Ocean Engineering 329 (2025) 121140

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

Team task complexity analysis in MASS operation using a fuzzy TOPSIS method

Juncheng Tao [a,d], Zhengjiang Liu [a,d], Xinjian Wang [a,b,c,d,*], Jiashi Wang [a,d], Shiyi Rao [a], Zaili Yang [b,e,**]

[a] Navigation College, Dalian Maritime University, Dalian, 116026, PR China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, L3 3AF, UK
[c] Seafarers Research Institute, Dalian Maritime University, Dalian, 116026, PR China
[d] Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Dalian, 116026, PR China
[e] Transport Engineering College, Dalian Maritime University, Dalian, 116026, PR China

A R T I C L E I N F O

Keywords:
Maritime safety
Maritime autonomous surface ships
Team task
Task complexity
Fuzzy TOPSIS

A B S T R A C T

Despite the growing interest in Maritime Autonomous Surface Ships (MASS), research on the task complexity in MASS operation remains limited, creating a gap in understanding how to optimize task allocation and system design. This study aims to develop a team task complexity evaluation model applicable to Degree of Autonomy 2 (DoA 2) operation of MASS. Firstly, in the qualitative analysis section, a task complexity framework tailored to the specific features of DoA 2 MASS operational teams is proposed. Secondly, in the quantitative analysis section, an innovative evaluation model based on fuzzy Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is developed. This model involves two main components. In the first part, four key characteristics are quantified to illustrate task complexity across four dimensions for MASS operational teams. In the second part, fuzzy TOPSIS is applied to aggregate and rank these characteristics, using a combined weighting approach based on subjective and entropy weights. Finally, based on the simplified structure and established task complexity framework, nine typical tasks for MASS operational teams are selected for case study to analysis task complexity. The results reveal that MASS operational team tasks involve extensive information exchange, requiring operators with professional knowledge. Among them, team collaboration tasks exhibit the highest complexity. It makes significant contributions to the optimization of task and system design for MASS operation.

1. Introduction

The maritime industry is undergoing a profound technological transformation, with the widespread application of Maritime Autonomous Surface Ships (MASS) introducing new operational modes and management systems that are propelling maritime transport and related industries toward an era of intelligent innovation. Although MASS shows considerable potential in enhancing automation, reducing human operational burdens, and improving shipping safety (Gan et al., 2025; Shu et al., 2024), potential safety issues arising from uncertainty remain paramount for stakeholders (Symes et al., 2024; Tao et al., 2024), particularly the transition to autonomous operations brings new challenges. With the transformation of MASS in operational mode, the traditional need for onboard watchkeeping officers is diminishing, giving rise to the increasing prominence of Remote Operation Centre (ROC), which alters task allocations and personnel configurations within operational teams (Palbar Misas et al., 2024). These changes are impacting workflows (Banda et al., 2018; Li et al., 2019; Storkersen, 2021), resource management (Goerlandt, 2020), and emergency operations (Zhang et al., 2020), requiring new studies to be conducted.

The term “autonomous” does not imply that there are no crew members on board for supervision or control, nor does it mean that no personnel are needed for necessary management or operations. According to the autonomous degree classification provided by the International Maritime Organization (IMO), Degree of Autonomy 2 (DoA 2) MASS is defined as equipping crew members on board but remotely

This article is part of a special issue entitled: Intell Ship Navi published in Ocean Engineering.

* Corresponding author. Navigation College, Dalian Maritime University, Dalian, 116026, PR China.
** Corresponding author. Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, L3 3AF, UK.
E-mail addresses: wangxinjian@dlmu.edu.cn (X. Wang), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.oceaneng.2025.121140
Received 24 November 2024; Received in revised form 24 March 2025; Accepted 1 April 2025
Available online 14 April 2025

controlling and operating from elsewhere, while still requiring humans for monitoring and intervention (IMO, 2018). This classification focuses on personnel configuration but lacks a detailed explanation from the perspective of operational autonomy levels. In this context, the degree of autonomy can be classified into two categories: "Decision Support" and "Automatic" (Rødseth and Nordahl, 2017). The "Decision Support" indicates direct control with no autonomy, while the "Automatic" refers to the ship's ability to perform certain complex operations without human intervention, with remote or direct control only required by the ROC and crew onboard when necessary. Moreover, the operational modes of these remotely controlled MASSs can also been classified as Manual Control (MC), Remote Control (RC), and Autonomous Control (AC) (Burmeister et al., 2014; Fan et al., 2022, 2024). These operational modes define how the ship is controlled during different stages of operation. Each of these modes of control contributes to the operational autonomy of MASS, with the level of human involvement decreasing as the autonomy degree increases. These modes are not mutually exclusive.

Regardless of the degree of autonomy in these ships, compared to traditional ship operational teams, MASS operational teams have undergone significant changes in both task structure and content. As these teams transition from traditional onboard roles to remote operation and monitoring, there have been profound changes in task allocation, operational modes, and team collaboration (Wu et al., 2024). In addition, the growing complexity of MASS operational systems has intensified cognitive load on human operators. While autonomous capabilities in MASS are advancing, critical information processing and decision-making responsibilities continue to rely on humans. This dependency underscores the importance of effective communication and coordination, which have become increasingly reliant on technological tools, adding complexity to task interactions and operational demands. Therefore, task complexity is widely recognized as a key determinant influencing team performance (Almaatouq et al., 2021; Pasarakonda et al., 2021). Whether DoA2 MASS operational teams can effectively manage these complexities and meet the associated challenges remains a critical and unresolved question (Feng et al., 2024; Song et al., 2024). In the meantime, the fast growth, trails and implementation of MASS technologies emphasize the urgency of producing the solution to the question.

Recent studies employing system-theoretic approaches, such as System Theoretic Process Analysis (STPA), have significantly contributed to the understanding of complex interactions within MASS teams and their operational environments. STPA, widely recognized for its role in safety design and hazard identification, has been particularly influential in the safety assessment of MASS (Chaal et al., 2020; Valdez Banda and Goerlandt, 2018; Zhou et al., 2021). Through focusing on system-level hazard identification, STPA has provided a comprehensive framework for analysing the risks associated with MASS operations (Li et al., 2023a). Dghaym et al. (2021) applied STPA to MASS systems with varying Degrees of Autonomy (DoAs), investigating how autonomy impacts operational hazards. Their findings highlighted that, while higher autonomy increases the available measures to mitigate hazards, many of these measures remain unknown, underscoring the complexities of autonomous operations. In addition to STPA, other methodologies, such as Bayesian Networks (BN) and network analysis, have been employed to analyse operational risks in MASS (Basnet et al., 2023; Han et al., 2024; Li et al., 2023b; Sezer et al., 2024). Notedly, the ground breaking work of Ramos et al. (2020a, 2020b) in the field of Human-Machine Interaction (HMI) laid a crucial foundation during the initial phases of MASS development. The Human-System Interaction in Autonomy (H-SIA) method provides a reliable and effective way to identify failure scenarios in MASS operations. From a microscopic perspective, Cheng et al. (2023) integrated human cognitive model with STPA to examine hazards in the human-machine-environment interaction across different operational modes. Cheng et al. (2024) further extended this by applying the H-SIA method to identify potential human errors, quantifying the likelihood of these errors using BN to model the Performance Shaping Factors (PSFs) that contribute to human error. Research shows that one of the central challenges human operators face in such environments is task complexity, a multidimensional concept widely used to describe the difficulty and complexity of tasks (Liu and Li, 2011). The definitions and interpretations of task complexity vary across fields, depending on task characteristics and research contexts. However, regardless of the definition, task complexity generally reflects actual human performance in specific tasks (Ham et al.; Schmidt and Hunter, 1998). Task complexity is typically divided into subjective and objective types based on the dimensions affecting the subject.

Subjective task complexity is tied to the task itself; but also closely related to an individual's situational awareness and workload. For example, the simulation experiment study by Li et al. (2021) demonstrated that as task complexity increases, situational awareness at both individual and team levels decreases, while workload gradually increases. Complex tasks typically require more cognitive resources, often leading to information overload issues. Das and Chernova (2020) pointed out that operators frequently rely on autonomous systems to alleviate workload or improve performance under complex task demands. However, over-reliance on these systems lead operators to overlook system errors, thereby compromising safety (Bansal et al; Poursabzi-Sangdeh et al., 2021). While subjective task complexity reflects situational awareness and workload, it is influenced by individual differences, making it challenging to avoid subjective bias towards objective evaluation. Therefore, this study focuses on the analysis of objective task complexity.

Objective task complexity is independent of the task executor and is quantified based on the inherent characteristics of the task itself. Research by Park et al. (2001) and Park and Jung (2007) introduced the concepts of task complexity, specifically Step Complexity (SC) and Task Complexity (TACOM). The concept of task complexity have been extensively validated in the field of nuclear power (Ham et al., 2011; Park and Jung, 2008; Podofillini et al., 2013) and extended to fields such as aerospace (Zhang et al., 2009) and aviation (Zheng et al., 2015). TACOM, an extension of SC, aims to quantitatively analyse the task complexity of Emergency Operating Procedures (EOP), optimizing procedures to mitigate increased workload, performance degradation, and cognitive overload in response to emergencies and procedural inaccuracies or gaps. Zhang et al. (2009) adapted the TACOM model to the aerospace field through proposing a four-element task complexity quantification method based on the entropy theory. This method extracts complexity elements from multiple dimensions, and applies the weighted Euclidean norms to enable comprehensive analysis. The entropy theory provides a hierarchical perspective for complexity evaluation, enabling the analysis of task characteristics and their interrelations layer by layer, thus revealing latent complexities within systems (Eskov et al., 2017).

Despite the above studies provide a significant theoretical foundation for task complexity analysis, most existing models are tailored to specific fields such as nuclear power, aviation, aerospace, and other industries characterized by high risk and high complexity. These models, which emphasize emergency operations, information processing, and dynamic changes inherent in complex systems, cannot fully capture the unique task complexity challenges faced by DoA 2 MASS operational teams, such as those arising from remote operation, system integration, and information dependency. Such uniqueness will also trigger necessary theoretical innovation of the associated task complexity analysis beyond the state of the art defined by the aforementioned methods. Additionally, this study aims to complement, rather than duplicate, existing studies on hazard identification and risk analysis in MASS. Through focusing on the task complexity of DoA 2 MASS teams, this study seeks to address the gap in understanding failure scenarios caused by a high level of task complexity, which could aid in the development of more effective operational strategies, enhance decision-making processes, and improve training programs for MASS teams. Through quantifying the various dimensions of task complexity, this study provides a systematic approach to identifying critical challenges in MASS operations. To fill this research gap, a complexity evaluation model applicable to MASS operational teams is developed, and structured in two primary sections:

(1) Identification and classification of task complexity characteristics. In this section, four key dimensions representing task complexity for MASS operational teams are extracted based on existing task complexity frameworks. These dimensions capture the task complexity in task scale, logical structure, information interaction, and operational instruments within the task execution process. Identifying and classifying these characteristics enables a comprehensive understanding of the complexity challenges faced by operational teams during task execution.
(2) Quantification of task complexity characteristics and ranking of task complexity. In this section, firstly, the four types of complexity characteristics are quantified based on entropy theory. Secondly, subjective scorings from experts on the weight of task complexity characteristics are collected through fuzzy logic reasoning. Thirdly, combining the subjective expert weights with entropy-based objective data yields a comprehensive weight with an objective basis for task complexity characteristics. Both the fuzzy logic reasoning and comprehensive weighting method contribute to mitigate the subjective bias introduced by the expert scorings and improve the applicability of the evaluation mode. Finally, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is applied to calculate the relative closeness of each task, providing a comprehensively ranking of task complexity for MASS operational teams.

The remainder of this study is as follows. Section 2 reviews existing task complexity frameworks and establishes a task complex framework applicable to MASS operational teams. Section 3 introduces a quantitative evaluation model for task complexity. Section 4 presents a case study with a detailed process for analysing task complexity in the typical tasks. Finally, Section 5 summarizes the overall study.

2. Task complexity framework for MASS operational teams

Existing studies suggest that multiple task complexity characteristics, which fall under the broader concept of task complexity, can potentially impact performance by influencing those executing the tasks (Rein et al.,). The literature that involves these concepts spans various fields, with some areas featuring prototypes that share similarities with MASS operational teams. Table 1 summarizes the representative classification of objective task complexity characteristics in current studies on task complexity.

Compared with other fields, MASS operational teams exhibit distinct characteristics. MASS is operated in complex, dynamic maritime environments for extended periods, aiming to ensure the safety of system and the effectiveness of task execution with minimal or no human intervention. MASS operational teams face significant task complexity at the human, machine and system levels, primarily manifested in remote collaboration requirements, extensive reliance on external information, and the high uncertainty of the operational environment. These tasks involve the coordination of multiple systems and sub-systems, requiring operators to possess advanced operational skills alongside outstanding task management and information processing capabilities. While some parallels can be drawn from task complexity studies in aviation, which typically operate in a three-dimensional airspace with relatively predictable patterns and minimal property damage, MASS operational teams face highly variable marine conditions. Moreover, the integration of multiple subsystems in MASS, ranging from autonomous navigation to cargo management, further amplifies task complexity. Traditional task complexity frameworks often lack comprehensive consideration of these aspects, especially regarding remote operations and multi-system

Table 1
Classification of task complexity characteristics.

