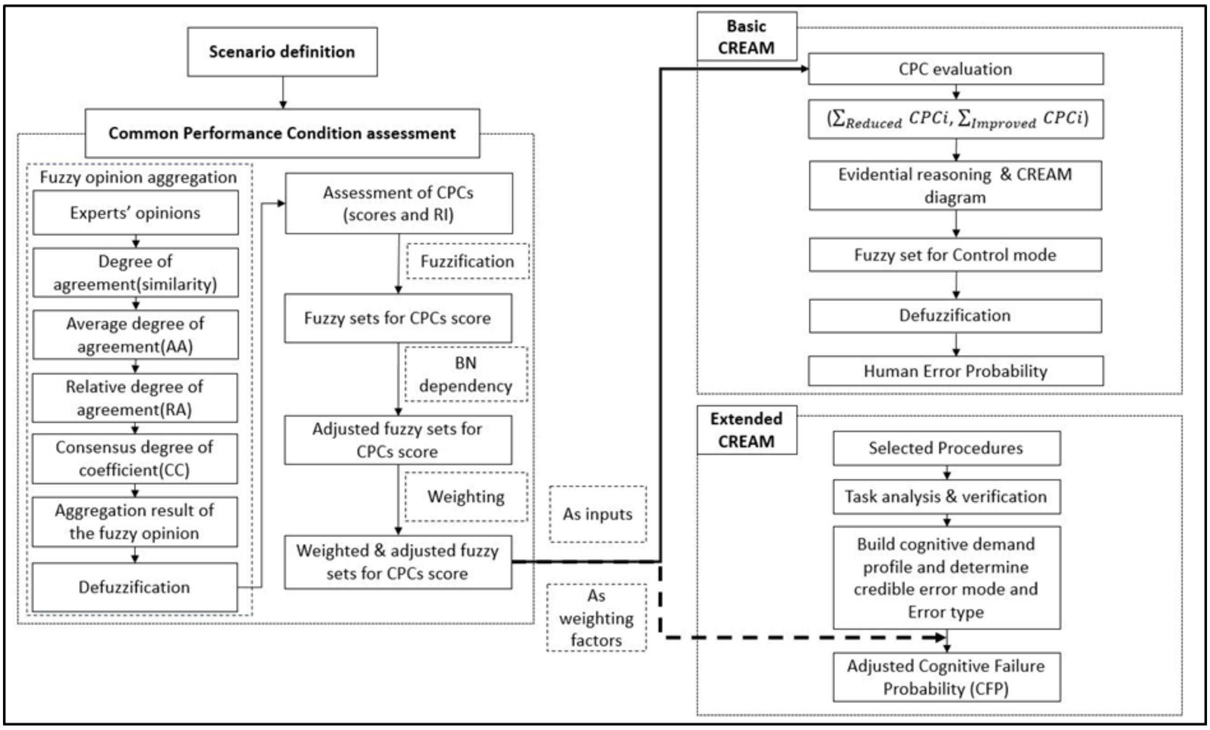


Figure 1 Flow chart of the proposed approach

3.1 Common Performance Condition Assessment

Individual CPCs have linguistic variables which indicate the level of CPC that addresses an expected effect on performance reliability in terms of negative or positive aspect. In the original CREAM, the only linguistic variable is decided with 100% degree of belief for an assessment of the concerned CPC. However, a limited number of linguistic variables is not sufficient to reflect CPC's impact on human reliabilities in a practical situation. In order to better depict the impact of CPC, fuzzy sets are employed because fuzzy sets are the best practice to tackle the ambiguity and vagueness in human error detection problem (Akyuz, 2016). Each CPC associates three or more fuzzy sets to describe the impact of each of the CPCs. In this paper, the trapezoidal fuzzy number is adopted, and the corresponding fuzzy numbers to each CPC level are developed and illustrated in Table *3*. The trapezoidal fuzzy number is selected since it is intuitively easy to be used by decision-maker (Ölçer and Odabaşi, 2005). For example, 'Adequacy of organisation' is assessed with four linguistic variables, namely 'Deficient', 'Inefficient', 'Efficient' and 'Very Efficient'. The horizontal axis represents a numerical score of this CPC varies from 0 to 100 where the most negative value is 0, and positive is 100, and Vertical axis represents a degree of membership from 0 to 1 in Figure *2*. Note that the fuzzy set for each CPC in this study is not an absolute value; it varies depending on the various situations and expert opinions. The method consists of three main steps as follows.

Table 3 CPCs and Performance reliability with fuzzy sets (Hollnagel, 1998)

CPC name	CPC level (L_{ij})	The expected effect on performance	Fuzzy sets
CPC$_1$ Adequacy of organisation	Very efficient ($L_{1,4}$)	Improved	(70, 90, 100, 100)
	Efficient($L_{1,3}$)	Not significant	(30, 70, 70, 90)
	Inefficient($L_{1,2}$)	Reduced	(10, 30, 30, 70)
	Deficient($L_{1,1}$)	Reduced	(0, 0, 10, 30)
CPC$_2$ Working condition	Advantageous ($L_{2,3}$)	Improved	(60, 90, 100, 100)
	Compatible ($L_{2,2}$)	Not significant	(20, 60, 60, 90)
	Incompatible($L_{2,1}$)	Reduced	(0, 0, 20, 60)
CPC$_3$ Adequacy of MMI and operational condition	Supportive ($L_{3,4}$)	Improved	(70, 90, 100, 100)
	Adequate ($L_{3,3}$)	Not significant	(30, 70, 70, 90)
	Tolerable ($L_{3,2}$)	Not significant	(10, 30, 30, 70)
	Inappropriate ($L_{3,1}$)	Reduced	(0, 0, 10, 30)
CPC$_4$ Availability of procedures / plan	Appropriate($L_{4,3}$)	Improved	(60, 90, 100, 100)
	Acceptable ($L_{4,2}$)	Not significant	(20, 60, 60, 90)
	Inappropriate ($L_{4,1}$)	Reduced	(0, 0, 20, 60)
CPC$_5$ Number of simultaneous goals	Fewer than capacity ($L_{5,3}$)	Not significant	(60, 90, 100, 100)
	Matching current capacity ($L_{5,2}$)	Not significant	(20, 60, 60, 90)
	More than capacity ($L_{5,1}$)	Reduced	(0, 0, 20, 60)
CPC$_6$ Available time	Appropriate ($L_{6,3}$)	Improved	(60, 90, 100, 100)
	Temporarily inadequate ($L_{6,2}$)	Not significant	(20, 60, 60, 90)
	Continuously inadequate ($L_{6,1}$)	Reduced	(0, 0, 20, 60)
CPC$_7$ Time of day	Day-time 8h to 17h ($L_{7,3}$)	Not significant	(5, 8, 17, 20)
	Night-time 0h to 8h ($L_{7,2}$)	Reduced	(0, 0, 5, 8)
	Night-time 17h to 24h ($L_{7,1}$)	Reduced	(17, 20, 24, 24)
CPC$_8$ Adequacy of training and experience	Adequate, high experience ($L_{8,3}$)	Improved	(60, 90, 100, 100)
	Adequate, limited experience ($L_{8,2}$)	Not significant	(20, 60, 60, 90)
	Inadequate ($L_{8,1}$)	Reduced	(0, 0, 20, 60)
CPC$_9$ Crew collaboration quality	Very efficient ($L_{9,4}$)	Improved	(70, 90, 100, 100)
	Efficient ($L_{9,3}$)	Not significant	(30, 70, 70, 90)
	Inefficient ($L_{9,2}$)	Not significant	(10, 30, 30, 70)
	Deficient ($L_{9,1}$)	Reduced	(0, 0, 10, 30)

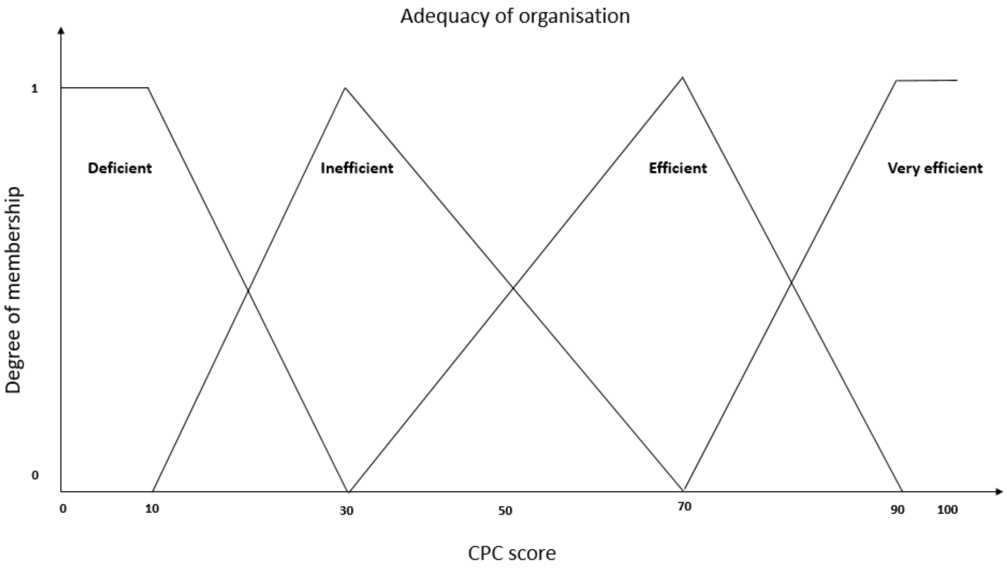


Figure 2 Membership functions for adequacy of organisation

3.1.1 Experts' judgement and fuzzy opinion aggregation

The experts are required to assess both each CPC score and their relative importance with corresponding linguistic terms. Linguistic scale for CPC level and their corresponding fuzzy set developed and provided in Table 3. For relative importance of CPCs, scale and standardised fuzzy sets are listed in Table *4*.

Table 4 Linguistic terms and their standardised fuzzy set

Linguistic terms	Standardised fuzzy sets
Very highly important	(0.8, 0.9, 1, 1)
Highly important	(0.6, 0.75, 0.75, 0.9)
Important	(0.3, 0.5, 0.5, 0.7)
Less important	(0.1, 0.25, 0.25, 0.4)
Not related	(0, 0, 0.1, 0.2)

The purpose of the application of the fuzzy opinion aggregation in Figure 1 is to translate the experts' multiple qualitative assessments of CPC score and relative importance into a single aggregated opinion with fuzzy opinion and convert it into a crisp value through defuzzification. The opinion aggregation procedure is made based on a fuzzy multiple attributive group decision making methodology by Ölçer and Odabaşi (2005) and modified as follows;

(a) Calculate the degree of agreement (Similarity)

Let's assume that A=(a_1, a_2, a_3, a_4), B=(b_1, b_2, b_3, b_4) and A and B are standardised fuzzy set. In here, S(A, B), which is the degree of similarity between A and B, is measured by the below equation;

$$S(A,B) = 1 - \frac{|a_1 - b_1| + |a_2 - b2_1| + |a_3 - b_3| + |a_4 - b_4|}{4} \qquad (1)$$

(b) Calculate the average degree of agreement (AA)

Let’s define $AA(Ex_i)$ as the i-th average degree of agreement and calculated by equation 2 as bellows;

$$AA(Ex_i) = \frac{1}{D-1}\sum_{\substack{i=1\\ i\neq j}}^{D} S(Ex_i, Ex_j) \qquad (2)$$

Where D is a number of experts

(c) Calculate the relative degree of agreement (RA)

Let’s define $RA(Ex_i)$ as the i-th relative degree of agreement and calculated by equation 3 as bellows;

$$RA(Ex_i) = \frac{AA(Ex_i)}{\sum_{i=1}^{D} AA(Ex_i)} \qquad (3)$$

(d) Calculate the consensus degree coefficient (CC)

Let’s define $CC(Ex_i)$ as the consensus degree coefficient for i-th expert and calculated by equation 4 as bellows;

$$CC(Ex_i) = \beta * w_i + (1-\beta) * RA(Ex_i) \qquad (4)$$

Where β is a relaxation factor between 0 and 1. A Homogeneous group of the expert is considered when β is 0 (Ölçer and Odabaşi, 2005). A coefficient w_i means the relative importance among the different experts.

(e) Calculate the aggregation result of the fuzzy opinion (R_{AG})

The aggregated result of the experts’ judgement R_{AG} can be obtained as

$$R_{AG} = \sum_{i=1}^{D} CC(Ex_i) * P(Ex_i) = (S_1, S_2, S_3, S_4) \qquad (5)$$

(f) Defuzzification

Finally, fuzzy opinions (R_{AG}) for each CPC and their relative importance are converted to crisp value by a centre of gravity (COG) method (Takagi and Sugeno, 1985) as

$$x = \frac{\int_{S_1}^{S_4} \mu(x) * x \, dx}{\int_{S_1}^{S_4} \mu(x) \, dx} \qquad (6)$$

Noted that defuzzified CPC scores need to be converted from standardised number to their original score with an interval between 0 and 100 and relative importance of CPC (RI_i) is a normalised number that means $\sum_{i=1}^{9} RI_i = 1$.

3.1.2 Fuzzification

Based on the defuzzified aggregated experts’ opinion for the level of the CPC, the scores for CPC are associated with a fuzzy set to the CPC level.

Let L_{ij}, μ_{ij} and CPC_i define as follows.

L_{ij} represents a j-th linguistic variable for i-th CPC.

μ_{ij} is a value of membership for L_{ij}.

CPC_i is a belief structure corresponding to i-th CPC score and expressed as follows.

$$CPC_i = ((\mu_{i1}, L_{i1}), (\mu_{i2}, L_{i2}), (\mu_{i3}, L_{i3}), (\mu_{ij}, L_{ij})), \text{ where } i = [1, 9] \text{ and } j = [1, 4] \qquad (7)$$

Trapezoidal fuzzy set expressed as (a,b,c,d) and membership function μ_{ij} for random score x is obtained as follows.

$$\mu_{ij} = \begin{cases} \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \quad \text{where } a \le b \le c \le d \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & \text{Otherwise} \end{cases} \tag{8}$$

3.1.3. Adjusted belief structure for CPC

In the previous step, each CPC is expressed by a belief structure. However, the relation of dependency among CPCs should be considered, and CPCs are to be adjusted because CPCs are not independent of the effect of other CPC. The rules for the mutual effects of CPCs are defined as shown in Table *5*. For example, Rule of 4th row indicates that 'Crew collaboration quality' depends on both 'adequacy of organisation' and 'adequacy of training and experience'. If 'crew collaboration of quality' is inefficient (Neutral) AND 'Adequacy of organisation' is very efficient (Positive) AND 'Adequacy of training and experience' is Adequate, high experience (Positive) then "Crew collaboration quality is adjusted to positive from neutral. Interactive relations can be modelled by a Bayesian network technique (Yang et al., 2013) and enable presenting rather complex systems (Hänninen, 2014). Bayesian network model based on Rules acquires four new adjusted CPCs from the nine original CPCs. Adjusted CPCs are also represented by a new belief structure as follows.

$$CPC_i' = ((\mu_{i1}', L_{i1}), (\mu_{i2}', L_{i2}), (\mu_{i3}', L_{i3}), (\mu_{ij}', L_{ij})), \text{ where } i = [1, 9] \text{ and } j = [1, 4] \tag{9}$$

Nine CPCs enter into a model as input variables with belief structures, and 4 CPCs are adjusted based on rules of dependency.

Table 5 Rules for adjusting CPCs (Hollnagel, 1998)

CPC					
Working Conditions(4/5)	Adequacy of organisation	Adequacy of MMI and operational support	Available time	Time of day	Adequacy of training and experience
Number of simultaneous goals(2/3)	Working Conditions	Adequacy of MMI and operational support	Availability of procedure and plans		
Available time(4/5)	Working Conditions	Adequacy of MMI and operational support	Availability of procedure and plans	Number of simultaneous goals	Time of day
Crew collaboration quality (2/2)	Adequacy of organisation	Adequacy of training and experience			

3.1.4 Weighted fuzzy set of CPC_i

Remained important issue regarding the model is, whether all input parameters have equal importance (Konstandinidou et al., 2006) because the distinction of CPCs is not assumed to be independent of one another (Fujita and Hollnagel, 2004). Therefore, the relative importance of CPCs is to be considered in the assessment process and decided carefully by expert judgement. This is the reason that the relative importance of each CPC is assigned by expert judgment in section 3.1.1. So, this section explains how to apply a relative importance value from the expert judgement to the proposed framework. For a calculation purpose, it is needed to define a weighting factor W_i which is calculated by multiplying the number of CPCs (i.e. 9) to RI_i. Then by multiplying weighting factors to adjusted CPC_i', the adjusted & weighted CPC_i'' from the original assessment of CPC score, is expressed as follows.

$$W_i = 9 \times RI_i \quad (10)$$

$$\mu_{ij}'' = W_i \times \mu_{ij}' \quad (11)$$

$$CPC_i'' = ((\mu_{i1}'', L_{i1}), (\mu_{i2}'', L_{i2}), (\mu_{i3}'', L_{i3}), (\mu_{ij}'', L_{ij})), \text{ where } i= [1, 9] \text{ and } j= [1, 4] \quad (12)$$

3.2 Human error quantification with the CREAM basic method

This section describes the process to determine the significant contextual control mode and predict overall human failure probability in the specific scenario by utilising nine fuzzy sets as a result of the context evaluation. The method consists of three main steps. Firstly, nine fuzzy sets are combined with positive and negative CPC score, respectively. This two crisp value indicates the point (sums of the reduced CPCs, sums of the improved CPCs) on two-dimensional CREAM Diagram of Control Mode in Figure *4*. Secondly, the control mode corresponding to the point of combined CPC score is determined with a form of the fuzzy set for four control modes through evidential reasoning. Finally, the human error probability is obtained through a defuzzification process by Weighted Mean of Maxima method from the fuzzy set of control mode.

3.2.1 CPC evaluation

Fuzzy sets of CPCs score can be quantified to a numerical value by defining a specific value as follows.

$$L_{ij} = \begin{cases} 1, & L_{ij} \text{ is 'Improved'.} \\ 0, & L_{ij} \text{ is 'Not significant'.} \\ -1, & L_{ij} \text{ is 'Reduced',} \end{cases} \quad (13)$$

$$CPC_i'' = \sum_{j=1}^{n} \mu_{ij}'' * L_{ij}, \text{ where } n= 3 \text{ or } 4 \quad (14)$$

CPC_i'' value has one of three values depending on the expected number: positive number, negative number, or zero. In order to combine CPC score, positive numbers are added between positive numbers and negative numbers are added between negative numbers separately. For not significant cases, i.e. $L_{ij}=0$, it is possible to assume $\sum_{Not\ significant} CPC_i''$ will not make a serious difference (Hollnagel, 1998) and does not need to be considered. The combined CPC score is finally represented on the Cartesian coordinate system in the form as $(\sum_{Reduced} CPC_i'', \sum_{Improved} CPC_i'')$

3.2.2 Fuzzification of combined CPC score

The Contextual Control Model (COCOM) is output for nine performance condition assessment. Human error probability concerning four control modes is defined with fuzzy triangular sets, as shown in Figure *3* based on Control modes and action probability in Table *6*. The human error probability is represented by the Napierian logarithm function.

Table 6 Control mode and action failure probability (Hollnagel, 1998)

Control mode	Action failure probability
Strategic	0.5E-5< p < 1.0E-2
Tactical	1.0E-3< p < 1.0E-1
Opportunistic	1.0E-2< p < 0.5E-0
Scrambled	1.0E-1< p < 1.0E-0

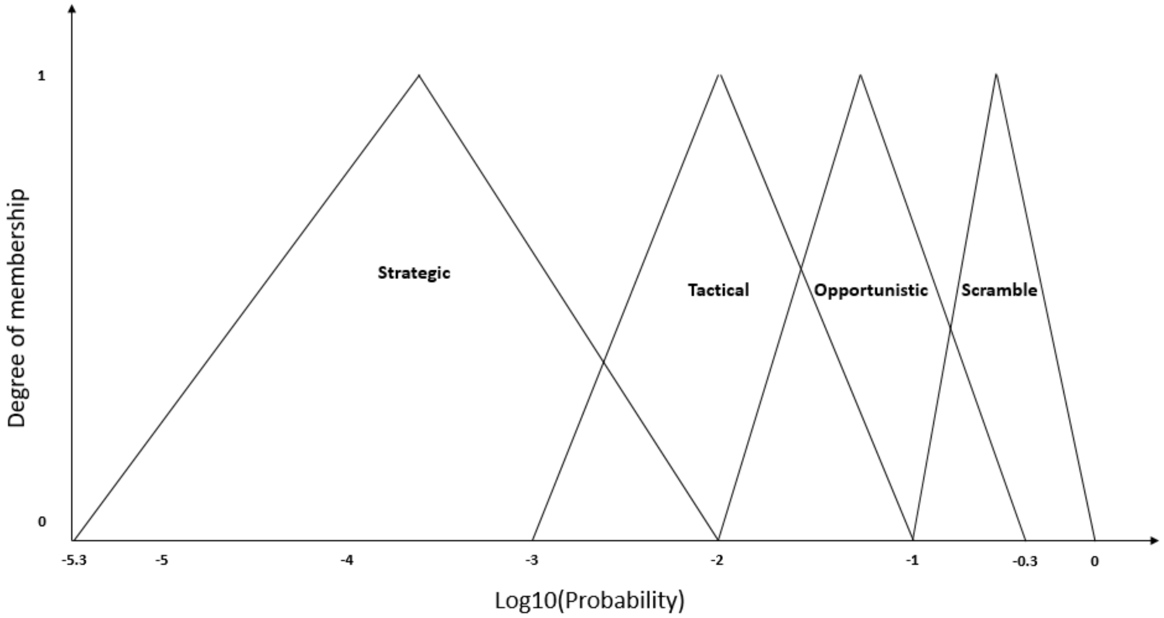


Figure 3 Membership functions for control modes

The combined CPC_i score is regarded as a point on the diagram of the CREAM methodology for operator control mode, as shown in Figure 4. However, the original diagram of CREAM provides four different control modes with their error probability interval in Table *6*. For the specific human error probability estimation corresponding to all different combined CPC_i scores, the approach introduced by Yang et al. (2013) based on the evidential reasoning algorism of Jian-Bo and Dong-Ling (2002) is employed to infer the distribution of degrees of belief to four control modes from a basic diagram of CREAM for operator control modes in this paper. This method enable to avoid a problem of incorporating fuzzy logic into CREAM is that too many IF-THEN rules need to be established in the inference engine(Wu et al., 2017). In the proposed method, control mode of the selected scenario is estimated by the distribution of degrees of belief to the four control modes instead of single control mode in a logical way. The algorithm of human error probability estimation to a point K of the combined CPC score can be analysed and explained by the following pathways.

