Sveučilište u Splitu
University of Split

FACULTY OF MARITIME STUDIES

Toni Meštrović

MODEL FOR EVALUATING THE PERFORMANCE OF SHIP OPERATORS AND COMBINATIONS OF FACTORS IN THE CONTROL OF AUTONOMOUS SHIPS USING SIMULATIONS AND BAYESIAN NETWORKS

DOCTORAL DISSERTATION

Split, 2026.

Sveučilište
u Splitu
University
of Split

FACULTY OF MARITIME STUDIES

Toni Meštrović

MODEL FOR EVALUATING THE PERFORMANCE OF SHIP OPERATORS AND COMBINATIONS OF FACTORS IN THE CONTROL OF AUTONOMOUS SHIPS USING SIMULATIONS AND BAYESIAN NETWORKS

DOCTORAL DISSERTATION

Supervisor: Associate professor Ivica Pavić, PhD

Co-supervisor: Assistant professor Andrej Androjna, PhD

Split, 2026

IMPRESSUM

This doctoral dissertation was submitted to the University of Split, Faculty of Maritime Studies, in partial fulfilment of the requirements for the academic degree of Doctor of Science (PhD) in the scientific area of technical sciences, the scientific field of Traffic and Transport Technology.

The dissertation has been written in British English.

Supervisor: Associate professor Ivica Pavić, PhD
University of Split, Faculty of Maritime Studies, Croatia

Co-supervisor: Assistant professor Andrej Androjna, PhD
University of Ljubljana, Faculty of Maritime Studies and Transport, Slovenia

COMMITTEES FOR EVALUATION AND DEFENCE OF THE DOCTORAL DISSERTATION

Doctoral Dissertation Evaluation Committee:

1. Chair: academic title, full name, PhD, full institutional affiliation, country
2. Member: academic title, full name, PhD, full institutional affiliation, country
3. Member: academic title, full name, PhD, full institutional affiliation, country
4. Member (opt.): academic title, full name, PhD, full institutional affiliation, country
5. Member (opt.): academic title, full name, PhD, full institutional affiliation, country

Doctoral Dissertation Defence Committee:

1. Chair: academic title, full name, PhD, full institutional affiliation, country
2. Member: academic title, full name, PhD, full institutional affiliation, country
3. Member: academic title, full name, PhD, full institutional affiliation, country
4. Member (opt.): academic title, full name, PhD, full institutional affiliation, country
5. Member (opt.): academic title, full name, PhD, full institutional affiliation, country

Date of defence: DD/MM/YYYY

ABSTRACT

The rapid development of Maritime Autonomous Surface Ships (MASS) is fundamentally transforming ship operations and redefining the role of seafarers within increasingly automated and remotely controlled environments. Although autonomous technologies promise improvements in efficiency, safety, and operational reliability, they also introduce new challenges related to human performance, decision-making, and training. Traditional maritime education and assessment frameworks, primarily designed for conventional ship operations, are inadequate to address the cognitive, operational, and organizational demands associated with MASS operations, particularly at higher levels of autonomy.

This doctoral dissertation proposes a simulation-based model for evaluating seafarer performance in MASS operations, integrating empirical data collection with Bayesian Network (BN) modelling. The research analyses how human, technical, and environmental factors interact and influence decision-making and operational performance under complex and time-critical conditions. High fidelity maritime simulation exercises are designed to replicate realistic operational scenarios, including system failures, emergency manoeuvres, collision avoidance, and navigation in congested waters. These scenarios are used to assess participant behavior under increasing levels of complexity and stress.

The study includes 65 participants with varying levels of maritime experience, such as nautical high school graduates, maritime faculty students, and experienced seafarers. Performance is assessed using a combination of behavioral indicators (reaction time, decision accuracy, situational awareness, adaptability, and help-seeking behavior) and physiological measurements, including heart rate variability and eye-tracking data, to capture cognitive workload and stress responses during simulation-based tasks.

The collected data are organized into a BN model to identify probabilistic relationships between key performance indicators and operational outcomes. Unlike existing approaches that rely mainly on expert judgement or accident data, this research emphasises empirically grounded modelling based on simulation-derived human performance data. The BN model enables the identification of critical combinations of factors that contribute to decision success or increased error probability in MASS operations and supports predictive analysis under varying operational conditions.

The developed BN model integrates behavioral, physiological, technical, and operational indicators into a unified probabilistic framework for evaluating MASS operator performance. The analysis identifies combinations of human, physiological, technical, and operational factors that most strongly influence Operational Effectiveness and quantifies their relative contributions through sensitivity analysis. These results provide an objective basis for performance assessment, support the development of evidence-based training programmes, and facilitate the design of decision-support tools for future MASS operators. The proposed model advances current approaches to MASS operator assessment by enabling objective, data-driven evaluation of operator performance under realistic, simulation-based conditions.

Keywords: Maritime Autonomous Surface Ships (MASS), simulator-based training, eye-tracking technology, BN model, ship operator performance

CONTENTS

1. INTRODUCTION

The shipping industry has continually evolved in response to technological advancements, economic pressures, and safety requirements. In recent years, rapid developments in automation, digitalisation, and artificial intelligence have accelerated the transition to MASS, fundamentally transforming ship operations and the role of human operators in the sector. Although autonomous technologies promise improvements in efficiency, safety, and sustainability, they also introduce new challenges related to human performance, decision-making, and training in highly automated and remotely operated environments.

As levels of autonomy increase, seafarers' responsibilities are shifting from direct ship handling to supervisory control, system monitoring, and intervention during abnormal or emergency situations. These changes place significant cognitive demands on operators, especially in complex navigational environments where time pressure, system failures, and uncertainty are present. Traditional maritime education and training systems, which are largely based on conventional ship operations, are increasingly insufficient to address the skills required for autonomous and remote ship control. Consequently, there is a growing need for evidence-based training and assessment frameworks that can capture the multifaceted nature of human performance in MASS operations.

Human factors remain a critical element in autonomous ship safety, despite increasing automation. Numerous studies indicate that human error, inadequate situational awareness, and ineffective human–machine interaction continue to play a decisive role in maritime incidents, even in highly automated systems. In the context of MASS, these issues are further amplified by reduced sensory input, reliance on complex interfaces, and the need to make rapid decisions under stress. Understanding how human capabilities, cognitive workload, and operational conditions interact is therefore essential for the safe and effective deployment of autonomous maritime technologies.

Simulation-based training has become a powerful tool for preparing seafarers for complex and safety critical operations. High fidelity maritime simulators enable the reproduction of realistic navigational scenarios, system failures, and emergency situations within a controlled environment, allowing systematic observation and assessment of operator behavior. However, while simulation-based exercises are widely used for training, their potential as a source of empirical data collection for modelling human performance and decision-making in MASS operations remains underexplored.

BN have attracted increasing attention as a probabilistic modelling approach capable of representing dependencies between multiple interacting factors under uncertainty. In maritime research, BNs have been successfully applied to risk assessment, accident analysis, and technical system reliability. However, there is a notable lack of empirically grounded BN models that explicitly incorporate human performance data from simulation-based experiments, particularly in the context of autonomous and remotely operated ships.

To address these challenges, this doctoral dissertation research develops a structured model for evaluating seafarer performance in MASS operations by integrating simulation-based exercises and BN modelling. By combining behavioral performance indicators, physiological measurements, and probabilistic analysis, the dissertation aims to provide a deeper understanding of how human, technical, and environmental factors jointly influence decision-making and operational outcomes in autonomous maritime environments. This approach supports the development of modern, evidence-based training programmes and contributes to the safe transition towards autonomous shipping.

1.1. Motivation

The shipping industry is undergoing a profound transformation driven by advances in autonomous systems, artificial intelligence, and digital technologies. The International Maritime Organization (IMO) has recognized these developments through the ongoing development of the MASS Code, with a binding regulatory framework expected by 2030 [1]. While increasing levels of autonomy offer significant opportunities for maritime transport, they also introduce new challenges related to human supervision, decision-making, and operational safety. Addressing these challenges through scientifically grounded methods for evaluating operator performance is the primary motivation for this dissertation.

Autonomy in shipping does not eliminate human involvement. Instead, the role of the human operator is being reshaped. Depending on the level of autonomy, ships may require remote monitoring, decision-support or emergency response from shore-based control centres. This new operational paradigm requires operators to maintain situational awareness and make effective decisions despite being physically separated from the vessel. Conventional onboard cues, such as sound, vibration and visual reference points, are less or non-existent, creating additional cognitive demands. Understanding and optimizing the interaction between human operators and autonomous ship systems is therefore a prerequisite for safe and reliable operation [2].

Despite technological advances, human factors remain central to maritime safety. As responsibility shifts from onboard crews to shore-based operators, the operational risk profile changes rather than disappears. Shore-based operators must maintain situational awareness, interpret multiple digital information sources, and respond effectively to operational situations despite the absence of direct sensory cues [3]. Existing approaches to assessing operator performance are largely based on conventional ship operations and do not provide an integrated evaluation of the cognitive, behavioral, and physiological aspects relevant to MASS environments. There remains a need for objective methodologies capable of capturing the complex interaction between human performance and operational conditions in autonomous ship operations.

Simulation-based assessment offers an effective way to investigate these challenges. High fidelity simulators enable the replication of challenging maritime scenarios such as equipment failures, collision avoidance in heavy traffic or sudden environmental changes. They provide a safe environment to measure operator performance under controlled conditions and capture metrics such as reaction time, decision accuracy and compliance with rules.

BN have shown significant potential for maritime risk assessment and operator error analysis [4],[5]. Most existing applications rely on expert judgement, accident reports, or predefined risk scenarios rather than empirical performance data from simulation-based experiments. Similarly, studies on human–machine interaction have emphasised the importance of interface design for maintaining situational awareness during remote ship operations [6], but these findings have rarely been integrated with objective behavioral and physiological measurements within a unified probabilistic assessment framework. A methodological gap remains in developing empirically validated BN models that combine simulator derived performance indicators, physiological responses, and human–machine interaction data for evaluating MASS operators. Addressing this gap is the primary motivation for this research.

1.2. Subject of Doctoral Dissertation

This doctoral dissertation concerns the development of a simulation-based model for evaluating operator performance and decision-making in autonomous ship operations. The research focuses on identifying, measuring, and modelling the key human and operational factors that influence the effectiveness and reliability of decisions in complex and High Stress maritime environments.

The core of the research is the assessment and modelling of operator performance in simulated MASS operations by integrating behavioral, physiological, technical, and operational indicators. Particular emphasis is placed on developing a BN model representing the probabilistic relationships between these factors and enabling analysis of their influence on decision-making and Operational Effectiveness.

The proposed BN model integrates simulation-derived performance indicators, physiological measurements, and expert knowledge to analyse interactions among human, technical, and operational factors that affect operator performance. The model offers a probabilistic framework for descriptive and predictive analysis of decision-making processes in MASS operations.