Reference	Research field	Task complexity characteristics	Description
Wood (1986)	The fields of human cognition, judgment, and decision-making	Component complexity	Component complexity accounts for the variety of task steps required to complete a task, focusing on the types rather than the quantity of actions. If an action repeats within a task, it is only counted once.
		Coordinative complexity	Coordination complexity highlights the interdependencies between steps, indicating that a longer priority relationship between actions correlates with higher task complexity.
		Dynamic complexity	Dynamic complexity demonstrates the impact of variations in actions and information flows required to complete a task on task complexity.
Campbell (1988) and Zigurs and Buckland (1998)	The fields of human cognition, judgment, and decision-making	Multiple paths	Multiple paths represent the existence of multiple possible paths to achieve the desired outcome during task execution.
		Multiple outcomes	Multiple outcomes represent that task execution involves multiple expected outcomes to be achieved.
		Conflicting interdependence among path	Conflicting interdependencies among paths represent the interdependence relationships that conflict within multiple paths to achieve the expected outcomes.
		Probabilistic linkages	Probabilistic linkages represent the existence of uncertain or probabilistic connections between multiple paths and expected outcomes.
Hendy et al. (1997), Harvey and Koubek (2000) and Rothrock et al. (2005)	The fields of transportation, engineering management, and human factors	Scope	Scope represents the breadth, extent, and overall scale of a task.
		Structurability	Structurability represents the impact of task sequencing and relationships among tasks on complexity.
		Uncertainty	Uncertainty represents the predictability or confidence associated with a task.

(continued on next page)

Table 1 (*continued*)

Reference	Research field	Task complexity characteristics	Description
Ham et al. (2012)	The fields of nuclear power	Size	Size represents the resources required to describe complexity.
		Variety	Variety represents the diversity closely related to the concepts of complexity and entropy.
		Order/ Organization	Order/Organization represents the constraints satisfied by a system.
Bonner (1994)	The field of auditing	Input complexity	Input complexity represents the amount of information input.
		Processing complexity	Processing complexity represents the steps and procedures involved in processing information.
		Output complexity	Output complexity represents the number of goals and solutions required.
Park et al. (2001), Park and Jung (2007), and Jang et al. (2021)	The fields of nuclear power	Step information complexity	Step information complexity represents the complexity arising from the amount of information that humans need to process.
		Step logic complexity	Step logic complexity represents the complexity arising from the sequence of actions that humans need to perform.
		Step size complexity	Step size complexity represents the complexity arising from the number of actions that humans need to complete.
		Abstraction hierarchy complexity	Abstraction hierarchy complexity represents the complexity arising from the knowledge resources required by humans.
		Engineering decision complexity	Engineering decision complexity represents the complexity arising from the cognitive resources required by humans.
Zhang et al. (2009)	The field of aerospace	Complexity of operation step size	Complexity of operation step size represents the amount of operations within an operation unit.
		Complexity of operation logic structure	Complexity of operation logic structure represents the logical sequence of actions within an operational activity.
		Complexity of operation instrument information	Complexity of operation instrument information represents the types and quantities of monitors and controllers in an operating unit.
		Complexity of space mission information	Complexity of space mission information represents the difficulty of the information required to complete an operational unit's task.
Zheng et al. (2015)	The field of aviation	Actions logic complexity of task	Actions logic complexity of task represents the complexity brought by the logical sequence required to perform actions.
		Actions size complexity of task	Actions size complexity of task represents the complexity brought by the number of actions required.
		Information control exchange complexity	Information control exchange complexity represents the complexity brought by information exchanges in human-machine interactions.
		Control mode complexity	Control mode complexity represents the complexity brought by control mode operations in human-machine interactions.

integration. Therefore, a redesigned task complexity framework is essential for MASS operational teams.

Ge et al. (2020) identified widely accepted task complexity characteristic as consisting of the number of components, inputs, products, and the relationships between these three involved in task performance. Firstly, based on this theoretical framework and referring to studies of Hærem et al. (2015); Liu and Li (2012), this study analyses the representative classification of task complexity characteristics in Table 1 to highlight those critical in MASS operational contexts. Secondly, the identified characteristics are refined and classified into five dimensions tailored to the specific needs of MASS operational teams, namely, action structure complexity, logic structure complexity, information resource complexity, instrument operation complexity, and dynamic variability complexity. Among them, action structure complexity is a concrete description of the number of components. Information resource complexity is a concrete description of the number of inputs. Instrument operation complexity and dynamic variability complexity is a concrete description of the number of products. Logic structure complexity is a concrete description of the interrelationships between components, inputs, and products. This classification approach aligns with the core dimensions proposed by Ge et al. (2020) and integrates the task complexity evaluation frameworks proposed by Park and Jung (2007); Park et al. (2001), enabling a classification system that captures the specific complexity requirements of MASS operational team tasks.

2.1. *Action structural complexity*

Action structural complexity stems from the structuralist perspective on task complexity, representing the complexity associated with the quantity or diversity of operations required in an independent event. Key attributes of action structural complexity include component complexity, multiple paths, multiple outcomes, size, step size complexity, complexity of operation step size, and actions size complexity.

In the perspective of structuralist, task complexity is understood through the structure of the task, which is measured by the number of actions it contains and their interrelations. As the number of required actions and sub-tasks grows, so does complexity, particularly when team collaborations are involved, such as in remote control or multi-system integration. For the DoA 2 MASS operational teams, the number of actions not only affects execution time and resource consumption, but also directly affects team workload and efficiency.

The Complexity of Action Size (CAS) is adopted as a primary characteristic to measure action structural complexity in MASS operational team tasks. Higher CAS levels correspond to increased coordination challenges and execution difficulty, requiring the team's enhanced planning and execution capacities to meet task deadlines. Additionally, CAS can be seen as another expression of the diversity of multiple paths or solution options. A higher number of required actions expands the range of potential action choices, which further intensifies complexity as teams select optimal strategies within complex scenarios.

2.2. Logical structural complexity

Logical structural complexity also stems from the structuralist perspective of task complexity, representing the logical sequence and dependencies between actions required to complete an independent event. Key attributes of logical structural complexity include coordinative complexity, conflicting interdependence among path, structurability, order/organization, step logic complexity, complexity of operation logic structure, and actions logic complexity.

The complexity of logical structure is closely related to the coordination requirements among team members during task execution. As logical structural complexity increases, so do the coordination demands, since tight dependencies among actions require team members to follow specific sequences and rules meticulously. For DoA 2 MASS operational teams, high autonomy within systems and complex environmental factors amplify the dependencies between actions. For instance, critical human operations often rely on feedback from multiple systems to make informed decisions. In this context, logical structural complexity not only impacts team coordination efficiency but also directly affects task success. Tasks with complex logical structures require teams with advanced coordination skills and precision in execution to ensure each action proceeds as planned. Additionally, logical structural complexity may also be influenced by external environmental variability. Uncertainties in the environment can further exacerbate the complexity, demanding heightened adaptability and decision-making capability. Therefore, the Complexity of Action Logic Structure (CALS) is adopted as a primary characteristic to measure the task complexity associated with logical structures in MASS team tasks.

2.3. Information resource complexity

Information resource complexity represents the complexity stemming from the information resources required to complete actions. Key attributes of information resource complexity include engineering decision complexity, abstraction hierarchy complexity, step information complexity, complexity of space mission information, and information control exchange complexity.

As the information resource increases, so does the challenge of information processing, especially when there are complex interrelations between information or when information requires frequent updates. However, the complexity of information resource depends not only on the quantity of information but also on dimensions such as diversity, uncertainty, and relevance. DoA 2 MASS operational teams, often dispersed across different geographical locations, rely on remote collaboration, adding layers of complexity to information exchanges during task execution. Team members need to handle substantial information inputs from multiple sources. For instance, DoA 2 MASS operational team tasks involve not only close cooperation among team members but also significant external data exchanges, including environmental sensor data, meteorological information, and feedback from communication systems. Moreover, information complexity is influenced by the uncertainty and complexity of information transmission pathways. On the one hand, various types of errors may occur in the process of information transmission, which may stem from deviations in information understanding, loss or miscommunication of content, or untimely information updates. On the other hand, the processing of tasks highly relies on accurate and timely information flow. However, as the complexity and length of transmission path increase, maintaining information accuracy and timeliness becomes more challenging, which in turn heightens task unpredictability and execution risk. Therefore, the Complexity of Action Information (CAI) is adopted as a primary characteristic to measure the information resource complexity.

2.4. Instrument operation complexity

Instrument operation complexity reflects the complexity of the instruments required to perform a task, encompassing factors such as the type and quantity of instrument, operational complexity, and interoperability between instruments. Key attributes of instrument operation complexity include control mode complexity and Complexity of Operation Instrument Information (COII).

As instrument operation complexity increases, the operational knowledge and skills required of team members also increase, especially when complex operational steps and dependencies exist among instruments. DoA 2 MASS operational team tasks involve multiple operational instruments and systems, which may need to be operated simultaneously. This complexity is evident in the diversity and type of instruments involved, and the dependencies among them, directly impacting task execution difficulty. Additionally, the collaborative complexity among instruments is also a crucial factor, requiring technical proficiency, effective coordination and communication in complex operational environments. Therefore, the COII is adopted as a primary characteristic to measure the instrument operation complexity.

2.5. Dynamic variability complexity

Dynamic variability complexity represents the complexity arising from changes in actions and information during task execution. Key attributes of dynamic variability complexity include dynamic complexity, uncertain or probabilistic linkages, uncertainty, and variety.

The concept of dynamic variability complexity originated from the task complexity framework developed by Wood (1986), which used dynamic complexity to reflect complexity from action variability. Subsequently, Hendy et al. (1997) and Harvey and Koubek (2000) focused on the inherent task attributes, using concept of uncertain or probabilistic linkages proposed by Campbell (1988) as a dimension to measure the complexity of team tasks. In these models, dynamic variability complexity emphasizes the operator's confidence or predictability regarding task outcomes, which is a subjective inference from results to causes. However, such subjective estimations can lack consistent evaluation criteria for the operational teams. While dynamic variability complexity is critical for understanding task execution, its quantification remains challenging due to the high variability and unpredictability associated with changes in task settings and external conditions. Consequently, this study focuses on the inherent structure, informational requirements, and operational logic of tasks, which are relatively stable and objective. This approach means that complexities arising from the fluctuations and uncertainties in task execution, which result in changes and interactions, are not included as a defining characteristic of task complexity for DoA 2 MASS operational teams. This limitation underscores one of the constraints of this study.

Task complexity for DoA 2 MASS operational teams encompasses not only fundamental elements like the number of actions and logical structures but also the complexity of remote collaboration, diverse operational instrument, and complex information transmission paths. Based on the above analysis, this study refines the existing task complexity framework as follows. Firstly, a classification and summary of traditional team task complexity frameworks categorizes current task complexity attributes into five types, detailing representative characteristics of each. Secondly, considering the unique characteristics of DoA 2 MASS operational teams and tasks, four task complexity indicators are identified, namely: CAS, CALS, CAI, COII. Finally, these four indicators are employed as four dimensions to measure task complexity, and the task complexity framework applicable to DoA 2 MASS operational teams is established.

3. Quantitative evaluation model for task complexity

Compared to team tasks in other fields, the complex and dynamic marine environment has led to the diversity and variability of MASS operational team tasks (Wang et al., 2025). This introduces two key requirements for the applicability of quantitative evaluation model. On the one hand, the impact of different task complexity characteristics on task performance varies. On the other hand, the influence of these characteristics on task complexity can shift across different navigation stages, weather conditions, or task types.

The proposed quantitative evaluation model is shown in Fig. 1. Firstly, different task complexity characteristics are quantified through the first-order and second-order entropy of Task Action Control Graph (TACG), the second-order entropy of Information Control Exchange Graph (ICEG), and the second-order entropy of Operation Instrument Information Graph (OIIG). Secondly, since the traditional TOPSIS heavily relies on the weighted Euclidean norm, which is difficult to apply to different scenarios and address the uncertainty and subjectivity inherent in expert judgments. To address these shortcomings, a new weighted approach is employed, assigning weights based on the relative impact of how task complexity characteristics contribute to overall task complexity. This approach includes three parts: subjective weighting, entropy weighting, and comprehensive weighting. Entropy weighting reflects the inherent variability in data, but its limited adaptability fails to account for the varying relative importance of different task complexity characteristics across scenarios. To address this, subjective weights are incorporated to represent expert knowledge and contextual relevance better. To minimize the uncertainty associated with subjective ratings, fuzzy logic reasoning is introduced, utilizing trapezoidal fuzzy numbers and belief degrees to enhance the precision and reliability of expert ratings. Comprehensive weighting combines the subjective weighs and entropy weights, results in the comprehensive weighted value for each task complexity characteristic. Finally, based on that, TOPSIS replaces the traditional Euclidean norm, using Euclidean distances to the defined Positive Ideal Solution (PIS) and Negative Ideal Solution (NIS) to more intuitively highlight differences in team task complexity.