Let point K to be corresponding to the combined CPC score, $(\sum_{Reduced} CPC_i'' , \sum_{Improved} CPC_i'')$, defined as the coordinates of x and y on the diagram, as shown in Figure *4*. The distribution of degrees of belief corresponding to four control modes consist of Strategic (D_1), Tactical (D_2), Opportunistic (D_3) and Scrambled (D_4) is defined by a set A^K and represented as follows.

$$A^K = ((A^k_1, D_1), (A^k_2, D_2), (A^k_3, D_3), (A^k_4, D_4)), \quad \text{where } \sum_{i=1}^{4} A_i^k = 1 \qquad (15)$$

The set of A^K can be obtained by synthesising two different subsets of the distribution of control mode, A^{K-} and A^{K+}, which are obtained by analysing the portion of squares of different control modes in each row and column about the point K as shown in Figure *4* and expressed as follows.

A^{K-}= ((A^{k-}_1, D_1), (A^{k-}_2, D_2), (A^{k-}_3, D_3), (A^{k-}_4, D_4))

A^{K+}= ((A^{k+}_1, D_1), (A^{K+}_2, D_2), (A^{k+}_3, D_3), (A^{k+}_4, D_4))

Where $\sum_{i=1}^{4} A_i^{K+} = 1,\ \sum_{i=1}^{4} A_i^{k-} = 1$ (16)

The difference between synthesising process introduced by Yang et al. (2013) and the proposed method is not to define the whole if-then rule, but to represent the selected CPC score into a distribution of belief degrees to the four control modes for quantification by defuzzification. The process to derive set A^K from A^+ and A^- is as follow.

Firstly, suppose coefficient values, θ^{K+} and θ^{K-}, represent a normalised number as equation (17) corresponding to X = ($\sum_{Reduced} CPC_i''$ +1) and Y = ($\sum_{Improved} CPC_i''$ +1) from point K. The reason for adding one respectively to the sum of positive and negative CPC is that the centre of the coordinates is moved parallel from (0,0) to (1,1) to prevent the normalised value θ from being zero when both $\sum_{Reduced} CPC_i''$ and $\sum_{Improved} CPC_i''$ are zero on the diagram.

$\theta^{K-} = \frac{X}{X+Y}$, $\theta^{K+} = \frac{Y}{X+Y}$ (17)

Then, assume that M^{K+} and M^{K-} are sets of belief degrees to support the hypothesis that the set A^{K+} and A^{K-} are identified in four control modes. It means a higher score of improved CPC increase value of θ^{K+} and a higher score of reduced CPC increases the value of θ^{K-}, thus sets M^{K+} and M^{K-} support hypothesis of set A^{K+} and A^{K-} respectively as weights.

M^{K-} = (($\theta^{K-}A^{k-}_1$, D_1), ($\theta^{K-}A^{K-}_2$, D_2), ($\theta^{K-}A^{k+}_3$, D_3), ($\theta^{K-}A^{k-}_4$, D_4))

M^{K+} = (($\theta^{K+}A^{k+}_1$, D_1), ($\theta^{K+}A^{K+}_2$, D_2), ($\theta^{K+}A^{k+}_3$, D_3), ($\theta^{K+}A^{k+}_4$ D_4)) (18)

Finally, an output of human error quantification model is represented as a set A^K = ($A^k_1D_1$, $A^k_2D_2$, $A^k_3D_3$, $A^k_4D_4$), it is a distribution of belief degrees to the four control modes for four control modes against a random point K which have $\sum_{Reduced} CPC_i''$ and $\sum_{Improved} CPC_i''$ in the selected scenario and relevant coefficients and equations are follow.

$A_i^{k\prime}$ = P(M_i^{K+} x M_i^{K-} + M_i^{K+} x θ^{K+} + M_i^{K-} x θ^{K-})

H = P (θ^{K+} x θ^{K-})

$P=\left|1-\sum_{T=1}^{4}\sum_{R=1,R\neq T}^{4}(M_T^{K+} * M_R^{K-})\right|^{-1}$

$A_i^k = \frac{A_i^{k\prime}}{1-H}$, (i =1, 2, 3, 4)

A^K = ((A^k_1, D_1), (A^k_2, D_2), (A^k_3, D_3), (A^k_4, D_4)) (19)

Where H is the non-normalised remaining belief unassigned after the commitment of belief to the four control modes as a result of the synthesis of A^+ and A^- and P is the normalising factor.

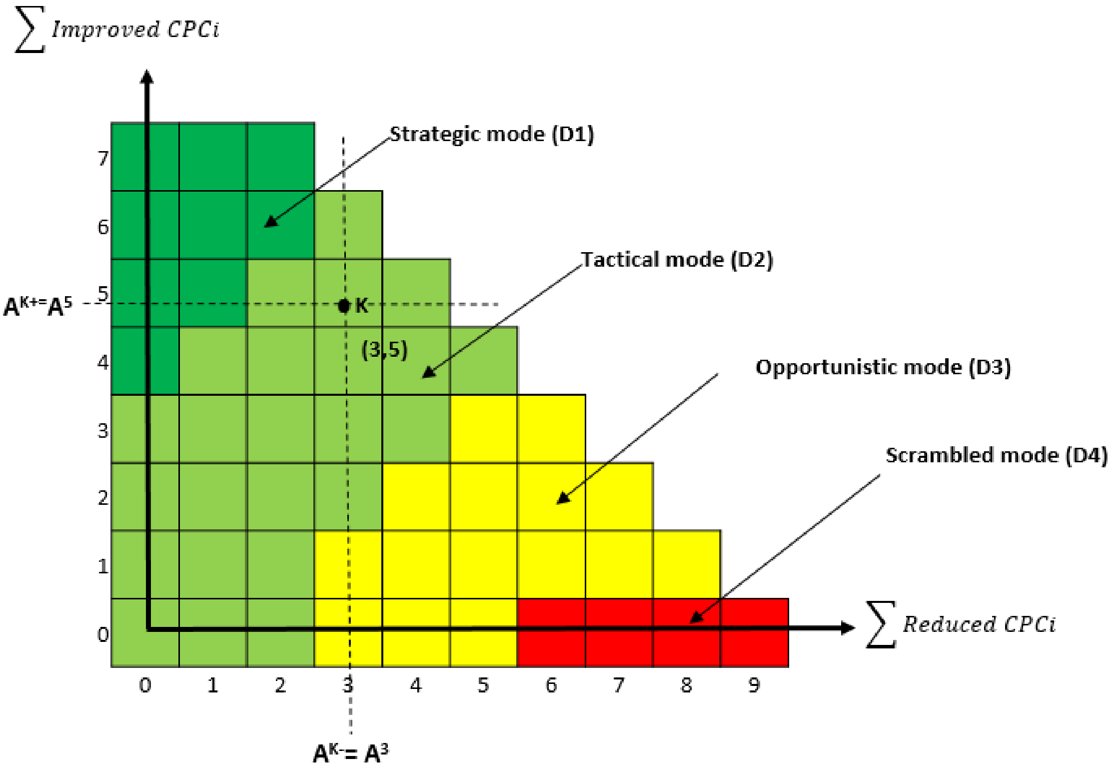


Figure 4 CREAM diagram of control mode

3.2.3 Defuzzification and Human error probability

The defuzzification is a process of converting a fuzzy conclusion to a crisp value. Weighted Mean of Maxima (WMoM) is selected for this defuzzification. A set of belief degrees to the four control modes is defuzzified into a crisp value as follow;

Crisp value (CV) = $\sum_{i=1}^{4} A_i^k * w_i$ (20)

Where w_i is the significant value of the i-th fuzzy membership function.

The weighted value of a fuzzy membership function is abscissa when fuzzy membership function is a maximum value. Membership functions have been developed based on human failure probability interval in CREAM, as shown in Figure *3*. The value w_i can be calculated as -3.651,-2,-1.151 and -0.5. The final step is to convert a crisp value to human error probability since the CV is a logarithm value of human failure probability as below;

HEP (human error probability) = 10^{CV} (21)

In the proposed method, all points on the surface can represent individual human error probability corresponding to the combined CPC scores, contrary to the conventional method addresses four modes for the 52 sets of CPC scores. This method makes the quantitative model much sensitive to the changes in the input value.

3.3 Human error quantification with the CREAM extended method

The purpose of the CREAM extended method is to produce specific action failure probabilities (Hollnagel, 1998), while the basic method does not consider specific human activities in predicting the action failure probability, but only through a context assessment. The CREAM extended method can be applied in the case that further analysis is required through the screening process using the human error probability obtained through the CREAM basic method, or when the analysis of individual event sequences is desired. In terms of risk assessment, this method can also be utilised for procedures review by identifying the delicate tasks that need risk control options or a task to revise from the whole task procedures. The CREAM extended method consists of three main steps, and the basic framework in this paper follows the original CREAM extended method introduced by Hollnagel (1998). The significant characteristic of the proposed method is that weighted and adjusted fuzzy sets for CPC scores are utilised to adjust a nominal cognitive failure probability. Therefore, this section summarises task analysis and verification in the step. 1, building the cognitive demand profile and determine the credit failure mode in step.2, then describes in detail how to use fuzzy sets to adjust the cognitive failure probabilities.

3.3.1 Task analysis and verification

Task analysis refers to methods of formally describing and analysing human-system interaction (Kirwan, 2017). Task analysis is conducted to define the steps which address the designated duties that the crew should complete successfully to achieve the main goal of the procedures with a hierarchical task analysis from the selected scenario. Then, the equipment or procedures of a vessel shall be evaluated to ensure that it satisfies the compulsory requirements of the domestic law or international convention according to the navigational area due to its operational characteristics. This process requires identifying the relevant requirements of the international Convention and domestic to verify the suitability of the procedures.

3.3.2 Build cognitive demand profile and determine credible error mode

The step starts by describing the scenario according to the event sequence and identify cognitive activities that characterise the activity of each work stage or event segment. The fifteen cognitive activity types are provided, and each cognitive activity is associated with one or more basic cognitive functions that consist of observation, interpretation, planning and execution by a generic cognitive-activity-by-cognitive-demand matrix as shown in Table *7*. Once cognitive demand is decided for task element, the next step is to identify the most likely generic failure type for the cognitive activity of the task element. The four basic cognitive functions are classified into 13 generic failure types, and the corresponding cognitive failure probability (CFP) for each generic failure type is given, as shown in Table *8*.

Table 7 Generic cognitive activity by cognitive demand matrix (Hollnagel, 1998)

Cognitive Activity type	Observation	Interpretation	Planning	Execution
Co-ordinate			V	V
Communicate				V
Compare		V		
Diagnose		V	V	
Evaluate		V	V	
Execute				V
Identify		V		
Maintain			V	V

Monitor	V	V		
Observe	V			
Plan			V	
Record		V		V
Regulate	V			V
Scan	V			
Verify	V	V		

Table 8 Nominal values and uncertainty bounds for cognitive function failures (Hollnagel, 1998)

Cognitive function	Generic failure type	Lower bound (5%)	Basic value	Upper bound (95%)
Observation	O1. Wrong object observed	3.00E-04	1.00E-03	3.00E-03
	O2. Wrong identification	2.00E-02	7.00E-02	1.70E-02
	O3. Observation not made	2.00E-02	7.00E-02	1.70E-02
Interpretation	I1. Faulty diagnosis	9.00E-02	2.00E-01	6.00E-01
	I2. Decision error	1.00E-03	1.00E-02	1.00E-01
	I3. Delayed interpretation	1.00E-03	1.00E-02	1.00E-01
Planning	P1. Priority error	1.00E-03	1.00E-02	1.00E-01
	P2. Inadequate plan	1.00E-03	1.00E-02	1.00E-01
Execution	E1. Action of wrong type	1.00E-03	3.00E-03	9.00E-03
	E2. Action at wrong time	1.00E-03	3.00E-03	9.00E-03
	E3. Action on wrong object	5.00E-05	5.00E-04	5.00E-03
	E4. Action out of sequence	1.00E-03	3.00E-03	9.00E-03
	E5. Missed action	2.50E-02	3.00E-03	4.00E-02

3.3.3 Adjusted CFP by weighting factors

The last step in the CREAM extended method is to adjust the nominal CFP with respect to the effect of the CPC. Nine fuzzy sets for all CPC scores are utilised in this step. For example, fuzzy set ((μ_{11}'', L_{11}), (μ_{12}'', $L_{12,}$), (μ_{13}'', $L_{13,}$), (μ_{14}'', L_{14})) represent a fuzzy score of CPC_1. Let define W_{ijn} as a weighting factor for the n-th generic failure type of the j-th CPC level at the i-th CPC and get data from the original CREAM by Hollnagel (1998). Then, let define W_{in} as a weighting factor for n-th cognitive function of CPC_i. The weighting factor, W_n, is acquired as follows;

$$W_{in} = \sum_{j=1}^{4} \mu_{ij} * W_{ijn} \quad (22)$$

$$W_n = \prod_{i=1}^{9} W_{in} \quad (23)$$

Where i= 1 to 9, j=1 to 3 or 4 and n= observation, Interpretation, planning and Execution

4. Case study on the engine room fire-fighting

According to Darbra and Casal (2004), accidents associated with fire and explosion at seaport account for 29% and 17% respectively. The statistical analysis for Maritime Accident Investigation Branch (MAIB) data by de Maya et al. (2019a) found fire and explosion accidents account for 6.78% of all marine accidents occurred from 1990 to 2016. Moreover, those incidents have a reputation of high mortality. Weng and Yang (2015) shows that fire and explosion related incidents result in 132% higher death tolls than other types of accidents. In particular, for passenger ships, fire/explosion accidents are the most frequent occurrence of total losses of ships compared to other accident types (Eliopoulou et al., 2016). According to Baalisampang et al. (2018), 48% of fire incidents in ships are related to human error, followed by mechanical failure 22% and temperature response 14%. In this context, this paper was motivated to apply the proposed method for potential fire incidents in engine room where majority of fire incidents take place.

For an illustration of the proposed approach, both of scenario and procedures for the engine room fire-fighting in general cargo ship have been selected since fire drill at sea is a critical situation in which the crews are required to complete tasks for fire-fighting with limited resources such as personnel, equipment and time. The scenario of an engine room fire-fighting is described in section 4.1 for the purpose to assess CPCs and predict overall HEP without considering specific human activity in selected control mode by the CREAM basic method. The procedures of the engine room fire drill are selected and described in section 4.3 to conduct task analysis and predict individual CFP to all tasks by the CREAM extended method.

The application of the proposed method to case study and data collection were conducted in the following ways;

Firstly, in order to develop an actual emergency response procedure, the existing fire-fighting procedures used in cargo ships were obtained from numerous companies. Developed final procedure was verified and enhanced by a group of experts to ensure compliance with SOLAS and STCW requirements. Next, the scenario was generated to reflect the nine CPC characteristics through meetings of the expert group. Also, a criterion was applied when selecting experts for evaluation stage. In other words, experts who have practical experience of fire-fighting drill on ship as a crew member or safety system auditor are selected for this evaluation. Then, the assessment was conducted independently by each expert to eliminate the group thinking bias. The procedures and scenarios of the fire-fighting were provided for evaluation by a questionnaire using linguistic terms on the relative importance of each CPC and CPC level.

4.1 Scenario definition

The scenario for engine room fire drill on a general cargo ship is described for illustration of the proposed method and focus on presenting CPCs for evaluation as follows.

On a hot summer day, a general cargo ship was waiting to departure at the anchoring position after finishing cargo loading. The temperature was 38 °C, and the humidity is 70 %. The sea conditions and winds were generally good. The vessel was five years old general cargo ship, G/T 5,000, and overall the vessel was in good condition. The ship's management company has managed a total of 30 vessels, holding both the company's DOC certificate and SMC certificates for individual ships in effect in accordance with an International Safety Management Code(ISM), and also obtained ISO certificates on the quality management system. Last month, an internal audit of the vessel was conducted by the company, and all three identified nonconformities have been rectified. A total of

20 crew members were on board and were made up of three different Nationalities. Six crew members were replaced the previous day and conducted familiarisation training in the afternoon of the previous day. Ship’s captain made a plan to conduct the fire drill and abandon ship today at 2 p.m. The fire extinguishing equipment consisted of a fixed CO2 gas system in the engine room; two main fire pumps located inside the main engine room, an emergency fire pump located in the steering gear room, portable fire extinguishers, two firemen's outfits, etc. All fire pumps were manually operated on-site and also remotely in the fire control room and bridge. All fire extinguishing equipment of ship has completed the periodical inspection in accordance with the SOLAS Convention. For communication during training, there were three portable communication devices. The company provided the Muster List to the vessel that consists of duties and responsibilities in case of such mishaps, designated and assigned to each person on the ship in case of emergency including fire and abandon ship. The captain had carried out a monthly fire-fighting and abandon ship drill three days ago, and the records were written in the ship's logbook. For six crews newly onboard, this drill is the first drill to be trained in the vessel, while the other 14 crews have all joined last month's training following the captain's training plan.

4.2 Common Performance Condition Assessment

The relative importance among experts is considered as a heterogeneous group depending on their background and assigned as 0.20, 0.18, 0.21, 0.20 and 0.21. For assessment, experts are asked to assign CPC scores and their relative importance as

Table 9 and Table 10. Then, opinion aggregation from CPC_1 to CPC_9 except the CPC_7 and relative importance for nine CPCs are done. A relaxation factor β is assumed to be 0.5. As an example, specific aggregation for CPC4 are illustrated in **Error! Reference source not found.**. Finally, aggregated fuzzy opinions are defuzzified and listed in **Error! Reference source not found.**. Once experts’ judgement and fuzzy opinion aggregation are completed, the next step is to convert the defuzzified CPC scores to fuzzy membership again for a human error quantification. Then adjust fuzzy sets by dependency relation a shown in **Error! Reference source not found.** which is illustrated by a Genie software. Finally, the weighted & adjusted fuzzy sets are obtained by multiplying

		E1	E2	E3	E4	E5
CPC1	Expert opinion	Efficient(3)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)
CPC2	Expert opinion	Incompatible(1)	Incompatible(1)	Compatible(2)	Compatible(2)	Incompatible(1)
	Fuzzy set	(0, 0, 0.2, 0.6)	(0, 0, 0.2, 0.6)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0, 0, 0.2, 0.6)
CPC3	Expert opinion	Adequate (3)	Adequate (3)	Tolerable(2)	Adequate (3)	Tolerable(2)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)
CPC4	Expert opinion	Appropriate(3)	Appropriate(3)	Acceptable(2)	Appropriate(3)	Acceptable(2)
	Fuzzy set	(0.6, 0.9, 1, 1)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)
CPC5	Expert opinion	Matching current capacity(2)	Matching current capacity(2)	Matching current capacity (2)	Matching current capacity (2)	Matching current capacity (2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC6	Expert opinion	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC8	Expert opinion	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC9	Expert opinion	Inefficient(2)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)

weighting factor to adjusted fuzzy sets. The fuzzy memberships are provided in **Error! Reference source not found.**.

		E1	E2	E3	E4	E5
CPC1	Expert opinion	Efficient(3)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)
CPC2	Expert opinion	Incompatible(1)	Incompatible(1)	Compatible(2)	Compatible(2)	Incompatible(1)
	Fuzzy set	(0, 0, 0.2, 0.6)	(0, 0, 0.2, 0.6)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0, 0, 0.2, 0.6)
CPC3	Expert opinion	Adequate (3)	Adequate (3)	Tolerable(2)	Adequate (3)	Tolerable(2)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)
CPC4	Expert opinion	Appropriate(3)	Appropriate(3)	Acceptable(2)	Appropriate(3)	Acceptable(2)
	Fuzzy set	(0.6, 0.9, 1, 1)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)
CPC5	Expert opinion	Matching current capacity(2)	Matching current capacity(2)	Matching current capacity (2)	Matching current capacity (2)	Matching current capacity (2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC6	Expert opinion	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC8	Expert opinion	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC9	Expert opinion	Inefficient(2)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)

Table 9 Experts' evaluations of CPCs and their standardised fuzzy set

Table 10 Experts' evaluation for the relative importance of CPCs

		E1	E2	E3	E4	E5
RI1	Expert opinion	Moderate	Highly important	Highly important	Moderate	Highly important
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)
RI2	Expert opinion	Highly important	Highly important	Highly important	Highly important	Highly important
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)
RI3	Expert opinion	Highly important	Moderate	Highly important	Moderate	Moderate
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)
RI4	Expert opinion	Moderate	Moderate	Very highly important	Moderate	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)	(0.8,0.9,1,1)	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)
RI5	Expert opinion	Moderate	Highly important	Moderate	Less important	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1, 0.25, 0.25, 0.4)	(0.3,0.5,0.5,0.7)
RI6	Expert opinion	Moderate	Highly important	Highly important	Highly important	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)
RI7	Expert opinion	Less important	Highly important	Moderate	Less important	Moderate
	Fuzzy set	(0.1,0.25,0.25,0.4)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1,0.25,0.25,0.4)	(0.3,0.5,0.5,0.7)
RI8	Expert opinion	Highly important	Very highly important	Very highly important	Very highly important	Highly important
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.8,0.9,1,1)	(0.8,0.9,1,1)	(0.8,0.9,1,1)	(0.6,0.75,0.75,0.9)
RI9	Expert opinion	Highly important	Highly important	Highly important	Moderate	Less important
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1,0.25,0.25,0.4)

Table 11 Aggregation under the CPC_4

Fuzzy sets for expert opinion					Relative degree of agreement(RA)	
Ex1	0.6	0.9	1	1	RA(Ex1)	0.207317
Ex2	0.6	0.9	1	1	RA(Ex2)	0.207317
Ex3	0.2	0.6	0.6	0.9	RA(Ex3)	0.189024
Ex4	0.6	0.9	1	1	RA(Ex4)	0.207317
Ex5	0.2	0.6	0.6	0.9	RA(Ex5)	0.189024
Degree of agreement(S)					Relative degree of agreement(RA)	
S12	1		S34	0.7	RA(Ex1)	0.207317
S23	0.7		S15	0.7	RA(Ex2)	0.207317
S13	0.7		S25	0.7	RA(Ex3)	0.189024
S14	1		S35	1	RA(Ex4)	0.207317
S24	1		S45	0.7	RA(Ex5)	0.189024
Average degree of agreement(AA)					Consensus degree coefficient(CC)	
AA(Ex1)				0.85	CC(Ex1)	0.202343
AA(Ex2)				0.85	CC(Ex2)	0.195764
AA(Ex3)				0.775	CC(Ex3)	0.199775
AA(Ex4)				0.85	CC(Ex4)	0.202343
AA(Ex5)				0.775	CC(Ex5)	0.199775
Rag(HT)	0.44018	0.780135	0.761874	0.960045		
Defuzzification				0.722883		
Normalised score				72.28828		