Accordingly, the central research question addressed in this dissertation is:

How can a model for evaluating operator performance and decision-making in MASS operations be developed by integrating human, physiological, technical and operational factors?

The results of this research are expected to offer new insights into the relationships among human, physiological, technical, and operational factors influencing operator performance and decision-making in MASS operations.

1.3. Purpose and Objectives of the Research

The purpose of this research is to develop and validate an approach for assessing and modelling operator performance in simulated MASS operations by integrating behavioral, physiological, technical, and operational indicators within a BN model. The research also aims to identify key factors influencing operator decision-making and Operational Effectiveness, and to analyse the probabilistic relationships among these factors.

The research aims to address the current gap between technological development and human-centred adaptation. While numerous studies have explored the technical feasibility of autonomous navigation, relatively few have provided empirical evidence on human decision-making under conditions of automation, uncertainty, and stress. This dissertation therefore places the human operator at the centre of the analysis, recognizing that the transition to autonomous operations will only be successful if human performance evolves in parallel with technological innovation.

To achieve this purpose, the research pursues four objectives, each addressing a key stage in the development of the proposed model for evaluating the performance of ship operators in the control of autonomous ships using simulations and BN:

1. To design and implement simulation-based navigation exercise that replicate realistic and challenging maritime scenario, including varying operational conditions and system failures. This simulation provides the empirical foundation for studying human behavior, situational awareness, and decision-making under controlled yet realistic conditions.
2. To collect and analyse multidimensional data on operator performance, including behavioral indicators (reaction time, decision accuracy, adaptability), physiological responses (heart rate variability), eye-tracking metrics (gaze behavior), and contextual factors (environmental and technological influences). This stage ensures objectivity and depth in performance assessment.
3. To develop a BN model that captures the probabilistic relationships among human, technical, and environmental variables influencing decision-making in autonomous ship operations. The BN model enables quantitative analysis of risk, performance dependencies, and causal links between factors affecting safe navigation.
4. To identify the factors that contribute to a higher probability of error in the operation of autonomous ships. The BN will be used to determine which variables and relationships most strongly influence human error, enabling the formulation of evidence-based measures to improve operator performance and reduce the likelihood of unsafe decisions.

Through these objectives, the research provides a structured and scientifically robust approach to linking simulation-based experimentation with probabilistic modelling. The expected outcome is a validated model that enhances understanding of human performance in autonomous navigation and supports the creation of adaptive training programmes and assessment tools aligned with the future requirements of maritime operations.

The dissertation seeks to bridge the gap between technological innovation and human adaptability, ensuring that the evolution of maritime autonomy remains both safe and human-centred.

1.4. Defining the Research Problem and Main Hypothesis

The ongoing development of MASS is transforming the traditional role of the seafarer and reshaping the competencies required for safe and efficient navigation. The implementation of autonomous systems has progressed more rapidly than the development of adequate models for assessing human performance in these new operational environments.

Current approaches to evaluating operator performance in MASS operations rely mainly on expert judgement, conventional competency assessment, or accident analysis, and do not integrate behavioral, physiological, technical, and operational indicators within a unified probabilistic framework. Consequently, there is no validated model capable of objectively assessing operator performance or identifying the key factors influencing decision-making in autonomous ship operations.

The key research problem addressed in this dissertation is therefore the absence of a model capable of evaluating human performance and identifying the factors that most strongly influence decision-making and human error during autonomous ship operations. Existing approaches do not sufficiently incorporate empirical data from simulation-based exercises or use probabilistic reasoning to model causal relationships between human, technical, and environmental variables.

To address this problem, the dissertation is based on the following main hypothesis:

Main Hypothesis:

The capability of ship operators can be determined through simulator-based training and Bayesian Networks to identify combinations of factors that have influence on performance during the operation of autonomous ships.

This hypothesis reflects the scientific objective of the research to combine experimental data obtained through simulation exercises with BN modelling to develop an evidence-based model for evaluating seafarer performance. Verification of this hypothesis will enable the identification of key variables that influence operational outcomes and provide a foundation for adapting maritime education and training systems to the requirements of autonomous and remotely operated vessels.

Based on the defined research problem and the main hypothesis, the scientific contributions of this dissertation can be summarized as follows:

- A comprehensive and systematic overview of the current state of research in the field of MASS, training, and human performance, identifying prevailing approaches, key challenges, and existing research gaps.
- Development of a structured and applicable approach that integrates simulation-based experimentation and BN modelling for analysing seafarer performance and risk in MASS operations.
- Empirical investigation of the effects of scenario-based simulation exercises on seafarer adaptability, reaction time, and situational awareness in the context of autonomous ship operations.
- Identification of key technical, cognitive, and problem-solving competencies that influence safe and effective performance in MASS operations.
- BN model that captures the interaction between human and operational risk factors and their influence on decision-making and performance in MASS Category 2 and 3 operations.

1.5. Methodology

The methodological framework of this doctoral dissertation is conceived as a simulation-based and probabilistic model designed to objectively evaluate operators' performance in MASS operations. The dissertation combines controlled simulator-based exercise with multidimensional data acquisition and BN modelling.

The research methodology comprises three interconnected phases. The first phase involves participant profiling and baseline assessment, using pre-simulation questionnaires and physiological measurements to establish initial cognitive, experiential, and stress-related parameters. The second phase consists of a structured simulator-based navigation exercise that exposes participants to realistic operational conditions with increasing navigational complexity and controlled system disturbances. The third phase integrates the collected data into a probabilistic Bayesian model, enabling the modelling of interdependencies among key performance variables.

The exercise requires participants to navigate through the coastal area along a designated route, encountering various navigational challenges typical of coastal navigation. The route includes a section of the fairway towards the northern port of Split, a common route for several merchant vessels entering the port. The purpose of the exercise is to simulate a realistic operational situation in which a remotely operated ship must safely navigate through the coastal area and reach the designated transfer point, where tugboats are positioned and control of the ship is handed over to the onboard crew. The scenario represents a critical phase of MASS operations, testing the operator's ability to manage complex navigation tasks, maintain situational awareness, and make timely decisions under realistic environmental and traffic conditions. This setup accurately reflects actual navigation conditions while providing a challenging environment for evaluating operator performance. The total duration of the exercise is limited to a maximum of 20 minutes, with the optimal completion time being approximately 18 minutes, representing the expected duration under normal navigational conditions. The exercise uses an 80-meter Offshore Support Vessel (OSV), whose characteristics closely resemble those of the existing autonomous ship Yara Birkeland.

The described simulation scenario provides a structured experimental platform for objective performance measurement. During the exercise, multiple synchronized data streams are collected, including simulator log data, eye-tracking recordings, and physiological indicators. This multidimensional approach enables a comprehensive evaluation of decision-making processes under conditions of uncertainty, workload, and partial system degradation.

Operational performance is assessed using measurable indicators such as route adherence, speed management, detection time for system failures, response to navigational problems, and occurrence of near miss situations. These behavioral metrics provide quantifiable evidence of decision quality and procedural compliance.

Physiological monitoring, including continuous heart rate tracking and pre- and post-session blood pressure measurements, allows assessment of stress reactivity and workload response during task execution.

Eye-tracking recordings are initially processed in the Pupil Core software environment to ensure calibration accuracy, fixation detection, and Area of Interest (AOI) segmentation. Following this initial processing stage, the extracted gaze data are further analysed using a custom Python programme developed specifically for this research. The Python-based analytical framework enables structured computation of fixation duration, transition frequencies between AOIs, temporal alignment with simulator events, and preparation of datasets suitable for statistical evaluation and probabilistic modelling.

Although the BN model was primarily developed using empirical data collected during the simulation exercises, some variables and conditional probability distributions could not be defined solely from the available dataset because of limited observations or the conceptual nature of specific constructs. In such cases, expert judgement from maritime specialists was incorporated to define the network structure, variable discretization, and conditional probability relationships. The expert-based development procedure is described in detail in Chapter 6.

The integration of simulator derived performance indicators, physiological measurements, computationally processed gaze metrics, and expert knowledge forms the empirical and methodological foundation for constructing the BN model. Through parameter estimation and sensitivity analysis, the model identifies combinations of factors that most strongly influence Operational Effectiveness and the probability of decision-related errors. The model was implemented using the Graphical Network Interface (GeNIe) software developed by the Decision Systems Laboratory, which enabled graphical construction of the BN structure, specification of conditional probability tables, and execution of probabilistic inference and sensitivity analysis.

analysis.

1.6. Review of Previous Research

This section summarizes current scientific findings from a systematic literature analysis conducted according to the method PRISMA (Preferred Reporting Items for Systematic

Reviews and Meta-Analyses) [7], [8].The analysis included relevant scientific works available in the Scopus, Web of Science, and Google Scholar databases, with searches conducted using combinations of key terms: autonomous ships, MASS, seafarer training, performance, simulation, human factors, and BNs. The literature selection process comprised four phases: identification, verification, inclusion, and synthesis. In the initial step, 260 papers related to MASS and seafarer education were identified. After removing duplicates and works not directly relevant to educational, security, or human aspects, 81 scientific works published between 2018 and 2023 were included in the shortlist. The selection process is shown in Figure 1.1, which illustrates the information flow of the systematic analysis according to the PRISMA protocol, including the number of excluded papers at each stage.

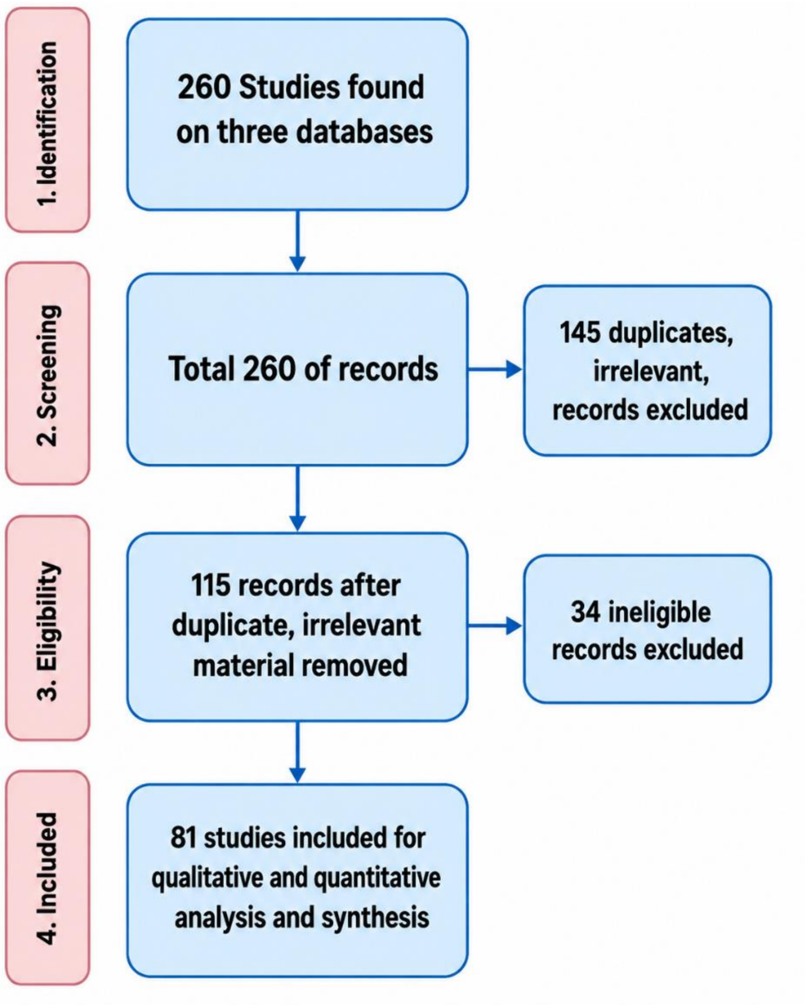

Figure 1.1 The information flow via the phases of systematic literature review [8]

Following the PRISMA based selection, a bibliometric analysis was conducted using VOS viewer to map the interconnections among the most frequently used keywords in the included studies. This graphical analysis reveals the structure of the research field and the main thematic clusters shaping education and training for autonomous shipping. The resulting map shows that technology-centric topics increasingly overlap with human factors, training, and operational safety, reflecting the interdisciplinary nature of research on MASS.

