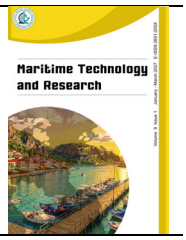
Maritime Technology and Research

https://so04.tci-thaijo.org/index.php/MTR



Research Article

A context-aware risk assessment framework for critical shipboard operations: Application to bunkering

Kadir Cicek[1,2,*], Metin Celik[2,3], Selcuk Cebi[2,4] and Gülşah Şahin[4]

[1]*Department of Marine Engineering, Istanbul Technical University, Istanbul, Turkiye*
[2]*Industrial Data Analytics and Decision Support Systems Center, Azerbaijan State University of Economics, Baku, Azerbaijan*
[3]*Department of Basic Science, Istanbul Technical University, Istanbul, Turkiye*
[4]*Department of Industrial Engineering, Yildiz Technical University, Istanbul, Turkiye*

[*]*Corresponding author's e-mail address: cicekk@itu.edu.tr*

Article information
Received: June 27, 2026
Revision: October 1, 2026
Accepted: October 10, 2026

Keywords
Bunkering;
Critical ship operations;
Maritime safety;
Operational context;
Risk assessment

Abstract

Critical shipboard operations such as bunkering, cargo handling, mooring, and enclosed-space entry are carried out under conditions that differ from one execution to the other, yet the risk score recorded for a hazard remains fixed for a set of procedures. This study develops a context-aware risk assessment framework in which the assessed severity of a hazard is conditioned on the circumstances of its execution. The framework is generic and integrates Delphi-based hazard identification, Safety and Critical Effect Analysis with fuzzy inference, and the Bayesian Best-Worst Method for weighting contextual factors, whereas the hazard inventory and the contextual dimensions are established for the operation under assessment. Bunkering is used as the demonstration case. Under adverse conditions the score of an unchanged hazard rose from 80 to 201, and the resulting ranking correlated strongly with conventional expert-based rankings, confirming that operational context can be quantified without discarding established practice.

1. Introduction

Maritime transport remains the backbone of global trade, carrying more than 80 % of world trade by volume. However, the sector now operates in a demanding risk environment shaped by supply-chain disruptions, chokepoint vulnerability, decarbonization pressure, tighter port turnaround expectations, and the need for resilient operations (United Nations Conference on Trade and Development, 2024). These pressures are reflected onboard ships, where critical operations are often performed under time pressure and variable environmental or organizational conditions. Accordingly, shipboard risk can no longer be treated solely as a static property of technical hazards; it must also be understood as a function of operational context.

Shipboard operations are treated as critical. In this study, these operations, which are performed repeatedly, involve the simultaneous participation of several parties or systems, and carry the potential for consequences that extend beyond the task itself, are considered. Bunkering, cargo handling and transfer operations, mooring and anchoring, enclosed-space entry, pilot transfer, navigation in restricted waters, ship-to-ship operations, hot work, and dry-docking preparation all fall within this class. Bunkering is used as the demonstration case, rather than as the limit of the scope, for three reasons. It is among the most frequently repeated critical operations on board; it is performed

on every category of vessel; and its risk profile will become considerably more demanding as alternative fuels enter general service. It is a routine but safety-critical activity involving fuel transfer, ship-shore or ship-ship coordination, manifold integrity, communication discipline, and strict pollution prevention. Although long-term tanker spill frequencies have declined, accidental spills have not disappeared. The International Tanker Owners Pollution Federation Limited (ITOPF) recorded 10 spills above 7 tons in 2024 and maintains records covering more than 10,000 tanker spills that have occurred since 1970 (ITOPF, 2025). Since bunkering failures have an ability to cause pollution, fire, operational disruption, legal liability, and economic loss, safety assessment should move beyond procedural compliance and reflect the conditions under which the operation is performed.

This requirement becomes more critical with the transition to alternative marine fuels. The *International Code of Safety for Ships Using Gases or Other Low-flashpoint Fuels* provides a goal-based safety framework for ships using gas and low-flashpoint fuels (International Maritime Organization, 2026). Recent technical literature also shows that alternative-fuel bunkering introduces fuel-specific hazards requiring detailed operational control. Ammonia bunkering requires dedicated risk assessment, gas dispersion analysis, trained personnel, and explicit treatment of toxicity, simultaneous operations (SIMOPS), and safety zones (American Bureau of Shipping, 2024a). Methanol bunkering similarly requires operation-specific risk assessment, formal training, and control of fire, explosion, toxicity, and vapor-related hazards (American Bureau of Shipping, 2024b). Recent ammonia dispersion research further indicates that release behavior may affect hazardous-zone formation under varying environmental conditions (Skarsvag et al., 2024). Thus, the fuel transition reinforces that bunkering risk is conditional on context, not merely on baseline hazard identity.

Recent maritime risk research has moved toward integrated and dynamic modeling strategies. Hybrid methods have been used to evaluate tanker ballasting/de-ballasting risks and cargo manifold failures under uncertainty (Elidolu et al., 2023; Sezer et al., 2023). At the system level, dynamic Bayesian modeling and the joint treatment of human, environmental, and technological factors have been emphasized for capturing evolving operational conditions in maritime decision-making (Tao et al., 2024; Wang et al., 2025). In the alternative-fuel domain, recent studies and technical guidance likewise show that ammonia-fuel adoption, methanol bunkering, and hydrogen bunkering require risk assessment structures that account for fuel-specific hazards, operational interfaces, and emergency preparedness (American Bureau of Shipping, 2024a, 2024b; Jang et al., 2023; Maritime Technologies Forum, 2024).

Despite these advances, a practical gap remains. Existing models identify and rank hazards effectively, but they are less explicit about how the severity of the same hazard changes across the settings in which it is encountered. A hazard that appears moderate under favorable conditions may become critical during night time operations, with limited crew experience, or aboard a vessel in suboptimal condition. This study is therefore guided by three questions. The first asks through which mechanisms operational conditions alter the severity of an unchanged hazard. The second asks how those conditions can be quantified and incorporated into a risk score without discarding the analytical structure that shipboard practice already relies upon. The third asks whether the resulting framework produces results that are consistent with established risk-ranking practice and stable under the uncertainty inherent in expert judgment. To answer these questions, the study proposes a context-aware risk assessment framework for critical shipboard operations and demonstrates it through bunkering, integrating Delphi-based hazard validation, Fuzzy Inference System (FIS)-based baseline risk estimation, Bayesian Best-Worst Method (BBWM)-based weighting of contextual factors, and scenario-based multiplier logic. The principal contribution is to link inherent hazard magnitude with operational context, which supports a more realistic and decision-oriented form of maritime risk assessment. The framework consists of two layers. The analytical procedure is generic and applies unchanged to any critical shipboard operation, whereas the hazard inventory and the contextual

dimensions on which that procedure operates are established for the operation being assessed. Transferring the framework to another operation therefore requires new content rather than a new method. The remainder of the paper is organized as follows. Section 2 establishes the regulatory basis of shipboard risk assessment and reviews the analytical approaches developed in the literature together with their limitations. Section 3 presents the four phases of the proposed methodology. Section 4 applies the framework to bunkering, reports the results obtained under three operational scenarios, validates the model against conventional rankings derived from expert judgment, and discusses the implications of the findings. Section 5 concludes with the principal findings and the directions in which the framework can be extended.

2. Literature review

2.1 Risk assessment in shipboard operations

Risk assessment on board merchant ships is a regulatory requirement rather than an optional practice. To identify the limitations of existing approaches, it is therefore necessary first to understand how this requirement is implemented in shipboard operations. The International Safety Management (ISM) Code at paragraph 1.2.2.2 requires that companies assess all identified risks to their ships, personnel, and the environment and establish appropriate safeguards, with this requirement being implemented through the Safety Management System (SMS) of each vessel. At the regulatory level, the International Maritime Organization has established Formal Safety Assessment (FSA) as a structured methodology that combines risk analysis with cost benefit assessment for use in rule-making (International Maritime Organization, 2018). This methodology has subsequently influenced both the terminology and the analytical logic adopted in shipboard risk management. Within operational practice, risk assessment is generally applied in three forms. Generic assessments are prepared in advance for routine operations and incorporated into the Safety Management System as standing documents. Task-specific assessments are conducted before non-routine activities, operations that depart from established procedures, and work governed by permit-to-work systems, such as enclosed-space entry and hot work. Dynamic assessments are undertaken by the officer in charge immediately before and, where necessary, during an operation, commonly through a toolbox talk, to account for the conditions prevailing at that time. Although these applications differ in timing, scope, and formality, they are based on a broadly common analytical procedure, the characteristics of which determine the extent to which current practice can represent operational risk.

The procedure itself is broadly consistent across these applications. Hazards associated with the operation are identified, after which the likelihood of occurrence and the severity of potential consequences are estimated, most commonly using ordinal scales. These estimates are then combined, typically through multiplication, to produce a risk score. The resulting scores are compared with predefined acceptance criteria, and additional control measures are specified where the assessed risk exceeds the acceptable level. Following the implementation of these measures, the remaining risk is reassessed and recorded as residual risk, while the completed assessment is subject to periodic review and reconsideration following an incident or a material change in circumstances. Two characteristics of this procedure are particularly relevant to the present study. The first concerns its temporal structure. Although dynamic checks may be performed immediately before or during an operation, the formally recorded assessment generally remains based on conditions anticipated during planning and is not continuously recalculated as the operation progresses. The second concerns the hazard-centered nature of the scoring process. The assigned score is primarily linked to the hazard itself and does not systematically vary according to the time of execution, the experience of the personnel involved, or the condition of the vessel. In practice, however, shipboard professionals recognize that these factors can materially influence the complexity and risk of an otherwise identical task. As a result, the risk documented in the SMS may not fully correspond to the risk encountered under actual operating conditions. These limitations define the central problem addressed in this study. Two objections may reasonably be raised against this position, and both merit a direct answer. The first is

that the competence of personnel is already regulated through the certification regime established by the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, so that any officer assigned to a task holds a certificate attesting to the required competence. The second is that the condition of the vessel is already verified through the survey and classification regime, so that a vessel in service has been confirmed to meet the applicable structural and machinery standards. Neither objection is disputed here. Certification and classification establish that a minimum standard has been met, and the present study neither questions that function nor proposes any substitute for it. What these regimes establish, however, is a threshold rather than a distribution. Two officers holding the same certificate may differ considerably in the number of times they have performed a particular operation, and two vessels holding valid certificates may differ considerably in age, in maintenance history, and in the condition of the specific equipment on which an operation depends. Shipboard practice already acknowledges this distinction. Tasks are assigned on the basis of experience rather than availability, and an officer who holds the required certificate and stands a watch has performed the operation before. The framework does not question this. What this represents is the depth of that experience and the degree of familiarity with the particular vessel. Certification establishes a minimum standard expressed in months of sea service, whereas the number of times an officer has carried out a given operation, and the period that officer has served on that specific ship, lie outside it. The ISM Code recognizes the same distinction at paragraph 6.3, which requires documented familiarization for personnel newly assigned to a ship. This judgment was elicited rather than assumed. Four of the five experts consulted in this study, all of whom had served in management-level positions at sea, identified the 0 - 2 year category as the most influential risk-amplifying condition and assigned it weights between 0.577 and 0.619. The same reasoning applies to a vessel that is fully compliant yet is older and more heavily worked than another. Neither condition can be changed on the day, and the framework therefore records them so that the assessed risk rises and indicates where supervision, briefing, or additional monitoring should be concentrated.

2.2 Analytical approaches in maritime risk research

The limitations of conventional shipboard risk assessment have encouraged the development and application of more formal analytical methods in maritime risk research. These methods seek to improve the representation of uncertainty, causal relationships, operational variability, and expert judgment within the assessment process. **Table 1** provides an overview of the principal approaches identified in the literature and highlights their main limitations. The following discussion then examines the relevant studies in greater detail.

Building on the methodological approaches summarized in **Table 1**, recent maritime risk research has increasingly adopted hybrid models that combine fuzzy logic, fault tree analysis, FMEA, Bayesian networks, rule-based systems, and evidential reasoning. The relevant literature can broadly be examined in terms of operational risk assessments and system-level analyses, although considerable variation exists in both the systems investigated and the analytical methods employed.

Within the operational risk literature, Gülen et al. (2026) integrated FMECA, evidential reasoning, and a rule-based Bayesian network to evaluate enclosed-space operations. Their analysis identified twenty-three failure modes and indicated that procedural non-compliance and inadequate atmospheric testing were among the most critical risk contributors. Sezer and Akyuz (2024a) examined cargo-tank fire and explosion risks on chemical tankers using an evidential-reasoning-based SLIM method and a Bayesian belief network. The findings highlighted the influence of human performance-shaping factors, particularly time pressure, on ignition risk during inert-gas purging. Sezer et al. (2024c) applied a fuzzy CREAM–Bayesian network framework to cargo oil pump operations and reported that conventional error-probability tables may not adequately represent cognitive failure probabilities. Similarly, Sezer et al. (2023) assessed cargo manifold operations through a hybrid framework combining FMECA, Dempster-Shafer theory, and Bayesian networks,

enabling the prioritization of specific failure modes, including valve misalignment, and the identification of corresponding preventive measures.

Table 1 Established approaches to maritime risk assessment and their principal limitations.