Table 12 Fuzzy sets for the CPCs assessment for fire-fighting scenario

				Fuzzy sets for CPC score				Adjusted fuzzy sets for CPC score				Weighted & adjusted fuzzy sets for CPC score			
CPC_i	score	RI_i	W_i	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i1}'	μ_{i2}'	μ_{i3}'	μ_{i4}'	μ_{i1}''	μ_{i2}''	μ_{i3}''	μ_{i4}''
CPC_1	57	0.12	1.05	0.00	0.33	0.68	0.00	0.00	0.33	0.68	0.00	0.00	0.34	0.71	0
CPC_2	35	0.13	1.20	0.63	0.38	0.00	-	0.63	0.38	0.00	-	0.75	0.45	0.00	-
CPC_3	51	0.11	0.95	0.00	0.48	0.53	0.00	0.00	0.48	0.53	0.00	0.00	0.45	0.50	0
CPC_4	72	0.10	0.89	0.00	0.60	0.40	-	0.00	0.60	0.40	-	0.00	0.53	0.36	-
CPC_5	55	0.09	0.80	0.13	0.88	0.00	-	0.13	0.88	0.00	-	0.10	0.70	0.00	-
CPC_6	55	0.12	1.05	0.13	0.88	0.00	-	0.15	0.85	0.00	-	0.16	0.89	0.00	-
CPC_7	14h	0.08	0.70	0.00	0.00	1.00	-	0.00	0.00	1.00	-	0.00	0.00	0.70	-
CPC_8	55	0.15	1.36	0.13	0.88	0.00	-	0.13	0.88	0.00	-	0.17	1.19	0.00	-
CPC_9	52	0.11	1.00	0.00	0.45	0.55	0.00	0.02	0.45	0.53	0.00	0.02	0.45	0.53	0
Total	-	1.00	9.00	9.00				9.00				9.00			

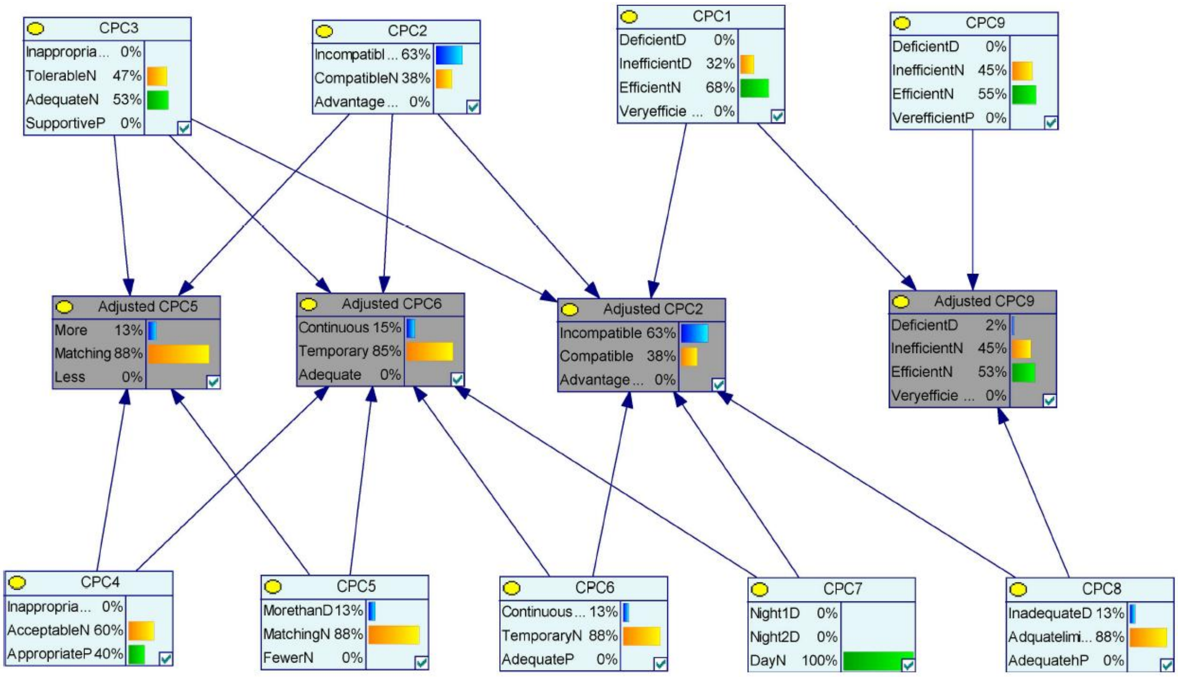


Figure 5 Bayesian presentation for the dependency of the performance condition

4.3 Human error quantification with the CREAM basic method

This section presents the process to calculate the overall human error probability from fuzzy memberships for CPCs by the proposed approach based CREAM basic method.

4.3.1 CPC evaluation

In this step, adjusted & weighted fuzzy sets of CPCs score is quantified to combined CPC score. The combined CPC score is calculated as reduced effect 1.54, improved effect 0.36 by multiplying expected effect in accordance with section 3.2.1.

4.3.2 Fuzzification of combined CPC score

This section describes the process to infer the distribution of belief degrees corresponding to four control modes consist of Strategic (D_1), Tactical (D_2), Opportunistic (D_3) and Scrambled (D_4) from the combined CPC score point K (1.54, 0.36). Subsets $A^{1.54}$ and $A^{0.36}$ are obtained by analysing the portion of squares of different control modes in each row and column to the point K as follows.

$$A^{K-} = A^{1.54} = ((\frac{2}{8}, D_1), (\frac{6}{8}, D_2), (0, D_3), (0, D_4))$$

$$A^{K+} = A^{0.36} = ((0, D_1), (\frac{3}{10}, D_2), (\frac{3}{10}, D_3), (\frac{4}{10}, D_4))$$

Normalised coefficient $\theta^{1.54}$ and $\theta^{0.36}$ are acquired after parallel movement of centre of coordinate from (0,0) to (1,1) by the equation (17) as follows.

$\theta^{1.54} = \frac{2.2.5418}{2.54+1.36}$ =0.65, $\theta^{0.36} = \frac{1.36}{2.54+1.36}$ =0.35

$M^{1.54}$ and $M^{0.36}$ are set of belief degrees to support the hypothesis that the subset A^{K-} and A^{K+} are identified in four control modes by the equation (18) as follows.

$M^{1.54}$ = $((0.65 *\frac{2}{8}, D_1), (0.65*\frac{6}{8}, D_2), (0, D_3), (0, D_4))$

$M^{0.36}$ = $((0, D_1), (0.35*\frac{3}{10}, D_2), (0.35*\frac{3}{10}, D_3), (0.35*\frac{4}{10}, D_4))$

Coefficients P, H and set of A^K are calculated by equation (19) and an output of human error quantification model is derived as follows.

P=1.21, H=0.27

$A^{(1.54,\ 0.36)}$ = ((0.18, D1), (0.68, D2), (0.06, D3), (0.08, D4))

4.3.3 Defuzzification and Human error probability

A set of belief degrees to the four control modes $A^{(1.54,\ 0.36)}$ is defuzzified into a logarithm number negative 2.12; then HEP is derived by equation (21) as follows.

HEP (human error probability) = 10^{CV} = 0.0076

4.4 Human error quantification with the CREAM extended method

In accordance with SOLAS Chapter3, Regulation 19.3.2, all crew members shall participate in at least one abandon ship and fire drill every month (IMO, 2001). Fire-fighting facilities in each ship vary depending on the requirement of fire detection and extinguish system as well as on the type of vessels and cargo. Therefore, fire drills for specific ships should be planned so that proper consideration of regular practice in various emergencies can be made. The procedures also have to consider an abandon-ship decision made by the ship's Master in case of fire-fighting failure.

4.4.1 Task analysis and verification

The hierarchical task analysis for the procedures of engine room fire-fighting is shown in Table *13*. The procedures are confirmed that all compulsory requirements by SOLAS* Chapter 3, Regulation 19.3.5.2 are included (IMO, 2001). The procedure consists of seven main tasks which are i) Fire detection and announcement, ii) Assembly at the muster station, iii) Check openings in the engine room area, iv) Preparation of the fireman, v) Preparation of the fire pump and water spray, vi) Fire-fighting, vii) Further actions and main tasks are divided to twenty-three subtasks as Table *13*.

*International Convention for the Safety of Life at Sea (SOLAS), 1974

Table 13 Procedures of the engine room fire-fighting in general ship

Engine room fire-fighting procedures
1. Fire detection and announcement
1.1 Detect fire in the engine room
1.2 Report to the wheelhouse
1.3 Push the fire alarm and make an announcement
1.4 Report to stations
2. Assembly at the muster station

2.1 Ensure all crew gathered at the muster station
2.2 Check fireman's outfit and other personal rescue equipment
2.3 Describe the fire-fighting procedures and duties to all crew members
2.4 Check communication equipment
3. Check openings in the engine room area
3.1 Stop all-electric ventilation fan
3.2 Close all air inlets and doors into the engine room
3.3 Ensure no air supply into the engine room
4. Preparation of the fireman
4.1 Wear fireman's outfit with equipment
4.2 Ensure all fireman's equipment good in order
5. Preparation of the fire pump and water spray
5.1 Open suction valve for the fire pump
5.2 Close main isolating valve
5.3 Connect at least two fire hoses to fire hydrants
5.4 Start the (emergency) fire pump
5.5 Check the water pressure
6. Fire fighting
6.1 Start water spray to engine room boundary for cooling
6.2 Fireman, access into fire site and fire fighting
7. Further actions
7.1 Ensure fire extinguished completely
7.2 Check the necessary of the fixed fire extinguisher system(e.g.CO2 gas)
7.3 Check the necessary of the abandon ship

4.4.2 Build Cognitive demand profile and determine credible error mode

All tasks from 1.1 to 7.3 matched to one of the cognitive activities associated with cognitive demand and credible failure mode. The most likely error mode to the cognitive activity of each task is decided carefully in Table 14. Nominal Cognitive Failure Probability (CFP_O) are provided from Table *8*.

4.4.3 Adjusted CFP by weighting factors

Weighting factor per cognitive demand is calculated by equation (22) and (23) for fire-fighting procedures and the adjusted CFP throughout the whole procedures is illustrated in Table 14.

Table 14 CREAM extended method analysis result for the engine room fire-fighting procedures

Tasks	Cognitive activity	Cognitive Demands	The most credible Error mode	CFP_O	W_n	Adjusted CFP
1.1	Observe	Observation	O3. Observation not made	7.00E-02	2.64	1.85E-01
1.2	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
1.3	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
1.4	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
2.1	Observe	Observation	O3. Observation not made	7.00E-02	2.64	1.85E-01
2.2	Verify	Observation	O2. Wrong	7.00E-02	2.64	1.85E-01

		Interpretation	identification			
2.3	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
2.4	Verify	Observation Interpretation	O3. Observation not made	7.00E-02	2.64	1.85E-01
3.1	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
3.2	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
3.3	Monitor	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
4.1	Execute	Execution	E1. Action of wrong type	3.00E-03	2.98	8.94E-03
4.2	Verify	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
5.1	Execute	Execution	E3. Action on wrong object	5.00E-04	2.98	1.49E-03
5.2	Execute	Execution	E3. Action on wrong object	5.00E-04	2.98	1.49E-03
5.3	Execute	Execution	E1. Action of wrong type	3.00E-03	2.98	8.94E-03
5.4	Execute	Execution	E4. Action out of sequence	3.00E-03	2.98	8.94E-03
5.5	Verify	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
6.1	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
6.2	Execute	Execution	E4. Action out of sequence	3.00E-03	2.98	8.94E-03
7.1	Observe	Observation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
7.2	Diagnose	Interpretation Plan	I2. Decision error	1.00E-02	3.84	3.84E-02
7.3	Diagnose	Interpretation Plan	I2. Decision error	1.00E-02	3.84	3.84E-02

5. Findings and discussion

The proposed approach presents individual human failure probabilities obtained by a proposed CREAM based method by separating the context assessment process and human error quantification process based on a particular maritime scenario; engine room fire-fighting procedures. From the result of the basic method, it is revealed that significant control mode is Tactical mode with 68 % belief and also have 18% belief of Strategic mode, 6 % belief of Opportunistic mode and 8% belief of Scrambled mode. The overall human failure probability indicates 0.0076, which can occur under the given circumstance described in the fire-fighting scenario. For the result of the extended method, the weighting factor per cognitive function shows the most significant adverse effect on the interpretation in a given scenario with 3.84, followed by 2.98 on an execution, 2.67 on planning and 2.64 on observation. For the comparison, the weighting factor in Tactical mode is 1.90 by a simple table in original CREAM. The range of weighting between 2.64 and 3.849 of the proposed approach is quite reasonable. The main finding is that the vulnerable subtasks with the higher failure probability are identified during the fire-fighting procedure, as shown in **Error! Reference source not found.**. The highest failure probability is task No. 1.1 (Detect fire in the engine room), 2.1 (Ensure all crew gathered at the muster station), 2.2 (Check fireman's outfit and other personal rescue equipment), 2.4 (Check communication equipment), 3.3 (Ensure no air supply into the engine room), 4.2 (Ensure all fireman's equipment good in order), 5.5 (Check the water pressure), and 7.1 (Ensure fire extinguished completely) with 0.185 failure probability. While Task No. 5.1 (Open suction valve for the fire pump) and 5.2 (Close main isolating valve) show the lowest HEP with 0.00149. This result means that simple physical activity has lower failure probability than complex cognitive activities which need the additional ability for an interpretation and decision. The study also found that 'Adequacy of training and experience' is recognized as the most significant CPC factor contributing to human error in fire-fighting scenarios with a weight of 1.36, followed by 'working conditions' with a weight of 1.20 times, 'the adequacy of organization' and ‘available time’ with a weight of 1.05. The weighting for nine CPCs is illustrated in Figure 6. For comparison, the original CREAM method is applied to the same assessment as Table 15. The overall results can be found to be within reasonable limits. The notable thing is that the proposed method can identify the effects of other control modes that are ignored by single control mode, and the quantified human failure probability can be obtained. The method allows the same analysis to be expressed in more detailed output. This research result can improve the fire-fighting procedures and also other critical operating’s procedures on the ship and finally contributes to safety at sea.

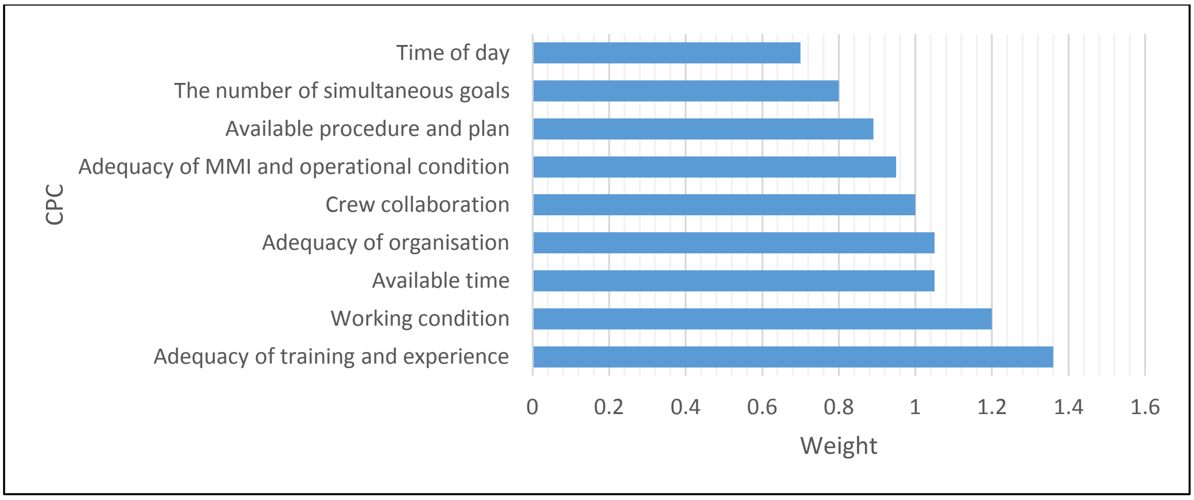


Figure 6 Factors contributing to human failure in fire-fighting

Table 15 Comparison result with the original CREAM method

	Original CREAM	Proposed Method
Combined CPC Score	Expert$_1$ (1,1) Expert$_2$ (1,1) Expert$_3$ (0,0) Expert$_4$ (1,1) Expert$_5$ (1,0)	(1.54,0.36)
Control Mode	Tactical (100%)	Tactical (68%) Strategic (18%) Scrambled (8 %) Opportunistic (6%)
Overall Human Error Probability	between 0.001 and 0.1	0.0076

6. Conclusion

This paper introduces a new framework based CREAM applicable to the maritime industry and illustrates practical fire-fighting scenario and procedures. The characteristics and expected advantages of the proposed method are: Firstly, the proposed method provides an independent process of Common Performance Condition (CPC) assessment from HEP quantification models. This structure is because to provide a simple way to reflect a change of parameters. For an example, when the concerned analysis is needed to change the type of CPCs and their linguistic terms with fuzzy sets to reflect characteristics of the context, the same HEP quantification model can be applied to various situations by separating quantification model from the CPCs assessment. Furthermore, the same quantification model can be applied to individual assessments by different experts, either with different weighting factors for the relative importance of CPC. This simple structure could be realised to get an instant estimation of human failure probability without adjusting the parameters of the HEP quantification model for assessing a specific task. Secondly, the output of CPC assessment can be utilised as an input value in the CREAM basic method and also weighting factors in the CREAM extended method, respectively. This method makes the whole procedures more useful by allowing the results of CPC assessment to be used not only in the basic method but also in the extended method. Finally, the proposed method can evaluate the context in a maritime scenario based on the CREAM basic method and illustrate practical application to onboard procedures in the context in vessels by using the CREAM extended method. The proposed framework also can be extended to apply to the other ship procedures with various scenarios. For a more convenient application, the quantification model does not require a rule-based inference system. Instead, it infers the distribution of belief for control modes from the specific combined score of CPC for human error quantification. In conclusion, the results of this study can make positive impact on the safety of shipping operations and the enhancement of safety at sea by providing a framework applicable to human error analysis.

REFERENCES

ABBASSI, R., KHAN, F., GARANIYA, V., CHAI, S., CHIN, C. & HOSSAIN, K. A. 2015. An integrated method for human error probability assessment during the maintenance of offshore facilities. *Process Safety and Environmental Protection,* 94**,** 172-179.

AKYUZ, E. 2015. Quantification of human error probability towards the gas inerting process on-board crude oil tankers. *Safety Science,* 80**,** 77-86.

AKYUZ, E. 2016. Quantitative human error assessment during abandon ship procedures in maritime transportation. *Ocean Engineering,* 120**,** 21-29.

AKYUZ, E., CELIK, E. & CELIK, M. 2018. A practical application of human reliability assessment for operating procedures of the emergency fire pump at ship. *Ships and Offshore Structures,* 13**,** 208-216.

AKYUZ, E. & CELIK, M. 2015. Application of CREAM human reliability model to cargo loading process of LPG tankers. *Journal of Loss Prevention in the Process Industries,* 34**,** 39-48.

AKYUZ, E. & CELIK, M. 2016. A hybrid human error probability determination approach: The case of cargo loading operation in oil/chemical tanker ship. *Journal of Loss Prevention in the Process Industries,* 43**,** 424-431.

ANTÃO, P. & SOARES, C. G. 2019. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks. *Accident Analysis & Prevention,* 133**,** 105262.

BAALISAMPANG, T., ABBASSI, R., GARANIYA, V., KHAN, F. & DADASHZADEH, M. 2018. Review and analysis of fire and explosion accidents in maritime transportation. *Ocean Engineering,* 158**,** 350-366.

BALMAT, J.-F., LAFONT, F., MAIFRET, R. & PESSEL, N. 2011. A decision-making system to maritime risk assessment. *Ocean Engineering,* 38**,** 171-176.

CELIK, M. & CEBI, S. 2009. Analytical HFACS for investigating human errors in shipping accidents. *Accident Analysis & Prevention,* 41**,** 66-75.

DARBRA, R.-M. & CASAL, J. 2004. Historical analysis of accidents in seaports. *Safety science,* 42**,** 85-98.

DE MAYA, B. N., AHN, S. & KURT, R. Statistical analysis of MAIB database for the period 1990-2016. 18th International Congress of the Maritime Association of the Mediterranean, 2019a.

DE MAYA, B. N., BABALEYE, A. O. & KURT, R. E. 2019b. Marine accident learning with fuzzy cognitive maps (MALFCMs) and Bayesian networks. *Safety in Extreme Environments***,** 1-10.

ELIOPOULOU, E., PAPANIKOLAOU, A. & VOULGARELLIS, M. 2016. Statistical analysis of ship accidents and review of safety level. *Safety science,* 85**,** 282-292.

FUJITA, Y. & HOLLNAGEL, E. 2004. Failures without errors: quantification of context in HRA. *Reliability Engineering & System Safety,* 83**,** 145-151.

HÄNNINEN, M. 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Accident Analysis & Prevention,* 73**,** 305-312.

HE, X., WANG, Y., SHEN, Z. & HUANG, X. 2008. A simplified CREAM prospective quantification process and its application. *Reliability Engineering & System Safety,* 93**,** 298-306.

HOLLNAGEL, E. 1998. *Cognitive reliability and error analysis method (CREAM)*, Elsevier.

IMO 2001. the Regulation 19 of SOLAS Chapter III.

ISLAM, R., ABBASSI, R., GARANIYA, V. & KHAN, F. 2017a. Development of a human reliability assessment technique for the maintenance procedures of marine and offshore operations. *Journal of Loss Prevention in the Process Industries,* 50**,** 416-428.

ISLAM, R., ABBASSI, R., GARANIYA, V. & KHAN, F. I. 2016. Determination of human error probabilities for the maintenance operations of marine engines. *Journal of Ship Production and Design,* 32**,** 226-234.

ISLAM, R., KHAN, F., ABBASSI, R. & GARANIYA, V. 2018. Human error probability assessment during maintenance activities of marine systems. *Safety and health at work,* 9**,** 42-52.

ISLAM, R., YU, H., ABBASSI, R., GARANIYA, V. & KHAN, F. 2017b. Development of a monograph for human error likelihood assessment in marine operations. *Safety science,* 91**,** 33-39.

JIAN-BO, Y. & DONG-LING, X. 2002. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* 32**,** 289-304.

KHAN, F. 2008. Human factors special issue. *Journal of Loss Prevention in the Process Industries,* 21**,** 225-226.

KIRWAN, B. 2017. *A guide to practical human reliability assessment*, CRC press.

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**,** 706-716.

KRISTIANSEN, S. 2013. *Maritime transportation: safety management and risk analysis*, Routledge.

KURT, R., ARSLAN, V., KHALID, H., COMRIE, E., BOULOUGOURIS, E. & TURAN, O. SEAHORSE procedure improvement system: development of instrument. International SEAHORSE Conference on Maritime Safety and Human Factors, 2016a. 1-8.