The keyword network centred on "autonomous ships" (Figure 1.2) highlights dominant concepts and their interrelations across the analysed literature. The largest clusters include human factors, safety, training, simulation, and situational awareness, frequently linked with artificial intelligence, risk assessment, and decision-making. These patterns indicate a shift from purely technical concerns towards human-centred and operational perspectives on autonomy, with growing attention to performance evaluation in realistic, safety critical environments.

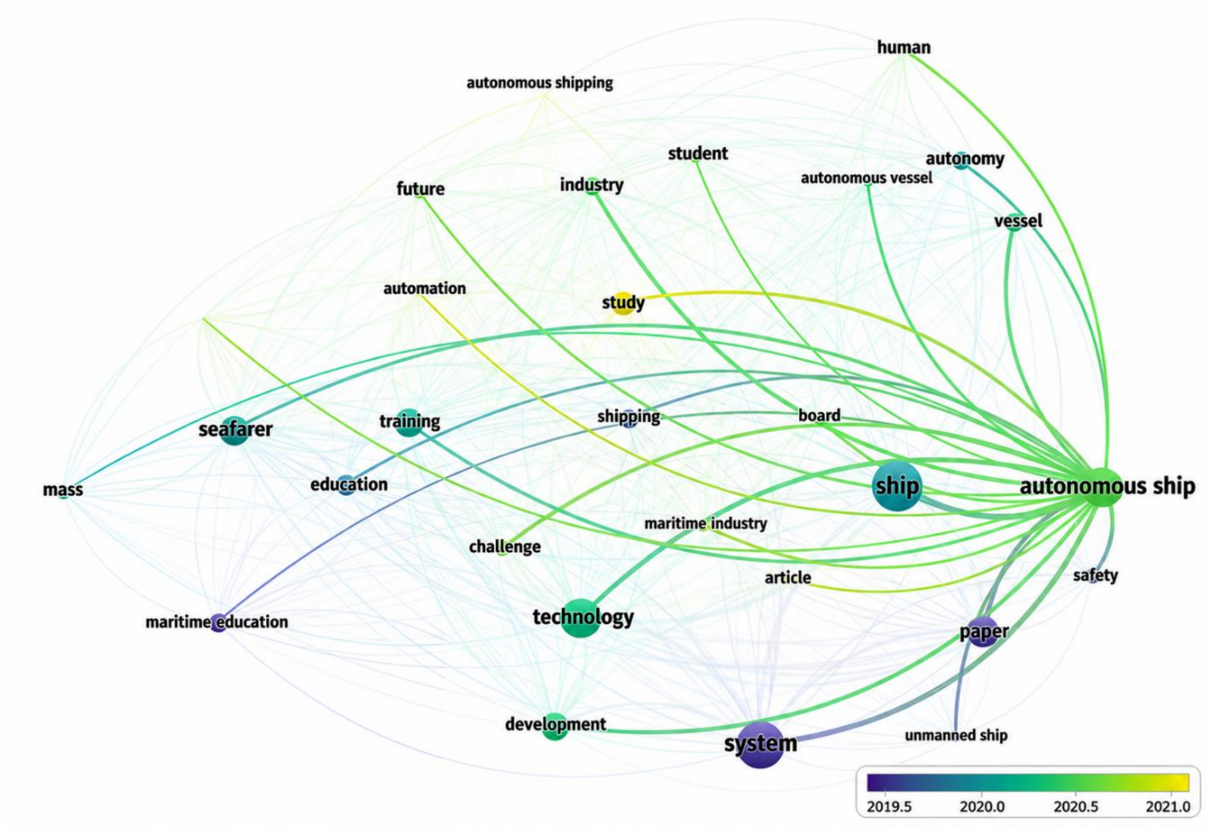

Figure 1.2 The link between the keyword "autonomous ships" and the 30 most frequently used keywords in the research (VOS viewer)[8]

Building on these bibliometric findings, the qualitative synthesis groups the literature into four coherent strands:

- maritime law and review studies,
- curriculum development and adaptation,
- training methods and techniques,
- survey studies and collaborative analyses.

Legal and policy oriented works clarify the evolving regulatory context and the implications of MASS for existing conventions, curriculum focused studies examine how emerging competencies (such as data literacy, remote operations, and cyber awareness) can be embedded into education and training, methodological contributions evaluate simulation-based training and objective assessment techniques (including eye-tracking and psychophysiological measures) and empirical surveys capture perceptions of autonomy, trust in technology, and organizational readiness across different stakeholder groups. Together, these strands delineate the current state of knowledge and reveal specific gaps notably the limited number of data-driven, experimental studies linking simulator derived behavioral and physiological indicators to probabilistic performance models which motivates the research framework developed in this dissertation.

The development of MASS has become one of the most intensively researched areas in the maritime sector in recent years. Digitalisation, artificial intelligence, and remote monitoring systems are changing how ships are operated and redefining the role of humans in the maritime ecosystem. The IMO is establishing a regulatory framework through the development of the MASS Code, which will enable the integration of autonomous ships into global maritime traffic, with a non-binding code planned for 2025 and a binding version for 2030 [9]. This transformation involves not only technological but also organizational and educational changes, as autonomous ships will not completely eliminate human presence but will alter how people monitor and make decisions about system operation [10]. The legal literature highlights those existing international conventions, including the International Convention for the Safety of Life at Sea (SOLAS) and the Convention on the International Regulations for Preventing Collisions at Sea (COLREG), were not originally designed for autonomous systems and therefore require reinterpretation of the terms “master”, “crew”, and “supervision” [11]. In addition to normative aspects, there is also an emphasis on the need for an ethical and responsible approach to implementing autonomy, particularly with regard to liability for human and environmental consequences of system errors. Hult *et al.* [12] emphasise that the development of autonomous ships cannot be considered separately from the broader technological ecosystem, including satellite communication, sensor networks, and artificial intelligence systems, while Fenton and Chapsos [13] note that international organizations, despite progress in defining autonomy, are still harmonising the terminology and criteria for classifying autonomous ships, which also presents a challenge for the educational system that must prepare future operators.

Recent developments in MASS have been significantly influenced by Japan's comprehensive approach to autonomous shipping. The IMO document Current Status of MASS Technological Development Project and Regulatory Framework in Japan (Maritime Safety Committee (MSC) 111-INF.12), submitted to the 111th session of the Maritime Safety Committee, provides an overview of the current state of MASS technological development and regulatory implementation in Japan. Japan treats MASS as an integrated combination of autonomous navigation technologies, operational safety, and regulatory development. The document highlights that the concept is based on the integration of Autonomous Navigation Systems (ANS), Remote Operation Centres (ROCs), and collaboration among more than 50 organizations from the shipbuilding, shipping and artificial intelligence sectors, demonstrating a multidisciplinary approach to the practical implementation of autonomous shipping.

The central element of this initiative is the MEGURI 2040 project, led by the Nippon Foundation carried out between 2022 and 2026. The project includes demonstration trials involving four commercial ships with different operational profiles: the newly built container ship Genbu; the passenger ship Olympia Dream Seto; the retrofitted container ship Mikage; and the ro-ro ship No. 2 Hokuren Maru. These demonstrations cover autonomous navigation, collision avoidance, automatic berthing and unberthing, autonomous manoeuvring in port areas, and remote operational support through dedicated ROCs. In parallel with technological development, Japan established one of the first comprehensive national regulatory frameworks for MASS in 2025, introducing the ANS, functional safety requirements, certification procedures, and survey methods. The framework emphasises that autonomous systems should demonstrate key operational capabilities, including situational awareness, collision avoidance, route execution, handover to human operators, alert management, and data recording, while maintaining appropriate human supervision. These developments represent an important step towards the commercial deployment of MASS and provide valuable practical experience for the ongoing development of the IMO International Code of Safety for MASS (IMO MASS Code) [14].

The adoption of the IMO MASS Code represents a significant milestone in the international regulation of autonomous shipping. Adopted through IMO Resolution MSC.595(111), the non-mandatory IMO MASS Code entered into effect on 1 July 2026, establishing the first international regulatory framework specifically developed for MASS. Prepared by the IMO MSC Working Group on MASS, the IMO MASS Code sets out goal-based principles and functional requirements for the safe, secure and environmentally sound operation of autonomous and remotely operated ships, while supporting the transition towards a future mandatory international regulatory framework [15].

Chapter 14 (Manning, Training and Watchkeeping) is of particular importance for this dissertation, as it addresses the training, familiarisation, competence and watchkeeping requirements for both onboard personnel and remote operators within ROCs. The IMO MASS Code recognizes that the existing STCW Convention may be applied to remote operational roles, while acknowledging the need for future adaptation to the specific operational requirements of MASS. The adoption of the non-mandatory IMO MASS Code marks the beginning of the Experience-Building Phase (EBP), during which operational experience will be collected to support the development of the mandatory IMO MASS Code, currently planned for adoption by 2030. The future mandatory IMO MASS Code is expected to introduce corresponding amendments to SOLAS and other relevant IMO instruments, providing a comprehensive international regulatory framework for the safe implementation and operation of autonomous ships [16].

Recent literature highlights the need to redefine the competencies required for future MASS operators. Emad and Ghosh [17] emphasise that new professional profiles must combine traditional maritime expertise with advanced digital, analytical, and communication skills. Similarly, Stefani and Apicella [18] propose an educational model linking modern technological competencies with cybersecurity awareness and sustainability principles, preparing seafarers for remote and automated operations. Fonseca *et al.* [19] stress the importance of developing a global framework for maritime education that integrates traditional navigational knowledge with data management, automation, and cyber-safety competencies.

Belev *et al*. [20] further support this view by proposing adaptations to the IMO Model Course 7.01 to accommodate the educational requirements of autonomous shipping, emphasising the need to incorporate remote operation, automation management, and digital competencies into future maritime curricula.