Approach	Analytical principle	Typical application	Principal limitation
Qualitative risk matrix	Ordinal likelihood and severity ratings combined multiplicatively	Shipboard SMS assessments and permit-to-work systems	Provides limited resolution, allows different likelihood-severity combinations to produce identical scores, and does not explicitly represent changing operational conditions
Failure mode and effects analysis (FMEA) / Failure mode, effects, and criticality analysis (FMECA)	Systematic identification and prioritization of failure modes, commonly using a risk priority number	Machinery systems, cargo operations and transfer processes	A low rating on one parameter may offset a high rating on another, while dependencies among failure modes are not explicitly represented
Fault and event tree analysis	Deductive or inductive decomposition of accident causes and consequences	Fire, explosion, collision, and system failure scenarios	Conventional Boolean logic has limited capacity to represent conditional dependence, partial failure, and dynamically changing conditions, while basic-event probabilities may be difficult to estimate
Bow-tie analysis	Integration of causal pathways, preventive barriers, consequence pathways, and mitigation barriers around a central event	Barrier management and major-accident hazards	Conventional applications are predominantly qualitative, and barrier performance is often represented without detailed probabilistic assessment
Bayesian networks	Probabilistic inference through dependencies represented in a directed acyclic graph	Collision, blackout, cyber-risk, and navigational safety assessment	Conditional probability tables are data-intensive and may depend heavily on expert judgment when empirical data are limited
Fuzzy inference systems	Rule-based transformation of linguistic inputs into a defuzzified risk output	Risk assessment under uncertainty and imprecise information	The number of rules may increase rapidly as the number of input variables grows, while the output is constrained by the structure and completeness of the rule base
Multi-criteria weighting methods	Pairwise or comparative evaluation used to derive criterion weights and priorities	Prioritization of hazards, risk factors, and control measures	The resulting weights reflect the judgments of a specific expert panel and are not always subjected to sensitivity or robustness analysis
Human reliability analysis	Estimation of human error probabilities using task characteristics and performance-shaping factors	Cargo operations, navigation, maintenance, and enclosed-space entry	The analysis may emphasize individual operator performance without fully representing the dynamic interaction among personnel, equipment, vessel condition, and operational context

Mooring and anchoring operations have received considerable attention in recent maritime risk research, particularly through the application of hybrid probabilistic and evidential methods. Tuncel et al. (2024) applied FMECA and Evidential Reasoning within a rule-based Bayesian network to anchoring operations, finding that rudder and thruster failures carry the highest aggregate Risk Priority Number and that strong currents in restricted areas act as systemic amplifiers. Demirel et al. (2024) assessed mooring risks at hydrocarbon berths using a fuzzy Best-Worst Method with a rule-based Bayesian network, concluding that organizational factors such as shift handover failures collectively outweigh individual mechanical faults. Tekeli et al. (2026) introduced a Z-number Bayesian network for mooring-line failure prediction during cargo operations, calculating a baseline failure probability of 0.015 that rises substantially under high sea-state and degraded line conditions. For Ship to Ship (STS) LNG bunkering, Feng et al. (2025) integrated fuzzy FMEA with TOPSIS to rank risk scenarios, identifying hose-connection failure and vapor cloud ignition as the highest priorities, while Choi et al. (2025) confirmed crew coordination failures and inadequate emergency drills as the dominant root level precursors across STS operations.

In the ship recycling and hazardous-material domain, Sezer et al. (2024b) examined risks across the ship dismantling lifecycle using FMECA and Dempster Shafer evidential theory, identifying heavy metal residues and toxic paint-stripping agents as critical hazards. Similarly, Khan et al. (2024) used Bayesian networks to assess hazardous cargo risks in seaport infrastructure, showing that cargo properties, berth limitations, and emergency response capability jointly influence accident probability.

In recent years, system-based maritime risk research has increasingly focused on Maritime Autonomous Surface Ships (MASS), navigation in harsh environments, and the reliability of safety-critical shipboard systems. Tao et al. (2024) systematically reviewed MASS risk factors, classifying them into human, environmental, and technological categories and identifying object detection failure and cyber interference as the most frequently cited high consequence modes. Li et al. (2024) translated this classification into a quantitative FTA Fuzzy Bayesian Network collision model for MASS, capturing conditional failure probabilities for sensor-fusion failures that Boolean FTA cannot express. Laakso et al. (2025) combined Systems Theoretic Process Analysis (STPA) with a Bayesian network for MASS autonomous navigation systems, capturing socio-technical hazards arising from shore operator and onboard algorithm interactions. Complementing these quantitative approaches, Na et al. (2025) proposed a qualitative risk assessment methodology for MASS based on cognitive model-driven functional analysis, demonstrating that structured hazard identification remains essential where probabilistic data are scarce. In Arctic navigation, Liu et al. (2024) developed a Dynamic Bayesian Network for the Northeast Passage across seven-time steps, finding that ice concentration and vessel icebreaking class determine total-loss probability far more strongly than crew experience. Wang et al. (2025) addressed long haul navigation risk in the North Pacific through an improved dynamic Bayesian network integrating hesitant cloud models, capturing temporal changes in environmental threats across seasonal variability.

At the sub-system level, Goksu et al. (2023) used a fuzzy Bayesian network to show that loose electrical connections and hydraulic seal degradation dominate steering gear failure risk and that current maintenance intervals are insufficient. Karanovic et al. (2024) applied fuzzy FMEA to a four-ram hydraulic steering system, quantifying seventeen failure modes to produce a priority maintenance schedule. Goksu et al. (2025) integrated FTA with Bayesian network inference for ship blackout scenarios, revealing that generator switchboard faults and fuel-supply interruptions interact non-independently, with their joint probability exceeding the product of their marginal probabilities by a factor of 2.7. Aydin et al. (2024) applied a holistic fuzzy Bayesian network to bulk carrier cargo hold fires, finding that human error during hot work permitting contributes more to initiating event probability than electrical faults. In the cyber dimension, Uflaz et al. (2024) combined FMECA, Dempster Shafer theory, and a rule-based Bayesian network to quantify maritime cyber-attack risks, identifying ECDIS manipulation and AIS spoofing as the highest-consequence attack vectors with

posterior failure probabilities approximately three times those of GPS jamming. Uğurlu et al. (2026) extended this line of inquiry using a Bayesian network with conditional probability tables to derive optimal redundancy architectures for ship navigational and communication systems under adversarial conditions. Regarding alternative fuels, Jang et al. (2023) identified critical IMO regulatory gaps for ammonia fueled ships, including deficiencies in fuel tank safety standards and crew training; subsequent studies have applied FSA-Bayesian network (Lankahaluge et al., 2025) and Noisy-OR gate Bayesian network approaches (Li et al., 2025) to ammonia fuel-storage and hydrogen fuel-system leakage risks aboard ships.

2.3 Research gap

Taken together, the studies reviewed above indicate a growing trend towards combining fuzzy logic, evidential reasoning, FMECA, and Bayesian networks within hybrid frameworks that can account for uncertainty in the available information and variations in operating conditions across both operational and system-level maritime risk assessments. These approaches have considerably advanced the identification of failure mechanisms, the representation of causal relationships, and the prioritization of preventive measures. Nevertheless, as summarized in **Table 1**, their principal limitations relate to analytical resolution, data requirements, model complexity, and the assumptions embedded in their internal structure. Comparatively less attention has been given to how the assessed risk changes according to the specific circumstances under which an operation is performed. In most of the models reviewed, the risk parameters assigned to a hazard are established during the assessment process and subsequently remain fixed, thereby reproducing the static characteristics of the conventional procedure discussed in Section 2.1. Consequently, an identified hazard may retain the same recorded score whether a bunkering operation is conducted during daylight by an experienced crew on a relatively new vessel or at night by less experienced personnel on an older vessel. Although these circumstances may materially influence the risk encountered during execution, existing models do not provide a simple and systematic mechanism for incorporating them into the final score. A further limitation concerns the treatment of the expert input on which many of these models depend, since the weights and probability estimates elicited from panels are frequently adopted without any examination of whether the resulting rankings are robust to the procedure used to aggregate them. A related limitation concerns operational usability, as many advanced models require extensive data, specialized expertise, or complex calculations that may restrict their routine application on board. The present study addresses these limitations by developing a framework in which the assessed risk is adjusted according to the conditions surrounding the execution of the operation, while the recurring requirement on board is limited to identifying those conditions rather than repeating the full analytical process.

3. Methodology

This section presents a new methodology including multi-stage decision support system to assess risks in ship operations under uncertainty. The proposed methodology integrates Safety and Critical Effect Analysis (SCEA), Bayesian Best-Worst Method (BBWM), and Fuzzy Inference System (FIS) to provide comprehensive performance assessment that considers uncertainty. In the first phase, an expert team identifies risks in critical ship operations. Then, the algorithm performs SCEA and FIS based algorithm to incorporate expert knowledge and evaluate subjective importance of the assessment criteria. Subsequently, the BBWM is applied to analyze the effects of varying operational conditions on risk outcomes. This approach enables the prediction of potential risks and their magnitudes, allowing ship personnel to implement appropriate preventive measures. The structure of the proposed method is depicted in **Figure 1**. The three methods are applied in sequence rather than as alternatives to one another, and each addresses a different stage of the assessment. SCEA supplies the parameter structure through which every hazard is characterized, namely probability, severity, frequency, and detectability. FIS converts those parameters into a baseline risk

magnitude by means of a rule base, which prevents a low value on one parameter from suppressing an extreme value on another. BBWM then determines the weights of the contextual dimensions that adjust the baseline magnitude, using a probabilistic hierarchical model that preserves the variability among individual expert judgments. SCEA and FIS, therefore, operate in Phase 2, and BBWM operates in Phase 3. The abbreviations used throughout the paper are listed in **Appendix Table 2**, and the mathematical symbols and variables are defined in **Appendix Table 4**.

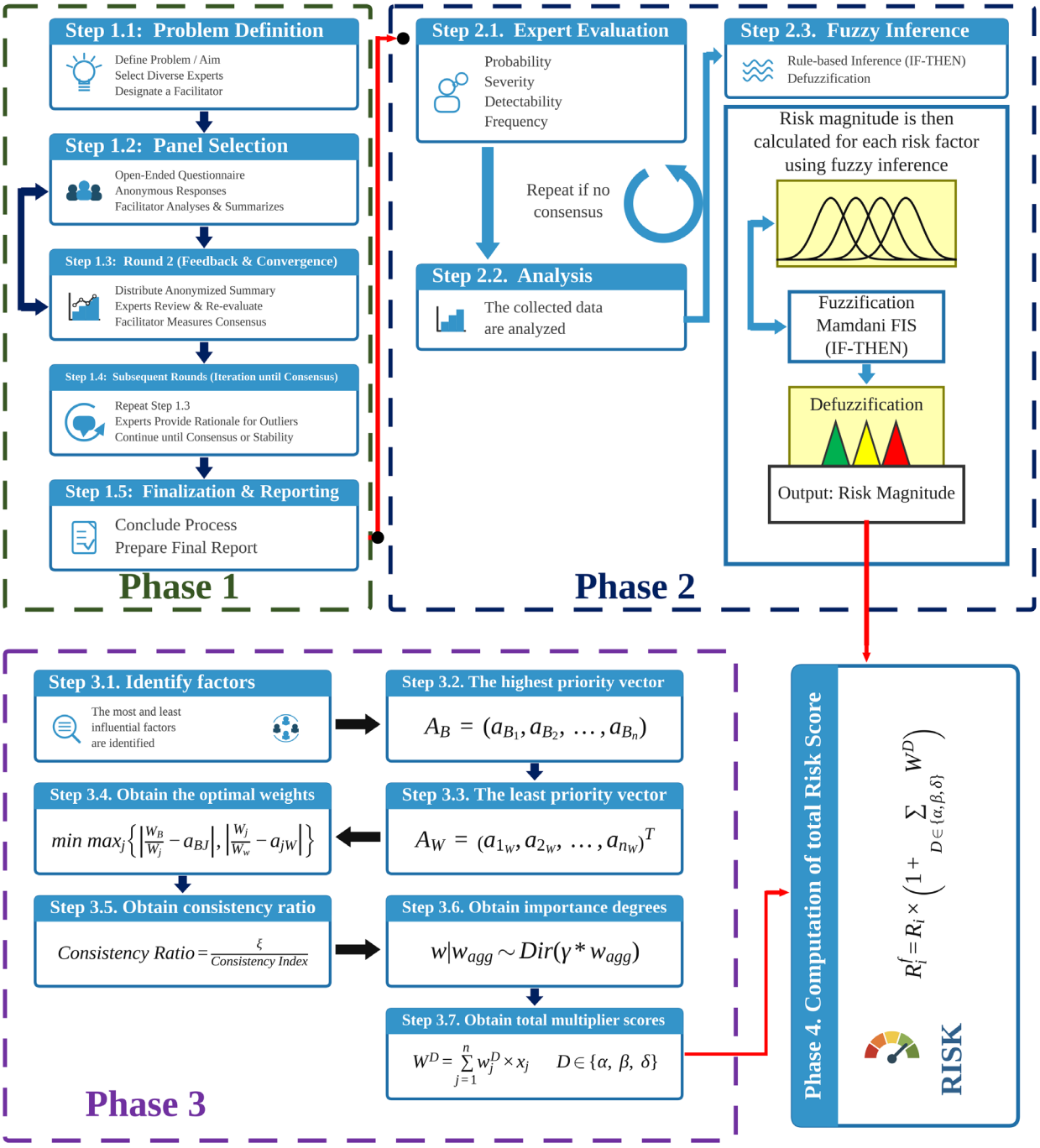


Figure 1 The main structure of the proposed methodology.

The framework distinguishes between its analytical procedure and the content on which that procedure operates, and this distinction determines how it may be transferred to operations other than the one examined here. The four phases, the steps they contain, the parameter set of SCEA, the

structure of the fuzzy rule base, the elicitation and aggregation procedure of BBWM, and the computational forms given in Eqs. (1) to (12) are generic. They are independent of the operation being assessed and are applied without modification. Three elements are specific to the operation. The first is the hazard inventory constructed in Phase 1, which must describe the failures that can occur during that particular operation. The second is the set of contextual dimensions and sub-factors established in Phase 3, since the conditions that amplify risk during bunkering are not necessarily those that amplify risk during mooring or enclosed-space entry. The third is the composition of the expert panel, whose members must have direct operational experience of the operation concerned. Applying the framework to a different operation therefore requires these three elements to be re-established, after which the procedure is executed unchanged. **Appendix Table 5** summarizes this distinction step by step.

Phase 1. Construction of the factor hierarchy: In this phase, all of the parameters that generate and influence risks in ship operations are defined and categorized, and their interrelationships visualized. The identification of risk factors was conducted using Delphi method (Gordon et al., 1969; Dalkey, 1969). The steps for this phase are given below:

Step 1.1 A panel of experts with in-depth knowledge and direct operational experience of the activity under assessment is established. Because the validity of a Delphi study rests on the qualification of the panel rather than on statistical representativeness, panelists are selected against explicit eligibility criteria rather than by random sampling (Okoli & Pawlowski, 2004). Three criteria are applied in the proposed methodology: (i) documented seagoing service in a management-level position on board vessels that routinely perform the operation under assessment, (ii) direct operational responsibility for that operation, whether in its planning or execution, and (iii) working familiarity with the applicable safety management and regulatory framework. The size of the panel follows from the homogeneity of the expert group and the specificity of the assessment task rather than from sampling considerations. Rowe and Wright (2001) suggest that Delphi panels may include between five and twenty participants. Smaller panels are generally appropriate when the experts have similar professional backgrounds and the evaluation focuses on a single, clearly defined operation. A larger panel may be needed when the participants represent diverse areas of expertise or when the scope of the assessment is broader. However, the adequacy of the panel should not be determined solely by the number of participants. It should also be evaluated based on the level of consensus achieved during the Delphi rounds and the consistency of the experts' individual pairwise comparisons. Therefore, the selected panel may be considered sufficient when the stopping criterion is met and the consistency ratios remain within acceptable limits.