KURT, R., ARSLAN, V., TURAN, O., DE WOLFF, L., WOOD, B., ARSLAN, O., KECECI, T., WINKELMAN, J., VAN WIJNGAARDEN, M. & PAPADAKIS, G. 2015. SEAHORSE project: Dealing with maritime workarounds and developing smarter procedures.

KURT, R. E., KHALID, H., TURAN, O., HOUBEN, M., BOS, J. & HELVACIOGLU, I. H. 2016b. Towards human-oriented norms: Considering the effects of noise exposure on board ships. *Ocean Engineering,* 120**,** 101-107.

LEE, S. M., HA, J. S. & SEONG, P. H. 2011. CREAM-based communication error analysis method (CEAM) for nuclear power plant operators' communication. *Journal of Loss Prevention in the Process Industries,* 24**,** 90-97.

MARSEGUERRA, M., ZIO, E. & LIBRIZZI, M. 2007. Human Reliability Analysis by Fuzzy "CREAM". *Risk Analysis,* 27**,** 137-154.

ÖLÇER, A. & ODABAŞI, A. 2005. A new fuzzy multiple attributive group decision making methodology and its application to propulsion/manoeuvring system selection problem. *European Journal of Operational Research,* 166**,** 93-114.

SALMON, P. M., CORNELISSEN, M. & TROTTER, M. J. 2012. Systems-based accident analysis methods: a comparison of Accimap, HFACS, and STAMP. *Safety science,* 50**,** 1158-1170.

SCHRÖDER-HINRICHS, J. U., BALDAUF, M. & GHIRXI, K. T. 2011. Accident investigation reporting deficiencies related to organizational factors in machinery space fires and explosions. *Accident Analysis & Prevention,* 43**,** 1187-1196.

SHAPPELL, S. A. & WIEGMANN, D. A. 2000. The human factors analysis and classification system--HFACS.

SHIRALI, G. A., HOSSEINZADEH, T. & KALHORI, S. R. N. 2019. Modifying a method for human reliability assessment based on CREAM-BN: A case study in control room of a petrochemical plant. *MethodsX,* 6**,** 300-315.

TAKAGI, T. & SUGENO, M. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics***,** 116-132.

UNG, S.-T. 2015. A weighted CREAM model for maritime human reliability analysis. *Safety science,* 72**,** 144-152.

UNG, S.-T. 2019. Evaluation of human error contribution to oil tanker collision using fault tree analysis and modified fuzzy Bayesian Network based CREAM. *Ocean Engineering,* 179**,** 159-172.

UNG, S. T. & SHEN, W. M. 2011. A novel human error probability assessment using fuzzy modeling. *Risk Analysis: An International Journal,* 31**,** 745-757.

VAGIAS, N. 2010. A bayesian network application for the prediction of human fatigue in the marine industry. *Unpublished dissertation, National Technical University of Athens*.

WENG, J. & YANG, D. 2015. Investigation of shipping accident injury severity and mortality. *Accident Analysis & Prevention,* 76**,** 92-101.

WU, B., YAN, X., WANG, Y. & SOARES, C. G. 2017. An evidential reasoning-based CREAM to human reliability analysis in maritime accident process. *Risk analysis,* 37**,** 1936-1957.

WU, B., YIP, T. L., YAN, X. & SOARES, C. G. 2019. Fuzzy logic based approach for ship-bridge collision alert system. *Ocean Engineering,* 187**,** 106152.

XI, Y., YANG, Z., FANG, Q., CHEN, W. & WANG, J. 2017. A new hybrid approach to human error probability quantification–applications in maritime operations. *Ocean Engineering,* 138**,** 45-54.

YANG, Z., ABUJAAFAR, K. M., QU, Z., WANG, J., NAZIR, S. & WAN, C. 2019. Use of evidential reasoning for eliciting bayesian subjective probabilities in human reliability analysis: A maritime case. *Ocean Engineering,* 186**,** 106095.

YANG, Z., BONSALL, S., WALL, A., WANG, J. & USMAN, M. 2013. A modified CREAM to human reliability quantification in marine engineering. *Ocean engineering,* 58**,** 293-303.

ZHANG, M., ZHANG, D., GOERLANDT, F., YAN, X. & KUJALA, P. 2019. Use of HFACS and fault tree model for collision risk factors analysis of icebreaker assistance in ice-covered waters. *Safety science,* 111**,** 128-143.

ZHOU, Q., WONG, Y. D., LOH, H. S. & YUEN, K. F. 2018. A fuzzy and Bayesian network CREAM model for human reliability analysis–The case of tanker shipping. *Safety science,* 105**,** 149-157.

ZHOU, Q., WONG, Y. D., XU, H., VAN THAI, V., LOH, H. S. & YUEN, K. F. 2017a. An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. *Safety science,* 95**,** 140-147.

ZHOU, T., WU, C., ZHANG, J. & ZHANG, D. 2017b. Incorporating CREAM and MCS into fault tree analysis of LNG carrier spill accidents. *Safety science,* 96**,** 183-191.

Application of a CREAM based framework to assess human reliability in emergency response to engine room fires on ships

Sung Il Ahn and **Rafet Emek Kurt***

Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde

ABSTRACT

For a human reliability assessment in the maritime domain, the main question is how we correctly understand the human factors in the maritime situation in a practical manner. This paper introduces a new approach based on Cognitive Reliability and Error Analysis Method (CREAM). The key to the method is to provide a framework for evaluating specific scenarios associated with maritime human errors and for conducting an assessment of the context, in which human actions take place. The output of the context assessment is, then, to be applied for the procedure assessment as model inputs for reflection of the context effect. The proposed approach can be divided into two parts: processing context assessment and modelling human error quantification. Fuzzy multiple attributive group decision-making method, Bayesian networks and evidential reasoning are employed for enhancing the reliability of human error quantification. Fuzzy conclusion of the context assessment is utilised by the model input in CREAM basic method and weighting factors in CREAM extended method respectively for considering human failure probability which varies depending on external conditions. This paper is expected to contribute to the improvement of safety by identifying frequently occurred human errors during the maritime operating for minimising of human failures.

Keywords: CREAM, Human Reliability Assessment, Maritime, Ship, Fire Fighting, Safety, Human Factors

1. Introduction

Safety is a critical issue in maritime, but it is still a challenge to predict and prevent accident occurrences because the cause of the accident consists of a variety of factors. Notably, the human factors aspects of ship operation in maritime is one of the significant contributions to the accident. The past studies show that human error is deeply related to accidents, ranging from 65 to 90 per cent. (Kristiansen (2013); Ung (2015); Akyuz et al. (2018); Kurt et al. (2016b); Antão and Soares (2019)). However, the terms of human factors and human error are often used without a clear understanding (Khan, 2008). It is due to the fact that the seafarers face many hazardous situations since they should not only carry out the navigation of ship but also have to conduct other responsibilities such as cargo loading and discharging, ballasting and de-ballasting, bunkering and maintenance work including hot and closed space work mostly independently in space away from land. Specific parts of the ship's functions have been automated, but a human still controls or interacts with most of the work. Therefore, in order to ensure safety at sea human factors, specifically Human Reliability Analysis (HRA) needs to be considered at the core of safety assessments. However, HRA has always been a concern for safety engineers and risk assessment analysts due to the fundamental limitations such as insufficient data, methodological limitations related to subjectivity of analysts and expert judgment, and uncertainty concerning the actual behaviour of people during accident conditions (Konstandinidou et al., 2006). According to Schröder-Hinrichs et al. (2011), it is more difficult to collect reliable data because human and organisational factors related to accident development and response to emergency situations are not reported enough. In this context, prospective methods for quantifying human reliability across the first generation and over the third generation HRA methods have been proposed through the nuclear and aviation sectors and recently applied to the marine sector, but the third generation methods are still in the development stage. As a representative method, cognitive reliability and error analysis method (CREAM) was first developed by Hollnagel (1998) and can be considered as one of the most popular and commonly used second-generation HRA method.

According to studies conducted by Hollnagel (1998) and later by Fujita and Hollnagel (2004), to predict human performance reliability, a context description must be provided because a discussion of what is likely to happen in a given situation must be based on a description of the specific circumstances or conditions. It is reasonable that human error probability can be determined directly from a characterisation of the context. This condition is described in terms of the degree of control presented by four characteristic control modes consist of Strategic, Tactical, Opportunistic and Scrambled mode, which identify different reliability of performance.

The CREAM can be used as both retrospective and prospective purposes and CREAM can apply to qualitative and quantitative analysis. The quantitative CREAM consists of basic and extended methods. Firstly, the CREAM basic method is a human failure probability quantification process that defines nine conditions, such as working conditions, crew collaborations, called Common Performance Conditions (CPCs) affecting human performance. In a basic predictive CREAM, it evaluates CPCs to predict human error probability concerning the contextual control modes with four different failure probability interval corresponding to a value of combined CPC scores by using mapping in the diagram of control mode. This method mainly used as screening purpose in HRA and also can be used to identify conditions that may reduce or improve the human reliability aspects of risk assessment. While subsequent and more detailed analyses of human interactions can be acquired by the CREAM extended method (He et al., 2008), the combined score of the CPCs for context assessment derived from the basic method can be an essential parameter for the extended method. The extended method will be necessary to obtain more accurate results for designated tasks of the procedures.

According to Kurt et al. (2015) and Kurt et al. (2016a), their research conducted in the EU funded SEAHORSE Project concluded 20-30% of standard operating procedures are ineffective hence not being followed strictly during operations. This means we need to bring more attention to review procedures on board with a specific focus on human performance in order to achieve safer operations.

In this regard, this paper provides a framework for estimating human error probabilities through scenario description and procedure analysis based on the CREAM method and illustrates the practical application by proposing a way to transform human activities on board and their contextual conditions into analytical forms for HRA. With this objective, the paper is organised as follows: This section introduces HRA in the maritime and CREAM overview. The second section is a literature review, and the third section presents the proposed method based on CREAM. The case study for the procedures of the engine room fire-fighting on the ship is presented in section four. The fifth section gives the finding and discussion, followed by a conclusion in the sixth section.

2. Literature Review

Over the decades, there have been vigorous efforts to understand the mechanism of human error and to prevent maritime incidents caused by human through utilising various human reliability assessment (HRA) techniques; such as, Success Likelihood Index Method (SLIM), Human Error Assessment and Reduction Technique (HEART), Technique of Human Error Rate Prediction (THERP), Human Factors Analysis and Classification System (HFACS), Cognitive reliability and error analysis method (CREAM).

Hence, researchers put a lot of effort to condense the complex circumstances, under which ship crews are highly likely to make mistakes, into simple descriptive numbers known as Human Error Probability (HEP) by means of several uncertainty treatment methods, such as fuzzy logic, Bayesian networks, evidential reasoning, Event tree, Fault tree, and other integrated methods.

Fuzzy logic has been successfully applied in maritime context to wide range of topics concerning maritime safety and risk. For example Balmat et al. (2011) presented a fuzzy approach in order to evaluate the maritime risk assessment to pollution prevention on the open sea while Wu et al. (2019) utilised fuzzy Multiple Attribute Decision Making for ship-bridge collision alert system. Fuzzy logic has also been utilised in numerous studies related to human reliability analysis to improve the reliability and reduce uncertainty in generated results.

In following paragraphs, the details of previous maritime research studies that are conducted by using aforementioned methods (known as the first generation HRAs) are shared:

Akyuz (2016) applied the concept of the SLIM for estimating HEP when conducting the abandon-ship procedures. The fuzzy sets were used to improve the reliability of the analysis against the vagueness of expert judgments and the arbitrary measure of performance shaping factors (PSFs). Based on the SLIM, Islam et al. (2016) determined the HEPs related to marine engine maintenance tasks, where in another study Islam et al. (2017b) developed a monograph for assessing the likelihood of human error in marine operations that could be applicable for instant decision making. It was identified that with SLIM method, it is possible to estimate not only general HEPs in a given context but also HEPs in specific activities by adding particular PSFs, such as training, experience, fatigue level of a seafarer, etc. However, SLIM is overly relying on expert judgment, which makes the analysis results highly subjective and less reliable; it is because the scope of PSFs is limited to certain contexts rather than fully reflective to every aspect that affects human performance. In particular, they are weak in dealing with social and organisational aspects. To remedy the challenges posed in the SLIM, Abbassi et al. (2015) proposed the integration of SLIM with the THERP to investigate PSFs related to an offshore condensate pump maintenance task. The SLIM was used to estimate the human errors that were not covered by THERP.

On the other hand, Akyuz and Celik (2016) applied the HEART in combination of AHP to predict human errors associated with cargo operation on oil/chemical tankers. Islam et al. (2017a) developed an operational specific methodology based on the HEART in order to capture unique features of maritime environment and operation, and applied the method to the maintenance procedures of a marine engine exhaust turbocharger and also a condensate pump fitted to offshore oil and gas facilities. The HEART has a similar nature as the SLIM but it provides nominal probabilities for generic HEART tasks. Thereafter, the overall HEPs are adjusted by evaluating Error Producing Conditions (EPCs) and the proportion of effect defined by experts' judgment. As a result, like the SLIM, the multiplier values are highly relied on experts' knowledge, which leaves uncertainties in analysis results.

Comment [SA1]: For a comment #1 by reviewer 2
The whole part of literature review was amended from page 4 to 9

Comment [SA2]: For a comment #12 by reviewer3

The HFACS is firstly proposed by Shappell and Wiegmann (2000). As a qualitative method, it adopts a taxonomic nature for better understanding of human behaviour. To obtain quantified outcomes, some researchers proposed the combination of the HFACS with a Fuzzy Analytical Hierarchy Process (FAHP) or Fault Tree Analysis (FTA). Celik and Cebi (2009) generated an analytical HFACS with the concept of the FAHP, in order to identify the role of human errors in boiler explosions onboard bulk carrier. This study provides an analytical foundation and group decision-making functionality in order to achieve a quantitative assessment of shipping accidents. Zhang et al. (2019) introduced a modified model of the HFACS for collision accidents between a ship and an icebreaker. Then, the FTA model was utilized to analyse the fundamental collision risk factors according to the statistical analysis of accident reports and experts' judgment based on the HFACS-SIBCI model. Collision risk factors during icebreaker assistance were identified and classified under the initial HFACS framework. However, the past research showed HFACS would not fully address the specifics of marine incidents. For example, Salmon et al. (2012) explained the main problems to apply HFACS to the outside of aviation is that it was developed specifically for aviation, a number of the error and failure modes are aviation specific.

Comment [SA3]: For a comment #10 by reviewer3

Furthermore, de Maya et al. (2019b) proposed MALFCM approach incorporated with BNs which is based on the concept and principles of fuzzy cognitive maps (FCMs) to represent the interrelations amongst accident contributor factors. As a weakness, although this database-driven research has led to successful results, the applicable range of the database is far limited to some specific cases rather than general ones.

Unlike the HRA studies mentioned above, Vagias (2010) investigated specific factors relating to human fatigue. BNs were utilised to predict fatigue prevalence and its importance, given the information regarding workload, environment, and ergonomic factors, prior to the occurrence of the accident. This study also provides comprehensive information about Human Factors and human error.

There have also been attempts to develop models that could directly estimate overall HEPs using BNs. Islam et al. (2018) introduced a BN model to estimate HEP by using priority probability and CPT (conditional probability table) from expert groups. In aforementioned study the impact of internal and external factors on human performance were defined in a case study for ship maintenance activities. The BN model provides flexible HEPs that could be obtained based on new information inputted to variables. As such, it is capable to predict HEPs across various maritime scenarios effectively. Despite its effectiveness on HEPs, the BN models may be subjected to produce uniform results against dissimilar activities. Hence the direct inference logic model is hard to consider the significant differences among subtasks under the similar situations. This is because contributing factors does not fully address the characteristics of the different level of tasks.

Comment [SA4]: For a comment #5 by reviewer2

According to the past research presented above, it can be concluded that the first generation HRA methods have relied on context assessment to estimate HEP and/or to determine performance shaping factors that may cause human errors or misbehaviours against certain features of the maritime tasks. However, those tools are less considerate for organisational factors and their interaction among PSFs.

To remedy the weakness of the first generation methods, cognitive reliability and error analysis method (CREAM) has been introduced as the second HRA generation where the individual events and their success or failures are further detailed and examined. The CREAM provides a framework of the subjective HEP estimation from expert judgement by evaluating PSFs in basic method and also provide a nominal probability for each subtask provided that subtask is converted to one of the cognitive activities. This means CREAM makes it possible to estimate overall HEP by evaluating context with PSFs. At the same time, CREAM provides nominal probabilities for cognitive activities. This makes it possible to generate more reliable data especially useful when there is unavailability of past data.

Comment [SA5]: For a comment #4 by reviewer2
Regarding the point of using CREAM

~~Since the introduction of CREAM, numerous follow-up studies have been conducted by researchers from different disciplines to provide a much-advanced CREAM method.~~

Fujita and Hollnagel (2004) introduced systematic procedures for calculating mean failure rates as a function of the CPC, without making any assumptions about individual human actions by establishing a simple mathematical manipulation. Konstandinidou et al. (2006) have developed a fuzzy modelling system for the estimation of the probability of erroneous human action in specific industrial and working contexts based on CREAM methodology. The developed fuzzy logic consists of 9 input variables similar to CPCs and if-then knowledge-based fuzzy inference system to predict a crisp value that is a failure probability of human operation. He et al. (2008) provided a simplified CREAM prospective quantification process to provide an easily practicable process to get the numeric results, and it can apply to both the basic method and extended method.

Since the introduction of the initial concept of the CREAM, numerous follow-up studies have been conducted at different disciplines to achieve highly advanced CREAM methods through which HEPs could be combined in different ways such as giving customised changes to reflect characteristics of the specific industry and its application to critical operations.

~~The HRA methods developed, including the CREAM method, have recently been working on customised changes to reflect characteristics of the specific industry and its application to critical operations.~~

Yang et al. (2013) proposed a modified CREAM to facilitate human reliability quantification in marine engineering by incorporating fuzzy evidential reasoning and Bayesian network based on inference logic. They extend the traditional CREAM method to a fuzzy environment to quantify human failure probabilities by incorporating Bayesian reasoning to model the dependency among CPCs. The multiple-input multiple-output rule concept, together with evidential reasoning, estimates human failure probabilities reasonable in the way of being sensitive to the minor changes of fuzzy input. It also makes it possible to realise the instant calculation of human failure probabilities in specific task analysis on-board ships. The developed method was demonstrated through an illustrative example where an oil tanker's Cargo Oil Pumps (COPs) shutdown scenario was analysed.

Ung and Shen (2011) proposed a systematic procedure to compute probabilities of operator action failure in CREAM, then in a further study Ung (2015) developed a weighted fuzzy CREAM method. The features of aforementioned model include; the consideration of the weight of each CPC, refinement of the logicality between the CPCs and Contextual Control Modes (COCOM) and the deliberations of useful information from each input for the oil tanker's COPs shutdown scenario same with the scenario of Yang et al. (2013). Furthermore, Zhou et al. (2017a) adopted the eight customised CPCs to better capture the essential aspects of the work situations and conditions for on-board tankers with the weighting of the CPCs by employing Fuzzy Analytical Hierarchy Process (FAHP). Lee et al. (2011) suggested a customised CPC called Cognitive Speaking Process (CSP) which focus on communication error in a nuclear plant.

~~Meanwhile, studies that more focus on reflecting the specific features of the Maritime Operation in the HRA include the following: Akyuz (2016) introduced other HRA technique application, the~~

~~Success Likelihood Index Method(SLIM), to the abandon-ship procedures in maritime transportation to estimate Human Error Probability (HEP) with the fuzzy sets deal with the vagueness of expert judgments and expression in decision-making during the weighting process of Performance Shaping Factors (PSF). Akyuz and Celik (2016) also introduced the application of Human Error Assessment and Reduction Technique (HEART) combining Analytic Hierarchy Process (AHP) to a case of cargo loading operation in oil/chemical tanker ship for human error probabilities estimation. Islam et al. (2017a) developed an operational specific methodology based on HEART in order to capture unique features of maritime environment and operation, and applied to the maintenance procedures of a marine engine exhaust turbocharger and also a condensate pump on an offshore oil and gas facilities.~~

Some studies illustrated a risk assessment combining the CREAM method. For example, Zhou et al. (2017b) utilised the CREAM method with a modified fault tree model for LNG spill accident during LNG carriers' handling operations for risk assessment Ung (2019) demonstrated risk assessments of human error contribution to oil tanker collision by using the Fault Tree Analysis (FTA) structure under which a modified Fuzzy Bayesian network which is also based on Cognitive Reliability Error Analysis Method (CREAM) . ~~Ung (2019) applied Fault Tree Analysis (FTA) structure under which a modified Fuzzy Bayesian network which based Cognitive Reliability Error Analysis Method (CREAM) to a risk assessment of human error contribution to oil tanker collision.~~

Even though newly developed CREAM methods can be considered as more reliable and sensitive quantification models, most of the advanced and modified CREAM methods focused on CREAM basic method to predict overall HEPs by evaluating contexts. Hence they would fail to utilise the extended CREAM method, which can predict individual cognitive failure probability for each task in operating procedures.

Meanwhile, a simplified CREAM method introduced by He et al. (2008) provided a different view to the CREAM basic and extended method. Akyuz (2015) and Akyuz and Celik (2015) analysed the critical maritime operating procedures by adopting both simplified CREAM basic and extended methods. Xi et al. (2017) introduced a modified CREAM methodology utilising an Evidential Reasoning (ER) approach and a Decision Making Trial and Evaluation Laboratory (DEMATEL) technique to make human error probability quantification in CREAM rational which applies to the CREAM basic and extended method. A simplified CREAM method is an easily accessible process to obtain the numeric results, but numerous assumptions were inevitably made to estimate the uncertainties posed in the over-simplification idea. For example, it is possibly misrepresented as two different scenarios, which may have an identical level of negative and positive impacts, will have the same failure probabilities. ~~Akyuz, Celik and Xi utilised a simplified CREAM method which is developed to provide an easily practicable process to get the numeric results but numerous assumptions were made to estimate these numerical results which may introduce uncertainty. For example, it assumed that if different scenarios have an equal difference of negative and positive impacts then they will have the same failure probabilities.~~

Finally, the previous research studies on CREAM which focus on maritime sector are summarised in **Error! Reference source not found.**~~Table 1~~. The commonly used advanced CREAM methods are evaluated with 5 criteria to describe the characteristic of the proposed method in **Error! Reference source not found.**~~*Table 2*~~.