3.1. Quantitative steps for task complexity characteristics

For the events involving human participation, four task complexity characteristic indicators are employed to represent the task complexity of independent events within the team tasks, namely: CAS, CALS, CAI, COII. These indicators are derived from three different graphs using the entropy theory based on the studies of Egbueri (2022); Zheng et al. (2015).

The entropy theory is a method used to assess complexity, commonly applied to quantify the structural and informational complexity within a system. In information theory, entropy is used to measure the disorder or uncertainty within the system. The entropy theory extends the concept of entropy by introducing the first-order entropy and the second-order entropy, allowing for the analysis of structural or informational interactions at different layers within a system. The first-order entropy measures the complexity of basic elements in the system, such as the in-degree and out-degree of an event node. The second-order entropy considers the relationships or interactions among nodes, measuring the complexity of the connections or interactions among nodes. Two different kinds of order entropy can be calculated as in Equation (1):

$$\omega = -\sum_{i=1}^{x} p(A_i)\log_2 p(A_i) \tag{1}$$

where x is the number of node types, $p(A_i)$ is the frequency of node type A_i.

To clarify potential confusion, it is important to distinguish the difference between the first-order, second-order entropy and entropy weights. The first-order and second-order entropy are used to measure the task complexity characteristics. The entropy weights are used to determine the relative importance of different indicators by analysing the variation or dispersion of data. While these methods both originate from the entropy theory, they serve distinct purposes in the evaluation process.

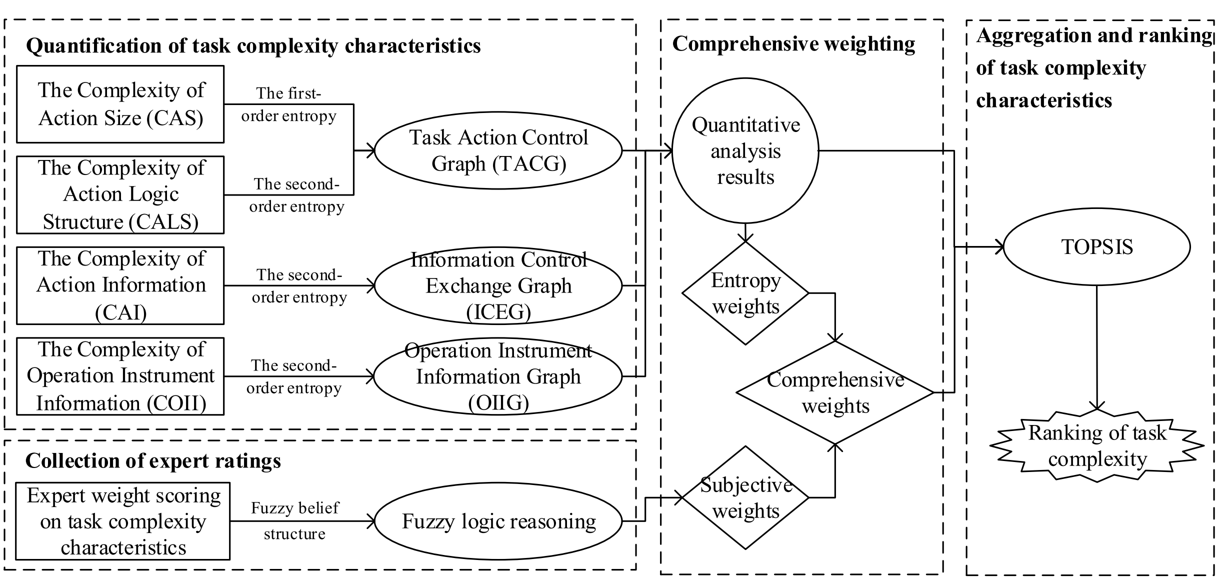

Fig. 1. The proposed quantitative evaluation model for task complexity.

3.1.1. CAS and CALS

CAS refers to the number of actions required to complete a task, while CALS represents the logical sequence and interdependencies among these actions. To quantify both, the TACG is employed in the analysis of the selected task nodes.

TACG is a graphical tool designed to visualize the relationships between the actions involved in task execution. Through outlining the specific steps and operations required in the task, it facilitates the analysis of action sequences, dependencies, and potential interactions. TACG not only captures all the actions necessary for task completion, but also illustrates the logical chain between these actions. Specifically, each node in TACG represents an independent action, with connections between nodes indicate the dependencies or sequential requirements between actions. The generation of TACG involves three steps: task decomposition, action dependency analysis, and visualization. In the task decomposition phase, complex tasks are broken down into multiple independent action nodes, with each node representing a specific operational step or decision point. In the action dependency analysis phase, the logical sequence and dependencies between actions are identified and linked with directed connections. Finally, the visualization phase displays the action nodes and their dependencies in a graphical format, resulting in a complete TACG.

In the analysis of TACG, the first-order entropy measures the number and distribution of actions, reflecting the scale of basic operations required to complete a task. The second-order entropy measures the logical dependencies and interactive complexities between the actions, reflecting the difficulty and coordination requirements of task execution. For example, certain tasks may involve a large number of action steps, yet the logical relationships between these steps may be relatively simple, resulting in a lower value of the second-order entropy. Conversely, tasks with fewer action steps may exhibit high interdependencies, leading to a higher value of the second-order entropy. In the analysis of task complexity for MASS operational teams, CAS and CALS are quantified using the first-order entropy and second-order entropy of TACG, respectively.

3.1.2. CAI

CAI measures the amount and complexity of information required to complete a task. To quantify CAI, the ICEG is employed, which evaluates the complexity of information transmission within the task by calculating its second-order entropy.

ICEG is a tool used to describe the flow and exchange of information during task execution. Through ICEG, it is possible to identify the sources of information, the paths through which information is transmitted, and the final recipients of this information. The generation of an ICEG involves three steps. The first is to identify the various types of information required to perform actions. The second step is to identify the source and recipient of the information. This step is crucial for ensuring the accuracy of information flow visualization. Finally, the specific paths of information flow need to be established to understand how information is transmitted from the source to the recipient.

The information flow path —whether unidirectional or bidirectional—directly impacts the overall complexity of the task. In the analysis of ICEG, the first-order entropy measures the number and distribution of information sources and recipients, while the second-order entropy further focuses on the dependencies and interaction complexity between nodes during information flow. When information transmission paths are numerous and exhibit high interdependencies or frequent interactions, the second-order entropy of ICEG is higher, indicating a higher CAI. Conversely, if the information transmission paths are relatively simple with fewer interactions, the second-order entropy of ICEG is lower, indicating a more straightforward information flow. In the analysis of task complexity for MASS operational teams, CAI is quantified using the second-order entropy of ICEG.

3.1.3. COII

COII focuses on two main aspects: the number and diversity of instruments required for an independent event, and the specific actions and interactions required to operate these devices. To quantify COII, the OIIG is employed which evaluates the complexity of instrument operation through calculating its second-order entropy.

OIIG is a graphical tool designed to analyse the various instruments involved in task execution and their interrelationships. It provides a clear visualization of the types, quantities, and distribution of instruments within the task process. The generation of OIIG involves two steps. In the first step, all operational instruments involved in an independent event are identified which may include autonomous operation systems, various information monitoring systems, communication devices, and other types of tools and equipment. Each instrument may have different operational requirements, technical demands, and procedural steps. In the second step, through outlining the specific steps required for task execution, the interdependencies and interactions among instruments are identified.

In the analysis of OIIG, the second-order entropy quantifies the complexity of interactions between the required information and instruments. If an independent event involves a large amount of information and a high degree of interdependence between instruments, COII for this task increases significantly. The presence of multiple types of information and a high degree of interdependence between instruments involved in an independent event result in a significant increase in COII, which indicates the considerable level of coordination and technical ability required by the operator. The MASS operational team may simultaneously operate multiple complex devices to gather multi-layered information. The complexity of operation instrument not only affects the efficiency of task execution, but may also directly affect the success of the task itself. In the analysis of task complexity for MASS operational teams, COII is quantified using the second-order entropy of OIIG.

3.2. The calculation of comprehensive weighting based on subjective and entropy weights

In this sub-section, the comprehensive weights are obtained through three steps. In the first step, fuzzy logic reasoning is employed to process expert scores on the weights assigned to each task complexity characteristic within a fuzzy belief structure. The subjective weighting ensures experts' fuzzy understanding of task complexity, balancing the quantitative data with experiential insights. In the second step, the entropy weights are calculated based on the quantified objective results gathered in sub-section 3.1. The objective weighting ensures a data-driven perspective, quantifying the range of complexity across tasks. In the third step, the subjective weights are comprehensive weighted with the entropy weights to obtain the comprehensively weighted weights of task complexity characteristics.

3.2.1. The calculation of subjective weights

Fuzzy belief structure is the core of fuzzy logic reasoning, collecting scores on the weights of task complexity characteristics in a fuzzy manner. This structure consists of two components: the linguistic rating assigned by experts to task complexity characteristics and the belief degree of this linguistic rating. The fuzzy belief structure is represented as $\{(H_i, U_i)\}$ (Yang et al., 2006). Based on expert scoring practices, the linguistic ratings H_i are set as $\{H_1, H_2, H_3, H_4, H_5\} = \{\mathrm{VL}, \mathrm{L}, \mathrm{M}, \mathrm{H}, \mathrm{VH}\}$. U_i represents the belief degree. The subjective rating can be expressed as in Equation (2):

$$P_{sj} = \left\{ \left(H_i, U_{sji} \right) \right\} \tag{2}$$

where P_{sj} is the fuzzy belief structure of subjective rating for j^{th} task complexity characteristic from the s^{th} expert. U_{sji} is the belief degree of the linguistic ratings H_i for the j^{th} task complexity characteristic from the

s^{th} expert. The integration of expert ratings in the form of fuzzy belief structure can be calculated as in Equation (3):

$$Q_j = \sum_{s=1}^{m} P_{sj} = \left\{ \left(H_i, \sum_{s=1}^{m} U_{sji} \right) \right\} \tag{3}$$

where Q_j is the subjective score of j^{th} task complexity characteristic in the form of fuzzy belief structure, and m is the total number of experts.

The integration of subjective scores are calculated in the form of fuzzy belief structure to avoid the loss of useful data, ensuring the accuracy of evaluation results. The process of defuzzification is divided into two parts.

In the first part, trapezoidal fuzzy numbers are used to express the linguistic ratings, where VL represents Very Low, L represents Low, M represents Medium, H represents High, and VH represents Very High. Trapezoidal fuzzy numbers allow for a clear representation of the range of values while avoiding the overly complex calculations associated with other forms of fuzzy numbers (Deli and Karaaslan, 2021). The affiliation of linguistic ratings is shown in Fig. 2.

The linguistic ratings H_i is represented by trapezoidal fuzzy numbers which can be expressed as $(a_{i0}, a_{i1}, b_{i0}, b_{i1})$, representing the rating of four vertices of trapezoidal fuzzy number. The defuzzification of trapezoidal fuzzy numbers is shown as Equation (4) (Kutlu and Ekmekçioğlu, 2012):

$$h_i = \frac{(b_{i0} - c) + (b_{i1} - c)}{[(b_{i0} - c) + (b_{i1} - c)] - [(a_{i0} - d) + (a_{i1} - d)]} \tag{4}$$

where h_i is the defuzzification value of linguistic ratings H_i, $i = 1, 2, \cdots, 5$, $c = 0$, $d = 10$ and the specific definition and defuzzification values for linguistic ratings are shown in Table 2.

In the second part, all experts have been assigned the same weight. The defuzzification value of fuzzy belief structure can be calculated as in Equations (5) and (6):

$$\varepsilon_j' = \frac{Q_j}{m} = \frac{\sum_{i=1}^{5} h_i \sum_{s=1}^{m} U_{sji}}{m} \tag{5}$$

$$\varepsilon_j = \frac{\varepsilon_j'}{\sum_{j=1}^{4} \varepsilon_j'} \tag{6}$$

where ε_j' is the defuzzification value of subjective scores before normalization. ε_j is the subjective weight of j^{th} task complexity characteristic after normalization.

3.2.2. The calculation of entropy weighting

The entropy value of the j^{th} task complexity characteristic can be calculated as in Equation (7):

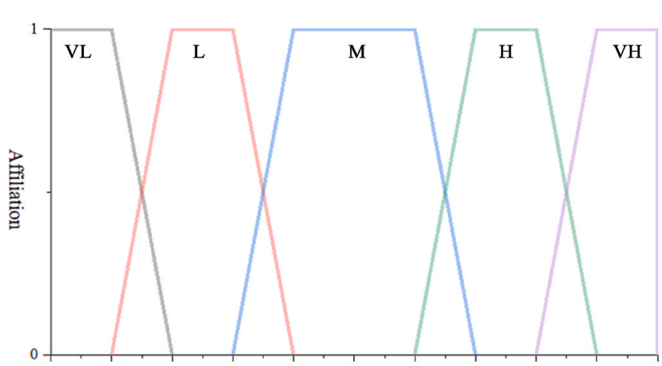

Fig. 2. The affiliation of linguistic rating.

Table 2
Schematic diagram of fuzzy belief structure.