Step 1.2 A questionnaire, which consists of open-ended questions, is distributed to the experts. The main aim is to gather all possible ideas and perspectives related to ship operation risks. The questionnaire can be administered electronically via an online form (Round 1).

Step 1.3 The collected expert opinions are analyzed by the research team to identify common themes and divergences. Irrelevant or duplicate information are eliminated, and a detail list of potential risks is compiled. Then, after results are analyzed and summarized, it is sent back to the experts. Experts are asked to rate or revise their initial opinions based on the aggregated feedback (Round 2).

Step 1.4 The revised expert evaluations are analyzed again by the research team, and the results are shared with expert team. The process of survey distribution, analysis, and feedback continues until a consensus is achieved.

Step 1.5 Once a majority consensus among the experts is obtained, the final opinions are classified by the research team, and a hierarchical structure of risk factors is established.

Phase 2. Calculation of risk magnitudes using the safety and critical effect analysis (SCEA) method: In this phase, the risk magnitudes for the factors determined in Phase 1 are calculated using the Safety and Critical Effect Analysis (SCEA) method. SCEA provides a holistic multi-parameter evaluation of risks within the scope of occupational health and safety. It has been proposed to address the limitation of conventional risk analysis approaches that often fail to consider the combined effects of multiple parameters (Karasan et al., 2018). Unlike traditional methods, SCEA simultaneously evaluates probability (P), severity (S), frequency (F), and detectability (D) parameters. Expert opinions for each of these parameters are collected as in Phase 1.

- Probability (P) represents the likelihood of a risk occurring.
- Severity (S) reflects the potential impact if the risk occurs.
- Frequency (F) indicates how often the risk-triggering condition may repeat.
- Detectability (D) measures the likelihood of identifying the risk before it occurs.

Step 2.1 Experts are asked to rate the probability, severity, frequency, and detectability of each risk using the linguistic scale shown in **Table 2** (Karasan et al., 2018). The abbreviations of the linguistic terms used in this scale are given in **Appendix Table 3**.

Table 2 Linguistic scale for SCEA.

Probability (P)	**Severity (S)**	**Frequency (F)**	**Detectability (D)**	**Value**
Very Low (VL)	Trivial (T)	Very Rare (VR)	Absolutely Possible (AP)	1
Low (L)	Significant (S)	Rare (R)	Strongly Possible (SP)	3
Medium (M)	Moderate (Mo)	Occasional (O)	Possible (P)	5
High (H)	High (H)	Regular (Re)	Impossible (I)	7
Very High (VH)	Catastrophic (C)	Permanent (Pe)	Strongly Impossible (SI)	9

Step 2.2 The collected results are analyzed and sent back to the experts for review and possible revision. The process is repeated until the change in median value is less than 1, which indicates consensus. For each risk, the consensus values of probability, severity, frequency, and detectability are expressed as fuzzy numbers given by Eqs. (1) - (4):

$$\tilde{P}_i = \left(Q_{i1}^p, M_i^p, Q_{i3}^p\right) \tag{1}$$

$$\tilde{S}_i = \left(Q_{i1}^s, M_i^s, Q_{i3}^s\right) \tag{2}$$

$$\tilde{F}_i = \left(Q_{i1}^f, M_i^f, Q_{i3}^f\right) \tag{3}$$

$$\tilde{D}_i = \left(Q_{i1}^d, M_i^d, Q_{i3}^d\right) \tag{4}$$

Step 2.3 Risk magnitude is then calculated using fuzzy inference. Fuzzy Inference Systems (FIS) map inputs to outputs through fuzzy logic, and the Mamdani FIS is one of the most widely used models. Inputs can be crisp or fuzzy, while the outputs are fuzzy sets that are later defuzzified to obtain crisp values (Mamdani, 1974; Mamdani & Gaines, 1981). The rules take the form of "If ... then ..." statements. In this study, the fuzzy values obtained in Step 2.2 are used as inputs to the FIS. A rule base linking the input parameters to risk magnitudes is shown in **Table 3** (Karasan et al., 2018). An example of the rule base is given below:

*IF Frequency (**F**) is very high (**VH**) AND Severity (**S**) is very high (**VH**) AND Detectability (**D**) is very high (**VH**) AND Probability (**P**) is very high (**VH**) THEN the risk magnitude is Catastrophic (**Cr**).*

Table 3 Fuzzy rule base for SCEA.

		D																									
		VL					L					M					H					VH					
		P					**P**					**P**					**P**					**P**					
F	**S**	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH	
VL	VL	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	
	L	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	
	M	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mn	
	H	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mj	
	VH	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	
L	VL	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mn	
	L	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mj	
	M	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mj	Mj	Mj	
	H	Ng	Ng	Ng	Ng	Mn	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mj	Mj	Mj	Mj	
	VH	Ng	Ng	Ng	Mn	Mn	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Mj	Cr	
M	VL	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	
	L	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Mn	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	
	M	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	
	H	Ng	Ng	Ng	Mn	Mn	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Cr	Mn	Mj	Mj	Mj	Cr	
	VH	Ng	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mj	Mj	Cr	Mn	Mj	Mj	Cr	Cr	Mj	Mj	Mj	Cr	Cr	
H	VL	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	
	L	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	
	M	Ng	Ng	Ng	Mn	Mn	Ng	Mn	Mn	Mn	Mn	Ng	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Mj	Cr	
	H	Ng	Ng	Mn	Mn	Mn	Mn	Mn	Mn	Mn	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Mj	Cr	Mj	Mj	Mj	Cr	Cr	
	VH	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mj	Mj	Mj	Cr	Mj	Mj	Mj	Cr	Cr	Mj	Mj	Cr	Cr	Cr	
VH	VL	Ng	Ng	Ng	Ng	Mn	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mj	Mn	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Cr	
	L	Ng	Ng	Ng	Mn	Mn	Ng	Ng	Mn	Mn	Mj	Ng	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Cr	Cr	
	M	Ng	Ng	Mn	Mn	Mn	Ng	Mn	Mn	Mj	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Mj	Cr	Mj	Mj	Cr	Cr	Cr	
	H	Ng	Mn	Mn	Mn	Mj	Mn	Mn	Mj	Mj	Mj	Mn	Mj	Mj	Mj	Cr	Mj	Mj	Mj	Cr	Cr	Mj	Cr	Cr	Cr	Cr	
	VH	Mn	Mn	Mn	Mj	Mj	Mn	Mj	Mj	Mj	Cr	Mj	Mj	Mj	Cr	Cr	Mj	Mj	Cr	Cr	Cr	Cr	Cr	Cr	Cr	Cr	

The linguistic scale for risk magnitude $\left(\tilde{R}^{k}_{t\in\{Ng,Mn,Mj,Cr\}}\right)$ is defined as *Negligible* (Ng), *Minor* (Mn), *Major* (Mj), and *Catastrophic* (Cr). The membership degree of each output is determined by taking the minimum value among the membership degrees of the four parameters, as given by Eq. (5):

$$\mu_{R^{k}_{t\in\{Ng,Mn,Mj,Cr\}}}(x) = min\{\mu_p(x),\ \mu_s(x),\ \mu_f(x),\ \mu_d(x)\} \tag{5}$$

where $\mu_p(x)$, $\mu_s(x)$, $\mu_f(x)$, and $\mu_d(x)$ presents the probability, severity, frequency, and detectability, respectively. The conversion of the fuzzy set output obtained from fuzzy inference into a crisp value is achieved through defuzzification. Common defuzzification techniques include the weighted average, centroid, and maximum membership methods. Due to the symmetric structure of the output membership function, the weighted average method given by Eq. (6) and the linguistic scale for risk magnitude given in **Figure 2** are used in the proposed approach.

$$R_i = \frac{\sum_{t\in\{Ng,Mn,Mj,Cr\}} \max\left(\mu_{R_t^k}\right) \times (R_t^*)}{\sum_{t\in\{Ng,Mn,Mj,Cr\}} \max\left(\mu_{R_t^k}\right)} \tag{6}$$

where R_i is the output of the rule base for the i^{th} factor and R_t^* gives the center of the corresponding fuzzy set.

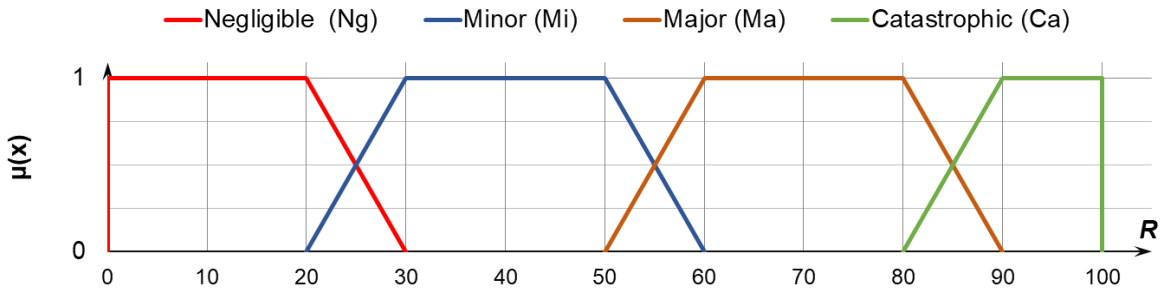


Figure 2 Linguistic scale for risk magnitude.

Phase 3. Identification of multiplier effect scenarios: In this step, scenario-based factors that increase the impact of risks in operations are identified by the expert team. In the proposed study, *environmental factors, expertise level,* and *ship condition* have been defined as the main factors that have the potential to influence the magnitude of the risk. The main and sub-factors related to the multiplier effects are detailed in **Table 4**. The three dimensions adopted here were identified by the expert panel for bunkering and are not fixed properties of the framework. The procedure through which dimensions are identified, weighted and aggregated is generic, whereas the dimensions themselves are established for the operation under assessment. An application to enclosed-space entry, for instance, would be expected to include atmospheric condition and rescue readiness among its dimensions, and an application to mooring would be expected to include sea state and line condition.

Table 4 Proposed taxonomy for multiplier risk dimensions.

Factors	Sub-factors	Code	Definition & Scope
Environmental Factors	Morning	EC1	Operational conditions during post-dawn hours (06:00 - 12:00)
	Afternoon	EC2	Peak operational and thermal exposure intervals (12:00 - 18:00)
	Evening	EC3	Transitional visibility and light degradation shift (18:00 - 22:00)
	Night	EC4	Critical visibility impairment and high-risk operation (22:00 - 06:00)
Expertise	0 - 2 years	EX1	Junior officer operational threshold representing baseline exposure
	2 - 5 years	EX2	Intermediate experience level with competent technical autonomy
	5 - 10 years	EX3	Advanced professional proficiency and seasoned tactical familiarity
	>10 years	EX4	Senior commanding officer level with long-term field mastery
Vessel Condition	0 - 12 Years	VC1	Young fleet status with optimal structural integrity
	12 - 24 Years	VC2	Mid-life vessel cohort requiring intensified technical maintenance
	>25 Years	VC3	Critical cohort prone to fatigue risks and structural wear

Contextual factors do not create new hazards; they alter the conditions under which an existing hazard develops into a consequence. Three mechanisms are distinguished in the proposed framework. First, environmental conditions affect the detectability of a developing failure and the time available to intervene: reduced visibility, fatigue associated with night watches, and the smaller number of personnel on duty outside working hours all lengthen the interval between the onset of an abnormal condition and its recognition. Second, the expertise of the personnel involved affects both the likelihood of an error being committed and the likelihood of it being recovered before it propagates, since experienced officers recognize deviations from the expected course of the operation earlier and respond to them more reliably. Third, the condition of the vessel affects the reliability of the technical barriers on which the operation depends, as aging valves, seals, gauges, and alarm circuits degrade the equipment that would otherwise contain or signal a failure. These mechanisms act on detection, on error recovery, and on barrier reliability, respectively, and they therefore modify the severity of an unchanged hazard rather than introducing additional hazards into the hierarchy.

In this phase, the Bayesian Best-Worst Method (BBWM) (Mohammadi & Rezaei, 2020) is applied to evaluate the relative importance of these main and sub factors. BBWM is an extension of the Best-Worst Method (BWM) proposed to obtain a consensus decision matrix in an environment with multiple decision-makers. The BWM, introduced by Rezaei (2015), is a method that allows pairwise comparisons to be conducted in a structured manner, requires less information, and simultaneously ensures greater consistency in the comparisons (Rezaei, 2015). BWM requires the decision-maker to identify the best (most important) and worst (least important) criteria and then perform pairwise comparisons between these and all other criteria. The optimal weights are then determined by solving a max–min optimization model, and a consistency ratio is computed to assess reliability. A consistency ratio is also calculated to check the reliability of the comparisons. The BBWM steps are summarized as follows (Rezaei, 2015, 2016). As stated at the beginning of this section, BBWM is applied only to the weighting of the contextual dimensions. The baseline risk magnitude to which those weights are subsequently applied is produced by SCEA and FIS in Phase 2, so the three methods act on different quantities and do not overlap.

Step 3.1 The most and least influential factors are identified. Experts are asked, through a web-based survey, to determine which factor has the highest and lowest influence at each hierarchical level.

Step 3.2 The highest priority vector is obtained. Using a web-based questionnaire, participants are requested to determine the most important factor relative to all other factors by using an integer between [1, 9]. The Best-to-Others vector $A_B = (a_{B_1}, a_{B_2}, \dots, a_{B_n})$ is created, where a_{B_j} presents the influence of the best factor (B) relative to factor j. In the evaluation, A_{BB} is equal to one.