Table 1 CREAM studies for the maritime application

CREAM studies for the maritime application	Case study	Methods

Comment [SA6]: For a comment #8 by reviewer 3

Yang et al. (2013)	Oil tanker's Cargo Oil Pumps shutdown scenario	CREAM incorporated with fuzzy evidential reasoning and Bayesian network based on inference logic
Ung (2015)	Oil tanker's Cargo Oil Pumps shutdown scenario	Weighted fuzzy CREAM
Akyuz and Celik (2015)	Cargo loading process of LPG tanker	Quantified CREAM utilising a context influence index
Akyuz (2015)	Gas inerting process in LPG tanker	Quantified CREAM utilising a context influence index
Wu et al. (2017)	Ship capsizing accident	CREAM incorporated with fuzzy evidential reasoning
Xi et al. (2017)	The collision avoidance of a particular scenario in Shanghai coastal waters	Modified CREAM based on an Evidential Reasoning (ER) approach and a Decision Making Trial and Evaluation Laboratory (DEMATEL) technique
Zhou et al. (2017a)	The general seafarers' human reliability when performing tasks under the operation circumstance in tanker shipping	Quantified CREAM incorporated with fuzzy analytical hierarchy process (FAHP) for the weighting of the CPCs
Zhou et al. (2017b)	LNG carrier spill accidents	Incorporating CREAM and MCS into fault tree analysis
Zhou et al. (2018)	The general seafarers' human reliability when performing tasks under the operation circumstance in tanker shipping	A fuzzy and Bayesian network CREAM model
Yang et al. (2019)	Drilling rig crew's actions in monitoring the Macondo well and managing the well control event on 20 April 2010	CREAM based on an Evidential reasoning for eliciting Bayesian subjective probabilities
Shirali et al. (2019)	Petrochemical plant control room	CREAM-Bayesian network
Ung (2019)	Oil tanker collision	Fault tree analysis and modified fuzzy Bayesian Network based CREAM

Table 2 Comparison table with other CREAM methods

	Model Input	Model output	Characteristic	Applicability	Weighting for CPCs
Fuzzy CREAM	Multi inputs for CPCs level	Single output (overall HEP based on single control mode)	If-then knowledge-based fuzzy inference system , instant calculation model	Basic CREAM	Not applicable

Comment [SA7]: For a comment #11 by reviewer3
Also comment #2 by reviewer2

Simplified CREAM	Multi inputs for CPCs level	Single output (overall HEP in basic and HEP per cognitive activities in extended method based on mathematical equation)	Simple mathematical manipulation, easy for user but many assumptions applied	Both basic and extended CREAM	Not applicable
Evidential Reasoning CREAM	Multi inputs for CPCs level	Multiple outputs (overall HEP based on distribution of control modes in basic and HEP per cognitive activities in extended method)	Evidential reasoning to establish fuzzy IF–THEN rule based Bayesian network, instant calculation model	Basic CREAM	Not applicable
Proposed CREAM	Multi inputs for CPCs level and relative importance	Multiple outputs (overall HEP based on distribution of control modes in basic and HEP per cognitive activities in extended method)	Fuzzy logic, Bayesian network and Evidential reasoning applied for uncertainty treatment to distribute fuzzy conclusion for context assessment, not for HEP estimation. Then fuzzy conclusion is utilised to basic method for overall HEP and extended method for weighting respectively.	Both basic and extended CREAM	Applicable

In this respect, this research aims to develop a framework combining the CREAM applicable to entire system process in practice during maritime on-board procedures in various scenarios. In order to achieve the research objective, independent CPCs assessment process is designed from quantification models. Then results of context, fuzzy CPCs score, can be fed into quantification models for CREAM basic and CREAM extended methods, respectively.

Furthermore, this proposed method employed fuzzy theory with multiple experts with the fuzzy opinion aggregation method, Bayesian network, evidential reasoning to realise the detailed analysis close to realistic HRA outcomes. With those combined methods, the procedures of engine room fire-fighting on a general cargo ship in a specific context defined by a scenario could be evaluated to present cognitive failure probabilities per duty under the multiple contextual control modes.

~~for a more precise and sensitive result and make these techniques apply to the human reliability analysis for scenario and procedures throughout both CREAM basic and extended method by multiple experts with the fuzzy opinion aggregation method. As an illustration purpose, the procedures of engine room fire-fighting on a general cargo ship in a specific context defined by a scenario is illustrated to present cognitive failure probabilities per duty under the multiple contextual control modes.~~

3. Methodology

This section proposes a hybrid approach combining fuzzy theory, Bayesian network and evidential reasoning to CREAM in order to predict human error probability in maritime on-board procedures. Also, a fuzzy multiple attributive group decision making methodology by Ölçer and Odabaşi (2005) is employed and customised for the opinion aggregation to minimise the subjectivity of experts' judgment. According to Marseguerra et al. (2007), human performance in accidents has shown that the influence of the contextual conditions to the task is actually greater than the characteristics of the task itself. The context of a critical maritime scenario which may include factors such as time management, the external environment, proper procedures and training level of crews, is more important and safety-critical in an emergency when compared to typical operating situations. Therefore, the effect of the context should be taken into account when predicting human error. In this respect, the CREAM method is selected as an appropriate framework for the evaluation of maritime emergency procedures on ships. The reasons are that firstly, CREAM can be used to evaluate the context assessment and also apply to an analysis of cognitive activities required for individual tasks, respectively. Secondly, CREAM is a convenient structure to employ other techniques for developing an advanced approach. The flow chart of the proposed approach is shown in Figure 1~~*Figure 1*~~.

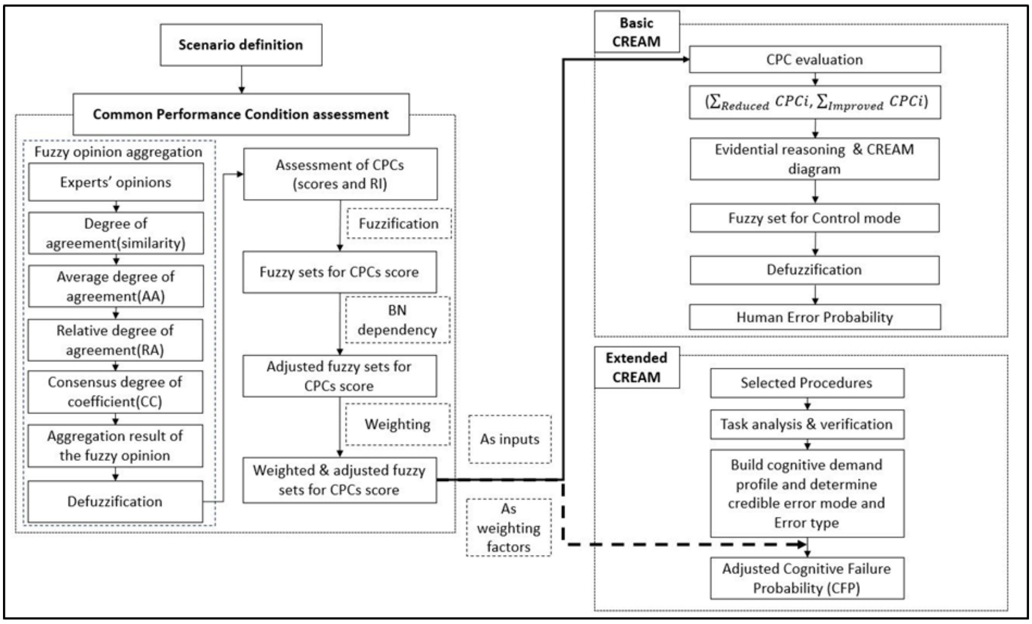


Figure 1 Flow chart of the proposed approach

3.1 Common Performance Condition Assessment

Individual CPCs have linguistic variables which indicate the level of CPC that addresses an expected effect on performance reliability in terms of negative or positive aspect. In the original CREAM, the only linguistic variable is decided with 100% degree of belief for an assessment of the concerned CPC. However, a limited number of linguistic variables is not sufficient to reflect CPC's impact on human reliabilities in a practical situation. In order to better depict the impact of CPC, fuzzy sets are employed because fuzzy sets are the best practice to tackle the ambiguity and vagueness in human error detection problem (Akyuz, 2016). Each CPC associates three or more fuzzy sets to describe the impact of each of the CPCs. In this paper, the trapezoidal fuzzy number is adopted, and the corresponding fuzzy numbers to each CPC level are developed and illustrated in Table 3~~Table 3~~. The trapezoidal fuzzy number is selected since it is intuitively easy to be used by decision-maker (Ölçer and Odabaşi, 2005). For example, 'Adequacy of organisation' is assessed with four linguistic variables, namely 'Deficient', 'Inefficient', 'Efficient' and 'Very Efficient'. The horizontal axis represents a numerical score of this CPC varies from 0 to 100 where the most negative value is 0, and positive is 100, and Vertical axis represents a degree of membership from 0 to 1 in Figure 2~~Figure 2~~. Note that the fuzzy set for each CPC in this study is not an absolute value; it varies depending on the various situations and expert opinions. The method consists of three main steps as follows.

Table 3 CPCs and Performance reliability with fuzzy sets (Hollnagel, 1998)

CPC name	CPC level (L_{ij})	The expected effect on performance	Fuzzy sets
CPC$_1$ Adequacy of organisation	Very efficient ($L_{1,4}$)	Improved	(70, 90, 100, 100)
	Efficient($L_{1,3}$)	Not significant	(30, 70, 70, 90)
	Inefficient($L_{1,2}$)	Reduced	(10, 30, 30, 70)
	Deficient($L_{1,1}$)	Reduced	(0, 0, 10, 30)
CPC$_2$ Working condition	Advantageous ($L_{2,3}$)	Improved	(60, 90, 100, 100)
	Compatible ($L_{2,2}$)	Not significant	(20, 60, 60, 90)
	Incompatible($L_{2,1}$)	Reduced	(0, 0, 20, 60)
CPC$_3$ Adequacy of MMI and operational condition	Supportive ($L_{3,4}$)	Improved	(70, 90, 100, 100)
	Adequate ($L_{3,3}$)	Not significant	(30, 70, 70, 90)
	Tolerable ($L_{3,2}$)	Not significant	(10, 30, 30, 70)
	Inappropriate ($L_{3,1}$)	Reduced	(0, 0, 10, 30)
CPC$_4$ Availability of procedures / plan	Appropriate($L_{4,3}$)	Improved	(60, 90, 100, 100)
	Acceptable ($L_{4,2}$)	Not significant	(20, 60, 60, 90)
	Inappropriate ($L_{4,1}$)	Reduced	(0, 0, 20, 60)
CPC$_5$ Number of simultaneous goals	Fewer than capacity ($L_{5,3}$)	Not significant	(60, 90, 100, 100)
	Matching current capacity ($L_{5,2}$)	Not significant	(20, 60, 60, 90)
	More than capacity ($L_{5,1}$)	Reduced	(0, 0, 20, 60)
CPC$_6$ Available time	Appropriate ($L_{6,3}$)	Improved	(60, 90, 100, 100)
	Temporarily inadequate ($L_{6,2}$)	Not significant	(20, 60, 60, 90)
	Continuously inadequate ($L_{6,1}$)	Reduced	(0, 0, 20, 60)
CPC$_7$ Time of day	Day-time 8h to 17h ($L_{7,3}$)	Not significant	(5, 8, 17, 20)
	Night-time 0h to 8h ($L_{7,2}$)	Reduced	(0, 0, 5, 8)
	Night-time 17h to 24h ($L_{7,1}$)	Reduced	(17, 20, 24, 24)
CPC$_8$ Adequacy of training and experience	Adequate, high experience ($L_{8,3}$)	Improved	(60, 90, 100, 100)
	Adequate, limited experience ($L_{8,2}$)	Not significant	(20, 60, 60, 90)
	Inadequate ($L_{8,1}$)	Reduced	(0, 0, 20, 60)
CPC$_9$ Crew collaboration quality	Very efficient ($L_{9,4}$)	Improved	(70, 90, 100, 100)
	Efficient ($L_{9,3}$)	Not significant	(30, 70, 70, 90)
	Inefficient ($L_{9,2}$)	Not significant	(10, 30, 30, 70)
	Deficient ($L_{9,1}$)	Reduced	(0, 0, 10, 30)

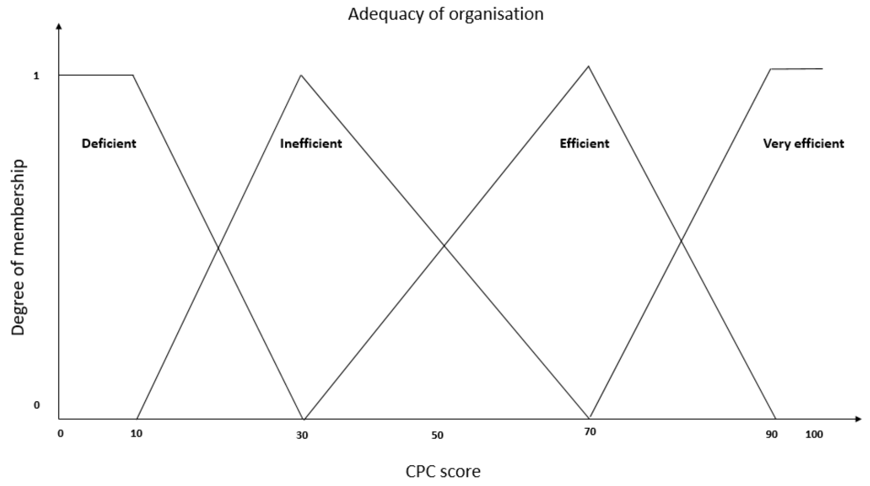


Figure 2 Membership functions for Adequacy of organisation

3.1.1 Experts' judgement and fuzzy opinion aggregation

The experts are required to assess both each CPC score and their relative importance with corresponding linguistic terms. Linguistic scale for CPC level and their corresponding fuzzy set developed and provided in Table 3~~*Table 3*~~. For relative importance of CPCs, scale and standardised fuzzy sets are listed in Table *4*~~Table 4~~.

Table 4 Linguistic terms and their standardised fuzzy set

Linguistic terms	Standardised fuzzy sets
Very highly important	(0.8, 0.9, 1, 1)
Highly important	(0.6, 0.75, 0.75, 0.9)
Important	(0.3, 0.5, 0.5, 0.7)
Less important	(0.1, 0.25, 0.25, 0.4)
Not related	(0, 0, 0.1, 0.2)

The purpose of the application of the fuzzy opinion aggregation in Figure 1~~Figure 1~~ is to translate the experts' multiple qualitative assessments of CPC score and relative importance into a single aggregated opinion with fuzzy opinion and convert it into a crisp value through defuzzification. The opinion aggregation procedure is made based on a fuzzy multiple attributive group decision making methodology by Ölçer and Odabaşi (2005) and modified as follows;

Formatted: Font: Not Italic

(a) Calculate the degree of agreement (Similarity)

Let's assume that A=(a_1, a_2, a_3, a_4), B=(b_1, b_2, b_3, b_4) and A and B are standardised fuzzy set. In here, S(A, B), which is the degree of similarity between A and B, is measured by the below equation;

$$S(A,B) = 1 - \frac{|a_1 - b_1| + |a_2 - b2_1| + |a_3 - b_3| + |a_4 - b_4|}{4} \qquad (1)$$

(b) Calculate the average degree of agreement (AA)

Let's define AA(Ex_i) as the i-th average degree of agreement and calculated by equation 2 as bellows;

$$AA(Ex_i) = \frac{1}{D-1} \sum_{\substack{i=1 \\ i \neq j}}^{D} S(Ex_i, Ex_j) \qquad (2)$$

Where D is a number of experts

(c) Calculate the relative degree of agreement (RA)

Let's define RA(Ex_i) as the i-th relative degree of agreement and calculated by equation 3 as bellows;

$$RA(Ex_i) = \frac{AA(Ex_i)}{\sum_{i=1}^{D} AA(Ex_i)} \qquad (3)$$

(d) Calculate the consensus degree coefficient (CC)

Let's define CC(Ex_i) as the consensus degree coefficient for i-th expert and calculated by equation 4 as bellows;

$$CC(Ex_i) = \beta * w_i + (1 - \beta) * RA(Ex_i) \qquad (4)$$

Where β is a relaxation factor between 0 and 1. A Homogeneous group of the expert is considered when β is 0 (Ölçer and Odabaşi, 2005). A coefficient w_i means the relative importance among the different experts.

(e) Calculate the aggregation result of the fuzzy opinion (R_{AG})

The aggregated result of the experts' judgement R_{AG} can be obtained as

$$R_{AG} = \sum_{i=1}^{D} CC(Ex_i) * P(Ex_i) = (S_1, S_2, S_3, S_4) \qquad (5)$$

(f) Defuzzification

Finally, fuzzy opinions (R_{AG}) for each CPC and their relative importance are converted to crisp value by a centre of gravity (COG) method (Takagi and Sugeno, 1985) as

$$x = \frac{\int_{S_1}^{S_4} \mu(x) * x \, dx}{\int_{S_1}^{S_4} \mu(x) \, dx} \qquad (6)$$

Noted that defuzzified CPC scores need to be converted from standardised number to their original score with an interval between 0 and 100 and relative importance of CPC (RI_i) is a normalised number that means $\sum_{i=1}^{9} RI_i = 1$.

3.1.2 Fuzzification

Based on the defuzzified aggregated experts' opinion for the level of the CPC, the scores for CPC are associated with a fuzzy set to the CPC level.

Let L_{ij}, μ_{ij} and CPC_i define as follows.

L_{ij} represents a j-th linguistic variable for i-th CPC.

μ_{ij} is a value of membership for L_{ij}.

CPC_i is a belief structure corresponding to i-th CPC score and expressed as follows.

$$CPC_i = ((\mu_{i1}, L_{i1}), (\mu_{i2}, L_{i2}), (\mu_{i3}, L_{i3}), (\mu_{ij}, L_{ij})), \text{ where } i = [1, 9] \text{ and } j = [1, 4] \qquad (7)$$

Trapezoidal fuzzy set expressed as (a,b,c,d) and membership function μ_{ij} for random score x is obtained as follows.

$$\mu_{ij} = \begin{cases} \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \qquad \text{where } a \le b \le c \le d \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & \text{Otherwise} \end{cases} \qquad (8)$$

3.1.3. Adjusted belief structure for CPC

In the previous step, each CPC is expressed by a belief structure. However, the relation of dependency among CPCs should be considered, and CPCs are to be adjusted because CPCs are not independent of the effect of other CPC. The rules for the mutual effects of CPCs are defined as shown in Table 5~~*Table 5*~~. For example, Rule of 4th row indicates that 'Crew collaboration quality' depends on both 'adequacy of organisation' and 'adequacy of training and experience'. If 'crew collaboration of quality' is inefficient (Neutral) AND 'Adequacy of organisation' is very efficient (Positive) AND 'Adequacy of training and experience' is Adequate, high experience (Positive) then "Crew collaboration quality is adjusted to positive from neutral. Interactive relations can be modelled by a Bayesian network technique (Yang et al., 2013) and enable presenting rather complex systems (Hänninen, 2014). Bayesian network model based on Rules acquires four new adjusted CPCs from the nine original CPCs. Adjusted CPCs are also represented by a new belief structure as follows.

Comment [SA8]: For a comment #4 by reviewer2
Regarding the point of map it on BN

$$CPC_i' = ((\mu_{i1}', L_{i1}), (\mu_{i2}', L_{i2}), (\mu_{i3}', L_{i3}), (\mu_{ij}', L_{ij})), \text{ where } i=[1, 9] \text{ and } j=[1, 4] \qquad (9)$$

Nine CPCs enter into a model as input variables with belief structures, and 4 CPCs are adjusted based on rules of dependency.

Table 5 Rules for adjusting CPCs (Hollnagel, 1998)

CPC					
Working Conditions(4/5)	Adequacy of organisation	Adequacy of MMI and operational support	Available time	Time of day	Adequacy of training and experience
Number of simultaneous goals(2/3)	Working Conditions	Adequacy of MMI and operational support	Availability of procedure and plans		
Available time(4/5)	Working Conditions	Adequacy of MMI and operational support	Availability of procedure and plans	Number of simultaneous goals	Time of day
Crew collaboration quality (2/2)	Adequacy of organisation	Adequacy of training and experience			

3.1.4 Weighted fuzzy set of CPC_i

Remained important issue regarding the model is, whether all input parameters have equal importance (Konstandinidou et al., 2006) because the distinction of CPCs is not assumed to be independent of one another (Fujita and Hollnagel, 2004). Therefore, the relative importance of CPCs is to be considered in the assessment process and decided carefully by expert judgement. This is the reason that the relative importance of each CPC is assigned by expert judgment in section 3.1.1. So, this section explains how to apply a relative importance value from the expert judgement to the proposed framework. For a calculation purpose, it is needed to define a weighting factor W_i which is calculated by multiplying the number of CPCs (i.e. 9) to RI_i. Then by multiplying weighting factors to adjusted CPC_i', the adjusted & weighted CPC_i'' from the original assessment of CPC score, is expressed as follows.

$$W_i = 9 \times RI_i \qquad (10)$$

$$\mu_{ij}'' = W_i \times \mu_{ij}' \qquad (11)$$

$$CPC_i'' = ((\mu_{i1}'', L_{i1}), (\mu_{i2}'', L_{i2}), (\mu_{i3}'', L_{i3}), (\mu_{ij}'', L_{ij})), \text{ where } i= [1, 9] \text{ and } j= [1, 4] \qquad (12)$$

3.2 Human error quantification with the CREAM basic method

This section describes the process to determine the significant contextual control mode and predict overall human failure probability in the specific scenario by utilising nine fuzzy sets as a result of the context evaluation. The method consists of three main steps. Firstly, nine fuzzy sets are combined with positive and negative CPC score, respectively. This two crisp value indicates the point (sums of the reduced CPCs, sums of the improved CPCs) on two-dimensional CREAM Diagram of Control Mode in Figure 4~~Figure 4~~. Secondly, the control mode corresponding to the point of combined CPC score is determined with a form of the fuzzy set for four control modes through evidential reasoning. Finally, the human error probability is obtained through a defuzzification process by Weighted Mean of Maxima method from the fuzzy set of control mode.

3.2.1 CPC evaluation

Fuzzy sets of CPCs score can be quantified to a numerical value by defining a specific value as follows.

$$L_{ij} = \begin{cases} 1, & L_{ij} \text{ is 'Improved'.} \\ 0, & L_{ij} \text{ is 'Not significant'.} \\ -1, & L_{ij} \text{ is 'Reduced',} \end{cases} \qquad (13)$$

$$CPC_i'' = \sum_{j=1}^{n} \mu_{ij}'' * L_{ij}, \text{ where } n= 3 \text{ or } 4 \qquad (14)$$

CPC_i'' value has one of three values depending on the expected number: positive number, negative number, or zero. In order to combine CPC score, positive numbers are added between positive numbers and negative numbers are added between negative numbers separately. For not significant cases, i.e. L_{ij}=0, it is possible to assume $\sum_{Not\ significant} CPC_i''$ will not make a serious difference (Hollnagel, 1998) and does not need to be considered. The combined CPC score is finally represented on the Cartesian coordinate system in the form as $(\sum_{Reduced} CPC_i'' , \sum_{Improved} CPC_i'')$

3.2.2 Fuzzification of combined CPC score

The Contextual Control Model (COCOM) is output for nine performance condition assessment. Human error probability concerning four control modes is defined with fuzzy triangular sets, as shown in Figure 3~~Figure 3~~ based on Control modes and action probability in Table 6~~Table 6~~. The human error probability is represented by the Napierian logarithm function.