Fuzzy belief structure	Linguistic ratings	Belief degree	Trapezoidal fuzzy numbers	Defuzzification value
$\{(H_1, 1.0)\}$	VL (Very Low)	100 %	(0, 0, 1, 2)	0.130
$\{(H_2, 0.5), (H_3, 0.5)\}$	L (Low) and M (Medium)	The belief degree of linguistic rating L is 50 %, and the belief degree of linguistic rating M is 50 %.	The trapezoidal fuzzy numbers for linguistic rating L are (1, 2, 3, 4), and the trapezoidal fuzzy numbers for linguistic rating M are (3, 4, 6, 7).	0.396
$\{(H_4, 0.4), (H_5, 0.5)\}$	H (High) and VH (Very High)	The belief degree of linguistic rating H is 40 %, the belief degree of linguistic rating VH is 50 %, and the remaining 10 % belief degree is evenly distributed among each linguistic rating.	The trapezoidal fuzzy numbers for linguistic rating H are (6, 7, 8, 9), and the trapezoidal fuzzy numbers for linguistic rating VH are (8, 9, 10, 10).	0.768

$$E_j = -\frac{1}{\ln(g)} \sum_{i=1}^{g} (r_{ij} \ln(r_{ij})) \tag{7}$$

where g is the total number of tasks, and r_{ij} is the quantified result of the j^{th} task complexity characteristic for the task i after normalization.

The entropy weight of each task complexity characteristic can be calculated as in Equation (8):

$$\alpha_j = \frac{1 - E_j}{\sum_{j=1}^{4} (1 - E_j)} \tag{8}$$

where α_j is the entropy weight of the j^{th} task complexity characteristic, E_j is the entropy value of the j^{th} task complexity characteristic.

3.2.3. The calculation of comprehensive weighting

Comprehensive weighting based on subjective weight collected from expert scoring and entropy weight can be calculated as in Equation (9):

$$\beta_j = \frac{\varepsilon_j \alpha_j}{\sum_{j=1}^{4} \varepsilon_j \alpha_j} \tag{9}$$

where β_j is the comprehensive weights of the j^{th} task complexity characteristic.

3.3. The aggregation and ranking of task complexity characteristics based on TOPSIS

In this sub-section, using TOPSIS, the relative closeness of each task is calculated to aggregate the four types of task complexity characteristics and rank task complexities accordingly. TOPSIS enables the evaluation of tasks in a manner that reflects both their proximity to an ideal solution and their performance relative to each other, providing a comprehensive ranking system that reflects the overall complexity of tasks within the operational framework of MASS teams. The specific calculation steps of aggregating and ranking CAS, CALS, CAI, and COII are as follows:

(1) Construct the decision matrix

In order to provide a comprehensive consideration of both subjective scores and objective data, a decision matrix is constructed. The decision matrix is composed of the quantified results in Section 3.1 which have been normalized, and is as demonstrated in Equation (10):

$$R = \begin{array}{c c} & \begin{array}{cccc} \omega_{CAS} & \omega_{CALS} & \omega_{CAI} & \omega_{COII} \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ \vdots \\ t_g \end{array} & \begin{bmatrix} (r_{11}) & (r_{12}) & \cdots & (r_{14}) \\ (r_{21}) & (r_{22}) & \cdots & (r_{24}) \\ \vdots & \vdots & (r_{ij}) & \vdots \\ (r_{g1}) & (r_{g2}) & \cdots & (r_{g4}) \end{bmatrix} \end{array} \tag{10}$$

where t_g is the g^{th} task within the analysis. For instance, r_{21} is the normalized value of ω_{CAS} for Task 2 which is obtained through the first-order entropy of TACG.

(2) Calculate Euclidean distance

The PIS A^+ and NIS A^- are used to calculate the Euclidean distance for each alternative. The PIS A^+ indicates the best possible value for each criterion, representing the optimal alternatives that achieve maximum benefit or minimum cost. Conversely, the NIS A^- indicates the worst possible value for each criterion, representing the alternatives that achieve minimum benefit or maximum cost. Therefore, the Euclidean distances to the PIS A^+ and NIS A^- can be calculated as in Equations (11) and (12) (Chen, 2000):

$$d_i^+ = \sqrt{\sum_{j=1}^{4} \left(r_{ij}\beta_j - A^+\right)^2} \tag{11}$$

$$d_i^- = \sqrt{\sum_{j=1}^{4} \left(r_{ij}\beta_j - A^-\right)^2} \tag{12}$$

where d_i^+ and d_i^- are the distance from the task i to the best and worst alternatives, respectively. $A^+ = (1,1,1,1)$ and $A^- = (0,0,0,0)$.

(3) Calculate the relative closeness

The core step of TOPSIS is used to calculate the relative closeness of each task as shown in Equation (13):

$$TC_i = \frac{d_i^-}{d_i^- + d_i^+} \tag{13}$$

where TC_i is the ratio of the distance to the NIS A^-, representing the relative task complex of MASS operational team tasks.

4. Case study

Whether or not crew members are onboard, the role of the ROC is unquestionable within the entire operational loop. Rødseth and Nordahl (2017) defined three key roles within the ROC: the master, the chief engineer, and the watchkeeping officer. They argued that this team configuration satisfies the equivalent requirements of current personnel allocation regulations and represents the minimum team configuration for MASS shore-based control. This discussion of the minimum configuration stems from academic concerns regarding Rule 5 of the 1972 International Regulations for Preventing Collisions at Sea (COLREG), which states that every ship must always maintain proper visual and auditory observation and take all appropriate available measures to assess the situation and collision risk under the current conditions. However, the personnel configuration modes vary for MASS depending on the operational model and the degree of autonomy.

The MUNIN project emphasizes that ROC is enabled to maintain centralized management of multiple autonomous vessels. A small number of remote operators manage several MASSs simultaneously (Fan et al., 2022; Magnhild and Braseth, 2020). This one-to-many personnel configuration mode has gained wide recognition and practical support in academia, particularly regarding how to effectively allocate tasks and ensure safe operations. Man et al. (2018) developed a remote monitoring simulation experiment based on a mature bridge system, investigating human factors during remote monitoring tasks for ship-to-shore systems. In this scenario, remote operators were required to monitor six autonomous vessels. Veitch et al. (2024) conducted an in-depth analysis of two simulated scenarios involving handover and takeover processes, evaluating the impact of the number of ships being monitored on human handover and takeover performance. The experiment creates a comparison to explore how monitoring multiple vessels affects human performance. It is noteworthy that the detailed descriptions of team configurations and the geographical distribution of teams and systems involved are not mentioned in most of the studies discussed above, and the role of human supervisor in these studies is to provide the overall plan to the systems on ship. This leads to significant differences in the constructions of MASS described in current studies, even when these studies focus on the MASS with the same DoA.

The Structure and Task (ST) diagrams for the departure, navigation, and berthing phases of MASS developed by Shiokari et al. (2024) are adopted in this study to extract typical tasks for the DoA 2 MASS operational team. The ST diagrams outline the overall structure of DoA 2 MASS systems, detailing each component's tasks, the information required for task execution, and the inter-component interactions, which support structure model–based hazard identification for MASS. In this study, the ST diagrams focus on four primary categories of static structural information, namely, the explicit conceptual model, operational team members, tasks required of operational team members, and the form of task execution. This ST framework aids in identifying typical tasks for the MASS operational team, as shown in Fig. 3.

The process of extraction is detailed as follows:

(1) Generalisation of ST diagrams

The original ST diagrams define two locations, namely the ship and the ROC. Thus, the framework is divided into two modules.

The ship module consists of the crew, Autonomous Operation Information Display System (AOIDS), Ship Manoeuvring System (SMS), and data acquisition systems. Among them, the ship is manned by only one crew members. AOIDS is responsible for displaying information to this crew member, such as approved manoeuvring plans, evasive action plans, and alert information. SMS executes the approved manoeuvring plans and makes diagnoses of actuators and itself. The data acquisition systems are responsible for collecting various types of information, such as own ship, obstacles, ship-land distance, weather, and cargo condition information, which interact solely with the SMS without human involvement, are excluded. Therefore, the simplified ST framework includes only the crew, AOIDS, and SMS components in the ship module, depicted in blue components in Fig. 3.

The ROC module consists of the master, Autonomous Operation System (AOS), Weather Information Display System (WIDS), Cargo Condition Monitoring System (CCMS), and Operational Design Domain-Monitoring System (ODD-MS). Among them, the ROC is manned by one master. AOS is responsible for integrating the received information. If necessary, AOS will modify the navigation plan and provide reasons for the modification to the master, send the approved navigation plan to AOIDS and SMS, and display integrated information, etc. As WIDS, CCMS, and ODD-MS all function as monitoring and display systems, they are further simplified into a single component, depicted in green components in Fig. 3. In this structure, the human team consists of the master in the ROC and the crew onboard the ship, depicted in elliptical

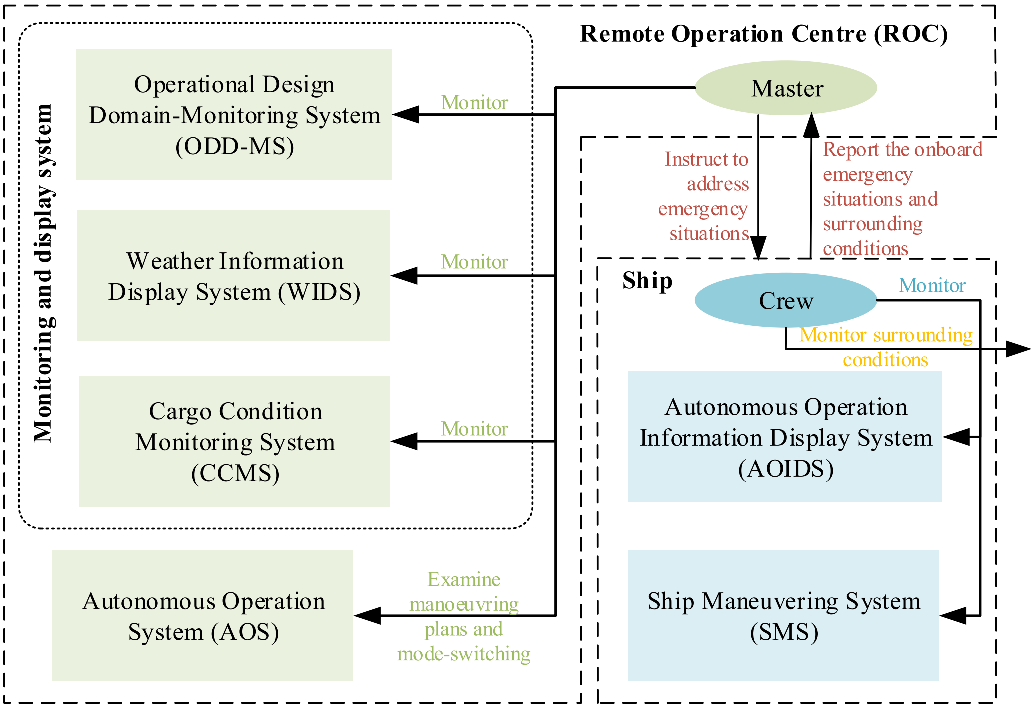

Fig. 3. The ST framework.

component in Fig. 3.

(2) Extraction of task interactions

The original ST diagrams identify three types of task interactions: manual operations or manual information transmissions, automatic information transmissions, and automatic operations. Among these, manual operations or manual information transmissions refer to task interactions with humans as the main body. Automatic information transmissions refer to task interactions with systems as the main body. Automatic operations refer to the control of engine and rudder actuators over the engine and rudder. Through the extraction of manual operations or manual information transmissions, human-centred task interactions are categorized into four types: internal ship interactions, internal ROC interactions, interactions between the ship and ROC, and ship-to-external interactions. These interactions are represented by the blue, green, red, and yellow arrows, respectively, in Fig. 3.

The task interactions within ROC involve the master gathering information from three monitoring and display systems, and approving and switching modes through AOS. The task interactions within ship involve the crew gathering information from the AOID and SMS. The two-way task interactions between the ship and ROC involve the crew reporting onboard emergencies and surrounding information to the master, while the master instructs the crew to address emergencies. The task interaction between ship and external tasks involves the crew gathering information from observing the surrounding environment of ship.

Based on the above analysis, nine typical tasks of a MASS operational team have been extracted, as shown in Table 3.

Table 3
Typical tasks of the MASS operational team.

Team members	Locations	No.	Describition
Master	ROC	Task 1	Monitoring CCMS.
		Task 2	Monitoring WIDS.
		Task 3	Monitoring ODD-MS.
		Task 4	Approving and switching modes through AOS.
		Task 5	Instructing to address emergency situations.
Crew	Ship	Task 6	Monitoring AOIDS.
		Task 7	Monitoring SMS.
		Task 8	Reporting onboard emergencies and surrounding information.
		Task 9	Observing the surrounding environment of ship.