Campos *et al*. [21] and Demirel [22] note that the current International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW) does not adequately cover the competencies required for monitoring and remote control of autonomous ships, calling for continuous revision and expansion to reflect technological advances. Sharma and Kim [23] and Kim and Mallam [24] underline the growing importance of digital literacy, cognitive adaptability, and leadership competencies, advocating adaptive and individualised training approaches within maritime education. Similarly, Emad *et al*. [25] argue that the transition towards autonomous maritime transport requires a comprehensive restructuring of maritime education and training systems, including the development of competency frameworks tailored to the requirements of remote and autonomous ship operations.

Beyond curriculum development, recent studies have increasingly emphasised the importance of stakeholder perceptions and Operator Readiness for the successful implementation of MASS. Bogusławski *et al*. [26] report that maritime cadets generally recognize the long-term benefits of autonomous shipping but also express concerns about whether existing education and training programmes adequately prepare future professionals for emerging operational roles. Similarly, Mallam *et al.* [27] show that maritime cadets anticipate substantial changes in traditional shipboard responsibilities, highlighting the growing importance of digital competencies, systems thinking, supervisory decision-making, and the ability to manage highly automated processes. These findings indicate that successful implementation of MASS depends not only on technological progress but also on the preparedness of future operators and the adaptation of maritime education.

Some studies have further expanded this perspective by examining the competencies and attitudes required for safe operation in autonomous maritime environments. Kim [28] identifies system monitoring, situation assessment, communication, and decision-making in remote operational settings as fundamental competencies for future MASS operators. Aalberg [29] shows that trust in autonomous technologies is strongly influenced by professional commitment and operational experience, suggesting that operator acceptance is a critical prerequisite for successful technology adoption. Complementing these findings, Belabyad *et al.* [30] argue that future maritime education should adopt an integrated approach, combining technical, cognitive, and non-technical competencies to prepare operators for increasingly complex human–automation collaboration. Collectively, these studies show that stakeholder perceptions, Operator Readiness, and competency development are essential foundations for the safe, effective, and sustainable implementation of autonomous maritime systems.

Building on these findings, increasing attention has been given to the objective assessment of competencies required for autonomous ship operations. Li *et al*. [31] propose a risk-based framework for evaluating MASS operator competencies by integrating technical knowledge, operational skills, and safety-related performance indicators into a structured assessment model. Similarly, Hwang *et al.* [32] show that simulation-based training provides an effective environment for objectively evaluating navigation performance and improving operational competencies under autonomous navigation conditions. These studies highlight the growing importance of evidence-based approaches to measuring operator performance beyond traditional knowledge-based assessment.

Some research has further emphasised the need to establish comprehensive competency frameworks that support both training and performance evaluation. Kim *et al.* [33] introduce a continuous learning framework that incorporates human decision-making into simulation-based autonomous ship operations, enabling systematic improvement of operator performance through iterative learning processes. Likewise, Kuntasa *et al.* [34] develop a conceptual competency model for remote MASS operators, identifying the knowledge, technical abilities, cognitive skills, and non-technical competencies required for effective supervisory control. These studies demonstrate a clear transition from conventional competency-based education to data-driven and performance-oriented assessment approaches, providing an important scientific foundation for the simulation-based BN model proposed in this dissertation.

Recent research increasingly recognizes that operator performance in MASS operations is strongly influenced by the quality of human–machine interaction and the establishment of appropriate trust in autonomous systems. Lützhöft *et al*. [35] advocate a human-centred approach to maritime autonomy, arguing that autonomous technologies should be designed around human cognitive capabilities and operational needs, rather than requiring operators to adapt to automated systems. This perspective highlights the importance of maintaining situational awareness, supporting effective supervisory control, and ensuring meaningful interaction between operators and autonomous technologies.

Existing studies further demonstrate that trust in automation is a dynamic factor influencing operator behavior and decision-making. Belabyad *et al*. [36] identify human–automation collaboration, communication, and adaptive interaction as essential components of future autonomous ship operations. Similarly, Aalberg *et al*. [37] show that trust in automated shipboard systems is closely associated with professional identity and previous operational experience, while Song *et al.* [38] propose an integrated framework in which safe and efficient MASS operations depend on continuous coordination between human operators and intelligent decision-support systems. Current studies indicate that effective operator performance in autonomous ship operations depends not only on technical performance but also on maintaining appropriate levels of trust, situational awareness, and human–machine collaboration throughout supervisory control tasks.

Simulation research is recognized as an effective tool for empirically studying human performance in complex operations. High level realism simulators enable repeatable experiments under controlled conditions, where reaction time, decision accuracy, and adaptation to changing circumstances can be measured. Minami *et al.* [39] highlight the importance of scenario-based safety assessments that include both normal and emergency situations, while Ahvenjärvi *et al.* [40] note the development of simulators as research and educational tools reflecting technological advances in maritime operations. Martelli *et al*. [41] and Smirnov and Tomforde [42] demonstrate the use of augmented and virtual reality and agent based simulations to evaluate operator behavior and assess navigational decisions in autonomous conditions. Ronca *et al.* [43] confirm the importance of neurophysiological and biomechanical measurements in analysing stress and attention during navigational tasks, which underpins the development of objective competency metrics.

The Warsash Maritime MASS Research Centre [44] stresses that integrating simulators into research and training allows detailed study of human–machine interaction and the development of customised training modules for remote ship control. Lokuketagoda *et al.* [45] emphasise the importance of linking theoretical knowledge and practical training using modern engine room simulators, which fosters critical thinking and the ability to solve complex operational problems. Emad and Kataria [46] show that simulation training can significantly reduce the gap between traditional and digital maritime education, particularly through quantitative learning models and the development of cognitive skills required for autonomous operations. Sandaruwan *et al.* [47] further note that ship motion simulations in six degrees of freedom (6DoF) enable deeper analysis of operator reactions in virtual conditions and a more realistic assessment of readiness for real situations. These approaches confirm that simulation results can be effectively used to predict performance in real circumstances and to develop new competency metrics based on objective indicators.

Wróbel *et al*. [48] warn that automation may reduce operational errors but increases the risk of system oversight and misinterpretation of information. These findings confirm that current methods of evaluating seafarers, based on subjective assessments by instructors, are no longer sufficient for an increasingly complex environment in which humans and machines complement each other.

Over the past decade, the use of quantitative and mixed methods to analyse seafarers' perceptions and experiences of autonomous systems has increased. Authors Chan *et al*. [49] examine how experience, previous training, and professional exposure influence perceptions of autonomy and trust in technology. Survey results show that younger generations of seafarers are more willing to accept digital tools, algorithmic decision-support, and remote monitoring systems, while experienced officers emphasise the importance of maintaining human supervision at all levels of autonomy. Similar findings were reported in studies by Kennard, Zhang and Rajagopal [50], and Sharma and Kim [23], who highlight that experience, situational awareness, and the ability to maintain a sense of safety are key competencies for future operators of autonomous ships. Jo *et al*.[51] and Nasur and Bogusławski [52] find that the level of technological knowledge and attitudes towards autonomy differ significantly between students and officers, with younger groups showing greater trust in automated processes. Research by Hannaford and Van Hassel [53] indicates that, despite positive attitudes towards technological progress, concerns remain about reduced human control and over-reliance on sensor systems. These findings confirm that technological solutions must be accompanied by psychological, organizational, and pedagogical adaptations within the maritime industry, with emphasis on building trust, transparency, and understanding in human–autonomous system interactions.

Building on insights from human factors and simulation approaches, an increasing body of research focuses on the application of BNs in the maritime domain as a tool for quantifying risk, reliability, and human error. BNs enable the modelling of uncertainty and interdependencies between human, technical, and environmental variables, providing a structured framework for assessing complex relationships in autonomous operations. Trucco *et al.* [54] were among the first to apply a BNs model to analyse organizational factors in maritime transport, while Kristensen *et al*. [55] used this approach for trajectory planning of autonomous surface ships, where risk is an input variable for navigational decision-making. Palbar Misas *et al.* [56] explore the connection between cybersecurity, trust, and situational awareness in the context of remote ship control, highlighting the need for new forms of training that reflect the risks of a digitalised environment.

Li *et al*. [57] develop a combined fault tree and Fuzzy Bayesian network (FTA-FBN) model for collision risk assessment, while Aydin *et al*.[58] confirm the effectiveness of the fuzzy-BN approach in assessing the risk of navigation in narrow waterways. Guo *et al*.[59] use BNs to quantify the probability of loss of navigational control in autonomous passenger ships, while Zhao *et al*. [60] analyse the causes of maritime accidents using the BN method and highlight the vulnerability of autonomous systems to external factors such as weather conditions and traffic density. Zhang *et al*.[61] model human error and collaborative decision-making between humans and autonomous systems, while Fan *et al*. [62] apply advanced BN risk analysis in inland navigation to identify key factors affecting safety. Sezer *et al*. [63] develop a hybrid human reliability model that integrates BNs with the Cognitive Reliability and Error Analysis Method (CREAM), emphasising the importance of structured training and clearly defined procedures in MASS operations at the third level of autonomy. Chang *et al*. [64] and Weber *et al*. [65] present advanced integrations of BNs with methods such as Failure Mode and Effects Analysis (FMEA), Evidential Reasoning (ER), and Recursive Bayesian Networks (RBN), enabling comprehensive quantification of risks and assessment of interactions between technical and human factors.

These studies confirm that BNs are a flexible and powerful tool that unifies subjective assessments, empirical measurements, and simulation data into a single probabilistic decision-making model. Their application in the maritime context enables the prediction of system behavior under uncertainty, estimation of error probabilities, and identification of key variables affecting the safety and efficiency of MASS operations. Such models form the theoretical and methodological basis of this dissertation, in which the BN approach is used to analyse human factors and evaluate performance in simulated autonomous navigation conditions.

The time trend analysis in Figure 1.3 shows a continuous increase in the number of scientific publications after 2017, confirming that education for autonomous navigation is a relatively new and fast-growing research area.