Step 3.3 The least priority vector is obtained. Similar to the previous step, the preference for the importance of all other multiplier factors relative to the multiplier factor with the least effect is determined using an integer between [1, 9]. The importance vector of each multiplier factor relative to the least effective multiplier factor $A_W = (a_{1_W}, a_{2_W}, \dots, a_{n_W})^T$ is created, where a_{jW} shows the importance of the j multiplier factor relative to the least important W multiplier factor. In the evaluation, A_{WW} is equal to one.

Step 3.4 The optimal weights of factors $(w_1^*, w_2^*, \dots, w_n^*)$ are calculated using the optimization model given by Eq. (7) (Rezaei, 2015):

$$min\ max_j \left\{ \left| \frac{W_B}{W_j} - a_{BJ} \right|, \left| \frac{W_j}{W_W} - a_{jW} \right| \right\},\ \sum_j w_j = 1\ ,\ w_j \geq 0\ for\ \forall j \tag{7}$$

The optimization model, which yields the significance degrees (or priority weights) of the factors, is transformed into the following model given by Eq. (8) and subsequently solved: (Rezaei, 2015).

$$\min \xi$$
$$\begin{Bmatrix} \left|\frac{W_B}{W_j} - a_{BJ}\right| \leq \xi, for\ \forall j \end{Bmatrix}$$
$$\begin{Bmatrix} \left|\frac{W_j}{W_w} - a_{jW}\right| \leq \xi, for\ \forall j \end{Bmatrix} \quad (8)$$
$$\sum_j w_j = 1\ ,\ w_j \geq 0\ for\ \forall j$$

The optimal weights, which represent the importance degrees of the multiplier factors for the relevant dimension, are obtained from the solution of the model as w_1^D , w_2^D ,..., w_n^D.

Step 3.5 The consistency ratio is calculated. The consistency of each expert's evaluation should be checked. Using the consistency index given in **Table 5**, the consistency ratio is calculated by Eq. (9). The lower the consistency ratio, the more consistent the comparisons are considered (Rezaei, 2015).

Table 5 Consistency Index.

a_{BW}	2	3	4	5	6	7	8	9
Consistency Index (max ξ)	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

$$\text{Consistency Ratio} = \frac{\xi}{Consistency\ Index}\ ,\ \in [0,1] \quad (9)$$

Step 3.6 The common importance degrees of the multiplier factors are determined. In literature, the aggregation process of importance degrees can be performed using the arithmetic or geometric mean methods. However, this process is highly sensitive to outliers and can cause significant loss of information since it ignores the individual distributions of experts. Therefore, Mohammadi and Rezaei (2020) proposed the BBWM method. BBWM is developed as a probabilistic group decision-making model. This method uses a Bayesian hierarchical model to calculate the criterion weights of a group of decision-makers. It determines the most appropriate weights for criteria based on the preferences of a group of decision-makers. As an extended version of BWM, BBWM calculates factor weights using a Bayesian hierarchical model to aggregate the preferences of a group of decision-makers. Compared to the original BWM, BBWM provides more information and is validated through a numerical diagram. The model, using the Dirichlet distribution, is shown in Eq. (10):

$$w|w_{agg} \sim Dir(\gamma * w_{agg}) \quad (10)$$

This model takes the Best-to-Others and Others-to-Worst vectors as input and produces the w_{agg} vector as output. The parameter γ is the concentration parameter and indicates the dispersion of the distribution. The value w is calculated separately for each dimension.

Step 3.7 The total multiplier score for each main dimension is obtained. In the previous step, the influence degrees of the multiplier factors under each dimension were obtained. The total multiplier factor for the relevant dimension is calculated using Eq. (11):

$$W^D = \sum_{j=1}^{n} w_j^D \times x_j \qquad D \in \{\alpha,\ \beta,\ \delta\} \quad (11)$$

where x_j indicates whether the factor is valid or not. If the factor is active, it is 1; otherwise, it is 0. α, β, δ correspond the environmental, expertise level, and vessel condition dimensions, respectively.

Phase 4. Computation of the total risk score: This step integrates operation-specific risk scores with multiplier effect factors to obtain the final weighted outcomes, as expressed in Eq. (12).

$$R_i^f = R_i \times \left(1 + \sum_{D\in\{\alpha,\beta,\delta\}} W^D\right) \tag{12}$$

The multiplier is applied to the baseline risk magnitude rather than added to it because contextual conditions do not create risk on their own. Instead, they increase or reduce the effect of an existing hazard. In other words, an unfavorable operating condition has no independent risk value when no hazard is present, and its influence should therefore vary with the magnitude of the hazard being assessed. The three contextual dimensions are combined additively within the multiplier. This approach assumes that each dimension contributes separately to the final score and also keeps the level of adjustment within a clear range. Since the weights within each dimension sum to one, the total multiplier cannot exceed three, which means that the final risk score cannot be more than four times the baseline value. Combining the dimensions multiplicatively would cause their effects to build on one another and could produce increases that were not reflected in the expert judgments and, therefore, would be more difficult to interpret in practice. Where two dimensions are known to interact, such as when less experienced personnel are regularly assigned to night watches, this relationship can be represented through a combined sub-factor without changing the general form of Eq. (12).

4. Application

4.1 Bunkering operation

Bunkering was selected as the application case because it is a routine yet safety-critical shipboard operation. Its risk profile is shaped by the interaction of technical, human, environmental, and organizational factors under operational time pressure. Failures during fuel transfer may lead to marine pollution, fire, operational interruption, legal liability, and economic loss. Recent tanker spill statistics also show that accidental spills continue to occur during operational activities, including cargo transfer and bunkering, despite the long-term decline in major tanker spill incidents (International Tanker Owners Pollution Federation Limited, 2025). In addition, the transition toward alternative marine fuels has increased the operational complexity of bunkering, since fuels such as LNG, ammonia, methanol, and hydrogen require fuel-specific risk assessment, crew competence, communication control, safety zones, and emergency preparedness (American Bureau of Shipping, 2021, 2024a, 2024b; Maritime Technologies Forum, 2024).

Bunkering also offers a practical advantage as a demonstration case, since its procedural structure does not vary materially with vessel category. Irrespective of whether the receiving vessel is a bulk carrier, a container ship, or a general cargo vessel, the operation follows the same sequence of preparation, ship-to-ship or ship-to-shore coordination, transfer, and monitoring and completion, and it is governed by the same pollution-prevention and safety-management requirements. The hazards examined in this study, which concern transfer planning, communication discipline, and the adequacy of safety arrangements, therefore arise in essentially the same form across vessel categories. This property was helpful in demonstrating the framework, because it allows the effect of the contextual dimensions to be observed without the confounding influence of vessel-type variation. It is a property of the case selected, however, and not a condition imposed by the framework. Where an operation is specific to a vessel category, such as cargo handling on a chemical tanker or cargo cooling on a gas carrier, the hazard inventory constructed in Phase 1 would be established for that category, and vessel type would therefore enter the assessment through the inventory while the procedure itself remained unchanged. The source of variation that is material to bunkering is the fuel rather than the vessel, and since the present application is confined to conventionally fueled bunkering, vessel category is not treated here as a source of hazard variation.

The case study builds upon the framework established in Cicek's (2013) doctoral dissertation, which systematically identified potential hazards associated with bunkering operations. As outlined in **Appendix Table 1**, these hazards were categorized under three fundamental categories: (i) Technical errors, (ii) Mechanical failures, and (iii) Procedural deficiencies. In the context of the application, particular emphasis was placed on the subcategory of technical errors, which served as the primary reference point for the development of the risk assessment model. By narrowing the analytical focus to this category, the study aims to provide a more in-depth and structured analysis of the operational vulnerabilities inherent to bunkering processes. The relationship between these three categories and the contextual dimensions of the framework should be stated explicitly, since the two are sometimes conflated. Technical errors, mechanical failures, and procedural deficiencies are hazard categories, and all three belong to the hazard inventory constructed in Phase 1. They are distinct from the contextual dimensions established in Phase 3, which are the time of execution, the experience of the personnel, and the condition of the vessel, and which act upon hazards rather than standing alongside them. Any hazard drawn from any of the three categories is assessed through the same procedure. Technical errors were selected for the present application because they are the most numerous categories in the inventory and because they depend most directly on how the operation is conducted, which makes the effect of the contextual dimensions clearest to observe. Extending the application to mechanical failures and procedural deficiencies would require no change to the procedure, only the elicitation of the corresponding parameter values.

In the first stage of the study, evaluations were collected from a panel of five experts regarding the identified factors in terms of probability, severity, detectability, and frequency. All five panelists had served on board seagoing vessels in management level positions and had personally supervised bunkering operations. The panel comprised two masters and three chief engineers, so that both the deck and the engine perspective of the bunkering process were represented, and each panelist had accumulated more than seven years of seagoing service, with total professional experience ranging from eight to twelve years, as reported in **Table 6**. The inclusion of masters reflects the division of responsibility during bunkering. Although the transfer itself is conducted under the supervision of the chief engineer, the operation is carried out under the overall responsibility of the master, who approves the bunkering plan, authorizes its commencement, and is accountable for the ship-shore safety checklist and for compliance with port and pollution-prevention requirements. The deck department is also directly involved in mooring the barge, in the access arrangements between the two vessels, and in watchkeeping during the transfer, and several of the hazards examined in this study fall within that scope. Two panelists remained in active seagoing service at the time of the study, while the remaining three held shore-based technical positions as a marine superintendent, a technical inspector, and a technical manager. This composition was deliberate: the seagoing members contribute direct familiarity with the execution of the operation, whereas the shore-based members contribute the fleet-level perspective from which vessel condition and crew competence are actually managed. The anonymized profile of each panelist is given in **Table 6**.

The panel size was set at five in accordance with the criteria given in Step 1.1. The experts formed a homogeneous group in terms of professional background, all having held management-level positions at sea and all having supervised the operation under assessment, and the scope of the assessment was confined to a single operation performed on conventionally fueled vessels. Under these conditions, the lower end of the range recommended in the literature is appropriate (Rowe & Wright, 2001). The third eligibility criterion given in Step 1.1 was also satisfied, since all five experts worked within certified safety management systems and were responsible, in their respective positions, for preparing or reviewing the risk assessments and checklists through which bunkering is controlled.

Table 6 Composition of the expert panel.

Expert	Highest rank held at sea	Seagoing service (years)	Current position	Total professional experience (years)
E1	Master	> 7	Active seagoing service	10
E2	Master	> 7	Marine superintendent	8
E3	Chief Engineer	> 7	Active seagoing service	12
E4	Chief Engineer	> 7	Technical inspector	9
E5	Chief Engineer	> 7	Technical manager	10

Each expert independently assessed all factors using the predefined ordinal scales given in **Table 2** for the four dimensions. This step was intended to capture the individual judgments, and the dispersion among them, before any consensus was sought. The Delphi method was then applied to build consensus while preserving the ordinal nature of the assessment scales: the aggregated results of each round were returned to the panel, and experts were invited to reconsider their ratings in the light of the collective response without knowing which rating belonged to which colleague. The ratings collected in the final round are given in **Table 7**.

Table 7 Final expert preferences.

	Probability (P)					Severity (S)					Detectability (D)					Frequency (F)				
	E1	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**	**E1**	**E2**	**E3**	**E4**	**E5**
T11	L	M	VL	VL	VL	H	H	C	C	Mo	P	SI	SI	P	SI	O	Re	VR	O	O
T12	M	L	VL	VL	L	H	H	S	H	H	I	SP	AP	AP	SP	O	O	VR	O	O
T13	L	M	L	H	M	Mo	Mo	C	Mo	Mo	I	I	SI	I	I	Re	Re	Pe	Re	Re
T14	L	L	L	VL	VL	H	H	C	H	H	P	SI	I	SI	P	O	O	R	O	O
T21	M	M	M	M	M	H	H	H	Mo	H	P	I	P	P	SP	Re	O	VR	Re	Re
T22	M	L	VH	L	M	H	H	C	H	H	P	SP	P	P	SP	Re	Re	Pe	Re	Re
T231	L	VL	VL	VL	VL	C	C	C	C	C	SP	SP	SP	SP	AP	O	O	R	O	O
T232	M	VL	VL	M	VL	H	C	H	H	H	SP	P	AP	SP	AP	O	O	R	O	O
T233	L	VL	VL	L	L	C	C	C	C	C	I	I	SI	P	P	R	R	VR	R	R
T234	M	H	VL	VL	M	Mo	Mo	C	Mo	Mo	SP	SP	AP	SP	SP	O	O	VR	O	O
T235	M	L	VL	VL	VL	Mo	Mo	C	H	Mo	SP	SP	AP	SP	SP	O	O	VR	O	O
T236	M	M	VL	VL	VL	C	C	C	C	C	P	I	I	P	P	O	O	VR	O	O
T237	M	VL	M	M	VL	S	S	T	T	S	SP	SP	P	SP	SP	O	O	VR	O	O
T24	VH	VH	VH	VH	H	H	H	C	C	H	P	AP	P	P	SP	Re	Re	O	Re	Re
T31	L	VL	VL	VL	M	H	H	C	H	H	SP	SP	AP	SP	SP	O	O	R	O	O
T32	M	L	VL	L	M	H	C	C	H	H	P	P	SP	P	SP	Re	Re	VR	Re	Re
T33	M	L	VL	M	M	H	C	C	H	H	P	P	I	P	SP	Re	Re	VR	Re	Re
T34	L	VL	VL	VL	L	H	H	Mo	H	H	P	I	I	SI	P	O	O	VR	O	O

In this approach, expert preferences have been statistically summarized using quartile-based measures rather than arithmetic means. For each factor and evaluation dimension, the first quartile (Q1), median (M), and third quartile (Q3) were calculated. The aggregated values represent the central tendency and dispersion of expert judgments and provide a robust measure of consensus. The use of

quartiles also enables the identification of agreement levels among experts, where narrower interquartile ranges indicate higher consensus. The final Delphi outcomes, representing the consensus values for probability, severity, frequency, and detectability, have been expressed as fuzzy numbers (Q1, M, Q3) calculated via Eqs. (1) - (4). **Table 8** summarizes each factor for all dimensions. These values constitute the finalized expert-based assessments used in the subsequent stages of analysis.

Table 8 Final Delphi results based on quartile statistics.