4.1. *Quantification process of task complexity characteristics*

Expert knowledge is gathered through a two-stage process. In the first stage, a panel of eight experts is established, comprising professionals with extensive maritime and operational experience. Each expert boasts a minimum of 10 years of professional experience, including maritime regulatory officials, experts in maritime safety, and senior officers with practical operational experience. This expert panel plays a crucial role in the construction of TACG, ICEG, and OIIG, ensuring a well-founded foundation for quantifying task complexity in MASS operations.

(1) Quantification of CAS and CALS

For Task 5, “Master: Instructing to address emergency situations”, the process of constructing the TACG is as follows:

Step 1: Collecting task information, physical media, and operation details.

This step aims to collect all relevant information required for task execution. The necessary task information, instrument operation details, and specific operation details are obtained through expert panel discussion and relevant literature.

For task 5, the physical media used by the master include monitoring and display systems, AOS, and communication device. The detailed description of task 5 is as follows. Firstly, the master receives an alert, which may originate from either the AOS or the monitoring and display systems. Secondly, the master confirms this alert information and reviews the ship's status. Finally, with the support of AOS, the master instructs the crew to respond and performs any required actions on AOS.

Step 2: Decomposing the task into action nodes.

This step aims to transform a complex, multi-step task into a series of simple, executable action nodes. This process involves a thorough analysis of the task flow and operational logic. Firstly, the overarching goal of Task 5 is clarified, which is to ensure that the master effectively instructs and assists the crew in responding to emergency situations. This goal provides clear guidance for task decomposition, ensuring the design and execution of each action to revolve around this central purpose. Secondly, the task is broken down into a series of logical steps. According to the detailed description of Task 5, it is decomposed into six logical steps, namely: receiving alerts, reviewing alerts, confirming status, obtaining suggestions, coordinating, and implementing. Finally, as each action node involves a specific input and output of information, the direction and content of the information flow determine the nature of the task. Through comparing the logical steps with their associated information inputs and outputs, a complete set of action nodes is constructed, as shown in Table 4.

Step 3: Visualizing the TACG based on action logic

The information inputs and outputs serve as the basis for determining the logical relationships between action nodes. Based on the information in Table 4, these relationships are visualized by constructing the TACG, as shown in Fig. 4.

Table 4
Action decomposition of independent events.

No.	Action nodes	Description	Physical media
1	Confirm receipt of alarms.	Receive alerts from monitoring and display systems or other sources, and click to confirm receipt of the alert.	Monitoring and display systems and AOS
2	Check the alarm information.	View the specific information of the alarm.	Monitoring and display systems
3	Confirm the ship's navigation information.	Confirm whether the alarm information corresponds to the ship information.	Monitoring and display systems
4	Check the surrounding environment, traffic conditions, and the condition of the vessel itself.	Confirm the status and relevant conditions of the current emergency situation.	Monitoring and display systems
5	Get advice on AOS.	Check whether the AOS considers it a threat, whether it is necessary to change the manoeuvring plan, and obtain suggestions for emergency response plans.	Monitoring and display systems, AOS, and communication device
6	Communicate with crew.	Coordinate crew and prepare to implement emergency plans.	Communication device
7	Implementing the tasks.	Choose the appropriate action according to the emergency situation	Monitoring and display systems and AOS

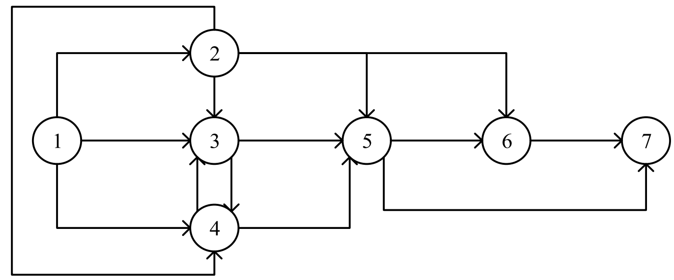

Fig. 4. Task action control graph for task 5.

CAS and CALS are derived from the TACG of the corresponding task, using the entropy theory. Among them, the first-order entropy of TACG represents the CAS of task, which is based on the in-degree a^{in} and out-degree a^{out} of each node. Nodes with the same in-degree and out-degree are grouped into the same category. As shown in Fig. 4, nodes 3, 4, and 5 share an in-degree of 3 and an out-degree of 2, categorizing them into the same node class. The remaining nodes, each with unique in-degrees and out-degrees, form individual classes. The in-degree and out-degree of nodes in TACG are shown on the left side of Table 5. Therefore, the TACG of task 5 comprises five categories of action node classes. The CAS of task 5 can be calculated using Equation (1) as follows:

$$\omega_{CAS} = -\sum_{i=1}^{7} p(A_i)\log_2 p(A_i) = -\left[4*\left(\frac{1}{7}\log_2\frac{1}{7}\right) + 1*\left(\frac{3}{7}\log_2\frac{3}{7}\right)\right] = 2.128$$

The second-order entropy of TACG represents the CALS of task, analysing the connections between every pair of nodes in the network. It determines the frequency or probability distribution of these connections, assessing the interactions or connection types and their frequency among actions. The second-order entropy is typically related to the first-order entropy, where the first-order entropy evaluates the structural characteristics, while the second-order entropy focuses on the relational patterns among nodes. Nodes are considered equivalent if they share the same type and number of neighbouring nodes. The statistical results for these relational patterns are shown on the right side of Table 5. The CALS of task 5 can be calculated using Equation (1) as $\omega_{CALS} = 2.585$.

(2) Quantification of CAI

The ICEG, constructed with physical media as nodes and based on information input and output relationships, is shown in Fig. 5. The CAI of task 5 can be calculated as $\omega_{CAI} = 2.75$.

(3) Quantification of COII

Table 5
The statistical results for the first and second order entropy of TACG.

The first-order entropy of TACG				The second -order entropy of TACG		
Category	Nodes	In-degree	Out-degree	Category	Nodes	Neighbouring nodes
1	Node 1	0	3	1	Node 1	Node 2, 3, and 4
2	Node 2	1	4	2	Node 2	Node 1, 3, 4, 5, and 6
3	Node 3	3	2	3	Node 3	Node 1, 2, 4, and 5
	Node 4			4	Node 4	Node 1, 2, 3, and 5
	Node 5			5	Node 5	Node 2, 3, 4, 6, and 7
4	Node 6	2	1	6	Node 6	Node 2, 5, and 7
5	Node 7	2	0	7	Node 7	Node 5 and 6

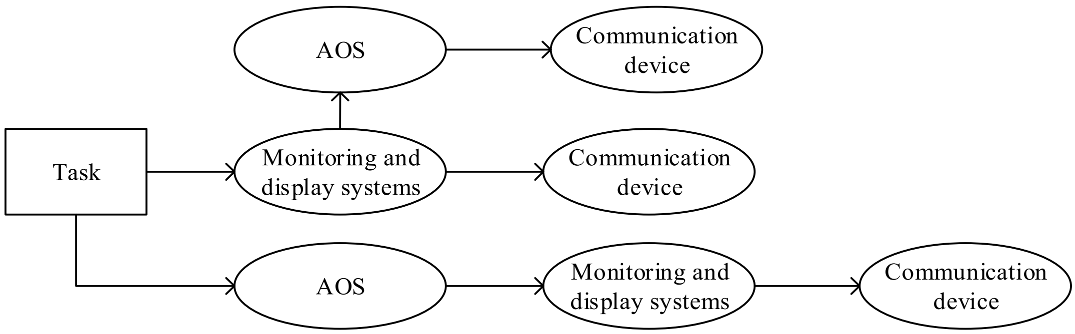

Fig. 5. Information control exchange graph for task 5.

The OIIG for Task 5, shown in Fig. 6, outlines the required operational instruments, encompassing three distinct types of physical media. The monitoring and display systems are involved in action Nodes 1–5 and 7, supporting the reception and transmission of information provided by the autonomous system. The communication devices are involved in action Nodes 5 and 6 to facilitate communication between the master and crew. The AOS is involved in action Nodes 1, 5, and 7, providing reference suggestions for situational response.

The COII of task 5 can be calculated as $\omega_{COII} = 2.252$.

Based on the above calculation process, the value of four task complexity characteristics across all typical tasks are obtained, as listed in Table 6.

Table 6
The quantified results of task complexity characteristics for all typical tasks.

No.	CAS	CALS	CAI	COII
Task 1	2.000	1.500	3.122	2.642
Task 2	2.000	1.500	2.585	2.236
Task 3	2.252	2.585	3.122	2.725
Task 4	2.642	2.948	2.807	2.550
Task 5	2.128	2.585	2.750	2.252
Task 6	1.585	1.585	2.000	1.880
Task 7	0.750	2.000	2.322	2.252
Task 8	2.000	2.000	2.522	2.252
Task 9	1.922	2.322	2.406	2.725

4.2. *Comprehensive weighting based on subjective and entropy weights*

Although objective data on task complexity has been obtained, integrating expert subjective ratings remains essential. Objective data, derived from quantifiable metrics such as task structure, information flow, and operational steps, may not fully capture cognitive workload, decision uncertainty, or situational complexity. Task complexity is context-dependent, varying with operational environments, DoAs, and human factors. Expert scoring provides critical insights into implicit factors such as decision difficulty, team coordination challenges, and scenarios. Moreover, a weighted combination improves the robustness and generalizability of the model.

4.2.1. *Subjective weighting of task complexity characteristics*

In the second stage of expert knowledge collection, the survey questionnaire is initially distributed to a panel of eight experts, then extended to a broader group of 23 experts renowned for their extensive academic and professional expertise in maritime operations and human factors. This survey seeks to solicit expert opinions on validating the constructions of TACG, ICEG, and OIIG, and to assess the four task complexity characteristics within the fuzzy belief structure through an online questionnaire.

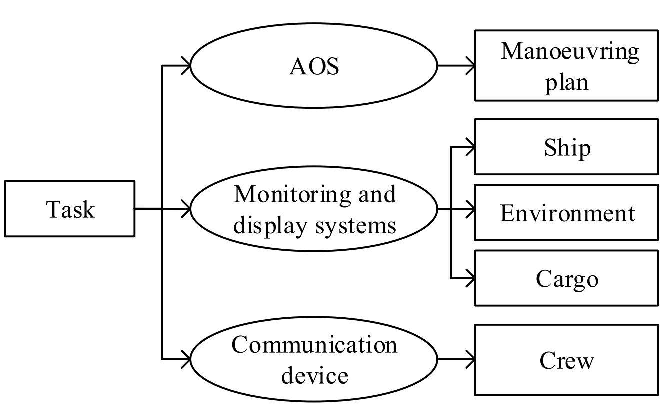

Fig. 6. Operation instrument information graph for task 5.

The judgements of experts for each task complexity characteristic in the fuzzy belief structure are shown in Table A1 of Appendix A.

Taking the subjective ratings of CAS as an example, using Equation (5), the subjective ratings provided by experts are defuzzied as follows, yielding quantitative scores reflecting their assessments of the four task complexity characteristics.

$$\varepsilon_1' = \frac{h_1 \times 0.6 + h_2 \times 7.9 + h_3 \times 14.25 + h_4 \times 7.05 + h_5 \times 1.2}{31} = 0.50146$$

After normalization, the experts' subjective weights for each characteristic are 0.261, 0.324, 0.225, and 0.190.

4.2.2. *Entropy weighting of task complexity characteristics*

Through the normalization of the quantitative analysis results in Table 6, the objective scores of task complexity characteristics for each task form the decision matrix, as listed in Table 7.

Table 7
The quantified results of task complexity characteristics after normalization.

No.	Description	CAS	CALS	CAI	COII
Task 1	Master monitors CCMS.	0.216	0.162	0.337	0.285
Task 2	Master monitors WIDS.	0.240	0.180	0.311	0.269
Task 3	Master monitors ODD-MS.	0.211	0.242	0.292	0.255
Task 4	Master approves and switches modes through AOS.	0.242	0.269	0.256	0.233
Task 5	Master instructs to address emergency situations.	0.219	0.266	0.283	0.232
Task 6	Crew monitors AOIDS.	0.225	0.225	0.284	0.266
Task 7	Crew monitors SMS.	0.102	0.273	0.317	0.308
Task 8	Crew reports onboard emergencies and surrounding information.	0.228	0.228	0.287	0.257
Task 9	Crew observes the surrounding environment of ship.	0.205	0.247	0.257	0.291

Taking the entropy weighting of CAS as an example. The entropy of CAS can be calculated using Equation (8) as follows:

$$E_1 = -\frac{1}{\ln(9)}(0.216\ln(0.216) + 0.240\ln(0.240) + \cdots + 0.205\ln(0.205)) = 1.323$$

The entropy values for those four task complexity characteristics are 1.323, 1.377, 1.467, and 1.439, respectively.

The entropy weight of CAS can be calculated using Equation (9) as follows:

$$\alpha_1 = \frac{1-E_1}{\sum_{j=1}^{4}(1-E_j)} = \frac{1-1.323}{(1-1.323)+(1-1.377)+(1-1.467)+(1-1.429)} = 0.202$$

The corresponding entropy weights for those four task complexity characteristics are 0.202, 0.236, 0.293, and 0.269, respectively.