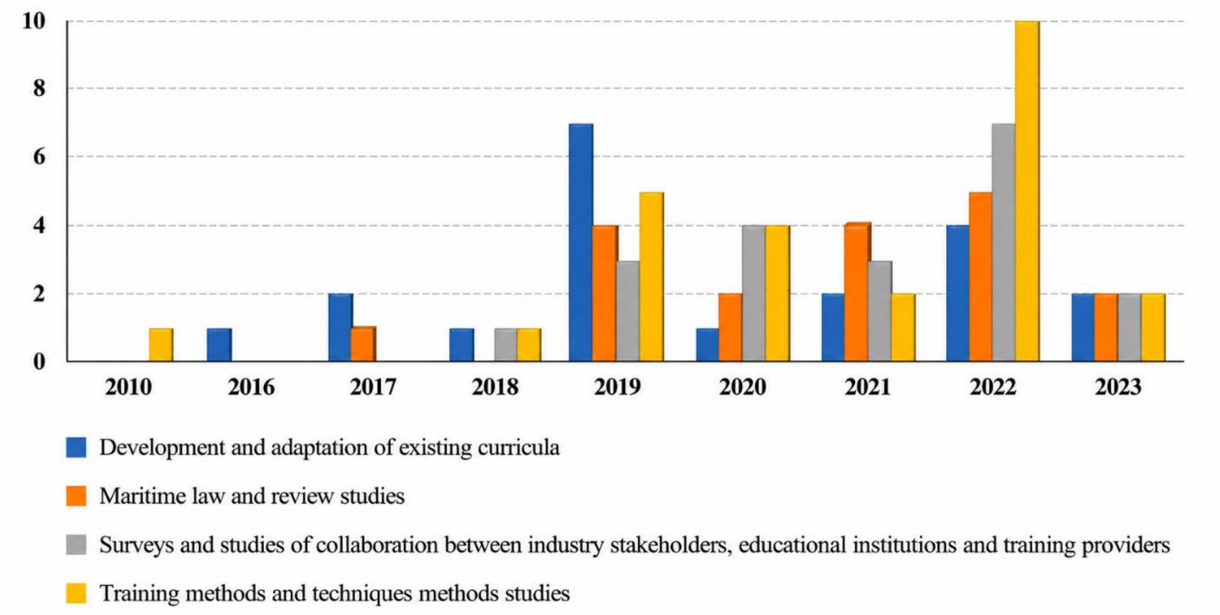

Figure 1.3 Temporal trend of publications related to MASS (2010–2023) [8]

Despite this growth, analysis of existing works reveals several key gaps. Most research to date is based on theoretical models, surveys, or expert assessments, while experimental studies combining physiological measurements, objective performance indicators, and integrated BN models remain very rare.

The results of the bibliometric analysis indicate that research interest is increasingly directed towards interdisciplinary approaches connecting technical and human factors in autonomous ship operations. However, most available literature still considers the competencies and training of future MASS operators from a descriptive and conceptual perspective, with a lack of empirical data collected in real or simulated conditions. This creates new research opportunities linking maritime expertise, data analytics, and probabilistic modelling to quantitatively understand human performance under uncertainty and autonomous decision-making.

This dissertation contributes to this goal by developing an empirically based model that integrates data from simulation exercises, biometric measurements, and BNs to assess and improve the competencies of future MASS operators. In this way, the research builds on the existing scientific foundation, provides a methodological model for objective evaluation of human performance, and lays the groundwork for the development of new educational standards and training programmes in maritime autonomy.

1.7. The Structure of the Doctoral Dissertation

Based on the gaps identified in the review of previous research, particularly the lack of empirically grounded, simulation-based studies integrating behavioral, physiological, and probabilistic modelling approaches, this dissertation is structured to systematically address these shortcomings.

Following the Introduction and Review of Previous Research, which establish the scientific context and identify unresolved challenges related to human performance in MASS operations, the dissertation first presents the experimental infrastructure used in the research. This includes the navigational simulator, the selected ship model, eye-tracking equipment, and physiological measurement devices. The next chapter provides a detailed description of the simulation exercise, including participant preparation, the operational area, the navigational route, traffic conditions, and the system failures introduced during the scenario.

Subsequent chapters introduce the theoretical foundations of BN, followed by the data collection and processing framework and the definition of variables used in the BN model. This structure enables a clear transition from the identified research gap, through experimental design and data acquisition, to probabilistic modelling, validation, and interpretation of results.

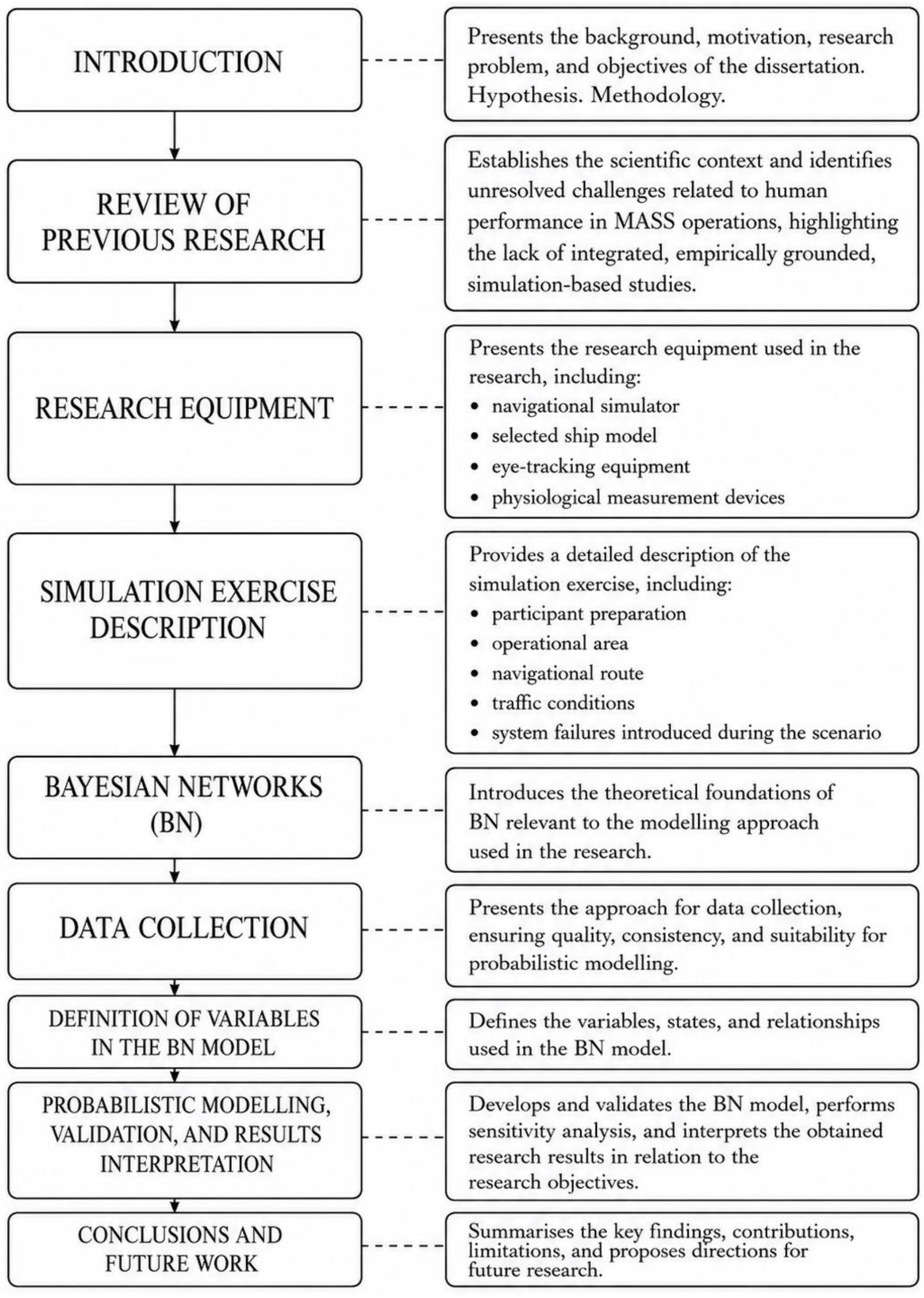

Figure 1.4 Structure of the Doctoral Dissertation

2. RESEARCH EQUIPMENT FOR DATA ACQUISITION AND BN MODEL DEVELOPMENT

This chapter provides an overview of the equipment used in the research and explains its complementary role in the experimental setup and data collection. The selected equipment provides synchronized navigational, visual, physiological, and behavioral data, enabling a comprehensive assessment of operator performance during MASS simulation exercise. Particular attention is given to the navigational simulator, which is the central element of the experimental environment and the platform for integrating all measurements.

2.1. Navigational Simulator

The navigational simulator is a key tool for conducting experimental exercises focused on human factors and decision-making in the context of ship operations. In this research, the Wärtsilä Navigation simulator NTPRO 5000 (Navi-Trainer Professional 5000), located in the Electronic Chart Display and Information System (ECDIS) classroom at the Faculty of Maritime Studies in Split, was used (Figure 2.1) The simulator was configured with a „Tug console", enabling the execution of complex manoeuvres and emergency scenarios in a realistic maritime environment.

The simulator reproduces navigational conditions through advanced hardware and software components, including an ECDIS, radar, Automatic Identification System (AIS), autopilot, engine control system, and various sensors equivalent to those used on real ships. Its high level of physical, functional, and psychological fidelity realistic conditions for studying operator behavior and decision-making processes during navigation. The concept of simulator fidelity in maritime training comprises these three interrelated dimensions, which collectively determine the realism and overall effectiveness of simulation-based learning and assessment [66].

Another important feature of the simulator is its ability to synchronously record user actions and navigational performance throughout the experimental session. Every task performed by participants, including radar activation, course changes, speed adjustments, and transitions between manual and automatic steering modes, is automatically logged and subsequently analysed. The simulator also provides objective indicators such as response time to navigation system failures, deviations from the planned route, vessel speed profile, arrival time at the assigned position, and near miss events. These simulation records were used to identify key indicators of navigational performance and operator behavior during task execution, which were then incorporated as operational variables in the BN model.

The Wärtsilä Navigation NTPRO 5000 simulator has been widely used in maritime research and training, demonstrating its ability to provide a reliable and realistic environment for investigating navigational performance and human factors in ship operations[67], [68].

The increasing development of e-navigation, digitalisation, and unmanned ships requires a redefinition of simulator-based education and training to ensure that scenarios accurately reflect the challenges of modern and future maritime operations [69]. In addition, the integration of biometric data collection tools into simulation environments represents a significant enhancement of maritime research. Simulator-based navigation exercises combined with physiological measurements, such as heart rate and gaze tracking enable the evaluation of operator workload and situational performance during realistic maritime tasks [70]. This approach allows researchers to link behavioral actions with physiological indicators, providing a comprehensive understanding of operator capability and stress levels during critical operations.

Simulator-based exercises allow complex and high risk scenarios to be conducted safely, without endangering ships, crews, or the environment. This ability to test decision-making, performance, and resilience under controlled yet realistic conditions is one of the main advantages of using simulation in maritime autonomy research.

Figure 2.1 Wärtsilä Navigation simulator NTPRO 5000 (Navi-Trainer Professional 5000), "Tug console"

2.2. Reference Ship and Simulation Model

For the experimental part of this research, the OSV9 ship model, with an overall length of 80.4 m, was selected as the simulation platform. This model was chosen because its principal dimensions and operational characteristics closely correspond to those of the reference autonomous ship Yara Birkeland, allowing a realistic representation of autonomous ship operations within the navigational simulator.

The Yara Birkeland is recognized as the world's first commercial autonomous container ship and represents a significant milestone in the practical implementation of MASS. Due to its dimensions, operational profile, and level of automation, it was adopted as the reference ship for this research. The Yara Birkeland was not available as a predefined ship model within the

Wärtsilä NTPRO 5000 simulator environment. Therefore, the OSV9 model was selected as a substitute, as its principal dimensions and operational characteristics closely resemble those of the reference ship and provide a realistic platform for investigating operator performance during autonomous ship operations.