	Probability (Q1, M, Q3)	**Severity (Q1, M, Q3)**	**Detectability (Q1, M, Q3)**	**Frequency (Q1, M, Q3)**
T11	(1,1,3)	(7,7,9)	(5,9,9)	(5,5,5)
T12	(1,3,3)	(7,7,7)	(1,3,3)	(5,5,5)
T13	(3,5,5)	(5,5,5)	(7,7,7)	(7,7,7)
T14	(1,3,3)	(7,7,7)	(5,7,9)	(5,5,5)
T21	(5,5,5)	(7,7,7)	(5,5,5)	(5,7,7)
T22	(3,5,5)	(7,7,7)	(3,5,5)	(7,7,7)
T231	(1,1,1)	(9,9,9)	(3,3,3)	(5,5,5)
T232	(1,1,5)	(7,7,7)	(1,3,3)	(5,5,5)
T233	(1,3,3)	(9,9,9)	(5,7,7)	(3,3,3)
T234	(1,5,5)	(5,5,5)	(3,3,3)	(5,5,5)
T235	(1,1,3)	(5,5,7)	(3,3,3)	(5,5,5)
T236	(1,1,5)	(9,9,9)	(5,5,7)	(5,5,5)
T237	(1,5,5)	(1,3,3)	(3,3,3)	(5,5,5)
T24	(9,9,9)	(7,7,9)	(3,5,5)	(7,7,7)
T31	(1,1,3)	(7,7,7)	(3,3,3)	(5,5,5)
T32	(3,3,5)	(7,7,9)	(3,5,5)	(7,7,7)
T33	(3,5,5)	(7,7,9)	(5,5,5)	(7,7,7)
T34	(1,1,3)	(7,7,7)	(5,7,7)	(5,5,5)

Then, the risk magnitude is calculated through FIS, utilizing the fuzzy numbers provided in **Table 8** and the rule base defined in **Table 3**. Initially, the membership degrees corresponding to the evaluations in **Table 8** must be determined. For this purpose, the membership functions illustrated in **Figure 3** are employed. The membership degrees representing each risk's affiliation with the respective sets are presented in **Table 9**. Based on these degrees and the rules activated within the rule base defined in **Table 3**, the risk magnitudes for each individual risk have been calculated and are detailed in **Table 10**.

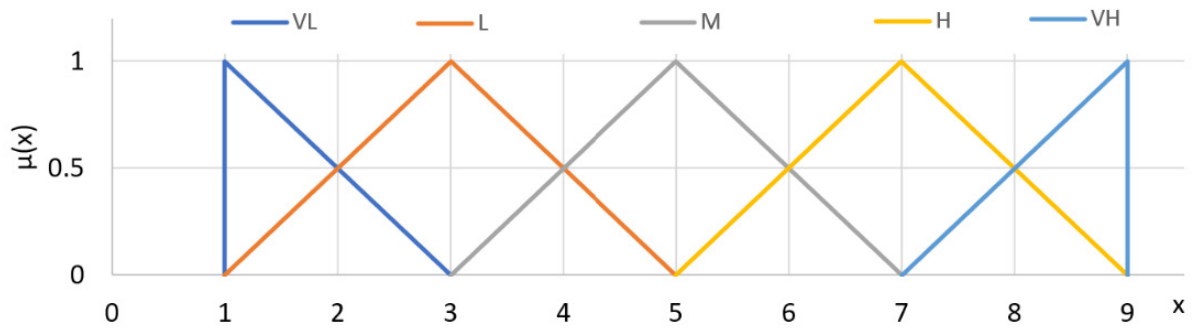


Figure 3 Membership value for the linguistic scale.

Table 9 Membership level of risk factors.

	Probability (P)				**Severity (S)**					**Detectability (D)**					**Frequency (F)**				
	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH
T11	1.00	0.50							1.00	0.50			0.33	0.67	1.00			1.00		
T12	0.50	1.00							1.00		0.50	1.00						1.00		
T13		0.50	1.00					1.00						1.00					1.00	
T14	1.00	0.50							1.00				0.50	1.00	0.50			1.00		
T21			1.00						1.00				1.00					0.50	1.00	
T22		0.50	1.00						1.00			0.50	1.00						1.00	
T231	1.00									1.00		1.00						1.00		
T232	1.00	0.67	0.33						1.00		0.50	1.00						1.00		
T233	1.00	0.50								1.00			0.50	1.00			1.00			
T234		0.67	1.00					1.00				1.00						1.00		
T235	1.00	0.50						1.00	0.50			1.00						1.00		
T236	1.00	0.67	0.33							1.00			1.00	0.50				1.00		
T237		0.67	1.00			0.50	1.00					1.00						1.00		
T24					1.00				1.00	0.50		0.50	1.00						1.00	
T31	1.00	0.50							1.00			1.00						1.00		
T32		1.00	0.50						1.00	0.50		0.50	1.00						1.00	
T33		0.50	1.00						1.00	0.50			1.00						1.00	
T34	1.00	0.50							1.00				0.50	1.00				1.00		

Table 10 Baseline risk magnitude for each risk factor.

	N_g	M_i	M_j	C	Risk magnitude %
T11	0.00	1.00	0.50	0.00	50.00
T12	0.50	1.00	0.00	0.00	33.33
T13	0.00	0.50	1.00	0.00	60.00
T14	0.00	1.00	0.50	0.00	50.00
T21	0.00	0.50	1.00	0.00	60.00
T22	0.00	0.50	1.00	0.00	60.00
T231	0.00	1.00	0.00	0.00	40.00
T232	1.00	0.70	0.00	0.00	28.02
T233	0.00	1.00	0.00	0.00	40.00
T234	0.70	1.00	0.00	0.00	31.98
T235	1.00	0.50	0.00	0.00	26.67
T236	0.00	1.00	0.50	0.00	50.00
T237	1.00	0.00	0.00	0.00	20.00
T24	0.00	0.00	1.00	0.50	80.00
T31	1.00	0.50	0.00	0.00	26.67
T32	0.00	1.00	0.50	0.00	50.00
T33	0.00	0.50	1.00	0.00	60.00
T34	0.00	1.00	0.00	0.00	40.00

As shown in **Table 10**, T24 received the highest risk magnitude (80 %) due to its combination of very high probability, high severity, moderate detectability, and repetitive occurrence frequency.

Following the identification of the risk magnitude of each risk, the multiplier effect dimensions are obtained given in **Table 4**. For this purpose, five domain experts participated in the

evaluation process to determine the relative importance of the environmental factors, expertise levels, and vessel condition categories (**Figure 4**). The assessments were conducted through a web-based questionnaire designed according to the Bayesian Best-Worst Method (BBWM) framework.

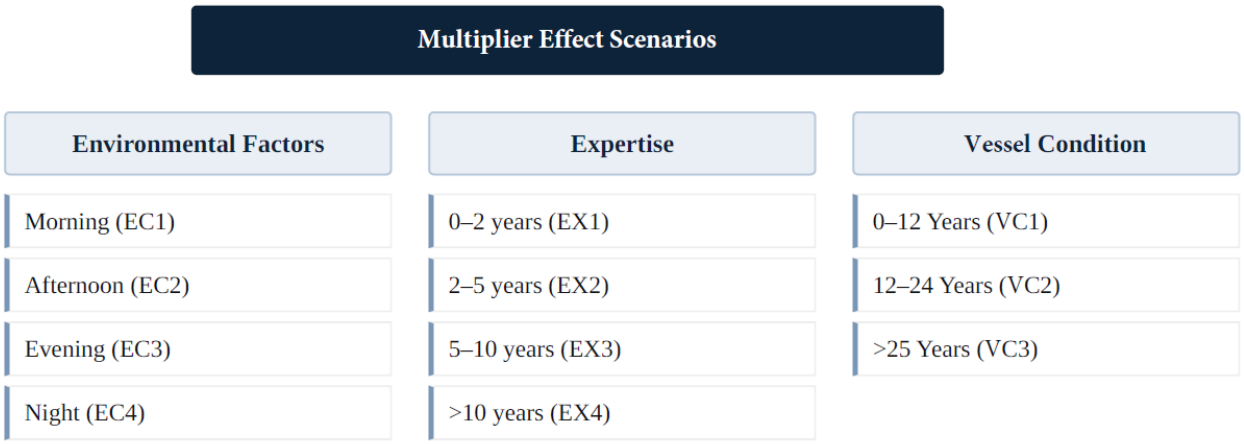


Figure 4 Structural decomposition of the multiplier effect scenarios.

For each dimension, experts first identified the most influential (best) and least influential (worst) factor. Subsequently, Best-to-Others and Others-to-Worst pairwise comparisons were obtained using a nine-point linguistic preference scale. The consistency of each expert's judgments was evaluated using the consistency ratio (CR) proposed by Rezaei (2015). As presented in **Table 11**, all CR values were below the corresponding threshold values, indicating that the collected judgments were sufficiently consistent and suitable for further analysis.

Table 11 Expert preferences on the multiplier effect dimensions.

Main Factors	Expert	Critical factor		Best to Others	Others to Worst	CR
		The most	The least			
Environmental Factors	E1	Night	Morning	[1, 3, 3, 9]	[9, 5, 5, 1]	0.264
	E2	Night	Morning	[1, 9, 5, 7]	[8, 1, 5, 4]	0.083
	E3	Night	Morning	[1, 9, 3, 7]	[9, 1, 7, 3]	0.167
	E4	Evening	Afternoon	[5, 7, 1, 9]	[5, 3, 7, 1]	0.222
	E5	Night	Afternoon	[1, 5, 3, 7]	[7, 3, 5, 1]	0.191
Expertise Level (Years)	E1	(0-2)	(>10)	[1, 3, 4, 9]	[7, 5, 4, 1]	0.097
	E2	(2-5)	(5-10)	[3, 1, 7, 5]	[3, 5, 1, 3]	0.191
	E3	(0-2)	(>10)	[1, 3, 5, 9]	[9, 7, 5, 1]	0.222
	E4	(0-2)	(>10)	[1, 4, 6, 8]	[5, 4, 3, 1]	0.179
	E5	(0-2)	(>10)	[1, 2, 4, 8]	[8, 4, 2, 1]	0.001
Vessel Condition (Years)	E1	(>25)	(0 - 12)	[9, 5, 1]	[1, 4, 8]	0.097
	E2	(>25)	(0 - 12)	[9, 4, 1]	[1, 3, 8]	0.083
	E3	(>25)	(0 - 12)	[9, 4, 1]	[1, 4, 8]	0.097
	E4	(>25)	(0 - 12)	[9, 5, 1]	[1, 3, 5]	0.083
	E5	(>25)	(0 - 12)	[9, 5, 1]	[1, 3, 7]	0.083

Subsequently, the traditional BWM was implemented for each expert individually to calculate the local weights of the multiplier factors and to evaluate the consistency of the judgments. The results obtained from the individual BWM analyses are summarized in **Table 12**.

Table 12 Weights obtained by individual expert preferences.

	Environmental Factors				**Expertise Level (Years)**				**Vessel Condition (Years)**		
	Night	**Morning**	**Evening**	**Afternoon**	**0 - 2**	**2 - 5**	**5 - 10**	**>10**	**0 - 12**	**12 - 24**	**>25**
E1	0.659	0.063	0.162	0.116	0.616	0.190	0.127	0.067	0.077	0.176	0.747
E2	0.543	0.202	0.202	0.053	0.225	0.553	0.086	0.135	0.077	0.176	0.747
E3	0.601	0.053	0.242	0.104	0.577	0.233	0.140	0.051	0.077	0.176	0.747
E4	0.165	0.118	0.650	0.068	0.597	0.189	0.126	0.087	0.102	0.184	0.714
E5	0.572	0.136	0.226	0.066	0.619	0.194	0.129	0.058	0.091	0.169	0.740
W*	0.445	0.125	0.319	0.112	0.423	0.307	0.166	0.104	0.108	0.245	0.647

*The multiplier factor weights obtained BBWM.

Regarding the environmental factors dimension, most experts identified *Night* as the most influential factor affecting the magnitude of operational risks, whereas *Morning* or *Afternoon* periods were generally considered the least influential. Differing from the other, E4 selected *Evening* as the most influential factor. The calculated local weights indicate that *Night* received the highest importance from the other experts, with individual weights ranging between 0.543 and 0.659. These present that risk severity is perceived to increase considerably during night time operations due to reduced visibility and operational challenges.

For the expertise level, most of the experts determined the "*0 - 2 years*" experience category as the most influential risk-enhancing factor and more than 10 years of experience as the least influential. Only E2 identified "*2 - 5 years*" experience category as the most influential factor. The final weights assigned the highest importance to the lower experience groups, particularly "*0 - 2 years"* category, of which weights are between 0.577 and 0.619 for the others. This result highlights the significant contribution of the limited operational experience to the amplification of the risk exposure.

In the vessel condition factor, all of the experts determined vessels older than 25 years as the most influential factor and vessels aged between 0-12 years as the least influential factor. For this, similar weight distributions were obtained among them. The weight assigned to the "*>25 years*" category varied between 0.714 and 0.747, indicating a strong consensus regarding impact of aging vessels on operational risk levels. Taken together, these results indicate that the panel converged despite its limited size. Agreement on the most influential vessel condition category was unanimous, and in each of the remaining two dimensions four of the five experts identified the same most influential factor. The divergences were isolated rather than systematic: they occurred in different dimensions and originated from different experts, and no dimension produced a split corresponding to the deck and engine composition of the panel. All consistency ratios remained below the corresponding threshold values, confirming that the panel size adopted in accordance with Step 1.1 was sufficient for the assessment task.

After the consistency of the individual evaluations have been checked, BBWM method is employed to obtain consensus weights for each factor. Unlike conventional arithmetic or geometric averaging methods, BBWM utilizes a Bayesian hierarchical framework to preserve the uncertainty and variability inherent in expert judgments while generating a collective preference structure. The method was implemented in MATLAB using the package developed by Mohammadi and Rezaei (2020). The aggregated weights present the relative contribution of each multiplier factor within its

respective dimension and constitute the basis for calculating the multiplier scores defined in Eq. (11). The aggregated BBWM results are given in **Table 12**. The obtained weights demonstrate consensus among experts regarding the factors that most significantly increase operational risk severity. Specifically, nighttime operations, low levels of experience, and older vessels emerged as dominant risk conditions. Then, these findings were incorporated into the multiplier scoring framework to quantify the cumulative impact of the operational conditions on risk magnitude.