4.2.3. Comprehensive weighting of task complexity characteristics

Taking the comprehensive weighting of CAS as an example. The comprehensive weight of CAS can be calculated using Equation (10) as follows:

$$\beta_1 = \frac{\varepsilon_1\alpha_1}{\sum_{j=1}^{4}\varepsilon_j\alpha_j} = \frac{0.261\times0.202}{0.261\times0.202+0.324\times0.236+0.225\times0.293+0.190\times0.269} = 0.214$$

The comprehensive weights for those four task complexity characteristic are 0.214, 0.311, 0.268, and 0.207, respectively.

4.3. Aggregation and ranking of task complexity characteristics

Finally, using Equations (11)–(13), along with the decision matrix presented in Table 7 and the comprehensive weights in Section 4.2.3, the relative closeness for each task is calculated, resulting in a comprehensive ranking as shown in Table 8.

Table 8
Comprehensive ranking of task complexity.

No.	Description	d_i^-	d_i^+	TC_i	Ranking
Task 1	Master monitors CCMS.	0.1277	1.8774	0.063688	5
Task 2	Master monitors WIDS.	0.1258	1.8770	0.062812	9
Task 3	Master monitors ODD-MS.	0.1289	1.8745	0.064341	4
Task 4	Master approves and switches modes through AOS.	0.1293	1.8741	0.064540	3
Task 5	Master instructs to address emergency situations.	0.1308	1.8735	0.065260	2
Task 6	Crew monitors AOIDS.	0.1266	1.8755	0.063234	8
Task 7	Crew monitors SMS.	0.1377	1.8730	0.068484	1
Task 8	Crew reports onboard emergencies and surrounding information.	0.1271	1.8752	0.063477	7
Task 9	Crew observes the surrounding environment of ship.	0.1273	1.8753	0.063567	6

4.4. Results and discussion

4.4.1. Research findings

The quantitative evaluation for the nine typical tasks of MASS operational teams shows that Task 7 (crew monitors SMS), Task 5 (master instructs to address emergency situations), and Task 4 (master approves and switches modes through AOS) have the highest task complexity scores. Task 7, involving extensive monitoring of the SMS, reflects higher action logic and information complexity due to its reliance on real-time data interpretation and coordination. Among these high-complexity tasks, two are executed by the master, requiring continuous and detailed monitoring of the operational processes, and decision making. Task 4, while involving critical decision-making and information flow, benefits from clear procedural guidelines that reduce its task complexity, supporting findings that the effectiveness and reasonableness of such procedural guidelines effectively reduces task complexity (Ge et al., 2020). However, this constant need for data integration and real-time decision-making significantly increases the action logic complexity, as the operator must make immediate judgments based on fluctuating conditions. The complexity of communication and coordination between different system components contributes to the task's information complexity. Furthermore, the established guidelines or automated response in Task 5 constitutes a clear response procedure, which studies have shown that may reduce physical and cognitive load on operators (Jang et al., 2021). However, the increased action logic and information flow within these procedures continue to contribute to higher task complexity.

In comparison, Task 3 (master monitors ODD-MS), Task 1 (master monitors CCMS), and Task 9 (crew observes the surrounding environment of ship) have moderate complexity scores. Intuitively, the complexity of these tasks likely stems from the integration of data from multiple subsystems and complex environment, which requires operators to continuously process, update, and analyse information from various sources, as well as to make dynamic adjustments. For Task 3 and 1, such tasks not only require operators processing significant data, but also require dealing with real-time changes and potential abnormal situations. However, as these tasks typically do not demand immediate decision-making or continuous adjustments unless hazards are detected, their complexity is often not immediately apparent. Therefore, the complexity may come from responding appropriately to potential hazards. For complex tasks, existing study suggests that in more complex tasks, humans are more inclined to rely on autonomous systems due to information overload (Das and Chernova, 2020). However, the increase in information content itself is a critical factor contributing to the increase in task complexity (Chan et al., 2015). This dual relationship indicates that while autonomous systems can alleviate cognitive load of operators, the growth in information content still poses substantial challenges to operators.

Task 8 (crew reports onboard emergencies and surrounding information), Task 6 (crew monitors AOIDS), and Task 2 (master monitors WIDS) have the lowest complexity scores. In these tasks, complexity may arise from requirements to process, analyse, and forecast information. For Task 8, the role of crew is largely restricted to observing the situation and conveying information to the appropriate channels. Since the crew does not need to engage in complex judgment calls or make real-time decisions unless an emergency situation arises, the complexity is relatively low. Compared to other tasks involved monitoring, the CAS of Task 6 and 2 show minimal differences. The substantial disparity in task complexity may stem from the limited information conveyed by systems like AOIDS and WIDS, or from the ease of understanding this information. Specifically, these tasks require relatively less professional knowledge and are more straightforward to understand. In addition, for the same reasons, involving passive monitoring rather than active decision-making leads to low task complexity.

4.4.2. Discussion

The task complexity evaluation model in this study consists of both qualitative and quantitative analyses.

In the qualitative analysis, well-defined task complexity frameworks are used to establish the task complexity framework applicable to DoA 2 MASS operational teams. Firstly, five dimensions representing task complexity characteristics are extracted through referencing frameworks from other fields. These dimensions characterize the complexity in terms of action structure, logical structure, information resources, operational instrument, and dynamic variabilities. Secondly, specific task complexity characteristics are selected based on the characteristics of MASS operational teams and their tasks. Finally, a task complexity framework applicable to DoA 2 MASS operational teams is developed. This framework represents task complexity across four dimensions to fill the gap in task complexity analysis within the MASS field, contributing to the development of more nuanced guidelines for task allocation.

In the quantitative analysis, a quantitative evaluation model for task complexity is constructed to objectively evaluate the complexity of typical tasks using objective measurement criteria. Traditional methods for quantifying task complexity characteristics are applied to obtain objective scores for these characteristics. To accommodate various contexts in task complexity analysis, a comprehensive weighting approach based on subjective and entropy weights is employed to assign weights to each of task complexity characteristic. One the one hand, entropy weights reflect the degree of dispersion of these characteristics, and larger entropy weights usually mean higher uncertainty and more significant complexity of tasks. On the other hand, expert scoring for each characteristic reflects their experience, expertise, and understanding of task complexity, which is essential for capturing the nuanced aspects of the tasks, particularly regarding future, unknown aspects of tasks. While expert scoring provides valuable insights into task complexity, it also introduces bias into the results. To mitigate this issue, fuzzy logic reasoning is applied to derive expert scores, thereby reducing the potential for subjective bias. Furthermore, the comprehensive weighting approach balance the influence of subjective and objective information and improves applicability across various contexts. The TOPSIS further accommodates the ambiguity and uncertainty in the evaluation process, providing a more detailed and intuitive evaluation of task complexity across multiple dimensions. However, it must be acknowledged that the two-stage collection of expert knowledge inevitably influenced the results subjectively.

The typical tasks in the case study are extracted from the ST diagrams in existing literature (Shiokari et al., 2024). The original ST diagrams includes a comprehensive description of the internal components and interactions within the MASS system. To facilitate the analysis of task complexity, the extensive and complex structure of the original ST diagrams is simplified, focusing on components and task interactions relevant to human roles. This simplification process identifies nine typical tasks for the DoA 2 MASS operational team. These tasks, grounded in a specific operational context, are a macro-level representation of the possible behaviours and actions that humans may perform within a MASS operational team.

The case study results reveal that tasks requiring collaboration between the master and crew generally exhibit higher complexity due to the significant reliance on real-time information exchange and synchronization. For complex tasks, team collaboration often proves more efficient than individual efforts (Almaatouq et al., 2021), as it helps distribute the information processing burden and improves resilience to uncertain situations.

Studies have shown that clear and effective response procedures can reduce cognitive load and simplify the logical structure of tasks in operations (Lee et al., 2005). The predefined procedures or guidelines play a crucial role in balancing system requirements with operator capabilities, as standardized processes or guidelines minimize deviations caused by individual differences (Cao et al., 2025). However, the high complexity of task 5 indicates that while the logical structure can mitigate task complexity to some extent, its impact is limited. Predefined procedures or guidelines are not universally applicable, especially given the inherently complex and dynamic nature of MASS. This necessitates adaptable response mechanisms, suggesting that predefined procedures should not be rigid. To address this, operators need to retain a degree of flexibility during task execution (Storkersen, 2021), which inevitably increases the uncertainty and task complexity.

Additionally, specific monitoring tasks exhibits high complexity. The results show that the monitoring tasks of SMS, ODD-MS, and CCMS are more complex than those of AOIDS and WIDS. Intuitively speaking, these tasks are passive monitoring and involve less active decision-making. Although this factor reduces task complexity, the differences in monitored physical media lead to variations in task complexity. From a design perspective, monitoring tasks may require the system to undertake more workload to alleviate task complexity. The task complexity of continuous information-processing tasks is influenced by information load, comprehensibility of information, and the requirement for professional knowledge. These factors significantly impact the cognitive demands on operators, directly impacting task complexity (Campbell, 1988; Salimzadeh et al., 2023). When the information load is excessive or difficult to interpret, the processing efficiency of operators may decrease, leading to further increase in task complexity.

5. Conclusion

The existing models of task complexity often fail to capture the unique challenges posed by Degree of Autonomy 2 (DoA 2) maritime autonomous surface ships (MASS). Therefore, this study proposes a task complexity evaluation model applicable to DoA 2 MASS, aiming to optimize the task allocation and system design for MASS operations.

The case study confirms the effectiveness and applicability of analysing task complexity of DoA 2 MASS operational teams and offers insights into how task complexity impacts both human performance and the overall effectiveness of systems. The results indicate that the specific monitoring tasks expose heightened complexity due to the requirements for information resources and professional knowledge, which should be paid to reducing cognitive load and enhancing operator support. Tasks that require direct human intervention are naturally more complex than those related to routine monitoring or reporting, emphasizing the importance of focusing on the more complex, decision-intensive tasks for training and system support. While clear and effective procedures or guidelines have a limited impact on reducing task complexity, they are crucial for ensuring performance stability among operators. The task complexity analysis proposed in this study offers insights into the areas where DoA 2 MASS operational teams should focus attention, providing a basis for optimizing task allocation and system design to enhance operational efficiency and safety.

The core of this research explores the framework of task complexity in MASS operations, including how to quantify task complexity and its impact on MASS operations. However, there are limitations in both the qualitative and quantitative analysis processes. Firstly, while objective task complexity captures the inherent difficulty of tasks from a design perspective and provides strong applicability for task analysis, there are differences in the perception of task difficulty and the capacity to handle tasks among individuals. Meanwhile, dynamic variability complexity, which represents the complexity arising from changes in actions and information, are not considered. Secondly, the two-stage process of gathering expert knowledge also introduces the risk of potential inconsistencies between the initial qualitative insights and the final quantitative weighting results. Although this study employs a comprehensive weighting method to reduce the impact of subjective biases, the evaluation process cannot fully eliminate the interference of subjective factors. Finally, the development of the task complexity framework and different graphs still relies on existing theoretical frameworks, which may not fully encompass all potential dimensions of complexity. Future work will incorporate subjective perceptions through simulation experiments and introduce the dimension of dynamic complexity to refine the classification of task complexity characteristics. This will enhance the framework's adaptability and accuracy in complex environments, providing more comprehensive support for task design and optimization of MASS operational teams.

CRediT authorship contribution statement

Juncheng Tao: Writing – original draft, Validation, Methodology, Conceptualization. **Zhengjiang Liu:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **Xinjian Wang:** Writing – original draft, Methodology, Funding acquisition, Conceptualization. **Jiashi Wang:** Writing – review & editing, Visualization, Validation, Formal analysis. **Shiyi Rao:** Writing – review & editing, Visualization, Methodology, Data curation. **Zaili Yang:** Writing – original draft, Validation, Investigation, Funding acquisition.

Data availability

Data will be made available on request.

Declaration of competing interest

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

Acknowledgements

The authors gratefully acknowledge support from the National Natural Science Foundation of China [Grant No. 52301409; 52101399], and the MarRI-UK Core Project [Grant Agreement No. MarRI-UK_CORE_2024_02]. This work has also received funding from the European Research Council Consolidator Grant program [TRUST Grant No.864724] under EU Horizon 2020 program.

Appendix A

Table A1
The scores for each task complexity characteristic.