As shown in Figure 2.2 and Table 2-1, both ships exhibit highly comparable principal dimensions and operational speed within the experimental scenario. Although differences exist in ship type and propulsion concept, these characteristics do not significantly affect the objectives of this research, which focuses on operator performance, decision-making, and human interaction with autonomous ship systems rather than on detailed ship hydrodynamics or propulsion efficiency.

(a) Yara Birkeland *(b) OSV9 simulator model*

Figure 2.2 Reference ship (a) and simulator model (b) used in the exercise [71]

Table 2-1. Comparison of the principal technical and operational characteristics of the reference vessel Yara Birkeland and the OSV9 simulator model

Characteristic	Yara Birkeland	OSV9 Simulator Model
Ship type	Autonomous container ship	Offshore Support Vessel
Length overall (m)	80.0	80.4
Beam (m)	15.0	18.0
Draft (m)	6.0	6.6
Maximum speed (kn)	13	16.2

Operating speed in the exercise (kn)	10	10
Number of propellers	2	2
Propeller type	Azimuth propellers	Controllable Pitch Propellers (CPP)
Bow thruster	Yes	1 × 883 kW
Stern thruster	Yes	1 × 883 kW

2.3. Eye-Tracking and Physiological Measurement Devices

Objective assessment of human performance during maritime simulation exercises requires the integration of behavioral and physiological measurement technologies. In this dissertation, participants were monitored using Pupil Core eye-tracking glasses and the Huawei Watch D2 smartwatch, enabling simultaneous collection of visual attention and physiological response data throughout the simulation exercise.

The eye-tracking system records participants' gaze behavior using two infrared eye cameras and one outward facing scene camera, with a sampling rate of up to 200 Hz. The recorded data enable precise determination of where, when, and for how long participants direct their attention to navigational instruments such as radar, ECDIS, and the visualization display. Based on the gaze data, several quantitative indicators were calculated, including the Gaze Dispersion Index (GDI), which describes the spatial distribution of visual attention; the Gaze Movement Score (GMS), which reflects gaze transition dynamics; and the Pupil Entropy Index (PEI), which represents changes in pupil diameter associated with cognitive workload. These indicators were integrated into the Stress-induced Gaze Pattern Score (SGPS), from which the Visual Attention Quality (VAQ) variable was derived. This variable was incorporated into the BN model as an indicator of visual attention and used in the assessment of Cognitive Overload and operator performance during autonomous ship operations. This approach enables objective evaluation of visual scanning strategies and situational awareness, providing a quantitative basis for analysing human performance in complex navigational environments [72],[73].

Continuous physiological monitoring was performed using the smartwatch. Heart rate was recorded throughout the simulation exercise, providing objective information on participants' physiological responses to increasing workload and stress. Blood pressure was measured immediately before and after the simulation using the smartwatch's integrated oscillometric wrist-based measurement system. The measurement principle implemented in the Huawei Watch D platform has been clinically validated and provides reliable results suitable for research applications [74]. The smartwatch was worn on the participant's non-dominant wrist and did not interfere with interaction with the simulator controls.

Recent study has shown that combining eye-tracking metrics with physiological indicators provides a more comprehensive assessment of operator workload, stress, and decision-making performance than either source alone [43]. The open-source architecture of the Pupil Core system enabled synchronization of gaze recordings with simulator outputs and biometric measurements through the Pupil Capture and Pupil Player software packages, resulting in a unified dataset linking visual attention, physiological responses, and operational performance. Such multisensory integration represents an important methodological approach for human factors research and provides a robust basis for evaluating the readiness of future autonomous ship operators.

Figure 2.3 Pupil Core eye-tracking glasses and Huawei Watch D2 smartwatch [75] [76]

3. SIMULATION EXERCISE

As previously stated, the purpose of the simulation exercise is to assess participants' performance under conditions reflecting autonomous ship operations, with particular focus on decision-making, situational awareness, and response to system failures. The simulation scenario was designed to reflect the operational principles of IMO Degrees of Autonomy 2 and 3, in which ship operations are conducted under remote supervision and control, while the human operator is responsible for monitoring autonomous navigation and intervening in response to unexpected situations or system failures. The following section describes the preparation and familiarisation of participants, as well as the simulation scenario, including its general characteristics, exercise area, and navigational route.

3.1. Preparation and Familiarisation

Before starting the simulation exercise, all participants were required to read an instructional document describing the scenario, navigational equipment, and operational tasks. This familiarisation step ensured a consistent baseline understanding and reduced perceptual discrepancies between participants, which is important given evidence from eye-tracking studies showing that insufficient familiarisation can affect the gap between perceived and actual navigational behavior [77].

To ensure methodological consistency, all participants underwent the same familiarisation procedure following a standardised protocol. The familiarisation lasted approximately five minutes and included an introduction to the simulator layout, navigational displays, control interfaces, and operational procedures relevant to the experimental scenario. The familiarisation phase was completed only after each participant had demonstrated adequate understanding of the simulator functions and confirmed readiness to perform the assigned tasks. This ensured that all participants began the experimental exercise with equivalent preparation, minimizing the influence of differences in prior simulator experience.

The simulation also included a structured watch handover procedure, reflecting real navigational practice and enhancing the methodological validity of the experiment. The handover was conducted in two phases: initial familiarisation with bridge equipment and vessel status, followed by a formal transfer of responsibility between the outgoing and incoming officer of the watch. This protocol aligns with the recommendations of the International Chamber of Shipping (ICS) Bridge Procedures Guide [78], which emphasises that an effective handover must provide the incoming officer with a complete and accurate operational picture to maintain situational awareness.

3.2. General Overview of the Exercise

The simulation exercise was designed to test the skills and reactions of operators when operating an autonomous ship in a coastal area. Prior to involving research participants, the exercise scenario was preliminarily tested by experienced faculty colleagues and professional seafarers to confirm its realism, safety, and appropriate level of complexity.

Ship handling in coastal waters is more complex than on the open sea due to the higher traffic density, the proximity to the coast and the presence of different types of ships as well as navigational hazards. The combination of these factors creates additional pressure on the operator and increases the need for timely and accurate decision-making. It is therefore assumed that an operator who successfully completes the exercise in such an environment will also be able to effectively manage operations in the simpler conditions of the open sea, where the challenges are primarily related to open sea navigation, meteorological conditions and possible technical errors, but without concurrent navigational limitations.

The main objective of the exercise is to assess the skills of the different categories of participants according to their level of training and previous experience in navigation. The exercise takes place at dusk, which makes the scenario even more realistic. Under these conditions, the navigation lights, lighted aids to navigation, shore background lighting and the basic contours of the coastline are visible, while the visibility of finer details is reduced, making orientation more difficult and requiring greater attention from the operator.

The simulation scenario followed the principle of progressive complexity, presenting participants with increasingly challenging navigational situations throughout the exercise. The combination of normal traffic encounters and deliberately introduced system disturbances was intended to create realistic operational conditions that required continuous monitoring, situational awareness, and timely decision-making. This approach enabled a comprehensive assessment of operator performance under conditions representative of remotely autonomous ship operations.

3.3. Description of the Navigation Area

For the purposes of this research, a coastal area in the immediate vicinity of Split was selected, which includes the entrance to Kaštela Bay on the Split side (Figure 3.1). This area represents one of the most complex maritime zones in the Croatian part of the Adriatic Sea, as different types of maritime traffic and activities converge here. Due to this complexity and dynamics, it is well suited to test the operator's skills in controlling an autonomous vessel.

Kaštela Bay itself is home to a number of important economic and transport facilities. On the northern shore there are terminals for handling bulk cargo, general cargo and petroleum products. These terminals regularly receive ships of various sizes, including international merchant ships, and thus contribute to the traffic volume in the region. In addition to the commercial terminals, Kaštel Gomilica also has a large marina with a considerable number of yachts and sailing boats, which are mainly used for tourist purposes.

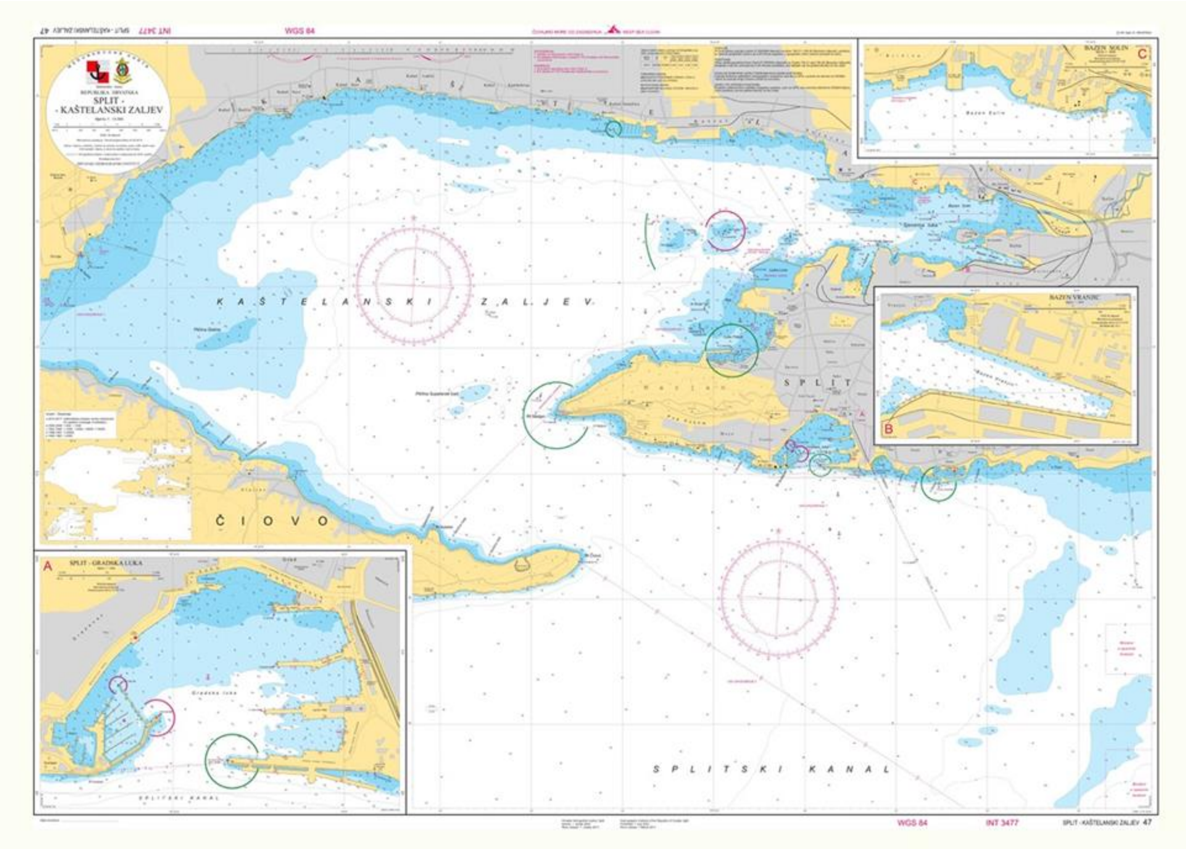

Figure 3.1 Harbour chart of the Kaštela Bay[79]

A special role for security and traffic is played by the Lora naval base, from which naval ships and other public ships regularly depart for military operations and exercises. In addition, there is a air base in the Kaštela area that specialises in helicopter operations, which further increases the strategic importance of this sea area. There is also a fishing harbour in the immediate vicinity, which serves as a base for a large number of fishing vessels, most of which are small but very active in coastal fishing. Additional complexity is added by the shipyard in Kaštela Bay and the Plovput base (responsible for the maintenance and placement of navigational aids).