Following the determination of multiplier factor weights through BBWM, the final risk scores have been calculated by integrating the baseline risk scores with the operational conditions under which bunkering activities are performed. Differing from the conventional risk assessment approaches that assume identical operating conditions for all hazards, the proposed method considers the contextual characteristics of the operation through multiplier factors representing environmental conditions, personnel expertise, and vessel condition.

To demonstrate the applicability of the proposed model, the following three representative operational scenarios were developed.

Scenario 1: The first scenario presents the most favorable operating conditions. The bunkering operation is conducted during the morning period, by personnel with more than ten years of experience, on a vessel aged between 0 and 12 years. These conditions present to the lowest multiplier values within each dimension and therefore represent the minimum amplification of the baseline risk scores.

Scenario 2: The second scenario reflects operational conditions encountered in daily bunkering activities. The operation is performed during the evening period by personnel with 2 - 5 years of experience on a vessel aged between 12 and 24 years.

Scenario 3: The last scenario represents the most adverse operating conditions. The bunkering operation is conducted at night by personnel with 0 - 2 years of experience on a vessel older than 25 years. Since these factors received the highest importance degrees within their respective dimensions, this scenario produces a maximum multiplier effect and consequently the highest final risk scores.

The operational scenarios used in the application are summarized in **Table 13**.

Table 13 Operational scenarios used for the computation of final risk scores.

Scenario	Environmental Factor	Expertise Level	Vessel Condition
Scenario 1	Morning	>10 Years	0 - 12 Years
Scenario 2	Evening	2 - 5 Years	12 - 24 Years
Scenario 3	Night	0 - 2 Years	>25 Years

4.2 Results

For each scenario, the multiplier score has been calculated by activating the corresponding factors within each dimension according to Eq. (11). Subsequently, the final risk score of each technical hazard was calculated using Eq. (12). The final risk scores calculated under the three operational scenarios are given in **Table 14**. These results demonstrate the influence of the proposed multiplier framework on the magnitude of the baseline risk values. While the initial risk magnitudes remain unchanged, the final risk scores vary by a factor of 1.88 between the most and the least favorable operational scenarios.

The obtained results show a substantial increase in risk levels as operational conditions become less favorable. For all hazards, Scenario 1 produced the lowest final risk scores, which is reflecting the relatively safe operational environment characterized by experienced personnel, a newer vessel, and favorable environmental conditions. Scenario 2 generated moderate increases, representing typical operational circumstances encountered during routine bunkering activities. The highest risk values are observed under Scenario 3, which combines night time operations, limited

crew experience, and an aging vessel. According to these scenarios, several hazards exceeded 150 points, including T13, T21, T22, and T33, while hazard T24 reached the highest final risk score of 201.26. This indicates that operational context amplifies the consequences of the technical failures by up to 88 % (difference between Scenario 1 and Scenario 3), even when the underlying hazard remains unchanged.

Table 14 Final risk scores under different operational scenarios.

Risk	Magnitude %	Scenario 1 %	Scenario 2 %	Scenario 3 %
T11	50.00	66.86	93.50	125.79
T12	33.33	44.57	62.32	83.85
T13	60.00	80.23	112.19	150.94
T14	50.00	66.86	93.50	125.79
T21	60.00	80.23	112.19	150.94
T22	60.00	80.23	112.19	150.94
T231	40.00	53.48	74.80	100.63
T232	28.02	37.47	52.39	70.49
T233	40.00	53.48	74.80	100.63
T234	31.98	42.76	59.80	80.45
T235	26.67	35.66	49.87	67.09
T236	50.00	66.86	93.50	125.79
T237	20.00	26.74	37.40	50.31
T24	80.00	106.97	149.59	201.26
T31	26.67	35.66	49.87	67.09
T32	50.00	66.86	93.50	125.79
T33	60.00	80.23	112.19	150.94
T34	40.00	53.48	74.80	100.63

Furthermore, the results show that the proposed multiplier-based framework provides a more realistic representation of the severity of hazards under varying operational conditions while maintaining their relative ranking. For example, although T24 was already identified as the most critical technical hazard based on its baseline magnitude (80 %), its risk level increased by approximately 120 % under the worst-case scenario. Similarly, hazards with moderate baseline risk values, such as T231, T233, and T34, exceeded a final score of 100 when assessed under adverse operating conditions.

These findings highlight the importance of considering operational factors in maritime risk assessment. Traditional approaches relying solely on baseline risk scores may underestimate the actual risk exposure associated with bunkering operations. The proposed model offers more dynamic and operationally realistic assessment framework by incorporating environmental conditions, expertise, and vessel condition into the evaluation process. As a result, critical operational situations have been better identified, mitigating measures have been prioritized by decision-makers, and safety resources have been allocated more effectively.

4.3 Model validation

The validity of the proposed risk assessment framework was examined by comparing the risk prioritization obtained from the proposed model with the prioritization derived from the conventional SCEA risk assessment approach based directly on expert evaluations. Since the proposed framework incorporates operational context into the assessment of risk magnitude, the validation focused on whether the proposed model preserves the fundamental risk prioritization established by expert

judgment while allowing context-dependent changes in the relative importance of individual risk factors.

To obtain a reference ranking, the expert evaluations of the risk factors were transformed to numerical values using the scale given in **Table 2**, following which the expert evaluations were aggregated using two commonly applied aggregation approaches: the arithmetic mean and the geometric mean. The resulting aggregated assessments were then used to establish the conventional risk ranking of the 18 identified risk factors. The use of both aggregation procedures also enabled the robustness of the reference ranking to be examined with respect to the aggregation method. The obtained results are given in **Table 15**.

Table 15 Comparison of risk-factor prioritization obtained from conventional SCEA and the proposed model.

	Arithmetic mean aggregation						**Geometric mean aggregation**						**Proposed Model**	
	P	**S**	**D**	**F**	**RM**	**Rank**	**P**	**S**	**D**	**F**	**RM**	**Rank**	**RM%**	**Rank**
T11	2.20	7.40	7.40	4.60	554.17	8.00	1.72	7.24	7.11	3.88	342.99	9.00	50.00	6.00
T12	2.60	6.20	3.00	4.20	203.11	14.00	2.14	5.91	2.29	3.62	105.00	16.00	33.33	13.00
T13	4.60	5.80	7.40	7.40	1,461.00	2.00	4.36	5.62	7.36	7.36	1,328.42	2.00	60.00	2.00
T14	2.20	7.40	7.00	4.60	524.22	9.00	1.93	7.36	6.77	4.51	434.61	7.00	50.00	6.00
T21	5.00	6.60	5.00	5.40	891.00	4.00	5.00	6.54	4.83	4.43	700.68	4.00	60.00	2.00
T22	5.00	7.40	4.20	7.40	1,149.96	3.00	4.58	7.36	4.08	7.36	1,012.44	3.00	60.00	2.00
T231	1.40	9.00	2.60	4.60	150.70	16.00	1.25	9.00	2.41	4.51	121.89	15.00	40.00	10.00
T232	2.60	7.40	2.60	4.60	230.11	13.00	1.90	7.36	2.14	4.51	135.44	14.00	28.02	15.00
T233	2.20	9.00	6.60	2.60	339.77	10.00	1.93	9.00	6.43	2.41	269.58	10.00	40.00	10.00
T234	3.80	5.80	2.60	4.20	240.68	12.00	2.81	5.62	2.41	3.62	137.88	12.00	31.98	14.00
T235	2.20	6.20	2.60	4.20	148.95	17.00	1.72	6.02	2.41	3.62	90.23	17.00	26.67	16.00
T236	2.60	9.00	5.80	4.20	570.02	7.00	1.90	9.00	5.72	3.62	355.16	8.00	50.00	6.00
T237	3.40	2.20	3.40	4.20	106.81	18.00	2.63	1.93	3.32	3.62	61.14	18.00	20.00	18.00
T24	8.60	7.80	3.80	6.60	1,682.37	1.00	8.56	7.74	3.27	6.54	1,418.56	1.00	80.00	1.00
T31	2.20	7.40	2.60	4.60	194.71	15.00	1.72	7.36	2.41	4.51	137.54	13.00	26.67	16.00
T32	3.40	7.80	4.20	5.80	646.03	6.00	2.95	7.74	4.08	4.74	442.08	6.00	50.00	6.00
T33	3.80	7.80	5.00	5.80	859.56	5.00	3.27	7.74	4.83	4.74	580.05	5.00	60.00	2.00
T34	1.80	6.60	6.60	4.20	329.31	11.00	1.55	6.54	6.43	3.62	236.80	11.00	40.00	10.00

Table 15 compares the rankings obtained using the arithmetic mean, geometric mean, and the proposed context-aware model. The results demonstrate a high degree of consistency between the conventional rankings and the proposed model. In particular, T24 and T13 were consistently identified as the first and second most critical risk factors, respectively, by all three approaches. T33 also remained among the five highest-ranked risk factors in all approaches. Furthermore, several risk factors, including T14, T236, T233, T34, T234, and T237, remained in closely comparable positions across the conventional and proposed rankings.

The differences observed between the conventional and proposed rankings are primarily attributable to the incorporation of operational conditions in the proposed framework. Unlike the conventional approach, in which the risk characteristics assigned to a hazard remain fixed, the proposed model modifies the risk magnitude according to environmental conditions, personnel expertise, and vessel condition. Therefore, the observed changes in the relative positions of some risk factors do not indicate a contradiction with the conventional assessment; rather, they demonstrate the intended effect of the contextual adjustment mechanism. For example, T21 and T22 occupy the third and fourth positions, respectively, in the conventional rankings, whereas their positions are reversed

in the proposed model. Similarly, T11 and T33 change their relative positions after the contextual factors are incorporated. Such changes indicate that the proposed framework preserves the underlying hazard prioritization while allowing operational context to differentiate risks that may have similar baseline characteristics. The factors whose positions do not coincide fall into two groups, and each has an identifiable cause. The first group consists of factors that share an identical risk magnitude in the proposed model and are therefore assigned the same rank. T13, T21, T22, and T33 all reach 60 percent and jointly occupy the second position, whereas the conventional aggregation separates them into the second, fourth, third, and fifth positions. This difference reflects the resolution of the rule base rather than a disagreement about relative importance, since fuzzy inference produces a bounded set of output values and does not distinguish between hazards that activate the same rules. The second group consists of factors for which the rule base and the multiplicative aggregation treat extreme parameter values differently. T231 is the clearest example, rising from the sixteenth position under conventional aggregation to the tenth position in the proposed model. This factor combines catastrophic severity with a very low probability of occurrence, and the multiplicative form allows the low probability to suppress the severity, whereas the rule base does not permit a catastrophic-severity event to fall below a defined consequence class irrespective of its probability. The same mechanism operates in the opposite direction for T232 and T234, which combine moderate values across all four parameters and consequently move downward. These divergences are therefore attributable to the aggregation logic rather than to instability in the framework.

An additional validation was conducted to determine whether the choice of aggregation procedure used for the expert assessments affected the conventional ranking. Spearman's rank correlation coefficient was employed to evaluate the strength of the association among the three ranking approaches. As shown in **Table 16**, the rankings obtained using arithmetic and geometric mean aggregation exhibited a very strong positive correlation ($\rho = 0.983$, $p < 0.001$). This result indicates that the conventional risk prioritization is highly robust with respect to the aggregation operator. More importantly, the ranking generated by the proposed model was very strongly correlated with both conventional rankings. The correlation between the proposed model and the arithmetic-mean-based ranking was $\rho = 0.950$ ($p < 0.001$), while the correlation with the geometric-mean-based ranking was $\rho = 0.944$ ($p < 0.001$). These results demonstrate a high degree of consistency between the proposed model and the conventional expert-based risk prioritization. The strong correlations provide evidence of the convergent validity of the proposed framework. In particular, the proposed model preserves the fundamental risk prioritization established by the expert assessments, while allowing the relative positions of individual risk factors to change when contextual operating conditions are incorporated.

Table 16 Statistical validation of risk-factor rankings using Spearman's rank correlation and Wilcoxon signed-rank tests.

Comparison	Spearman's ρ	Spearman's p-value	Wilcoxon p-value
Arithmetic – Geometric	0.983	<0.001	0.931
Arithmetic – Proposed Model	0.950	<0.001	0.076
Geometric – Proposed Model	0.944	<0.001	0.077

Note: Spearman's ρ assesses the strength of the monotonic association between paired rankings, whereas the Wilcoxon signed-rank test examines whether the median paired difference in rank positions differs from zero. The significance level was set at $\alpha = 0.05$.

To further validate the consistency of the risk prioritization, the paired rankings obtained from the arithmetic mean aggregation, geometric mean aggregation, and the proposed model were

statistically compared using the Wilcoxon signed-rank test. The test was employed to examine whether the median difference between the paired rank positions was statistically different from zero. As shown in **Table 16**, no statistically significant difference was observed between the geometric-mean-based ranking and the proposed model (p = 0.077). Similarly, the comparison between the arithmetic-mean-based ranking and the proposed model yielded no statistically significant difference (p = 0.076). The comparison between the arithmetic- and geometric-mean-based rankings also indicated no statistically significant difference (p = 0.931). Since all p-values were greater than the predefined significance level of 0.05, the null hypotheses could not be rejected.

Taken together, the Spearman's rank correlation and Wilcoxon signed-rank results provide complementary statistical evidence supporting the consistency of the proposed model with the conventional expert-based risk prioritization. The proposed model demonstrates very strong positive associations with both conventional rankings while introducing no statistically significant systematic shift in the paired rank positions. These findings support the convergent validity of the proposed framework. At the same time, the changes observed in the relative positions of individual risk factors indicate that the proposed model does not merely reproduce the conventional SCEA results. Rather, it preserves the fundamental risk prioritization derived from expert judgment while refining the prioritization according to operational context, such as environmental conditions, personnel expertise, and vessel condition. This is consistent with the primary objective of the proposed framework, which is to provide a context-sensitive assessment of risk rather than a static risk ranking.

Overall, the validation results provide statistical evidence supporting the consistency of the proposed model with the conventional expert-based risk prioritization. The very strong rank associations and the absence of statistically significant systematic differences between the rankings indicate that the proposed framework preserves the fundamental risk structure established through expert assessment, while its contextual adjustment mechanism allows the relative importance of individual risk factors to vary according to operational conditions. The absence of a systematic shift in rank positions is the intended outcome, since the framework is designed to preserve the prioritization established by expert judgment while allowing the assessed magnitude of a hazard to respond to the conditions under which it is encountered.