Number of experts	CAS	CALS	CAI	COII
1	{(H_3, 0.45), (H_4, 0.55)}	{(H_2, 0.4), (H_3, 0.4), (H_4, 0.2)}	{(H_1, 0.1), (H_2, 0.55), (H_3, 0.35)}	{(H_1, 0.7), (H_2, 0.1), (H_3, 0.2)}
2	{(H_3, 0.2), (H_4, 0.8)}	{(H_3, 0.3), (H_4, 0.7)}	{(H_1, 0.2), (H_2, 0.4), (H_3, 0.4)}	{(H_3, 0.85), (H_4, 0.15)}
3	{(H_2, 0.35), (H_3, 0.2), (H_4, 0.45)}	{(H_4, 0.5), (H_5, 0.5)}	{(H_3, 0.2), (H_4, 0.3), (H_5, 0.5)}	{(H_1, 0.7), (H_2, 0.3)}
4	{(H_3, 0.6), (H_4, 0.3), (H_5, 0.1)}	{(H_4, 1)}	{(H_2, 0.6), (H_3, 0.2), (H_4, 0.2)}	{(H_2, 0.45), (H_3, 0.55)}
5	{(H_2, 0.6), (H_3, 0.4)}	{(H_4, 0.6), (H_5, 0.4)}	{(H_1, 0.3), (H_2, 0.3), (H_3, 0.4)}	{(H_2, 0.65), (H_3, 0.35)}
6	{(H_2, 0.2), (H_3, 0.8)}	{(H_3, 0.2), (H_4, 0.5), (H_5, 0.3)}	{(H_3, 0.7), (H_4, 0.3)}	{(H_1, 0.6), (H_2, 0.1), (H_3, 0.3)}
7	{(H_3, 0.6), (H_4, 0.4}	{(H_1, 0.25), (H_2, 0.25), (H_4, 0.5)}	{(H_2, 0.7), (H_3, 0.3)}	{(H_2, 0.5), (H_3, 0.35), (H_4, 0.15)}
8	{(H_2, 0.45), (H_3, 0.55)}	{(H_4, 0.7), (H_5, 0.3)}	{(H_2, 0.1), (H_3, 0.6), (H_4, 0.2), (H_5, 0.1)}	{(H_1, 0.5), (H_2, 0.3), (H_3, 0.2)}
9	{(H_3, 0.2), (H_4, 0.8)}	{(H_4, 0.1), (H_5, 0.9)}	{(H_2, 0.45), (H_3, 0.55)}	{(H_1, 0.7), (H_2, 0.3)}
10	{(H_2, 0.4), (H_3, 0.6)}	{(H_3, 0.3) (H_4, 0.7)}	{(H_2, 0.35), (H_3, 0.3), (H_4, 0.15), (H_5, 0.2)}	{(H_3, 0.9), (H_4, 0.1)}
11	{(H_2, 0.25), (H_3, 0.25), (H_4, 0.5)}	{(H_3, 0.5), (H_4, 0.5)}	{(H_2, 0.75), (H_3, 0.25)}	{(H_3, 0.7), (H_4, 0.3)}
12	{(H_3, 0.7), (H_4, 0.3)}	{(H_4, 0.75), (H_5, 0.25)}	{(H_3, 0.6), (H_4, 0.4)}	{(H_3, 1)}
13	{(H_2, 0.5), (H_3, 0.5)}	{(H_3, 0.9), (H_4, 0.1)}	{(H_2, 0.4), (H_3, 0.5), (H_4, 0.1)}	{(H_1, 0.1), (H_2, 0.3), (H_3, 0.6)}
14	{(H_1, 0.3), (H_2, 0.3), (H_3, 0.4)}	{(H_3, 0.2), (H_4, 0.7), (H_5, 0.1)}	{(H_1, 0.5), (H_2, 0.3), (H_3, 0.2)}	{(H_2, 0.7), (H_3, 0.3)}
15	{(H_2, 0.7), (H_3, 0.3)}	{(H_3, 0.6), (H_4, 0.2), (H_5, 0.2)}	{(H_2, 0.6), (H_3, 0.2), (H_4, 0.2)}	{(H_2, 0.9), (H_3, 0.1)}
16	{(H_3, 1)}	{(H_3, 0.5), (H_4, 0.5)}	{(H_2, 0.5), (H_3, 0.5)}	{(H_2, 1)}
17	{(H_3, 0.6), (H_4, 0.4)}	{(H_3, 1)}	{(H_3, 1)}	{(H_2, 0.8), (H_3, 0.2)}
18	{(H_2, 0.45), (H_3, 0.55)}	{(H_2, 0.3), (H_3, 0.7)}	{(H_3, 1)}	{(H_2, 0.8), (H_3, 0.2)}
19	{(H_2, 0.6), (H_3, 0.2), (H_4, 0.2)}	{(H_3, 0.6), (H_4, 0.2), (H_5, 0.2)}	{(H_3, 0.9), (H_4, 0.1)}	{(H_1, 0.5), (H_2, 0.4), (H_3, 0.1)}
20	{(H_3, 0.6), (H_4, 0.4)}	{(H_3, 0.45), (H_4, 0.55)}	{(H_3, 0.7), (H_4, 0.3)}	{(H_1, 0.2), (H_2, 0.6), (H_3, 0.2)}
21	{(H_2, 0.5), (H_3, 0.35), (H_4, 0.15)}	{(H_4, 0.5), (H_5, 0.5)}	{(H_2, 0.7), (H_3, 0.2), (H_4, 0.1)}	{(H_2, 0.75), (H_3, 0.25)}
22	{(H_2, 0.4), (H_3, 0.5), (H_4, 0.1)}	{(H_4, 0.75), (H_5, 0.25)}	{(H_2, 0.5), (H_3, 0.5)}	{(H_3, 0.7), (H_4, 0.3)}
23	{(H_2, 0.5), (H_3, 0.3), (H_4, 0.1), (H_5, 0.1)}	{(H_3, 0.9), (H_4, 0.1)}	{(H_1, 0.6), (H_2, 0.4)}	{(H_1, 0.6), (H_2, 0.2), (H_3, 0.2)}
24	{(H_1, 0.3), (H_2, 0.3), (H_3, 0.4)}	{(H_4, 0.5), (H_5, 0.5)}	{(H_2, 0.5), (H_3, 0.5)}	{(H_2, 0.7), (H_3, 0.3)}
25	{(H_3, 0.5), (H_4, 0.5)}	{(H_1, 0.4), (H_2, 0.5), (H_3, 0.1)}	{(H_4, 0.8), (H_5, 0.2)}	{(H_4, 0.85), (H_5, 0.15)}
26	{(H_2, 0.6), (H_3, 0.4)}	{(H_2, 0.2), (H_3, 0.2), (H_4, 0.6)}	{(H_2, 0.6), (H_3, 0.2), (H_4, 0.2)}	{(H_1, 0.2), (H_2, 0.3), (H_3, 0.5)}
27	{(H_4, 0.5), (H_5, 0.5)}	{(H_3, 0.5), (H_4, 0.5)}	{(H_1, 0.35), (H_2, 0.65)}	{(H_2, 0.5), (H_3, 0.5)}
28	{(H_2, 0.8), (H_3, 0.2)}	{(H_3, 0.2), (H_4, 0.6), (H_5, 0.2)}	{(H_2, 0.45), (H_3, 0.55)}	{(H_2, 0.5), (H_3, 0.5)}
29	{(H_3, 1)}	{(H_3, 0.2), (H_4, 0.8)}	{(H_2, 0.8), (H_3, 0.2)}	{(H_1, 0.4), (H_2, 0.4), (H_3, 0.2)}
30	{(H_3, 0.2), (H_4, 0.3), (H_5, 0.5)}	{(H_2, 0.4), (H_3, 0.6)}	{(H_2, 0.6), (H_3, 0.4)}	{(H_2, 0.7), (H_3, 0.2), (H_4, 0.1)}
31	{(H_3, 0.7), (H_4, 0.3)}	{(H_3, 0.9), (H_4, 0.1)}	{(H_2, 0.5), (H_3, 0.4), (H_4, 0.1)}	{(H_2, 0.5), (H_3, 0.5) }
Tota l*	{(H_1, 0.6), (H_2, 7.9), (H_3, 14.25), (H_4, 7.05), (H_5, 1.2)}	{(H_1, 0.65), (H_2, 2.05), (H_3, 10.25), (H_4, 13.45), (H_5, 4.6)}	{(H_1, 2.05), (H_2, 11.7), (H_3, 12.8), (H_4, 3.45), (H_5, 1)}	{(H_1, 5.2), (H_2, 12.75), (H_3, 10.95), (H_4, 1.95), (H_5, 0.15)}
Defuzzification value	0.50146	0.62364	0.43211	0.36726

* Total represents the aggregation of subjective ratings in the fuzzy belief structure.

References

Almaatouq, A., Alsobay, M., Yin, M., Watts, D.J., 2021. Task complexity moderates group synergy. Proc. Natl. Acad. Sci. 118 (36), e2101062118. https://doi.org/10.1073/pnas.2101062118.

Banda, O.A.V., Kujala, P., Goerlandt, F., Bergstrom, M., Ahola, M., van Gelder, P., Sonninen, S., 2018. The need for systematic and systemic safety management for autonomous vessels. Jun 10-14 13th International Marine Design Conference (IMDC). FINLAND, Espoo, pp. 853–859.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W.S., Weld, D.S., Horvitz, E. Beyond accuracy: the role of mental models in human-AI team performance. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp. 2-11.

Basnet, S., BahooToroody, A., Chaal, M., Lahtinen, J., Bolbot, V., Valdez Banda, O.A., 2023. Risk analysis methodology using STPA-based Bayesian network- applied to remote pilotage operation. Ocean Eng. 270, 113569. https://doi.org/10.1016/j.oceaneng.2022.113569.

Bonner, S.E., 1994. A model of the effects of audit task complexity. Account. Org. Soc. 19 (3), 213–234. https://doi.org/10.1016/0361-3682(94)90033-7.

Burmeister, H.-C., Bruhn, W., Rødseth, Ø.J., Porathe, T., 2014. Autonomous unmanned merchant vessel and its contribution towards the e-navigation implementation: the MUNIN perspective. Int J e-Navigation Maritime Economy 1, 1–13. https://doi.org/10.1016/j.enavi.2014.12.002.

Campbell, D.J., 1988. Task complexity: a review and analysis. Acad. Manag. Rev. 13 (1), 40–52. https://doi.org/10.5465/amr.1988.4306775.

Cao, Y., Iulia, M., Majumdar, A., Feng, Y., Xin, X., Wang, X., Wang, H., Yang, Z., 2025. Investigation of the risk influential factors of maritime accidents: a novel topology and robustness analytical framework. Reliab. Eng. Syst. Saf. 254, 110636. https://doi.org/10.1016/j.ress.2024.110636.

Chaal, M., Valdez Banda, O.A., Glomsrud, J.A., Basnet, S., Hirdaris, S., Kujala, P., 2020. A framework to model the STPA hierarchical control structure of an autonomous ship. Saf. Sci. 132, 104939. https://doi.org/10.1016/j.ssci.2020.104939.

Chan, S.H., Song, Q., Yao, L.J., 2015. The moderating roles of subjective (perceived) and objective task complexity in system use and performance. Comput. Hum. Behav. 51, 393–402. https://doi.org/10.1016/j.chb.2015.04.059.

Chen, C.-T., 2000. Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Set Syst. 114 (1), 1–9. https://doi.org/10.1016/S0165-0114(97)00377-1.

Cheng, T., Utne, I.B., Wu, B., Wu, Q., 2023. A novel system-theoretic approach for human-system collaboration safety: case studies on two degrees of autonomy for autonomous ships. Reliab. Eng. Syst. Saf. 237, 109388. https://doi.org/10.1016/j.ress.2023.109388.

Cheng, T., Veitch, E.A., Utne, I.B., Ramos, M.A., Mosleh, A., Alsos, O.A., Wu, B., 2024. Analysis of human errors in human-autonomy collaboration in autonomous ships operations through shore control experimental data. Reliab. Eng. Syst. Saf. 246, 110080. https://doi.org/10.1016/j.ress.2024.110080.

Das, D., Chernova, S., 2020. Leveraging rationales to improve human task performance. Proceedings of the 25th International Conference on Intelligent User Interfaces. Association for Computing Machinery, Cagliari, Italy, pp. 510–518.

Deli, I., Karaaslan, F., 2021. Generalized trapezoidal hesitant fuzzy numbers and their applications to multi criteria decision-making problems. Soft Comput. 25 (2), 1017–1032. https://doi.org/10.1007/s00500-020-05201-2.

Dghaym, D., Hoang, T.S., Turnock, S.R., Butler, M., Downes, J., Pritchard, B., 2021. An STPA-based formal composition framework for trustworthy autonomous maritime systems. Saf. Sci. 136, 105139. https://doi.org/10.1016/j.ssci.2020.105139.

Egbueri, J.C., 2022. Incorporation of information entropy theory, artificial neural network, and soft computing models in the development of integrated industrial water quality index. Environ. Monit. Assess. 194 (10), 693. https://doi.org/10.1007/s10661-022-10389-x.

Eskov, V.M., Eskov, V.V., Vochmina, Y.V., Gorbunov, D.V., Ilyashenko, L.K., 2017. Shannon entropy in the research on stationary regimes and the evolution of complexity. Moscow Univ. Phys. Bull. 72 (3), 309–317. https://doi.org/10.3103/S0027134917030067.

Fan, C., Montewka, J., Zhang, D., 2022. A risk comparison framework for autonomous ships navigation. Reliab. Eng. Syst. Saf. 226, 108709. https://doi.org/10.1016/j.ress.2022.108709.

Fan, C., Montewka, J., Bolbot, V., Zhang, Y., Qiu, Y., Hu, S., 2024. Towards an analysis framework for operational risk coupling mode: a case from MASS navigating in restricted waters. Reliab. Eng. Syst. Saf. 248, 110176. https://doi.org/10.1016/j.ress.2024.110176.