In the wider area, there is also a harbour of the Institute of Oceanography and Fisheries at the tip of the Marjan peninsula in Split, which is used for the deployment of specialised research vessels. In addition, Split Airport, located on the northern part of Kaštela Bay, is an important mode of transport. Next to the airport, there is a special pier for ships that bring passengers from Split and the surrounding islands directly to the terminal, creating a link between sea and air transport.

In addition to the larger terminals and harbours, there are more than ten smaller marinas and harbours along the coast of Kaštela Bay that serve as berths for small private boats and ships. Altogether, these facilities form a maritime zone with pronounced traffic heterogeneity, in which commercial, military, fishing, tourist and leisure shipping take place simultaneously.

Due to the high level of activity, the area is categorized as a relatively dense traffic area. In addition to the international commercial ships entering Split harbour, there are also fishing vessels, naval ships, tourist yachts, other non-SOLAS ships and passenger ships connecting the airport with the surrounding areas. Such a mix of traffic creates a highly demanding navigational environment, particularly for autonomous vessel operations, where it is essential to rapidly and accurately determine a vessel's intentions, her navigational status, and the applicable COLREGs responsibilities.

Alongside this, there are zones with navigation restrictions as well as prohibited areas in Kaštela Bay that further limit the manoeuvring space for vessels. The existence of such zones requires operators to be constantly vigilant and adapt their navigational decisions to the spatial restrictions. The approach to the international harbour of Split is specially regulated and marked with the IALA system, with buoys and lights indicating safe passage. The approach markings also indicate shoals, reefs and small islands in the vicinity of the navigation route, all of which are potential dangers to safe navigation.

The selection of this area is based on the fact that it simultaneously presents a multitude of challenges: a high traffic density, the heterogeneity of vessels, the presence of several specialised terminals and ports, a military component, fishing activities, leisure and tourist shipping and a direct link to air traffic. This diversity and complexity make the area an ideal testing ground for simulating and assessing the performance of autonomous ship operators, as it reflects almost all aspects of maritime transport that can be encountered on a global scale in a relatively limited area.

3.4. Description of the Navigation Route

The exercise begins at position 43°29.92'N, 016°24.06'E (Figure 3.2) and consists of three segments. The first segment is designed to gradually introduce participants to the scenario, with medium level navigational requirements. The ship enters the navigation channel between the

Marjan peninsula and the island of Čiovo, the main approaches to Kaštela Bay. This segment is characterized by increased traffic activity, but situations are structured so that the operator has sufficient time to make decisions and perform basic manoeuvres. In this segment, the ship proceeds on a course of 289.5° over a distance of 1.7 nautical miles, representing the initial and navigationally simplest part of the exercise.

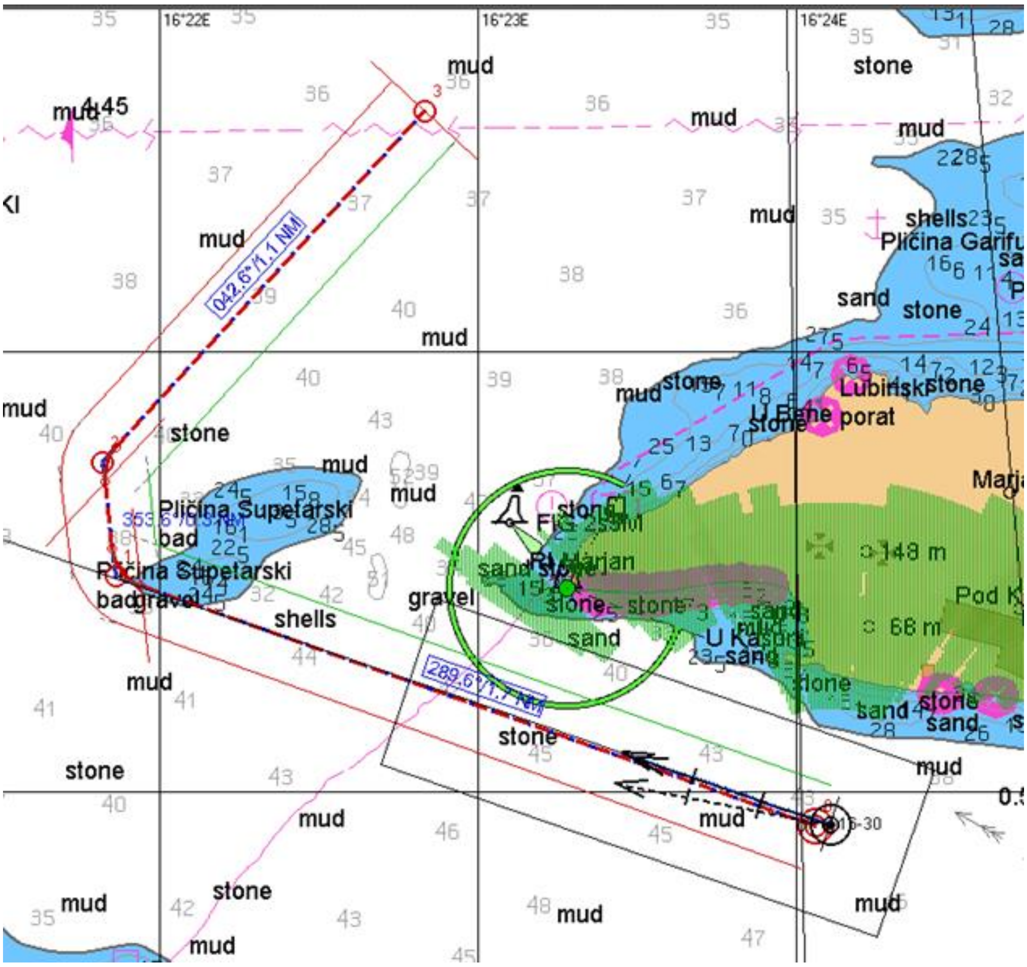

Figure 3.2 Navigation route used in the simulator exercise (ECDIS)

At the entrance to the channel, the ship passes the motor yacht, proceeding at 8.7 knots on the starboard side of the fairway, while simultaneously the super yacht approaches on the port side and passes the autonomous ship. These situations require participants to apply the basic rules of COLREGs and correctly assess the right of way in conditions of simultaneous encounters on both sides. Additional complexity is introduced by the motor yacht, travelling at 15 knots, which, although at a safe distance, crosses the course of the autonomous vessel. Although this situation does not require an immediate reaction, the operator must monitor’s movement and evaluate its potential development to maintain safe navigation. At the same time, a helicopter overflight from a nearby air base occurs above the navigation area, further enhancing the realism and dynamic nature of the traffic environment. Although the helicopter does not directly affect the navigation route, its presence symbolises the complexity and multi-layered nature of activities in Kaštela Bay.

At a greater distance, ships approaching head on become visible, increasing the sense of traffic density and requiring continuous visual and radar monitoring. In addition to larger ships, smaller sport fishing boats are observed on the starboard side, operating in shallower waters. Their presence further complicates navigation, as participants must ensure a safe distance and, after passing these boats, execute a timely turn to starboard to remain within the designated route.

This first segment of the route is designed to require operators to correctly identify and coordinate with different types of ships, from merchant ships and tugs to smaller sport fishing boats, when entering a relatively congested area. The objective is to test their ability to monitor multiple situations simultaneously, correctly apply basic navigation rules, and make decisions that ensure safe entry into the channel.

The second segment of the route begins at position 43°30.25’N, 016°22.78’E. This part of the route is structured to further increase the complexity of traffic situations compared to the initial segment, requiring greater concentration and simultaneous monitoring of multiple elements in the navigational environment. During the second segment, the ship procced in the previous course of 289.5° before executing a starboard turn to 353°, covering a total distance of 1.0 nautical mile before reaching the next waypoint.

At the start of this segment, the ship passes motor yacht on the port side. A high-speed ferry, travelling at 21.5 knots, also passes on the port side of the autonomous ship. Almost simultaneously, a rigid hull inflatable boat approaches from starboard and crosses the course of the autonomous ship. This combination of situations creates additional pressure on the operator, who must monitor two different encounters with ships approaching from opposite sides.

Participants are expected to correctly assess the right of way in accordance with COLREGs and make decisions that allow safe manoeuvring through the channel.

As the route progresses, a starboard turn must be executed at the waypoint located at 43°30.49'N, 016°21.86'E. During this manoeuvre, an additional challenge arises from sport fishing boats operating in the shallower part of the channel on the starboard side. Their presence requires increased caution to ensure the safety of the turn.

After the turn, a naval ship appears ahead. This ship is not visible on the ECDIS and can only be detected by radar and visual means. Its presence requires operators to maintain heightened situational awareness and make precise use of radar data, as the identity and intentions of the naval ship are unknown. This situation is designed to assess how participants react under conditions of partially limited information and the need to rely on all available navigational means.

This second segment of the route tests participants' capabilities under conditions of increased traffic density, gradual reduction of manoeuvring space, and uncertainty when support systems (such as AIS) are unavailable. In this way, the simulation accurately reflects the complex challenges that operators of autonomous ships may encounter in real world conditions. The third segment of the route is conducted on a course of 042° over a distance of 1.1 nautical miles, leading to the final waypoint of the exercise at position 43°31.55'N, 016°22.83'E.

The third segment of the route begins at 43°30.75'N, 016°21.82'E, which also serves as a waypoint for another starboard turn. This phase further increases scenario complexity, as the operator must make decisions under conditions of both reduced technical information and intensified traffic activity.

After the turn, the route leads to the waypoint at 43°30.89'N, 016°22.00'E, where a sudden radar blackout occurs. This element of the simulation tests participants' ability to maintain control of the situation during a failure of a key navigational system. The absence of radar data significantly complicates monitoring of the navigation environment and requires operators to rely on visual observations and any remaining available systems.

At the same time, two small sailing boats, sailing at speeds between 5 and 6 knots, appear in the area, one on the starboard side and one on the port side, both intending to cross the course of the autonomous ship. Additional complexity arises from the fact that these boats are not displayed on ECDIS, highlighting the limited reliability of the system in detecting smaller sports and recreational boats. Participants must therefore ensure timely assessment and decision-making regarding avoidance manoeuvres without the support of two key systems radar and ECDIS.