4.4 Extended critical discussions

The findings demonstrate that operational risk during critical shipboard operations is dynamic and varies with the operational conditions. Unlike conventional maritime risk assessments that assign a fixed value to each hazard, the results demonstrate that operational risk varies with the operational context. Since shipboard operations are performed under continuously changing environmental, technical, and human conditions, static risk assessments are insufficient for representing real operational complexity. Hence, the proposed framework is targeted to overcome the operational behavior limitations.

The observed findings are consistent with previous studies that emphasize the dynamic and multidimensional nature of maritime risk. Wang et al. (2025) demonstrated that navigation risks evolve according to changing operational conditions, while Tao et al. (2024) highlighted the necessity of simultaneously considering human, technological, and environmental factors within autonomous shipping. Likewise, Sezer et al. (2023) and Elidolu et al. (2023) showed that structured analytical models improve the evaluation of tanker operations, whereas Goksu et al. (2023) and Jang et al. (2023) emphasized the influence of maintenance conditions, organizational preparedness, and crew competence on operational safety. Recent technical guidance published for alternative fuel bunkering by American Bureau of Shipping (2021, 2024a, 2024b) and the Maritime Technologies Forum (2024) similarly stresses that operational conditions and personnel competency substantially influence safe execution. The present study extends the ongoing research focus by introducing a context-aware assessment framework that quantitatively incorporates operational conditions into the final risk calculation.

The principal contribution of the proposed model lies in changing the philosophy of operational risk assessment rather than introducing another scoring methodology. Existing approaches such as FMEA, FTA, Bow-Tie analysis, conventional risk matrices, and many hybrid models primarily evaluate the intrinsic characteristics of hazards. Although these methods effectively identify failure mechanisms and prioritize preventive actions, they generally assume that operational conditions remain relatively stable throughout the execution of an activity. However, real shipboard operations rarely satisfy this assumption. The proposed framework uses contextual multipliers to integrate operational dynamics (i.e., changing weather, personnel behaviors, and fleet conditional variations) into baseline hazard severity. Hereby, the methodology represents a transition from hazard-oriented assessment toward context-aware operational risk management. This conceptual shift has significant implications for shipboard operation management. Operational risk assessments prepared under the Safety Management System (SMS) required by the International Safety Management Code are typically produced before an operation begins and are reviewed periodically rather than revised as the operation unfolds. The Code obliges companies to assess identified risks and establish safeguards, yet it prescribes neither the analytical form of the assessment nor any mechanism through which a recorded score should respond to the conditions actually encountered. As a result, risk assessments frequently function as compliance-oriented documents rather than as active decision-support instruments, a limitation that has also been noted in maritime risk research, where the transition from academic models to shipboard application remains incomplete. The proposed methodology enables operational risk to evolve simultaneously with real-time operational conditions. For instance, deterioration of weather conditions, assignment of less experienced personnel, or execution of the same task onboard an aging vessel immediately changes the overall operational risk without requiring the entire assessment process to be repeated. The improvements enable a dynamic monitoring tool capable of supporting remedial actions along with the critical operations. The practical value of this capability becomes evident through the application results. Under identical technical hazards, substantial differences were observed between the three operational scenarios. The outcomes combine the focus both on critical hazards and operational conditions simultaneously. This proactive perspective enables considerably more effective allocation of supervision, experienced personnel, maintenance resources, and relevant operational control measures. The proposed framework provides a structured tool for ship operators, port managers, and bunker suppliers to track real-time operations and to enhance ship-shore communication.

A recurring concern in expert-based risk models is the practical availability of suitably qualified participants. The proposed framework addresses this issue through the distinction it establishes between model calibration and operational application. Phases 1 to 3 constitute a one-time calibration process through which the hazard inventory, baseline risk values, and multiplier weights are determined for a particular operation and fleet profile. Phase 4, in contrast, represents the routine operational use of the model and does not require any further expert involvement. At this stage, the officer in charge identifies only the three prevailing conditions, namely the relevant time band, the experience level of the assigned personnel, and the age category of the vessel, after which the final risk score is calculated directly using Eq. (12). The recurring requirement on board is therefore limited to three operational selections rather than a repeated expert-panel exercise. Because expert participation is confined to the calibration stage, companies of different sizes can meet this requirement through alternative arrangements. Large operators may form the panel internally, as marine superintendents, technical managers, inspectors, and senior seagoing officers generally satisfy the eligibility criteria defined in Step 1.1. Smaller operators with more limited shore-based resources may obtain equivalent expertise through the designated person ashore, classification societies, P&I correspondents, marine surveyors, or consultants already engaged for statutory and commercial purposes. A further option is sector-wide calibration. Since the calibration is operation-specific rather than company-specific, a common set of weights developed by an industry association, flag administration, professional body, or training institution may be adopted by multiple operators and

subsequently adjusted where fleet characteristics differ. The weights presented in **Table 12** illustrate this possibility, as they were derived for conventional bunkering operations and are not specific to a single company. This result suggests that operators using published weights instead of conducting an independent elicitation process would experience only a limited reduction in model precision while facing a substantially lower barrier to implementation.

Beyond the bunkering case examined in this study, the proposed methodology could be adapted to a wider range of shipboard operations, including cargo handling, mooring, enclosed-space entry, pilot transfer, navigation in restricted waters, ship-to-ship operations, ballast water operations, and dry-docking activities. These operations are repeatedly performed under conditions that may vary considerably between individual executions. Application to a different operation would require the development of an operation-specific hazard inventory and the recalibration of the relevant multiplier weights, while the underlying computational logic of the framework would remain largely unchanged. A similar approach could be applied to the bunkering of alternative marine fuels, including LNG, methanol, ammonia, and hydrogen, by incorporating additional factors related to fuel-specific hazards such as flammability, toxicity, leakage, and containment. The index structure defined in Eq. (11) also allows further contextual dimensions to be introduced, including specialized crew competence, sensor reliability, emergency preparedness, equipment condition, and regulatory compliance. In addition, the framework has the potential to be integrated with digital monitoring systems available on modern vessels. Data obtained from weather sensors, machinery condition-monitoring systems, crew workload and fatigue indicators, inspection records, and operational performance systems could provide the contextual inputs required by the multiplier dimensions, thereby enabling more frequent or automated updating of the final risk score. Although further validation would be required, the underlying principle may also be relevant to other safety-critical sectors in which comparable tasks are repeatedly undertaken under changing operational conditions, such as offshore energy, aviation, railway maintenance, chemical processing, and complex construction activities. One point concerning the extension of the framework requires explicit statement, since it bears on the role of the expert panel. The panel assembled for this study was selected for its experience of bunkering, and its judgments are therefore valid for bunkering alone. The weights reported in **Table 12** are not claimed to hold for anchoring, mooring, or enclosed-space entry, and no such claim is made here. What transfers to those operations is the methodology rather than the numerical results. An application to a different operation would begin by convening a panel with direct experience of that operation, which would construct the corresponding hazard inventory and identify the contextual dimensions that amplify risk within it, after which the same four phases would be executed to produce a calibrated matrix for that operation. The intended outcome is therefore a family of operation-specific matrices generated by a common procedure, each calibrated by practitioners of the operation concerned and each available for routine use on board once established. The present study provides that procedure and demonstrates it in full for one operation, which is the necessary first step before comparable matrices can be developed for others.

5. Conclusions

Bunkering operations remain among the most safety-critical shipboard activities because technical failures, human errors, and unfavorable operating conditions may lead to pollution, operational disruption, and economic loss. This study proposed an integrated and context-aware risk assessment framework that evaluates not only the inherent characteristics of bunkering hazards but also the operational conditions that may amplify their consequences.

The proposed framework combines expert-based hazard validation, FIS-based risk magnitude calculation, BBWM-based weighting of multiplier factors, and scenario-based final risk scoring. Its application demonstrated that operational context has a decisive influence on final risk levels. Night time operation, limited crew experience, and aging vessel condition emerged as the most influential risk-amplifying factors. Under the most adverse scenario, final risk scores increased substantially

even though baseline risk magnitudes remained unchanged, confirming that static risk assessment may underestimate actual exposure during bunkering operations.

The main contribution is a practical decision-support structure that helps operators identify when a hazard becomes critical rather than only which hazard is critical, enabling more targeted allocation of supervision, crew planning, training and maintenance resources. The analytical procedure is generic. Transferring it to another critical shipboard operation requires a hazard inventory, a set of contextual dimensions, and an expert panel appropriate to that operation, while the four phases and the computational structure remain unchanged. Although demonstrated here on technical-error hazards in conventional bunkering, the framework extends to procedural deficiencies, mechanical failures, and alternative-fuel bunkering. Establishing calibrations for mooring, cargo handling, and enclosed-space entry is the principal direction for further work.

Declaration of generative AI in scientific writing

The authors used generative artificial intelligence tools (ChatGPT and Claude) for language refinement of the manuscript, specifically grammar, punctuation, and style. These tools were not used for scientific analysis, data interpretation, methodological development, or the generation of research findings. With or without AI applications, the authors are fully accountable for every process of the work and manuscript writing. All final edits, interpretations, and conclusions are the sole responsibility of the authors. The authors academically and ethically confirmed that the stated AI statements are true, and the authors are ultimately responsible and accountable for the contents of the published work, including the correctness of the references in the reference list.

CRediT contribution statement

Kadir Cicek: Conceptualization; Methodology; Investigation; Data curation; Writing - Original Draft; Visualization; Project administration. **Metin Celik:** Conceptualization; Supervision; Writing - Review & Editing. **Selcuk Cebi:** Methodology; Software; Formal analysis; Validation; Writing - Review & Editing. **Gülşah Şahin:** Software; Formal analysis; Data curation; Writing - Review & Editing.

References

American Bureau of Shipping. (2021). *LNG bunkering: Technical and operational advisory.* American Bureau of Shipping.

American Bureau of Shipping. (2024a). *Ammonia bunkering: Technical and operational advisory*. American Bureau of Shipping.

American Bureau of Shipping. (2024b). *Methanol bunkering: Technical and operational advisory*. American Bureau of Shipping.

Aydin, M., Akyuz, E., & Boustras, G. (2024). A holistic safety assessment for cargo holds and decks fire & explosion risks under fuzzy Bayesian network approach. *Safety Science, 176*, 106555. https://doi.org/10.1016/j.ssci.2024.106555

Choi, Y. S., Listan Bernal, M., Krivoshapkina, M., & Yeo, G. T. (2025). Analyzing risk factors in ship-to-ship liquefied natural gas bunkering operations. *Asian Journal of Shipping and Logistics*, *41*(1), 52-60. https://doi.org/10.1016/j.ajsl.2025.01.003

Cicek, K. (2013). *An analytical approach to operational risk management in ship safety management* (Doctoral dissertation). Istanbul Technical University. Thesis No. 333021 Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Dalkey, N. (1969). An experimental study of group opinion: The Delphi method. *Futures, 1*(5), 408-426. https://doi.org/10.1016/S0016-3287(69)80025-X

Demirel, H., Başhan, V., Yücesan, M., & Gül, M. (2024). A comprehensive risk assessment framework for mooring risks at hydrocarbon berths using fuzzy rule-based Bayesian

network and multi-attribute decision-making. *Stochastic Environmental Research and Risk Assessment, 38*(11), 4393-4414. https://doi.org/10.1007/s00477-024-02809-w

Elidolu, G., Sezer, S. I., Akyuz, E., Arslan, O., & Arslanoglu, Y. (2023). Operational risk assessment of ballasting and de-ballasting on-board tanker ship. *Reliability Engineering & System Safety, 231*, 108975. https://doi.org/10.1016/j.ress.2022.108975

Feng, W., Wang, Z., Dai, X., Dong, S., Qiao, W., & Ma, X. (2025). Ship-to-ship liquefied natural gas bunkering risk assessment by integrating fuzzy failure mode and effect analysis and the technique for order preference by similarity to an ideal solution. *Journal of Marine Science and Engineering, 13*(4), 710. https://doi.org/10.3390/jmse13040710

Goksu, B., Sakar, C., & Yuksel, O. (2025). A probabilistic assessment of ship blackout incident with Fault Tree Analysis into (FTA) Bayesian Network (BN). *Journal of Marine Engineering & Technology, 24*(1), 54-69. https://doi.org/10.1080/20464177.2024.2423425

Goksu, B., Yuksel, O., & Sakar, C. (2023). Risk assessment of the ship steering gear failures using fuzzy-Bayesian networks. *Ocean Engineering, 274*, 114064. https://doi.org/10.1016/j.oceaneng.2023.114064

Gordon, T. J., Wolfson, R. J., & Sahr, R. C. (1969). Methodology: A method for rapid reproduction of group data and individual estimates in the use of the Delphi method. *Futures, 1*(6), 552. https://doi.org/10.1016/S0016-3287(69)80045-5

Gülen, M. F., Guedes Soares, C., & Arslan, O. (2026). A comprehensive risk assessment of enclosed space operations on ships. *Reliability Engineering & System Safety, 265*, 111543. https://doi.org/10.1016/j.ress.2025.111543

International Maritime Organization. (2018). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process* (MSC-MEPC.2/Circ.12/Rev.2). London. Retrieved from https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/MSC-MEPC%202-Circ%2012-Rev%202.pdf

International Maritime Organization. (2026). *International code of safety for ships using gases or other low-flashpoint fuels (IGF code)*. Retrieved from https://www.imo.org/en/OurWork/Safety/Pages/IGF-Code.aspx

International Tanker Owners Pollution Federation Limited. (2025). *Oil tanker spill statistics 2024*. ITOPF Ltd. Retrieved from https://www.itopf.org/knowledge-resources/documents-guides/oil-tanker-spill-statistics/

Jang, H., Mujeeb-Ahmed, M. P., Wang, H., Park, C., Hwang, I., Jeong, B., Zhou, P., & Mickeviciene, R. (2023). Regulatory gap analysis for risk assessment of ammonia-fuelled ships. *Ocean Engineering, 287*, 115751. https://doi.org/10.1016/j.oceaneng.2023.115751