Feng, Y., Wang, H., Xia, G., Cao, W., Li, T., Wang, X., Liu, Z., 2024. A machine learning-based data-driven method for risk analysis of marine accidents. J Marine Eng &; Technol 1–12. https://doi.org/10.1080/20464177.2024.2368914.

Gan, L., Li, X., Yan, T., Song, L., Xiao, J., Shu, Y., 2025. Intelligent ship path planning based on improved artificial potential field in narrow inland waterways. Ocean Eng. 317, 119928. https://doi.org/10.1016/j.oceaneng.2024.119928.

Ge, X., Zhou, Q., Liu, Z., 2020. Assessment of space station on-orbit maintenance task complexity. Reliab. Eng. Syst. Saf. 193, 106661. https://doi.org/10.1016/j.ress.2019.106661.

Goerlandt, F., 2020. Maritime autonomous surface ships from a risk governance perspective: interpretation and implications. Saf. Sci. 128, 104758. https://doi.org/10.1016/j.ssci.2020.104758.

Ham, D., Park, J., Jung, W. Multi-view based conceptual framework for identifying and organizing complexity factors of humanesystem interaction. Proceedings of the Eighth Symposium on Human Interaction with Complex Systems, pp.

Ham, D.-H., Park, J., Jung, W., 2011. Extension of TACOM to the complexity of tasks designed for abnormal situations in nuclear power plants. J. Loss Prev. Process. Ind. 24 (5), 601–611. https://doi.org/10.1016/j.jlp.2011.04.006.

Ham, D.-H., Park, J., Jung, W., 2012. Model-based identification and use of task complexity factors of human integrated systems. Reliab. Eng. Syst. Saf. 100, 33–47. https://doi.org/10.1016/j.ress.2011.12.019.

Han, Z., Zhang, D., Fan, L., Zhang, J., Zhang, M., 2024. A dynamic bayesian network model to evaluate the availability of machinery systems in maritime autonomous surface ships. Accid. Anal. Prev. 194, 107342. https://doi.org/10.1016/j.aap.2023.107342.

Hærem, T., Pentland, B.T., Miller, K.D., 2015. Task complexity: extending a core concept. Acad. Manag. Rev. 40 (3), 446–460. https://doi.org/10.5465/amr.2013.0350.

Harvey, C.M., Koubek, R.J., 2000. Cognitive, social, and environmental attributes of distributed engineering collaboration: a review and proposed model of collaboration. Human Factors Ergonomics Manufact &; Service Industries 10 (4), 369–393. https://doi.org/10.1002/1520-6564(200023)10:4<369::AID-HFM2>3.0.CO, 2-Y.

Hendy, K.C., Liao, J., Milgram, P., 1997. Combining time and intensity effects in assessing operator information-processing load. Hum. Factors 39 (1), 30–47. https://doi.org/10.1518/001872097778940597.

IMO, 2018. Maritime Safety Committee (MSC). 100th Session-Regulatory Scoping Exercise on Maritime Autonomous Surface Ships (MASS). International Maritime Organisation.

Jang, I., Kim, Y., Park, J., 2021. Investigating the effect of task complexity on the occurrence of human errors observed in a nuclear power plant full-scope simulator. Reliab. Eng. Syst. Saf. 214, 107704. https://doi.org/10.1016/j.ress.2021.107704.

Kutlu, A.C., Ekmekçioğlu, M., 2012. Fuzzy failure modes and effects analysis by using fuzzy TOPSIS-based fuzzy AHP. Expert Syst. Appl. 39 (1), 61–67. https://doi.org/10.1016/j.eswa.2011.06.044.

Lee, Y.-L., Hwang, S.-L., Min-Yang Wang, E., 2005. Reducing cognitive workload of a computer-based procedure system. Int. J. Hum. Comput. Stud. 63 (6), 587–606. https://doi.org/10.1016/j.ijhcs.2005.05.003.

Li, P.-c., Wang, Y.-x., Chen, J.-h., Luo, Z.-h., Dai, L.-c., 2021. An experimental study on the effects of task complexity and knowledge and experience level on SA, TSA and workload. Nucl. Eng. Des. 376, 111112. https://doi.org/10.1016/j.nucengdes.2021.111112.

Li, W., Chen, W., Hu, S., Xi, Y., Guo, Y., 2023a. Risk evolution model of marine traffic via STPA method and MC simulation: a case of MASS along coastal setting. Ocean Eng. 281, 114673. https://doi.org/10.1016/j.oceaneng.2023.114673.

Li, X., Oh, P., Zhou, Y., Yuen, K.F., 2023b. Operational risk identification of maritime surface autonomous ship: a network analysis approach. Transp. Policy 130, 1–14. https://doi.org/10.1016/j.tranpol.2022.10.012.

Li, Y., Duan, Z., Liu, Z., 2019. Study on risk-based operators' competence assessment of maritime autonomous surface ships. 5th International Conference on Transportation Information and Safety (ICTIS). ENGLAND, Liverpool, pp. 1412–1417.

Liu, P., Li, Z., 2011. Toward understanding the relationship between task complexity and task performance. Int, Design Global Dev 192–200. Berlin, Heidelberg.

Liu, P., Li, Z., 2012. Task complexity: a review and conceptualization framework. Int. J. Ind. Ergon. 42 (6), 553–568. https://doi.org/10.1016/j.ergon.2012.09.001.

Magnhild, K., Braseth, A.O., 2020. Operating Autonomous Ships Remotely from Land-Based Operation Centers: the Current State-Of-The-Art.

Man, Y., Weber, R., Cimbritz, J., Lundh, M., MacKinnon, S.N., 2018. Human factor issues during remote ship monitoring tasks: An ecological lesson for system design in a distributed context. Int. J. Ind. Ergon. 68, 231–244. https://doi.org/10.1016/j.ergon.2018.08.005.

Palbar Misas, J.D., Hopcraft, R., Tam, K., Jones, K., 2024. Future of maritime autonomy: cybersecurity, trust and mariner's situational awareness. J Marine Eng &; Technol 23 (3), 224–235. https://doi.org/10.1080/20464177.2024.2330176.

Park, J., Jung, W., 2007. A study on the development of a task complexity measure for emergency operating procedures of nuclear power plants. Reliab. Eng. Syst. Saf. 92 (8), 1102–1116. https://doi.org/10.1016/j.ress.2006.03.009.

Park, J., Jung, W., 2008. A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants—comparing task complexity scores with two sets of operator response time data obtained under a simulated SGTR. Reliab. Eng. Syst. Saf. 93 (4), 557–566. https://doi.org/10.1016/j.ress.2007.02.002.

Park, J., Jung, W., Ha, J., 2001. Development of the step complexity measure for emergency operating procedures using entropy concepts. Reliab. Eng. Syst. Saf. 71 (2), 115–130. https://doi.org/10.1016/S0951-8320(00)00087-9.

Pasarakonda, S., Grote, G., Schmutz, J.B., Bogdanovic, J., Guggenheim, M., Manser, T., 2021. A strategic core role perspective on team coordination: benefits of centralized leadership for managing task complexity in the operating room. Hum. Factors 63 (5), 910–925. https://doi.org/10.1177/0018720820906041.

Podofillini, L., Park, J., Dang, V.N., 2013. Measuring the influence of task complexity on human error probability: an empirical evaluation. Nucl. Eng. Technol. 45 (2), 151–164. https://doi.org/10.5516/NET.04.2013.702.

Poursabzi-Sangdeh, F., Goldstein, D.G., Hofman, J.M., Vaughan, J.W.W., Wallach, H., 2021. Manipulating and measuring model interpretability. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, Yokohama, Japan. Article 237.

Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2020a. A generic approach to analysing failures in human – system interaction in autonomy. Saf. Sci. 129, 104808. https://doi.org/10.1016/j.ssci.2020.104808.

Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2020b. Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliab. Eng. Syst. Saf. 195, 106697. https://doi.org/10.1016/j.ress.2019.106697.

Rein, P., Beckmann, T., Krebs, E., Mattis, T., Hirschfeld, R. Too simple? Notions of task complexity used in maintenance-based studies of programming tools. 2023 IEEE/ACM 31st International Conference on Program Comprehension (ICPC), Melbourne, Australia, May 15-16, 2023, pp. 254-265.

Rødseth, Ø., Nordahl, H., 2017. Definitions for Autonomous Merchant Ships.

Rothrock, L., Harvey, C.M., Burns, J., 2005. A theoretical framework and quantitative architecture to assess team task complexity in dynamic environments. Theor. Issues Ergon. Sci. 6 (2), 157–171. https://doi.org/10.1080/1463922042000295678.

Salimzadeh, S., He, G., Gadiraju, U., 2023. A missing piece in the puzzle: considering the role of task complexity in human-AI decision making. Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization. Association for Computing Machinery, Limassol, Cyprus, pp. 215–227.

Schmidt, F.L., Hunter, J.E., 1998. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychol. Bull. 124 (2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262.

Sezer, S.I., Ahn, S.I., Akyuz, E., Kurt, R.E., Gardoni, P., 2024. A hybrid human reliability analysis approach for a remotely-controlled maritime autonomous surface ship (MASS- degree 3) operation. Appl. Ocean Res. 147, 103966. https://doi.org/10.1016/j.apor.2024.103966.

Shiokari, M., Itoh, H., Yuzui, T., Ishimura, E., Miyake, R., Kudo, J., Kawashima, S., 2024. Structure model-based hazard identification method for autonomous ships. Reliab. Eng. Syst. Saf. 247, 110046. https://doi.org/10.1016/j.ress.2024.110046.

Shu, Y., Xiong, C., Zhu, Y., Liu, K., Liu, R.W., Xu, F., Gan, L., Zhang, L., 2024. Reference path for ships in ports and waterways based on optimal control. Ocean Coast Manag. 253, 107168. https://doi.org/10.1016/j.ocecoaman.2024.107168.

Song, R., Papadimitriou, E., Negenborn, R.R., van Gelder, P., 2024. Safety and efficiency of human-MASS interactions: towards an integrated framework. J Marine Eng &; Technol 1–20. https://doi.org/10.1080/20464177.2024.2414959.

Storkersen, K.V., 2021. Safety management in remotely controlled vessel operations. Mar. Pol. 130, 104349. https://doi.org/10.1016/j.marpol.2020.104349.

Symes, S., Blanco-Davis, E., Graham, T., Wang, J., Shaw, E., 2024. The survivability of autonomous vessels from cyber-attacks. J Marine Eng &; Technol 1–23. https://doi.org/10.1080/20464177.2024.2428022.

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 Eng. 307, 118174. https://doi.org/10.1016/j.oceaneng.2024.118174.

Valdez Banda, O.A., Goerlandt, F., 2018. A STAMP-based approach for designing maritime safety management systems. Saf. Sci. 109, 109–129. https://doi.org/10.1016/j.ssci.2018.05.003.

Veitch, E., Alsos, O.A., Cheng, T., Senderud, K., Utne, I.B., 2024. Human factor influences on supervisory control of remotely operated and autonomous vessels. Ocean Eng. 299, 117257. https://doi.org/10.1016/j.oceaneng.2024.117257.

Wang, X., Cao, W., Li, T., Feng, Y., Uğurlu, Ö., Wang, J., 2025. An integrated multidimensional model for heterogeneity analysis of maritime accidents during different watchkeeping periods. Ocean Coast Manag. 264, 107625. https://doi.org/10.1016/j.ocecoaman.2025.107625.

Wood, R.E., 1986. Task complexity: definition of the construct. Organ. Behav. Hum. Decis. Process. 37 (1), 60–82. https://doi.org/10.1016/0749-5978(86)90044-0.

Wu, W., Xiaolong, H., Shicai, C., Yuchi, C., Wang, X., 2024. A quantitative study on anchoring speed for Maritime Autonomous Surface Ships. J Marine Eng &; Technol 1–11. https://doi.org/10.1080/20464177.2024.2432776.

Yang, J.B., Wang, Y.M., Xu, D.L., Chin, K.S., 2006. The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties. Eur. J. Oper. Res. 171 (1), 309–343. https://doi.org/10.1016/j.ejor.2004.09.017.

Zhang, Y., Li, Z., Wu, B., Wu, S., 2009. A spaceflight operation complexity measure and its experimental validation. Int. J. Ind. Ergon. 39 (5), 756–765. https://doi.org/10.1016/j.ergon.2009.03.003.

Zhang, M., Zhang, D., Yao, H., Zhang, K., 2020. A probabilistic model of human error assessment for autonomous cargo ships focusing on human-autonomy collaboration. Saf. Sci. 130, 104838. https://doi.org/10.1016/j.ssci.2020.104838.

Zheng, Y., Lu, Y., Wang, Z., Huang, D., Fu, S., 2015. Developing a measurement for task complexity in flight. Aerospace Med Human Performance 86 (8), 698–704.

Zhou, X.-Y., Liu, Z.-J., Wang, F.-W., Wu, Z.-L., 2021. A system-theoretic approach to safety and security co-analysis of autonomous ships. Ocean Eng. 222, 108569. https://doi.org/10.1016/j.oceaneng.2021.108569.

Zigurs, I., Buckland, B.K., 1998. A theory of task/technology fit and group support systems effectiveness. MIS Q. 22 (3), 313–334. https://doi.org/10.2307/249668.