Table 6 Control mode and action failure probability (Hollnagel, 1998)

Control mode	Action failure probability
Strategic	0.5E-5< p < 1.0E-2
Tactical	1.0E-3< p < 1.0E-1
Opportunistic	1.0E-2< p < 0.5E-0
Scrambled	1.0E-1< p < 1.0E-0

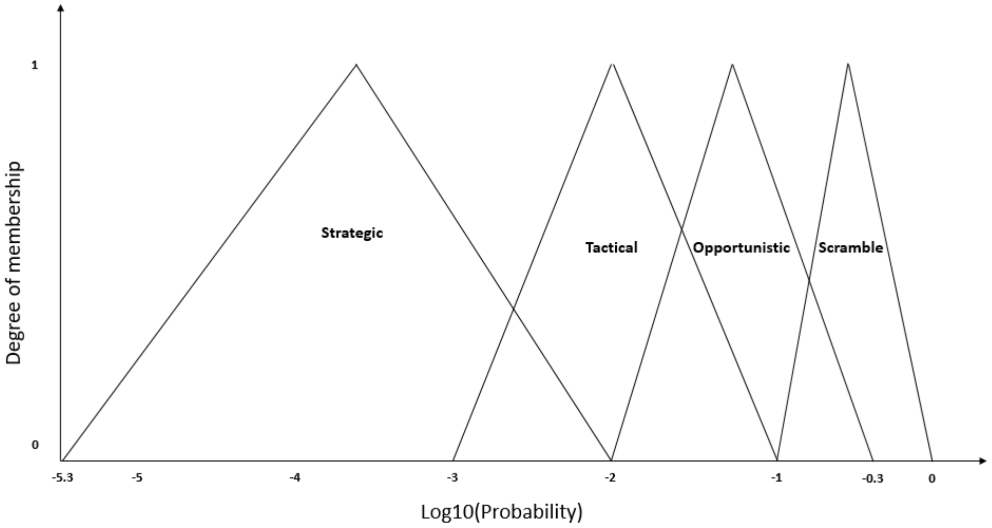


Figure 3 Membership functions for control modes

The combined CPC$_i$ score is regarded as a point on the diagram of the CREAM methodology for operator control mode, as shown in Figure 4~~Figure 4~~. However, the original diagram of CREAM provides four different control modes with their error probability interval in Table 6~~Table 6~~. For the specific human error probability estimation corresponding to all different combined CPC$_i$ scores, the approach introduced by Yang et al. (2013) based on the evidential reasoning algorism of Jian-Bo and Dong-Ling (2002) is employed to infer the distribution of degrees of belief to four control modes from a basic diagram of CREAM for operator control modes in this paper. This method enable to avoid a problem of incorporating fuzzy logic into CREAM is that too many IF-THEN rules need to be established in the inference engine(Wu et al., 2017). In the proposed method, control mode of the selected scenario is estimated by the distribution of degrees of belief to the four control modes instead of single control mode in a logical way. The algorithm of human error probability estimation to a point K of the combined CPC score can be analysed and explained by the following pathways.

Formatted: Font: Not Italic

Let point K to be corresponding to the combined CPC score, $(\sum_{Reduced} CPC_i'' , \sum_{Improved} CPC_i'')$, defined as the coordinates of x and y on the diagram, as shown in Figure 4~~Figure 4~~. The distribution of degrees of belief corresponding to four control modes consist of Strategic (D_1), Tactical (D_2), Opportunistic (D_3) and Scrambled (D_4) is defined by a set A^K and represented as follows.

$A^K = ((A^k_{1,}\, D_1), (A^k_{2,}\, D_2), (A^k_{3,}\, D_3), (A^k_{4,}\, D_4))$, where $\sum_{i=1}^{4} A_i^k = 1$ (15)

The set of A^K can be obtained by synthesising two different subsets of the distribution of control mode, A^{K-} and A^{K+}, which are obtained by analysing the portion of squares of different control modes in each row and column about the point K as shown in Figure 4~~Figure 4~~ and expressed as follows.

$A^{K-} = ((A^{k-}_1, D_1), (A^{k-}_2, D_2), (A^{k-}_3, D_3), (A^{k-}_4, D_4))$

$A^{K+} = ((A^{k+}_1, D_1), (A^{K+}_2, D_2), (A^{k+}_3, D_3), (A^{k+}_4, D_4))$

Where $\sum_{i=1}^{4} A_i^{K+} = 1,\ \sum_{i=1}^{4} A_i^{k-} = 1$ (16)

The difference between synthesising process introduced by Yang et al. (2013) and the proposed method is not to define the whole if-then rule, but to represent the selected CPC score into a distribution of belief degrees to the four control modes for quantification by defuzzification. The process to derive set A^K from A^+ and A^- is as follow.

Firstly, suppose coefficient values, θ^{K+} and θ^{K-}, represent a normalised number as equation (17) corresponding to X = ($\sum_{Reduced} CPC_i''$ +1) and Y =($\sum_{Improved} CPC_i''$ +1) from point K. The reason for adding one respectively to the sum of positive and negative CPC is that the centre of the coordinates is moved parallel from (0,0) to (1,1) to prevent the normalised value θ from being zero when both $\sum_{Reduced} CPC_i''$ and $\sum_{Improved} CPC_i''$ are zero on the diagram.

$\theta^{K-} = \frac{X}{X+Y}$, $\theta^{K+} = \frac{Y}{X+Y}$ (17)

Then, assume that M^{K+} and M^{K-} are sets of belief degrees to support the hypothesis that the set A^{K+} and A^{K-} are identified in four control modes. It means a higher score of improved CPC increase value of θ^{K+} and a higher score of reduced CPC increases the value of θ^{K-}, thus sets M^{K+} and M^{K-} support hypothesis of set A^{K+} and A^{K-} respectively as weights.

$M^{K-} = ((\theta^{K-}A^{k-}_1, D_1), (\theta^{K-}A^{K-}_2, D_2), (\theta^{K-}A^{k+}_3, D_3), (\theta^{K-}A^{k-}_4, D_4))$

$M^{K+} = ((\theta^{K+}A^{k+}_1, D_1), (\theta^{K+}A^{K+}_2, D_2), (\theta^{K+}A^{k+}_3, D_3), (\theta^{K+}A^{k+}_4\, D_4))$ (18)

Finally, an output of human error quantification model is represented as a set $A^K = (A^k_1D_1, A^k_2D_{2,}\, A^k_3D_{3,}\, A^k_4D_4)$, it is a distribution of belief degrees to the four control modes for four control modes against a random point K which have $\sum_{Reduced} CPC_i''$ and $\sum_{Improved} CPC_i''$ in the selected scenario and relevant coefficients and equations are follow.

$A_i^{k\prime} = P(M_i^{K+} \times M_i^{K-} + M_i^{K+} \times \theta^{K+} + M_i^{K-} \times \theta^{K-})$

$H = P\,(\theta^{K+} \times \theta^{K-})$

$P = \left|1 - \sum_{T=1}^{4} \sum_{R=1,R\neq T}^{4} (M_T^{K+} * M_R^{K-})\right|^{-1}$

$A_i^k = \frac{A_i^{k\prime}}{1-H}$, (i=1,2,3, 4)

$A^K = ((A^k_{1,}\, D_1), (A^k_{2,}\, D_2), (A^k_{3,}\, D_3)_, (A^k_{4,}\, D_4))$ (19)

Where H is the non-normalised remaining belief unassigned after the commitment of belief to the four control modes as a result of the synthesis of A^+ and A^- and P is the normalising factor.

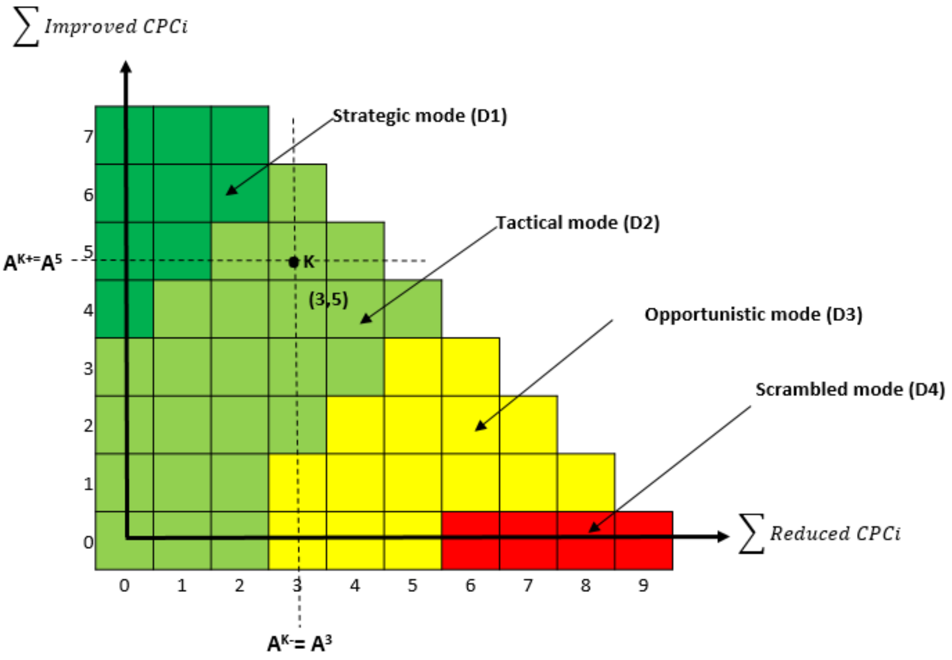


Figure 4 CREAM diagram of control mode

3.2.3 Defuzzification and Human error probability

The defuzzification is a process of converting a fuzzy conclusion to a crisp value. Weighted Mean of Maxima (WMoM) is selected for this defuzzification. A set of belief degrees to the four control modes is defuzzified into a crisp value as follow;

Crisp value (CV) = $\sum_{i=1}^{4} A_i^k * w_i$ (20)

Where w_i is the significant value of the i-th fuzzy membership function.

The weighted value of a fuzzy membership function is abscissa when fuzzy membership function is a maximum value. Membership functions have been developed based on human failure probability interval in CREAM, as shown in Figure 3~~Figure 3~~. The value w_i can be calculated as -3.651,-2,-1.151 and -0.5. The final step is to convert a crisp value to human error probability since the CV is a logarithm value of human failure probability as below;

HEP (human error probability) = 10^{CV} (21)

In the proposed method, all points on the surface can represent individual human error probability corresponding to the combined CPC scores, contrary to the conventional method addresses four modes for the 52 sets of CPC scores. This method makes the quantitative model much sensitive to the changes in the input value.

3.3 Human error quantification with the CREAM extended method

The purpose of the CREAM extended method is to produce specific action failure probabilities (Hollnagel, 1998), while the basic method does not consider specific human activities in predicting the action failure probability, but only through a context assessment. The CREAM extended method can be applied in the case that further analysis is required through the screening process using the human error probability obtained through the CREAM basic method, or when the analysis of individual event sequences is desired. In terms of risk assessment, this method can also be utilised for procedures review by identifying the delicate tasks that need risk control options or a task to revise from the whole task procedures. The CREAM extended method consists of three main steps, and the basic framework in this paper follows the original CREAM extended method introduced by Hollnagel (1998). The significant characteristic of the proposed method is that weighted and adjusted fuzzy sets for CPC scores are utilised to adjust a nominal cognitive failure probability. Therefore, this section summarises task analysis and verification in the step. 1, building the cognitive demand profile and determine the credit failure mode in step.2, then describes in detail how to use fuzzy sets to adjust the cognitive failure probabilities.

3.3.1 Task analysis and verification

Task analysis refers to methods of formally describing and analysing human-system interaction (Kirwan, 2017). Task analysis is conducted to define the steps which address the designated duties that the crew should complete successfully to achieve the main goal of the procedures with a hierarchical task analysis from the selected scenario. Then, the equipment or procedures of a vessel shall be evaluated to ensure that it satisfies the compulsory requirements of the domestic law or international convention according to the navigational area due to its operational characteristics. This process requires identifying the relevant requirements of the international Convention and domestic to verify the suitability of the procedures.

3.3.2 Build cognitive demand profile and determine credible error mode

The step starts by describing the scenario according to the event sequence and identify cognitive activities that characterise the activity of each work stage or event segment. The fifteen cognitive activity types are provided, and each cognitive activity is associated with one or more basic cognitive functions that consist of observation, interpretation, planning and execution by a generic cognitive-activity-by-cognitive-demand matrix as shown in Table *7*~~*Table 7*~~. Once cognitive demand is decided for task element, the next step is to identify the most likely generic failure type for the cognitive activity of the task element. The four basic cognitive functions are classified into 13 generic failure types, and the corresponding cognitive failure probability (CFP) for each generic failure type is given, as shown in Table *8*~~*Table 8*~~.

Table 7 Generic cognitive activity by cognitive demand matrix (Hollnagel, 1998)

Cognitive Activity type	Observation	Interpretation	Planning	Execution
Co-ordinate			V	V
Communicate				V
Compare		V		
Diagnose		V	V	
Evaluate		V	V	
Execute				V
Identify		V		
Maintain			V	V

Monitor	V	V		
Observe	V			
Plan			V	
Record		V		V
Regulate	V			V
Scan	V			
Verify	V	V		

Table 8 Nominal values and uncertainty bounds for cognitive function failures (Hollnagel, 1998)

Cognitive function	Generic failure type	Lower bound (5%)	Basic value	Upper bound (95%)
Observation	O1. Wrong object observed	3.00E-04	1.00E-03	3.00E-03
	O2. Wrong identification	2.00E-02	7.00E-02	1.70E-02
	O3. Observation not made	2.00E-02	7.00E-02	1.70E-02
Interpretation	I1. Faulty diagnosis	9.00E-02	2.00E-01	6.00E-01
	I2. Decision error	1.00E-03	1.00E-02	1.00E-01
	I3. Delayed interpretation	1.00E-03	1.00E-02	1.00E-01
Planning	P1. Priority error	1.00E-03	1.00E-02	1.00E-01
	P2. Inadequate plan	1.00E-03	1.00E-02	1.00E-01
Execution	E1. Action of wrong type	1.00E-03	3.00E-03	9.00E-03
	E2. Action at wrong time	1.00E-03	3.00E-03	9.00E-03
	E3. Action on wrong object	5.00E-05	5.00E-04	5.00E-03
	E4. Action out of sequence	1.00E-03	3.00E-03	9.00E-03
	E5. Missed action	2.50E-02	3.00E-03	4.00E-02

3.3.3 Adjusted CFP by weighting factors

The last step in the CREAM extended method is to adjust the nominal CFP with respect to the effect of the CPC. Nine fuzzy sets for all CPC scores are utilised in this step. For example, fuzzy set ((μ_{11}'', L_{11}), (μ_{12}'', $L_{12,}$), (μ_{13}'', $L_{13,}$), (μ_{14}'', L_{14})) represent a fuzzy score of CPC_1. Let define W_{ijn} as a weighting factor for the n-th generic failure type of the j-th CPC level at the i-th CPC and get data from the original CREAM by Hollnagel (1998). Then, let define W_{in} as a weighting factor for n-th cognitive function of CPC_i. The weighting factor, W_n, is acquired as follows;

$$W_{in} = \sum_{j=1}^{4} \mu_{ij} * W_{ijn} \quad (22)$$

$$W_n = \prod_{i=1}^{9} W_{in} \quad (23)$$

Where i= 1 to 9, j=1 to 3 or 4 and n= observation, Interpretation, planning and Execution

4. Case study on the engine room fire-fighting

According to Darbra and Casal (2004), accidents associated with fire and explosion at seaport account for 29% and 17% respectively. The statistical analysis for Maritime Accident Investigation Branch (MAIB) data by de Maya et al. (2019a) found fire and explosion accidents account for 6.78% of all marine accidents occurred from 1990 to 2016. Moreover, those incidents have a reputation of high mortality. Weng and Yang (2015) shows that fire and explosion related incidents result in 132% higher death tolls than other types of accidents. In particular, for passenger ships, fire/explosion accidents are the most frequent occurrence of total losses of ships compared to other accident types (Eliopoulou et al., 2016). According to Baalisampang et al. (2018), 48% of fire incidents in ships are related to human error, followed by mechanical failure 22% and temperature response 14%. In this context, this paper was motivated to apply the proposed method for potential fire incidents in engine room where majority of fire incidents take place.

Comment [SA9]: For a comment #9 by reviewer3

For an illustration of the proposed approach, both of scenario and procedures for the engine room fire-fighting in general cargo ship have been selected since fire drill at sea is a critical situation in which the crews are required to complete tasks for fire-fighting with limited resources such as personnel, equipment and time. The scenario of an engine room fire-fighting is described in section 4.1 for the purpose to assess CPCs and predict overall HEP without considering specific human activity in selected control mode by the CREAM basic method. The procedures of the engine room fire drill are selected and described in section 4.3 to conduct task analysis and predict individual CFP to all tasks by the CREAM extended method.

The application of the proposed method to case study and data collection were conducted in the following ways;

Firstly, in order to develop an actual emergency response procedure, the existing fire-fighting procedures used in cargo ships were obtained from numerous companies. Developed final procedure was verified and enhanced by a group of experts to ensure compliance with SOLAS and STCW requirements. Next, the scenario was generated to reflect the nine CPC characteristics through meetings of the expert group. Also, a criterion was applied when selecting experts for evaluation stage. In other words, experts who have practical experience of fire-fighting drill on ship as a crew member or safety system auditor are selected for this evaluation. Then, the assessment was conducted independently by each expert to eliminate the group thinking bias. The procedures and scenarios of the fire-fighting were provided for evaluation by a questionnaire using linguistic terms on the relative importance of each CPC and CPC level.

Comment [SA10]: For a comment #3 by reviewer2

4.1 Scenario definition

The scenario for engine room fire drill on a general cargo ship is described for illustration of the proposed method and focus on presenting CPCs for evaluation as follows.

On a hot summer day, a general cargo ship was waiting to departure at the anchoring position after finishing cargo loading. The temperature was 38 °C, and the humidity is 70 %. The sea conditions and winds were generally good. The vessel was five years old general cargo ship, G/T 5,000, and overall the vessel was in good condition. The ship's management company has managed a total of 30 vessels, holding both the company's DOC certificate and SMC certificates for individual ships in effect in accordance with an International Safety Management Code(ISM), and also obtained ISO certificates on the quality management system. Last month, an internal audit of the vessel was conducted by the company, and all three identified nonconformities have been rectified. A total of 20 crew members were on board and were made up of three different Nationalities. Six crew members were replaced the previous day and conducted familiarisation training in the afternoon of the previous day. Ship’s captain made a plan to conduct the fire drill and abandon ship today at 2 p.m. The fire extinguishing equipment consisted of a fixed CO2 gas system in the engine room; two main fire pumps located inside the main engine room, an emergency fire pump located in the steering gear room, portable fire extinguishers, two firemen's outfits, etc. All fire pumps were manually operated on-site and also remotely in the fire control room and bridge. All fire extinguishing equipment of ship has completed the periodical inspection in accordance with the SOLAS Convention. For communication during training, there were three portable communication devices. The company provided the Muster List to the vessel that consists of duties and responsibilities in case of such mishaps, designated and assigned to each person on the ship in case of emergency including fire and abandon ship. The captain had carried out a monthly fire-fighting and abandon ship drill three days ago, and the records were written in the ship's logbook. For six crews newly onboard, this drill is the first drill to be trained in the vessel, while the other 14 crews have all joined last month's training following the captain's training plan.