Karanovic, V., Ceylan, B. O., & Jocanović, M. (2024). Reliable ships: A fuzzy FMEA based risk analysis on four-ram type hydraulic steering system. *Ocean Engineering, 314*, 119758. https://doi.org/10.1016/j.oceaneng.2024.119758

Karasan, A., Ilbahar, E., Cebi, S., & Kahraman, C. (2018). A new risk assessment approach: Safety and Critical Effect Analysis (SCEA) and its extension with Pythagorean fuzzy sets. *Safety Science, 108*, 173-187. https://doi.org/10.1016/j.ssci.2018.04.031

Khan, R. U., Yin, J., Ahani, E., Nawaz, R., & Yang, M. (2024). Seaport infrastructure risk assessment for hazardous cargo operations using Bayesian networks. *Marine Pollution Bulletin, 208*, 116966. https://doi.org/10.1016/j.marpolbul.2024.116966

Laakso, A., Chaal, M., & Valdez Banda, O. A. (2025). A risk assessment of an autonomous navigation system for a maritime autonomous surface ship. *Journal of Marine Engineering & Technology, 24*(4), 253-269. https://doi.org/10.1080/20464177.2025.2460268

Lankahaluge, A. S., Graham, T., Wang, H., Bashir, M., Blanco-Davis, E., & Wang, J. (2025). Formal safety assessment for ammonia fuel storage onboard ships using Bayesian network. *Journal of Marine Science and Engineering, 13*(4), 768. https://doi.org/10.3390/jmse13040768

Li, G., Zhang, H., Li, S., & Zhang, C. (2025). Risk assessment of hydrogen fuel system leakage in ships based on Noisy-OR gate model Bayesian network. *Journal of Marine Science and Engineering, 13*(3), 523. https://doi.org/10.3390/jmse13030523

Li, P., Wang, Y., & Yang, Z. (2024). Risk assessment of maritime autonomous surface ships collisions using an FTA-FBN model. *Ocean Engineering, 309*, 118444. https://doi.org/10.1016/j.oceaneng.2024.118444

Liu, Y., Xue, Y., Lu, Y., Yuan, L., Li, F., & Li, R. (2024). A dynamic Bayesian network model for ship navigation risk in the Arctic Northeast Passage. *Ocean Engineering, 312*, 119024. https://doi.org/10.1016/j.oceaneng.2024.119024

Mamdani, E. H. (1974). Application of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the Institution of Electrical Engineers, 121*(12), 1585-1588. https://doi.org/10.1049/piee.1974.0328

Mamdani, E. H., & Gaines, R. R. (1981). *Fuzzy reasoning and its applications*. London: Academic Press.

Maritime Technologies Forum. (2024). *Guidelines for the development of liquefied hydrogen bunkering systems and procedures*. Maritime Technologies Forum. Retrieved from https://www.maritimetechnologiesforum.com/en/news-publications/guidelines-for-the-development-of-liquified-hydrogen-bunkering-systems-and-procedures

Mohammadi, M., & Rezaei, J. (2020). Bayesian best–worst method: A probabilistic group decision-making model. *Omega, 96,* 102075. https://doi.org/10.1016/j.omega.2019.06.001

Na, S., Lee, D., Baek, J., Kim, S., & Choung, C. (2025). Qualitative risk assessment methodology for maritime autonomous surface ships: Cognitive model-based functional analysis and hazard identification. *Journal of Marine Science and Engineering, 13*(5), 970. https://doi.org/10.3390/jmse13050970

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49-57. https://doi.org/10.1016/j.omega.2014.11.009

Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega, 64*, 126-130. https://doi.org/10.1016/j.omega.2015.12.001

Rowe, G., & Wright, G. (2001). *Expert opinions in forecasting: The role of the Delphi technique* (pp. 125-144). In Armstrong, J. S. (Ed.). Principles of forecasting. Boston, MA: Kluwer Academic.

Sezer, S. I., & Akyuz, E. (2024a). A conceptual risk modelling for cargo tank fire/explosion in chemical tanker by using Evidential Reasoning SLIM and Bayesian belief network approach. *Reliability Engineering & System Safety, 252*, 110455. https://doi.org/10.1016/j.ress.2024.110455

Sezer, S. I., Camlıyurt, G., Aydin, M., Akyuz, E., Boustras, G., & Park, S. (2024b). A holistic risk assessment under the D-S evidential theory and FMECA approach of ship recycling process hazards in the maritime environment. *Human and Ecological Risk Assessment, 30*(1-2), 201-216. https://doi.org/10.1080/10807039.2024.2312969

Sezer, S. I., Elidolu, G., Akyuz, E., & Arslan, O. (2023). An integrated risk assessment modelling for cargo manifold process on tanker ships under FMECA extended Dempster–Shafer theory and rule-based Bayesian network approach. *Process Safety and Environmental Protection, 174,* 340-352. https://doi.org/10.1016/j.psep.2023.04.024

Sezer, S. I., Elidolu, G., Aydin, M., Ahn, S. I., Akyuz, E., & Kurt, R. E. (2024c). Analyzing human reliability for the operation of cargo oil pump using fuzzy CREAM extended Bayesian Network (BN). *Ocean Engineering, 299*, 117345. https://doi.org/10.1016/j.oceaneng.2024.117345

Skarsvag, H. L., Fyhn, E. H., & Aasen, A. (2024). Influence of ammonia-water fog formation on ammonia dispersion from a liquid spill. *Journal of Loss Prevention in the Process Industries, 92*, 105446. https://doi.org/10.1016/j.jlp.2024.105446.

Tao, J., Liu, Z., Wang, X., Cao, Y., Zhang, M., Loughney, S., Wang, J., & Yang, Z. (2024). Hazard identification and risk analysis of maritime autonomous surface ships: A systematic review and future directions. *Ocean Engineering, 307*, 118174. https://doi.org/10.1016/j.oceaneng.2024.118174

Tekeli, M. M., Gülen, M. F., Akyuz, E., & Teixeira, A. P. (2026). Improved hybrid Z-number Bayesian network approach to predict mooring line failure during cargo operations in ships. *Journal of Ocean Engineering and Science, 11*(1), 141-153. https://doi.org/10.1016/j.joes.2025.10.007

Tuncel, A. L., Sezer, S. I., Elidolu, G., Uflaz, E., Akyuz, E., & Arslan, O. (2024). A rule-based Bayesian network modelling under evidential reasoning theory for risk analysis of anchoring operation in maritime transportation. *Ocean Engineering, 292*, 116521. https://doi.org/10.1016/j.oceaneng.2023.116521

Uflaz, E., Sezer, S. I., Tuncel, A. L., Aydin, M., Akyuz, E., & Arslan, O. (2024). Quantifying potential cyber-attack risks in maritime transportation under Dempster Shafer theory FMECA and rule-based Bayesian network modelling. *Reliability Engineering & System Safety, 243*, 109825. https://doi.org/10.1016/j.ress.2023.109825

Uğurlu, Ö., Aydın, B., Kaya, A. Y., & Tonoğlu, F. (2026). Analysis of cybersecurity threats to ships in maritime transportation and optimal system design on ships. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 240*(4), 805-825. https://doi.org/10.1177/14750902251408643

United Nations Conference on Trade and Development. (2024). *Review of maritime transport 2024: Navigating maritime chokepoints.* United Nations. Retrieved from https://unctad.org/system/files/official-document/rmt2024_en.pdf

Wang, Y., Qian, L., Hong, M., & Li, D. (2025). Navigation risk assessment for ocean-going ships in the north pacific ocean based on an improved dynamic Bayesian network model. *Ocean Engineering, 315*, 119804. https://doi.org/10.1016/j.oceaneng.2024.119804

Appendix Table 1.

T	Technical Errors	M	Mechanical Failures	P	Procedural Deficiencies
T1	Transfer planning error	M1	Level control system malfunction	P1	Ineffective safety meetings
T11	Fuel quantity miscalculation	M11	System calibration or display failure	P11	Incomplete attendance
T12	Incorrect sequencing of tanks	M12	Air or pressure fluctuations in the line	P12	Absence of a list of responsible personnel
T13	Improper sounding records	M2	Remote control system malfunction	P13	Lack of updated responsibility information
T14	Inaccurate trim data of the vessel	M21	Control unit failure	P2	Routine inspection deficiencies
T2	Administrative planning process error	M22	Control switch failure	P21	Non-compliance with the bunkering plan
T21	Personnel lack of awareness regarding responsibilities	M23	Power source failure	P22	Inadequate process reporting during adverse weather and sea conditions
T22	Incomplete review of the checklist	M3	Alarm/warning system malfunction	P3	Emergency response planning deficiencies
T23	Inadequate safety measures	M31	Cable failure	P31	Missing port authority contact details
T231	Barge not being properly moored to the vessel	M32	Level/gauge sensor malfunction	P32	Negligence in emergency shutdown procedures
T232	Unsafe access arrangement between ship and barge	M4	Fuel circuit valve and connection group failures	P33	Insufficient use of personal protective equipment
T233	Lack of static electricity insulation	M41	Worn-out valve seats		
T234	Insufficient lighting at the bunkering station	M42	Broken valve stem		
T235	Safety signs not positioned visibly	M43	Seal failure		
T236	Fuel hoses in poor condition	M44	Loosely tightened bolts and nuts		
T237	Sampling flange and container not properly secured				
T24	Failure to comply with crew working hours				
T3	General communication errors				
T31	Inadequate communication equipment				
T32	Poor coordination between barge and vessel				
T33	Lack of communication among responsible personnel				
T34	Unfamiliarity with emergency communication signals				

Appendix Table 2 Acronyms and abbreviations.

Abbreviation	Definition
ABS	American Bureau of Shipping
BBWM	Bayesian Best–Worst Method
BWM	Best-Worst Method
CR	Consistency Ratio
CREAM	Cognitive Reliability and Error Analysis Method
ETA	Event Tree Analysis
FFTA	Fuzzy Fault Tree Analysis
FIS	Fuzzy Inference System
FMEA	Failure Modes and Effects Analysis
FMECA	Failure Mode, Effects, and Criticality Analysis
FSA	Formal Safety Assessment
FTA	Fault Tree Analysis
IMO	International Maritime Organization
ISM	International Safety Management
ITOPF	International Tanker Owners Pollution Federation Limited
LNG	Liquefied Natural Gas
MASS	Maritime Autonomous Surface Ships
MTF	Maritime Technologies Forum
RPN	Risk Priority Number
SCEA	Safety and Critical Effect Analysis
SIMOPS	Simultaneous Operations
SLIM	Success Likelihood Index Method
SMS	Safety Management System
STPA	Systems Theoretic Process Analysis
STS	Ship-to-Ship
UNCTAD	United Nations Conference on Trade and Development

Appendix Table 3 Linguistic scale abbreviations.

Abbreviation	Definition	Abbreviation	Definition
VL	Very Low	AP	Absolutely Possible
L	Low	SP	Strongly Possible
M	Medium	P	Possible
H	High	I	Impossible
VH	Very High	SI	Strongly Impossible
T	Trivial	Ng	Negligible
S	Significant	Mn	Minor
Mo	Moderate	Mj	Major
C	Catastrophic	Cr	Catastrophic
VR	Very Rare	O	Occasional
R	Rare	Re	Regular
Pe	Permanent		

Appendix Table 4 Mathematical parameters and variables.

Parameter	Definition	Unit
P_i	Probability fuzzy number for factor *i*	Unitless
S_i	Severity fuzzy number for factor *i*	Unitless
F_i	Frequency fuzzy number for factor *i*	Unitless
D_i	Detectability fuzzy number for factor *i*	Unitless
Q_1	First quartile value in expert evaluation	Unitless
M	Median value in expert evaluation	Unitless
Q_3	Third quartile value in expert evaluation	Unitless
R_t	Linguistic scale for risk magnitude	Unitless
$\mu(x)$	Membership degree of the fuzzy variable	Unitless
R_i	Baseline risk magnitude/score for factor *i*	%
R_i^f	Final (adjusted) risk score for factor *i*	%
A^B	Best-to-Others priority vector	Unitless
A^W	Others-to-Worst priority vector	Unitless
a_{BJ}	Preference of the best factor over factor *j*	Unitless
a_{jw}	Preference of factor *j* over the worst factor	Unitless
w_j	Optimal priority weight for factor *j*	Unitless
w_{agg}	Aggregated weights vector in BBWM	Unitless
W^D	Total multiplier score for dimension *D*	Unitless
ξ	Consistency index / maximum error	Unitless
γ	Concentration parameter for Dirichlet distribution	Unitless
x_j	Binary validation indicator (1 if active, 0 otherwise)	Unitless
α	Environmental factors dimension index	Unitless
β	Expertise level dimension index	Unitless
δ	Vessel condition dimension index	Unitless
RM	Risk magnitude obtained from the aggregated expert evaluations	Unitless
ρ	Spearman's rank correlation coefficient	Unitless
p	Significance level of the statistical test	Unitless

Appendix Table 5 Generic and operation-specific elements of the proposed framework.

Phase	Element of the framework	Status
Phase 1	Eligibility criteria and rule governing panel size (Step 1.1)	Generic
Phase 1	Composition of the expert panel (Step 1.1)	Operation-specific
Phase 1	Delphi rounds, feedback procedure, and consensus criterion (Steps 1.2 - 1.5)	Generic
Phase 1	Hazard inventory and hierarchy of risk factors (Step 1.5)	Operation-specific
Phase 2	Parameter set of SCEA: probability, severity, frequency, detectability (Step 2.1)	Generic
Phase 2	Linguistic scales and input membership functions (**Table 2**, **Figure 3**)	Generic
Phase 2	Parameter ratings assigned to each individual hazard (Step 2.1)	Operation-specific
Phase 2	Quartile-based aggregation and consensus criterion (Step 2.2, Eqs. 1 - 4)	Generic
Phase 2	Fuzzy rule base, output scale, and defuzzification (Step 2.3, **Table 3**, **Figure 2**, Eqs. 5 - 6)	Generic
Phase 3	Definition of the contextual dimensions and their sub-factors (**Table 4**)	Operation-specific
Phase 3	BBWM elicitation, optimization, and aggregation (Steps 3.1 - 3.7, Eqs. 7 - 11)	Generic
Phase 3	Numerical values of the multiplier weights (**Table 12**)	Operation-specific
Phase 4	Computation of the final risk score (Eq. 12)	Generic
Phase 4	Operational scenarios selected for evaluation (**Table 13**)	Operation-specific