4.2 Common Performance Condition Assessment

The relative importance among experts is considered as a heterogeneous group depending on their background and assigned as 0.20, 0.18, 0.21, 0.20 and 0.21. For assessment, experts are asked to assign CPC scores and their relative importance as Table 9~~Table 9~~ and Table 10~~Table 10~~. Then, opinion aggregation from CPC$_1$ to CPC$_9$ except the CPC$_7$ and relative importance for nine CPCs are done. A relaxation factor β is assumed to be 0.5. As an example, specific aggregation for CPC4 are illustrated in ~~Table 11~~. Finally, aggregated fuzzy opinions are defuzzified and listed in ~~Table 12~~. Once experts' judgement and fuzzy opinion aggregation are completed, the next step is to convert the defuzzified CPC scores to fuzzy membership again for a human error quantification. Then adjust fuzzy sets by dependency relation a shown in ~~Figure 5~~ which is illustrated by a Genie software. Finally, the weighted & adjusted fuzzy sets are obtained by multiplying weighting factor to adjusted fuzzy sets. The fuzzy memberships are provided in ~~Table 12~~.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

Table 9 Experts' evaluations of CPCs and their standardised fuzzy set

		E1	E2	E3	E4	E5
CPC1	Expert opinion	Efficient(3)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)
CPC2	Expert opinion	Incompatible(1)	Incompatible(1)	Compatible(2)	Compatible(2)	Incompatible(1)
	Fuzzy set	(0, 0, 0.2, 0.6)	(0, 0, 0.2, 0.6)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0, 0, 0.2, 0.6)
CPC3	Expert opinion	Adequate (3)	Adequate (3)	Tolerable(2)	Adequate (3)	Tolerable(2)
	Fuzzy set	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)
CPC4	Expert opinion	Appropriate(3)	Appropriate(3)	Acceptable(2)	Appropriate(3)	Acceptable(2)
	Fuzzy set	(0.6, 0.9, 1, 1)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)	(0.6, 0.9, 1, 1)	(0.2, 0.6, 0.6, 0.9)
CPC5	Expert opinion	Matching current capacity(2)	Matching current capacity(2)	Matching current capacity (2)	Matching current capacity (2)	Matching current capacity (2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC6	Expert opinion	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)	Temporarily inadequate(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC8	Expert opinion	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)	Adequate, limited experience(2)
	Fuzzy set	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)	(0.2, 0.6, 0.6, 0.9)
CPC9	Expert opinion	Inefficient(2)	Efficient(3)	Efficient(3)	Inefficient(2)	Efficient(3)
	Fuzzy set	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)	(0.3, 0.7, 0.7, 0.9)	(0.1, 0.3, 0.3, 0.7)	(0.3, 0.7, 0.7, 0.9)

Table 10 Experts' evaluation for the relative importance of CPCs

		E1	E2	E3	E4	E5
RI1	Expert opinion	Moderate	Highly important	Highly important	Moderate	Highly important
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)
RI2	Expert opinion	Highly important	Highly important	Highly important	Highly important	Highly important

	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)
RI3	Expert opinion	Highly important	Moderate	Highly important	Moderate	Moderate
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)
RI4	Expert opinion	Moderate	Moderate	Very highly important	Moderate	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)	(0.8,0.9,1,1)	(0.3,0.5,0.5,0.7)	(0.3,0.5,0.5,0.7)
RI5	Expert opinion	Moderate	Highly important	Moderate	Less important	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1, 0.25, 0.25, 0.4)	(0.3,0.5,0.5,0.7)
RI6	Expert opinion	Moderate	Highly important	Highly important	Highly important	Moderate
	Fuzzy set	(0.3,0.5,0.5,0.7)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)
RI7	Expert opinion	Less important	Highly important	Moderate	Less important	Moderate
	Fuzzy set	(0.1,0.25,0.25,0.4)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1,0.25,0.25,0.4)	(0.3,0.5,0.5,0.7)
RI8	Expert opinion	Highly important	Very highly important	Very highly important	Very highly important	Highly important
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.8,0.9,1,1)	(0.8,0.9,1,1)	(0.8,0.9,1,1)	(0.6,0.75,0.75,0.9)
RI9	Expert opinion	Highly important	Highly important	Highly important	Moderate	Less important
	Fuzzy set	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.6,0.75,0.75,0.9)	(0.3,0.5,0.5,0.7)	(0.1,0.25,0.25,0.4)

Table 11 Aggregation under the CPC$_4$

Fuzzy sets for expert opinion					Relative degree of agreement(RA)	
Ex1	0.6	0.9	1	1	RA(Ex1)	0.207317
Ex2	0.6	0.9	1	1	RA(Ex2)	0.207317
Ex3	0.2	0.6	0.6	0.9	RA(Ex3)	0.189024
Ex4	0.6	0.9	1	1	RA(Ex4)	0.207317
Ex5	0.2	0.6	0.6	0.9	RA(Ex5)	0.189024
Degree of agreement(S)					Relative degree of agreement(RA)	
S12	1		S34	0.7	RA(Ex1)	0.207317
S23	0.7		S15	0.7	RA(Ex2)	0.207317
S13	0.7		S25	0.7	RA(Ex3)	0.189024
S14	1		S35	1	RA(Ex4)	0.207317
S24	1		S45	0.7	RA(Ex5)	0.189024
Average degree of agreement(AA)					Consensus degree coefficient(CC)	
AA(Ex1)				0.85	CC(Ex1)	0.202343
AA(Ex2)				0.85	CC(Ex2)	0.195764
AA(Ex3)				0.775	CC(Ex3)	0.199775
AA(Ex4)				0.85	CC(Ex4)	0.202343
AA(Ex5)				0.775	CC(Ex5)	0.199775
Rag(HT)	0.44018	0.780135	0.761874	0.960045		
Defuzzification				0.722883		
Normalised score				72.28828		

Table 12 Fuzzy sets for the CPCs assessment for fire-fighting scenario

				Fuzzy sets for CPC score				Adjusted fuzzy sets for CPC score				Weighted & adjusted fuzzy sets for CPC score			
CPC$_i$	score	RI$_i$	W$_i$	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i1}'	μ_{i2}'	μ_{i3}'	μ_{i4}'	μ_{i1}''	μ_{i2}''	μ_{i3}''	μ_{i4}''

CPC_1	57	0.12	1.05	0.00	0.33	0.68	0.00	0.00	0.33	0.68	0.00	0.00	0.34	0.71	0
CPC_2	35	0.13	1.20	0.63	0.38	0.00	-	0.63	0.38	0.00	-	0.75	0.45	0.00	-
CPC_3	51	0.11	0.95	0.00	0.48	0.53	0.00	0.00	0.48	0.53	0.00	0.00	0.45	0.50	0
CPC_4	72	0.10	0.89	0.00	0.60	0.40	-	0.00	0.60	0.40	-	0.00	0.53	0.36	-
CPC_5	55	0.09	0.80	0.13	0.88	0.00	-	0.13	0.88	0.00	-	0.10	0.70	0.00	-
CPC_6	55	0.12	1.05	0.13	0.88	0.00	-	0.15	0.85	0.00	-	0.16	0.89	0.00	-
CPC_7	14h	0.08	0.70	0.00	0.00	1.00	-	0.00	0.00	1.00	-	0.00	0.00	0.70	-
CPC_8	55	0.15	1.36	0.13	0.88	0.00	-	0.13	0.88	0.00	-	0.17	1.19	0.00	-
CPC_9	52	0.11	1.00	0.00	0.45	0.55	0.00	0.02	0.45	0.53	0.00	0.02	0.45	0.53	0
Total	-	1.00	9.00		9.00				9.00				9.00		

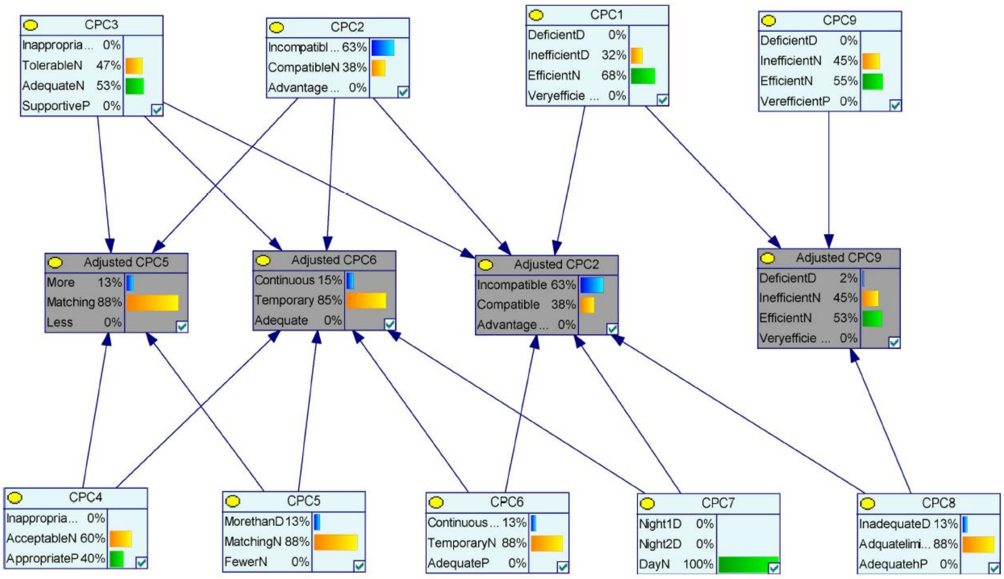


Figure 5 Bayesian presentation for the dependency of the performance condition

Comment [SA11]: For a comment #7 by reviewer2
Revised results by additional experts

~~Three marine experts are carefully selected for this assessment, and all experts have practical experience of fire-fighting drill on ship as a crew member or ISM auditor. The relative importance among experts is considered as a heterogeneous group depending on their background. For assessment, experts are asked to assign CPC scores and their relative importance as Table 7 and Table 8. Then, opinion aggregation from CPC_1 to CPC_9 except the CPC_7 and relative importance for nine CPCs are done. A relaxation factor β is assumed to be 0.5 and relative importance among experts are assigned as 0.33, 0.31 and 0.36 for three experts. As an example, specific aggregation for CPC4 are illustrated in Table 9. Finally, aggregated fuzzy opinions are defuzzified and listed in Table 10. Once experts' judgement and fuzzy opinion aggregation are completed, the next step is to convert the defuzzified CPC scores to fuzzy membership again for a human error quantification. Then adjust fuzzy sets by dependency relation a shown in Figure 5 which is illustrated by a Genie software. Finally, the weighted & adjusted fuzzy sets are obtained by multiplying weighting factor to adjusted fuzzy sets. The fuzzy memberships are provided in Table 10.~~

~~Table 7 Experts' evaluations of CPCs and their standardised fuzzy set~~

~~-~~	~~-~~	~~E1~~	~~E2~~	~~E3~~
~~CPC1~~	~~Expert opinion~~	~~Efficient~~	~~Efficient(3)~~	~~Efficient(3)~~
	~~St. fuzzy set~~	~~(0.3, 0.7, 0.7, 0.9)~~	~~(0.3, 0.7, 0.7, 0.9)~~	~~(0.3, 0.7, 0.7, 0.9)~~
~~CPC2~~	~~Expert opinion~~	~~Incompatible(1)~~	~~Incompatible(1)~~	~~Compatible(2)~~
	~~St. fuzzy set~~	~~(0, 0, 0.2, 0.6)~~	~~(0, 0, 0.2, 0.6)~~	~~(0.2, 0.6, 0.6, 0.9)~~
~~CPC3~~	~~Expert opinion~~	~~Tolerable(2)~~	~~Tolerable(2)~~	~~Tolerable(2)~~
	~~St. fuzzy set~~	~~(0.1, 0.3, 0.3, 0.7)~~	~~(0.1, 0.3, 0.3, 0.7)~~	~~(0.1, 0.3, 0.3, 0.7)~~
~~CPC4~~	~~Expert opinion~~	~~Appropriate(3)~~	~~Appropriate(3)~~	~~Acceptable(2)~~
	~~St. fuzzy set~~	~~(0.6, 0.9, 1, 1)~~	~~(0.6, 0.9, 1, 1)~~	~~(0.2, 0.6, 0.6, 0.9)~~
~~CPC5~~	~~Expert opinion~~	~~Matching current capacity~~	~~Matching current capacity~~	~~Matching current capacity (2)~~
	~~St. fuzzy set~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~
~~CPC6~~	~~Expert opinion~~	~~inadequate~~	~~inadequate(1)~~	~~inadequate(1)~~
	~~St. fuzzy set~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~
~~CPC8~~	~~Expert opinion~~	~~Adequate with limited experience~~	~~Adequate limited -experience~~	~~Adequate with -limited experience~~
	~~St.fuzzy set~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~	~~(0.2, 0.6, 0.6, 0.9)~~
~~CPC9~~	~~Expert opinion~~	~~Inefficient(2)~~	~~Inefficient(2)~~	~~Efficient(3)~~
	~~St. fuzzy set~~	~~(0.1, 0.3, 0.3, 0.7)~~	~~(0.1, 0.3, 0.3, 0.7)~~	~~(0.1, 0.3, 0.3, 0.7)~~

~~Table 8 Experts' evaluation for the relative importance of CPCs~~

~~-~~	~~-~~	~~E1~~	~~E2~~	~~E3~~
~~RI_1~~	~~Expert opinion~~	~~Moderate~~	~~Highly important~~	~~Highly important~~
	~~St. fuzzy set~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~
~~RI_2~~	~~Expert opinion~~	~~Highly important~~	~~Highly important~~	~~Highly important~~
	~~St. fuzzy set~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~
~~RI_3~~	~~Expert opinion~~	~~Highly important~~	~~Moderate~~	~~Highly important~~
	~~St. fuzzy set~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.6,0.75,0.75,0.9)~~
~~RI_4~~	~~Expert opinion~~	~~Moderate~~	~~Moderate~~	~~Very highly important~~
	~~St. fuzzy set~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.8,0.9,1,1)~~
~~RI_5~~	~~Expert opinion~~	~~Moderate~~	~~Highly important~~	~~Moderate~~
	~~St. fuzzy set~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.3,0.5,0.5,0.7)~~
~~RI_6~~	~~Expert opinion~~	~~Moderate~~	~~Highly important~~	~~Highly important~~
	~~St. fuzzy set~~	~~(0.3,0.5,0.5,0.7)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~
~~RI_7~~	~~Expert opinion~~	~~Less important~~	~~Highly important~~	~~Moderate~~
	~~St. fuzzy set~~	~~(0.1,0.25,0.25,0.4)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.3,0.5,0.5,0.7)~~
~~RI_8~~	~~Expert opinion~~	~~Highly important~~	~~Very highly important~~	~~Very highly important~~
	~~St. fuzzy set~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.8,0.9,1,1)~~	~~(0.8,0.9,1,1)~~
~~RI_9~~	~~Expert opinion~~	~~Highly important~~	~~Highly important~~	~~Highly important~~
	~~St. fuzzy set~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~	~~(0.6,0.75,0.75,0.9)~~

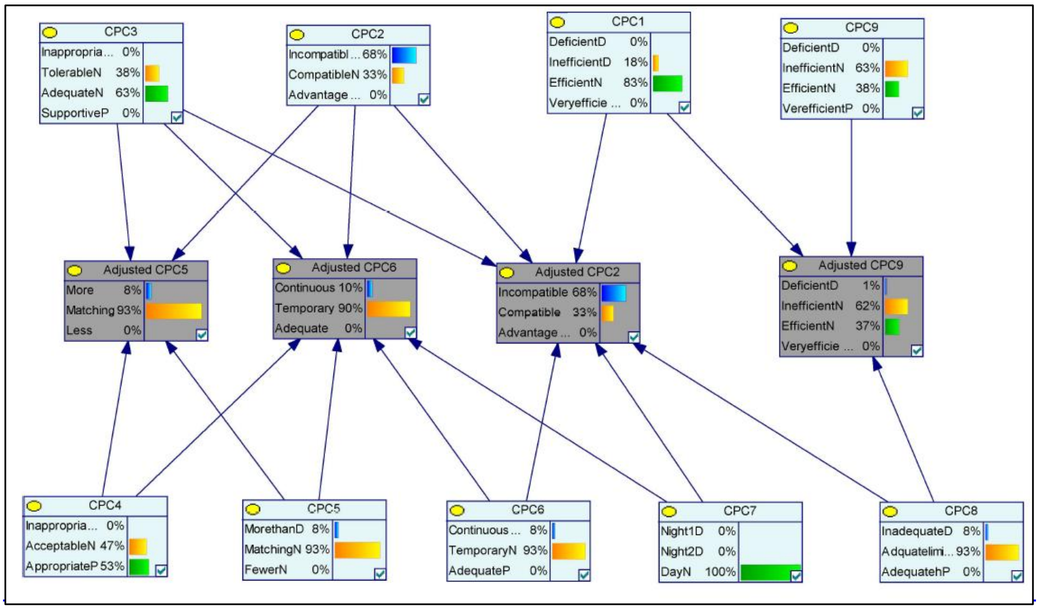
~~Table 9 Aggregation under the CPC$_4$~~

~~Experts' opinion~~	
~~Ex1~~	~~(0.6, 0.9, 1, 1)~~
~~Ex2~~	~~(0.6, 0.9, 1, 1)~~
~~Ex3~~	~~(0.2, 0.6, 0.6, 0.9)~~

~~The degree of agreement(S)~~	
~~S12~~	~~1.00~~
~~S23~~	~~0.70~~
~~S31~~	~~0.70~~
~~The average degree of agreement(AA)~~	
~~AA(Ex1)~~	~~0.85~~
~~AA(Ex2)~~	~~0.85~~
~~AA(Ex3)~~	~~0.70~~
~~The relative degree of agreement(RA)~~	
~~RA(Ex1)~~	~~0.35~~
~~RA(Ex2)~~	~~0.35~~
~~RA(Ex3)~~	~~0.29~~
~~The consensus degree coefficient(CC)~~	
~~CC(Ex1)~~	~~0.34~~
~~CC(Ex2)~~	~~0.33~~
~~CC(Ex3)~~	~~0.33~~
~~Aggregated result~~	
~~Rag(HT)~~	~~(0.47, 0.80, 0.87, 0.97)~~
~~Defuzzification~~	~~0.76~~
~~Normalised S.~~	~~76~~

~~Table 10 Fuzzy sets for the CPCs assessment for fire-fighting scenario~~

	~~Defuzzified Score and Relative importance~~			~~Fuzzy sets for CPC score~~				~~Adjusted fuzzy sets for CPC score~~				~~Weighted & adjusted fuzzy sets for CPC score~~			
~~CPC_i~~	~~score~~	~~RI_i~~	~~W_i~~	~~μ_{i1}~~	~~μ_{i2}~~	~~μ_{i3}~~	~~μ_{i4}~~	~~μ_{i1}'~~	~~μ_{i2}'~~	~~μ_{i3}'~~	~~μ_{i4}'~~	~~μ_{i1}''~~	~~μ_{i2}''~~	~~μ_{i3}''~~	~~μ_{i4}''~~
~~CPC_1~~	~~63~~	~~0.11~~	~~0.99~~	~~0.00~~	~~0.18~~	~~0.83~~	~~0.00~~	~~0.00~~	~~0.18~~	~~0.83~~	~~0.00~~	~~0.00~~	~~0.17~~	~~0.82~~	~~0~~
~~CPC_2~~	~~33~~	~~0.12~~	~~1.11~~	~~0.68~~	~~0.33~~	~~0.00~~	~~-~~	~~0.68~~	~~0.33~~	~~0.00~~	~~-~~	~~0.75~~	~~0.36~~	~~0.00~~	~~-~~
~~CPC_3~~	~~55~~	~~0.11~~	~~1.00~~	~~0.00~~	~~0.38~~	~~0.63~~	~~0.00~~	~~0.00~~	~~0.38~~	~~0.63~~	~~0.00~~	~~0.00~~	~~0.37~~	~~0.62~~	~~0~~
~~CPC_4~~	~~76~~	~~0.10~~	~~0.93~~	~~0.00~~	~~0.47~~	~~0.53~~	~~-~~	~~0.00~~	~~0.47~~	~~0.53~~	~~-~~	~~0.00~~	~~0.44~~	~~0.50~~	~~-~~
~~CPC_5~~	~~57~~	~~0.09~~	~~0.85~~	~~0.08~~	~~0.93~~	~~0.00~~	~~-~~	~~0.08~~	~~0.93~~	~~0.00~~	~~-~~	~~0.06~~	~~0.79~~	~~0.00~~	~~-~~
~~CPC_6~~	~~57~~	~~0.11~~	~~0.99~~	~~0.08~~	~~0.93~~	~~0.00~~	~~-~~	~~0.10~~	~~0.90~~	~~0.00~~	~~-~~	~~0.10~~	~~0.89~~	~~0.00~~	~~-~~
~~CPC_7~~	~~14h~~	~~0.08~~	~~0.74~~	~~0.00~~	~~0.00~~	~~1.00~~	~~-~~	~~0.00~~	~~0.00~~	~~1.00~~	~~-~~	~~0.00~~	~~0.00~~	~~0.74~~	~~-~~
~~CPC_8~~	~~57~~	~~0.14~~	~~1.28~~	~~0.08~~	~~0.93~~	~~0.00~~	~~-~~	~~0.08~~	~~0.93~~	~~0.00~~	~~-~~	~~0.10~~	~~1.18~~	~~0.00~~	~~-~~
~~CPC_9~~	~~55~~	~~0.12~~	~~1.11~~	~~0.00~~	~~0.38~~	~~0.63~~	~~0.00~~	~~0.00~~	~~0.38~~	~~0.63~~	~~0.00~~	~~0.00~~	~~0.42~~	~~0.69~~	~~0~~
~~Total~~	~~-~~	~~1.00~~	~~9.00~~	~~9.00~~				~~9.00~~				~~9.00~~			



~~Figure 5 Bayesian presentation for the dependency of the performance condition~~

4.3 Human error quantification with the CREAM basic method

This section presents the process to calculate the overall human error probability from fuzzy memberships for CPCs by the proposed approach based CREAM basic method.

4.3.1 CPC evaluation

In this step, adjusted & weighted fuzzy sets of CPCs score is quantified to combined CPC score. The combined CPC score is calculated as reduced effect 1.54, improved effect 0.36 by multiplying expected effect in accordance with section 3.2.1.

4.3.2 Fuzzification of combined CPC score

This section describes the process to infer the distribution of belief degrees corresponding to four control modes consist of Strategic (D_1), Tactical (D_2), Opportunistic (D_3) and Scrambled (D_4) from the combined CPC score point K (1.54, 0.36). Subsets $A^{1.54}$ and $A^{0.36}$ are obtained by analysing the portion of squares of different control modes in each row and column to the point K as follows.

$$A^{K-} = A^{1.54} = ((\frac{2}{8}, D_1), (\frac{6}{8}, D_2), (0, D_3), (0, D_4))$$

$$A^{K+} = A^{0.36} = ((0, D_1), (\frac{3}{10}, D_2), (\frac{3}{10}, D_3), (\frac{4}{10}, D_4))$$

Normalised coefficient $\theta^{1.54}$ and $\theta^{0.36}$ are acquired after parallel movement of centre of coordinate from (0,0) to (1,1) by the equation (17) as follows.

$\theta^{1.54} = \frac{2.2.5418}{2.54+1.36} = 0.65$, $\theta^{0.36} = \frac{1.36}{2.54+1.36} = 0.35$

$M^{1.54}$ and $M^{0.36}$ are set of belief degrees to support the hypothesis that the subset A^{K-} and A^{K+} are identified in four control modes by the equation (18) as follows.

$M^{1.54} = ((0.65 * \frac{2}{8}, D_1), (0.65 * \frac{6}{8}, D_2), (0, D_3), (0, D_4))$

$M^{0.36} = ((0, D_1), (0.35 * \frac{3}{10}, D_2), (0.35 * \frac{3}{10}, D_3), (0.35 * \frac{4}{10}, D_4))$

Coefficients P, H and set of A^K are calculated by equation (19) and an output of human error quantification model is derived as follows.

P=1.21, H=0.27

$A^{(1.54,0.36)}$ = ((0.18, D1), (0.68, D2), (0.06, D3), (0.08, D4))

4.3.3 Defuzzification and Human error probability

A set of belief degrees to the four control modes $A^{(1.54,0.36)}$ is defuzzified into a logarithm number negative 2.12; then HEP is derived by equation (21) as follows.

HEP (human error probability) = 10^{CV} = 0.0076

4.4 Human error quantification with the CREAM extended method

In accordance with SOLAS Chapter3, Regulation 19.3.2, all crew members shall participate in at least one abandon ship and fire drill every month (IMO, 2001). Fire-fighting facilities in each ship vary depending on the requirement of fire detection and extinguish system as well as on the type of vessels and cargo. Therefore, fire drills for specific ships should be planned so that proper consideration of regular practice in various emergencies can be made. The procedures also have to consider an abandon-ship decision made by the ship's Master in case of fire-fighting failure.

4.4.1 Task analysis and verification

The hierarchical task analysis for the procedures of engine room fire-fighting is shown in Table 13~~Table 13~~. The procedures are confirmed that all compulsory requirements by SOLAS* Chapter 3, Regulation 19.3.5.2 are included (IMO, 2001). The procedure consists of seven main tasks which are i) Fire detection and announcement, ii) Assembly at the muster station, iii) Check openings in the engine room area, iv) Preparation of the fireman, v) Preparation of the fire pump and water spray, vi) Fire-fighting, vii) Further actions and main tasks are divided to twenty-three subtasks as Table 13~~Table 13~~.

*International Convention for the Safety of Life at Sea (SOLAS), 1974

Table 13 Procedures of the engine room fire-fighting in general ship

Engine room fire-fighting procedures
1. Fire detection and announcement
1.1 Detect fire in the engine room
1.2 Report to the wheelhouse
1.3 Push the fire alarm and make an announcement
1.4 Report to stations
2. Assembly at the muster station
2.1 Ensure all crew gathered at the muster station
2.2 Check fireman's outfit and other personal rescue equipment

  2.3 Describe the fire-fighting procedures and duties to all crew members
  2.4 Check communication equipment
3. Check openings in the engine room area
  3.1 Stop all-electric ventilation fan
  3.2 Close all air inlets and doors into the engine room
  3.3 Ensure no air supply into the engine room
4. Preparation of the fireman
  4.1 Wear fireman's outfit with equipment
  4.2 Ensure all fireman's equipment good in order
5. Preparation of the fire pump and water spray
  5.1 Open suction valve for the fire pump
  5.2 Close main isolating valve
  5.3 Connect at least two fire hoses to fire hydrants
  5.4 Start the (emergency) fire pump
  5.5 Check the water pressure
6. Fire fighting
  6.1 Start water spray to engine room boundary for cooling
  6.2 Fireman, access into fire site and fire fighting
7. Further actions
  7.1 Ensure fire extinguished completely
  7.2 Check the necessary of the fixed fire extinguisher system(e.g.CO2 gas)
  7.3 Check the necessary of the abandon ship

4.4.2 Build Cognitive demand profile and determine credible error mode

All tasks from 1.1 to 7.3 matched to one of the cognitive activities associated with cognitive demand and credible failure mode. The most likely error mode to the cognitive activity of each task is decided carefully in Table 14~~Table 14~~. Nominal Cognitive Failure Probability (CFP_O) are provided from Table *8*~~*Table 8*~~.

4.4.3 Adjusted CFP by weighting factors

Weighting factor per cognitive demand is calculated by equation (22) and (23) for fire-fighting procedures and the adjusted CFP throughout the whole procedures is illustrated in Table 14~~Table 14~~.

Table 14 CREAM extended method analysis result for the engine room fire-fighting procedures

Tasks	Cognitive activity	Cognitive Demands	The most credible Error mode	CFP_O	W_n	Adjusted CFP
1.1	Observe	Observation	O3. Observation not made	7.00E-02	2.64	1.85E-01
1.2	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
1.3	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
1.4	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
2.1	Observe	Observation	O3. Observation not made	7.00E-02	2.64	1.85E-01
2.2	Verify	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
2.3	Communicate	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03

2.4	Verify	Observation Interpretation	O3. Observation not made	7.00E-02	2.64	1.85E-01
3.1	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
3.2	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
3.3	Monitor	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
4.1	Execute	Execution	E1. Action of wrong type	3.00E-03	2.98	8.94E-03
4.2	Verify	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
5.1	Execute	Execution	E3. Action on wrong object	5.00E-04	2.98	1.49E-03
5.2	Execute	Execution	E3. Action on wrong object	5.00E-04	2.98	1.49E-03
5.3	Execute	Execution	E1. Action of wrong type	3.00E-03	2.98	8.94E-03
5.4	Execute	Execution	E4. Action out of sequence	3.00E-03	2.98	8.94E-03
5.5	Verify	Observation Interpretation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
6.1	Execute	Execution	E5. Missed action	3.00E-03	2.98	8.94E-03
6.2	Execute	Execution	E4. Action out of sequence	3.00E-03	2.98	8.94E-03
7.1	Observe	Observation	O2. Wrong identification	7.00E-02	2.64	1.85E-01
7.2	Diagnose	Interpretation Plan	I2. Decision error	1.00E-02	3.84	3.84E-02
7.3	Diagnose	Interpretation Plan	I2. Decision error	1.00E-02	3.84	3.84E-02

~~Table 12 CREAM extended method analysis result for the engine room fire-fighting procedures~~

~~Tasks~~	~~Cognitive activity~~	~~Cognitive Demands~~	~~The most credible Error mode~~	~~CFP_0~~	~~W_n~~	~~Adjusted CFP~~
~~1.1~~	~~Observe~~	~~Observation~~	~~O3. Observation not made~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~1.2~~	~~Communicate~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~1.3~~	~~Execute~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~1.4~~	~~Communicate~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~2.1~~	~~Observe~~	~~Observation~~	~~O3. Observation not made~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~2.2~~	~~Verify~~	~~Observation~~ ~~Interpretation~~	~~O2. Wrong identification~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~2.3~~	~~Communicate~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~2.4~~	~~Verify~~	~~Observation~~ ~~Interpretation~~	~~O3. Observation not made~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~3.1~~	~~Execute~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~3.2~~	~~Execute~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~3.3~~	~~Monitor~~	~~Observation~~ ~~Interpretation~~	~~O2. Wrong identification~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~

~~4.1~~	~~Execute~~	~~Execution~~	~~E1. Action of wrong type~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~4.2~~	~~Verify~~	~~Observation Interpretation~~	~~O2. Wrong identification~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~5.1~~	~~Execute~~	~~Execution~~	~~E3. Action on wrong object~~	~~5.00E-04~~	~~2.27~~	~~1.14E-03~~
~~5.2~~	~~Execute~~	~~Execution~~	~~E3. Action on wrong object~~	~~5.00E-04~~	~~2.27~~	~~1.14E-03~~
~~5.3~~	~~Execute~~	~~Execution~~	~~E1. Action of wrong type~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~5.4~~	~~Execute~~	~~Execution~~	~~E4. Action out of sequence~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~5.5~~	~~Verify~~	~~Observation Interpretation~~	~~O2. Wrong identification~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~6.1~~	~~Execute~~	~~Execution~~	~~E5. Missed action~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~6.2~~	~~Execute~~	~~Execution~~	~~E4. Action out of sequence~~	~~3.00E-03~~	~~2.27~~	~~6.81E-03~~
~~7.1~~	~~Observe~~	~~Observation~~	~~O2. Wrong identification~~	~~7.00E-02~~	~~2.20~~	~~1.54E-01~~
~~7.2~~	~~Diagnose~~	~~Interpretation Plan~~	~~I2. Decision error~~	~~1.00E-02~~	~~2.99~~	~~2.99E-02~~
~~7.3~~	~~Diagnose~~	~~Interpretation Plan~~	~~I2. Decision error~~	~~1.00E-02~~	~~2.99~~	~~2.99E-02~~

5. Findings and discussion

The proposed approach presents individual human failure probabilities obtained by a proposed CREAM based method by separating the context assessment process and human error quantification process based on a particular maritime scenario; engine room fire-fighting procedures. From the result of the basic method, it is revealed that significant control mode is Tactical mode with 68 % belief and also have 18% belief of Strategic mode, 6 % belief of Opportunistic mode and 8% belief of Scrambled mode. The overall human failure probability indicates 0.0076, which can occur under the given circumstance described in the fire-fighting scenario. For the result of the extended method, the weighting factor per cognitive function shows the most significant adverse effect on the interpretation in a given scenario 3.84, followed by 2.98 on an execution, 2.67 on planning and 2.64 on observation. For the comparison, the weighting factor in Tactical mode is 1.90 by a simple table in original CREAM. The range of weighting between 1.62 and 2.99 of the proposed approach is quite reasonable. The main finding is that the vulnerable subtasks with the higher failure probability are identified during the fire-fighting procedure, as shown in ~~Table 12~~. The highest failure probability is task No. 1.1 (Detect fire in the engine room), 2.1 (Ensure all crew gathered at the muster station), 2.2 (Check fireman's outfit and other personal rescue equipment), 2.4 (Check communication equipment), 3.3 (Ensure no air supply into the engine room), 4.2 (Ensure all fireman's equipment good in order), 5.5 (Check the water pressure), and 7.1 (Ensure fire extinguished completely) with 0.185 failure probability. While Task No. 5.1 (Open suction valve for the fire pump) and 5.2 (Close main isolating valve) show the lowest HEP with 0.00149. This result means that simple physical activity has lower failure probability than complex cognitive activities which need the additional ability for an interpretation and decision. The study also found that 'Adequacy of training and experience' is recognized as the most significant CPC factor contributing to human error in fire-fighting scenarios with a weight of 1.36, followed by 'working conditions' with a weight of 1.20 times,

Field Code Changed

'the adequacy of organization' and 'available time' with a weight of 1.05. The weighting for nine CPCs is illustrated in Figure 6~~Figure 6~~. For comparison, the original CREAM method is applied to the same assessment as Table 15~~Table 15~~. The overall results can be found to be within reasonable limits. The notable thing is that the proposed method can identify the effects of other control modes that are ignored by single control mode, and the quantified human failure probability can be obtained. The method allows the same analysis to be expressed in more detailed output. This research result can improve the fire-fighting procedures and also other critical operating's procedures on the ship and finally contributes to safety at sea.

Comment [SA12]: For a comment #6 by reviewer2

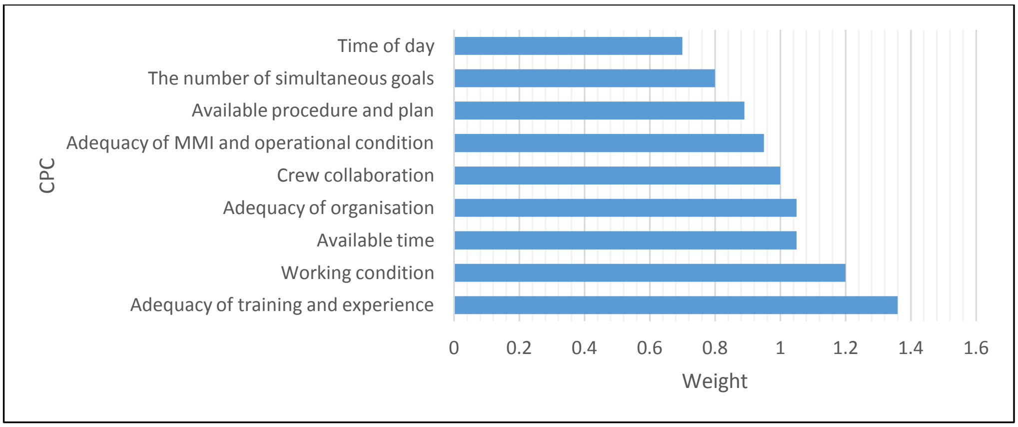


Figure 6 Factors contributing to human failure in fire-fighting

Table 15 Comparison result with the original CREAM method

-	Original CREAM	Proposed Method
Combined CPC Score	Expert$_1$ (1,1) Expert$_2$ (1,1) Expert$_3$ (0,0) Expert$_4$ (1,1) Expert$_5$ (1,0)	(1.54,0.36)
Control Mode	Tactical (100%)	Tactical (68%) Strategic (18%) Scrambled (8 %) Opportunistic (6%)
Overall Human Error Probability	between 0.001 and 0.1	0.0076

~~Table 13 Comparison result with the original CREAM method~~

~~-~~	~~Original CREAM~~	~~Proposed Method~~
~~Combined CPC Score~~	~~Expert$_1$ (1,1)~~ ~~Expert$_2$ (1,1)~~ ~~Expert$_3$ (0,0)~~	~~(1.18,0.50)~~
~~Control Mode~~	~~Tactical (100%)~~	~~Tactical (56%)~~ ~~Strategic (23%)~~

		~~Scrambled (12 %)~~ ~~Opportunistic (9%)~~
~~Overall Human Error Probability~~	~~between 0.001 and 0.1~~	~~0.007348~~

6. Conclusion

This paper introduces a new framework based CREAM applicable to the maritime industry and illustrates practical fire-fighting scenario and procedures. The characteristics and expected advantages of the proposed method are: Firstly, the proposed method provides an independent process of Common Performance Condition (CPC) assessment from HEP quantification models. This structure is because to provide a simple way to reflect a change of parameters. For an example, when the concerned analysis is needed to change the type of CPCs and their linguistic terms with fuzzy sets to reflect characteristics of the context, the same HEP quantification model can be applied to various situations by separating quantification model from the CPCs assessment. Furthermore, the same quantification model can be applied to individual assessments by different experts, either with different weighting factors for the relative importance of CPC. This simple structure could be realised to get an instant estimation of human failure probability without adjusting the parameters of the HEP quantification model for assessing a specific task. Secondly, the output of CPC assessment can be utilised as an input value in the CREAM basic method and also weighting factors in the CREAM extended method, respectively. This method makes the whole procedures more useful by allowing the results of CPC assessment to be used not only in the basic method but also in the extended method. Finally, the proposed method can evaluate the context in a maritime scenario based on the CREAM basic method and illustrate practical application to onboard procedures in the context in vessels by using the CREAM extended method. The proposed framework also can be extended to apply to the other ship procedures with various scenarios. For a more convenient application, the quantification model does not require a rule-based inference system. Instead, it infers the distribution of belief for control modes from the specific combined score of CPC for human error quantification. In conclusion, the results of this study can make positive impact on the safety of shipping operations and the enhancement of safety at sea by providing a framework applicable to human error analysis.

REFERENCES

ABBASSI, R., KHAN, F., GARANIYA, V., CHAI, S., CHIN, C. & HOSSAIN, K. A. 2015. An integrated method for human error probability assessment during the maintenance of offshore facilities. *Process Safety and Environmental Protection,* 94**,** 172-179.

AKYUZ, E. 2015. Quantification of human error probability towards the gas inerting process on-board crude oil tankers. *Safety Science,* 80**,** 77-86.

AKYUZ, E. 2016. Quantitative human error assessment during abandon ship procedures in maritime transportation. *Ocean Engineering,* 120**,** 21-29.

AKYUZ, E., CELIK, E. & CELIK, M. 2018. A practical application of human reliability assessment for operating procedures of the emergency fire pump at ship. *Ships and Offshore Structures,* 13**,** 208-216.

AKYUZ, E. & CELIK, M. 2015. Application of CREAM human reliability model to cargo loading process of LPG tankers. *Journal of Loss Prevention in the Process Industries,* 34**,** 39-48.

AKYUZ, E. & CELIK, M. 2016. A hybrid human error probability determination approach: The case of cargo loading operation in oil/chemical tanker ship. *Journal of Loss Prevention in the Process Industries,* 43**,** 424-431.

ANTÃO, P. & SOARES, C. G. 2019. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks. *Accident Analysis & Prevention,* 133**,** 105262.

BAALISAMPANG, T., ABBASSI, R., GARANIYA, V., KHAN, F. & DADASHZADEH, M. 2018. Review and analysis of fire and explosion accidents in maritime transportation. *Ocean Engineering,* 158**,** 350-366.

BALMAT, J.-F., LAFONT, F., MAIFRET, R. & PESSEL, N. 2011. A decision-making system to maritime risk assessment. *Ocean Engineering,* 38**,** 171-176.

CELIK, M. & CEBI, S. 2009. Analytical HFACS for investigating human errors in shipping accidents. *Accident Analysis & Prevention,* 41**,** 66-75.

DARBRA, R.-M. & CASAL, J. 2004. Historical analysis of accidents in seaports. *Safety science,* 42**,** 85-98.

DE MAYA, B. N., AHN, S. & KURT, R. Statistical analysis of MAIB database for the period 1990-2016. 18th International Congress of the Maritime Association of the Mediterranean, 2019a.

DE MAYA, B. N., BABALEYE, A. O. & KURT, R. E. 2019b. Marine accident learning with fuzzy cognitive maps (MALFCMs) and Bayesian networks. *Safety in Extreme Environments***,** 1-10.

ELIOPOULOU, E., PAPANIKOLAOU, A. & VOULGARELLIS, M. 2016. Statistical analysis of ship accidents and review of safety level. *Safety science,* 85**,** 282-292.

FUJITA, Y. & HOLLNAGEL, E. 2004. Failures without errors: quantification of context in HRA. *Reliability Engineering & System Safety,* 83**,** 145-151.

HÄNNINEN, M. 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Accident Analysis & Prevention,* 73**,** 305-312.

HE, X., WANG, Y., SHEN, Z. & HUANG, X. 2008. A simplified CREAM prospective quantification process and its application. *Reliability Engineering & System Safety,* 93**,** 298-306.

HOLLNAGEL, E. 1998. *Cognitive reliability and error analysis method (CREAM)*, Elsevier.

IMO 2001. the Regulation 19 of SOLAS Chapter III.

ISLAM, R., ABBASSI, R., GARANIYA, V. & KHAN, F. 2017a. Development of a human reliability assessment technique for the maintenance procedures of marine and offshore operations. *Journal of Loss Prevention in the Process Industries,* 50**,** 416-428.

ISLAM, R., ABBASSI, R., GARANIYA, V. & KHAN, F. I. 2016. Determination of human error probabilities for the maintenance operations of marine engines. *Journal of Ship Production and Design,* 32**,** 226-234.

ISLAM, R., KHAN, F., ABBASSI, R. & GARANIYA, V. 2018. Human error probability assessment during maintenance activities of marine systems. *Safety and health at work,* 9**,** 42-52.

ISLAM, R., YU, H., ABBASSI, R., GARANIYA, V. & KHAN, F. 2017b. Development of a monograph for human error likelihood assessment in marine operations. *Safety science,* 91**,** 33-39.

JIAN-BO, Y. & DONG-LING, X. 2002. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* 32**,** 289-304.

KHAN, F. 2008. Human factors special issue. *Journal of Loss Prevention in the Process Industries,* 21**,** 225-226.

KIRWAN, B. 2017. *A guide to practical human reliability assessment*, CRC press.

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**,** 706-716.

KRISTIANSEN, S. 2013. *Maritime transportation: safety management and risk analysis*, Routledge.

KURT, R., ARSLAN, V., KHALID, H., COMRIE, E., BOULOUGOURIS, E. & TURAN, O. SEAHORSE procedure improvement system: development of instrument. International SEAHORSE Conference on Maritime Safety and Human Factors, 2016a. 1-8.

KURT, R., ARSLAN, V., TURAN, O., DE WOLFF, L., WOOD, B., ARSLAN, O., KECECI, T., WINKELMAN, J., VAN WIJNGAARDEN, M. & PAPADAKIS, G. 2015. SEAHORSE project: Dealing with maritime workarounds and developing smarter procedures.

KURT, R. E., KHALID, H., TURAN, O., HOUBEN, M., BOS, J. & HELVACIOGLU, I. H. 2016b. Towards human-oriented norms: Considering the effects of noise exposure on board ships. *Ocean Engineering,* 120**,** 101-107.

LEE, S. M., HA, J. S. & SEONG, P. H. 2011. CREAM-based communication error analysis method (CEAM) for nuclear power plant operators' communication. *Journal of Loss Prevention in the Process Industries,* 24**,** 90-97.

MARSEGUERRA, M., ZIO, E. & LIBRIZZI, M. 2007. Human Reliability Analysis by Fuzzy "CREAM". *Risk Analysis,* 27**,** 137-154.

ÖLÇER, A. & ODABAŞI, A. 2005. A new fuzzy multiple attributive group decision making methodology and its application to propulsion/manoeuvring system selection problem. *European Journal of Operational Research,* 166**,** 93-114.

SALMON, P. M., CORNELISSEN, M. & TROTTER, M. J. 2012. Systems-based accident analysis methods: a comparison of Accimap, HFACS, and STAMP. *Safety science,* 50**,** 1158-1170.

SCHRÖDER-HINRICHS, J. U., BALDAUF, M. & GHIRXI, K. T. 2011. Accident investigation reporting deficiencies related to organizational factors in machinery space fires and explosions. *Accident Analysis & Prevention,* 43**,** 1187-1196.

SHAPPELL, S. A. & WIEGMANN, D. A. 2000. The human factors analysis and classification system--HFACS.

SHIRALI, G. A., HOSSEINZADEH, T. & KALHORI, S. R. N. 2019. Modifying a method for human reliability assessment based on CREAM-BN: A case study in control room of a petrochemical plant. *MethodsX,* 6**,** 300-315.

TAKAGI, T. & SUGENO, M. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics***,** 116-132.

UNG, S.-T. 2015. A weighted CREAM model for maritime human reliability analysis. *Safety science,* 72**,** 144-152.

UNG, S.-T. 2019. Evaluation of human error contribution to oil tanker collision using fault tree analysis and modified fuzzy Bayesian Network based CREAM. *Ocean Engineering,* 179**,** 159-172.

UNG, S. T. & SHEN, W. M. 2011. A novel human error probability assessment using fuzzy modeling. *Risk Analysis: An International Journal,* 31**,** 745-757.

VAGIAS, N. 2010. A bayesian network application for the prediction of human fatigue in the marine industry. *Unpublished dissertation, National Technical University of Athens*.

WENG, J. & YANG, D. 2015. Investigation of shipping accident injury severity and mortality. *Accident Analysis & Prevention,* 76**,** 92-101.

WU, B., YAN, X., WANG, Y. & SOARES, C. G. 2017. An evidential reasoning-based CREAM to human reliability analysis in maritime accident process. *Risk analysis,* 37**,** 1936-1957.

WU, B., YIP, T. L., YAN, X. & SOARES, C. G. 2019. Fuzzy logic based approach for ship-bridge collision alert system. *Ocean Engineering,* 187**,** 106152.

XI, Y., YANG, Z., FANG, Q., CHEN, W. & WANG, J. 2017. A new hybrid approach to human error probability quantification–applications in maritime operations. *Ocean Engineering,* 138**,** 45-54.

YANG, Z., ABUJAAFAR, K. M., QU, Z., WANG, J., NAZIR, S. & WAN, C. 2019. Use of evidential reasoning for eliciting bayesian subjective probabilities in human reliability analysis: A maritime case. *Ocean Engineering,* 186**,** 106095.

YANG, Z., BONSALL, S., WALL, A., WANG, J. & USMAN, M. 2013. A modified CREAM to human reliability quantification in marine engineering. *Ocean engineering,* 58**,** 293-303.

ZHANG, M., ZHANG, D., GOERLANDT, F., YAN, X. & KUJALA, P. 2019. Use of HFACS and fault tree model for collision risk factors analysis of icebreaker assistance in ice-covered waters. *Safety science,* 111**,** 128-143.

ZHOU, Q., WONG, Y. D., LOH, H. S. & YUEN, K. F. 2018. A fuzzy and Bayesian network CREAM model for human reliability analysis–The case of tanker shipping. *Safety science,* 105**,** 149-157.

ZHOU, Q., WONG, Y. D., XU, H., VAN THAI, V., LOH, H. S. & YUEN, K. F. 2017a. An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. *Safety science,* 95**,** 140-147.

ZHOU, T., WU, C., ZHANG, J. & ZHANG, D. 2017b. Incorporating CREAM and MCS into fault tree analysis of LNG carrier spill accidents. *Safety science,* 96**,** 183-191.

***Declaration of Interest Statement**

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

~~☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:~~

CRediT author statement

Sung Il Ahn: Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Visualization

Rafet Emek Kurt*: Conceptualization, Validation, Resources, Writing - Review & Editing, Supervision, Project administration

*Corresponding Author