This third segment of the route is conceived as the culmination of navigational complexity. The emphasis is on testing participants' ability to maintain situational awareness and make safe decisions despite the absence of complete informational support. Particular attention is given to their flexibility in transitioning from reliance on sophisticated systems to the use of basic navigational methods and visual observation.

In the immediate vicinity of the route, several other ships are present. Although they do not directly affect the planned track, provided the ship adheres to the designated route and complies with maritime navigation rules, their presence contributes to the realism of the simulation, ensuring the scenario more accurately reflects real navigation condition's. Sequential structure of the simulation exercise (Figure 3.3), showing three consecutive segments with progressively increasing navigational complexity and operator workload.

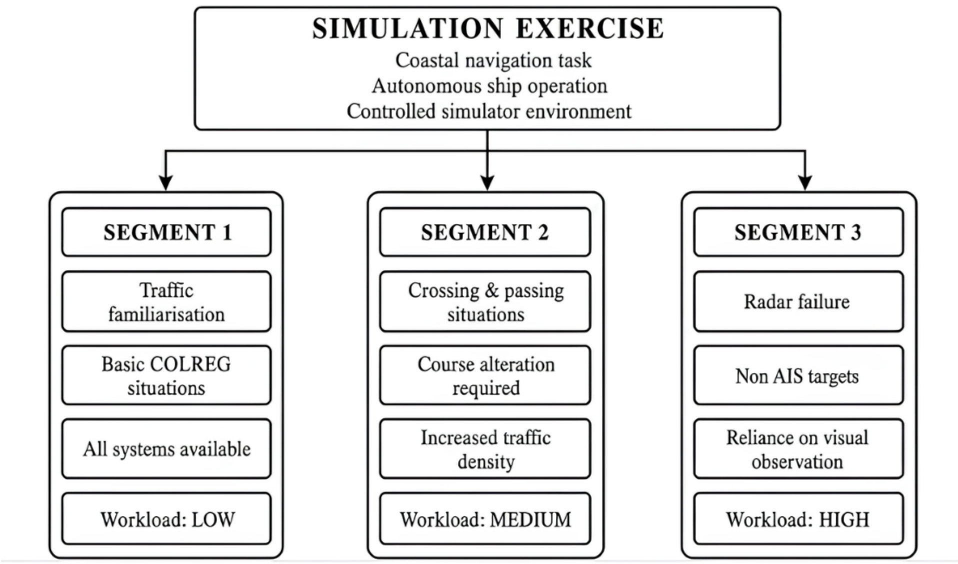

Figure 3.3 Sequential structure of the simulation exercise

The exercise ends at position 43°31.55’N, 016°22.83’E, where the ship enters the area where tugboats are positioned to provide assistance if required during the final approach to the port. At this point, control of the ship is transferred from the remote operator to the onboard crew, marking the end of the remote operation phase. The subsequent navigation towards the port, located in the north eastern part of Kaštela Bay, is conducted under direct human supervision due to the increased complexity of manoeuvring in confined waters. At this position, participants’ overall performance is evaluated, and their decisions and actions are compared against the defined objectives of the exercise. Segments of the route (in cream colour) are shown in Figure 3.4 together with the trajectories (in black) of crafts that affect the navigational situation in the vicinity of the autonomous ship’s route (shown in cream).

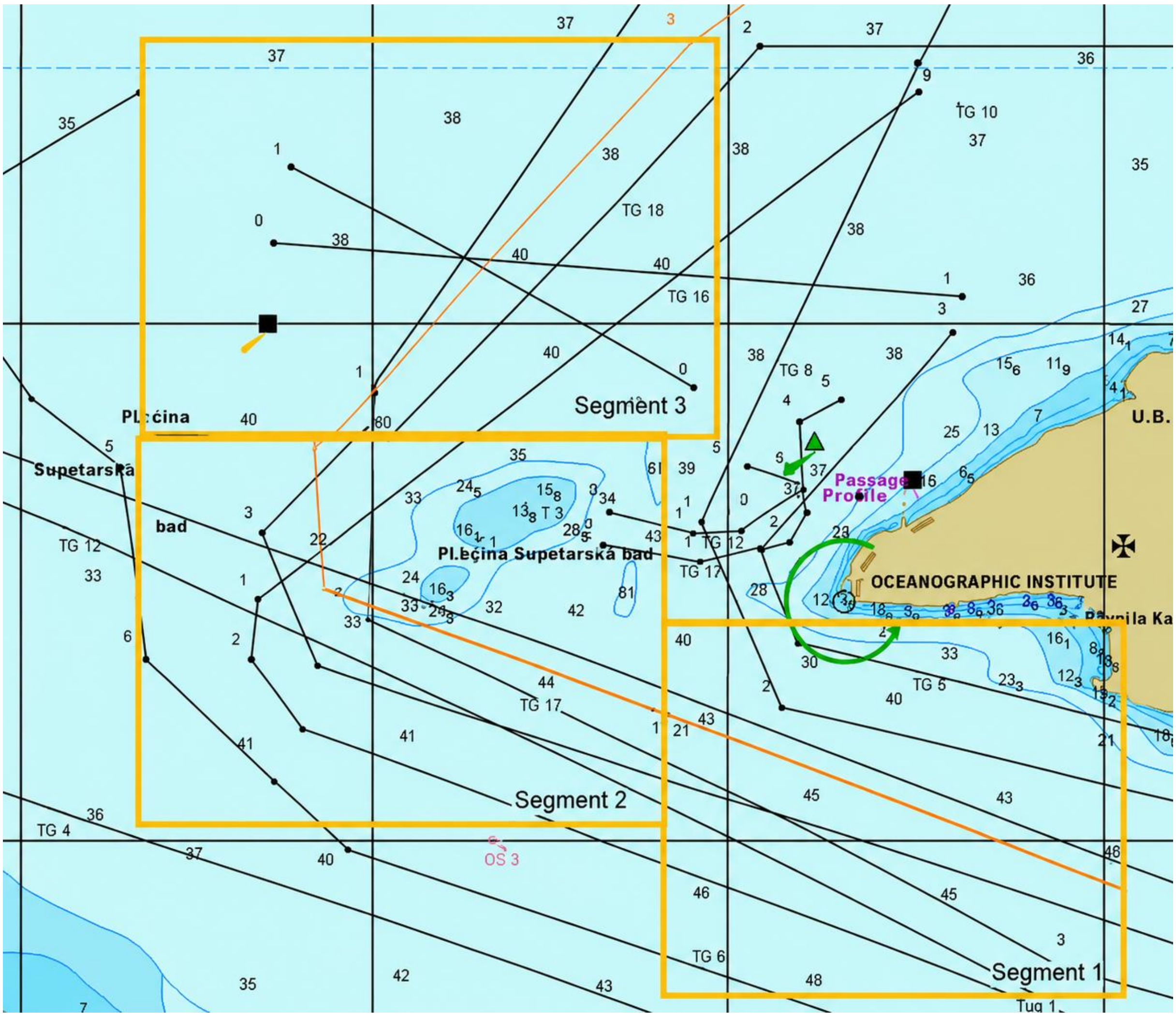

Figure 3.4 Segments of the route and trajectories of the crafts in the simulation exercise area (Simulator)

4. BAYESIAN NETWORKS

Bayesian Networks are named after the British mathematician Thomas Bayes (1702–1761), who introduced a mathematical formulation that is now a cornerstone of probability theory. Although Bayes developed this principle in the eighteenth century, its broader practical relevance emerged only in the twentieth century with the development of artificial intelligence and probabilistic reasoning methods. Bayes's contribution departed from the classical interpretation of probability by introducing a formal mechanism for updating probabilities as new information becomes available, enabling dynamic reasoning under uncertainty.

BN are graphical structures that represent conditional probabilities among multiple variables and support conditional inference based on their interdependencies [80]. They are probabilistic graphical models that capture conditional relationships and dependencies between random variables [81]. Formally, a BN consists of a directed acyclic graph (DAG) composed of nodes and directed edges [82]. Each edge has a defined direction, indicating the influence of one variable on another, while nodes represent random variables and their possible states. Unlike decision trees, nodes in BN do not necessarily correspond to discrete events, but rather to system states or influencing factors relevant to the problem under consideration [83]. Directed edges express direct influence and encode assumptions of conditional independence between variables [84].

BN have been increasingly adopted in maritime research for their ability to model uncertainty and complex interactions among technical, operational, environmental, and human factors. Their use has been demonstrated in several doctoral dissertations on maritime risk assessment, ship safety, human reliability, and autonomous ship operations, confirming the suitability of BN as a robust model for probabilistic modelling in the maritime domain [85], [86]. Building on these studies, this dissertation applies the BN approach to evaluate ship operator performance in simulated MASS operations by integrating behavioral, physiological, and navigational performance indicators.

BN can be used for both qualitative and quantitative analysis of complex systems. The variables in a BN typically take discrete values associated with either prior or conditional probabilities. A prior probability represents the probability distribution of a random variable before any additional evidence is observed, while a conditional probability describes the probability distribution of a random variable given the state of another related variable. Before calculating conditional or joint probabilities, it is necessary to determine whether random variables are dependent, independent, or mutually exclusive. Two random variables are independent if the state of one does not influence the probability distribution of the other. In such cases, their joint probability equals the product of their individual probabilities.

A node without parents is called a root node, while nodes receiving directed edges are called child nodes. As shown in Figure 4.1, the parent–child relationship means that the probability distribution of a child node is conditional on the probability distributions of its parent nodes.

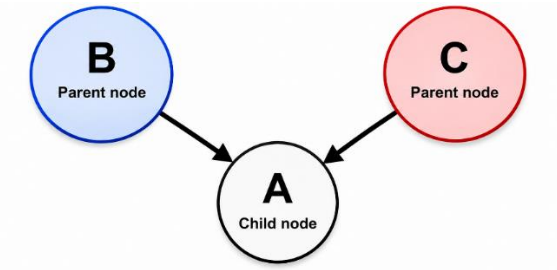

Figure 4.1 Basic structure of a BN illustrating parent–child relationships

The mathematical relationships governing this dependency are given in Equations (4.1) – (4.5), progressing from the basic definition of joint probability to Bayes' theorem and its generalised form.

In these expressions, P(A) denotes the probability of random variable A being in a particular state, P(A,B) represents the joint probability of random variables A and B being in their respective states, and P(A|B) denotes the conditional probability of random variable A given the state of random variable B..

The fundamental mathematical basis of BN lies in the rules of conditional probability. For two random variables A and B, the joint probability is expressed as:

$$P(A\backslash B)P(B)=P(A,B) \tag{4.1}$$

Equation (4.1) expresses the joint probability of variables A and B. It states that the probability of both variables occurring simultaneously equals the conditional probability of A given B multiplied by the probability of B. This relationship forms the basic mathematical representation of dependency between a parent and a child node in a BN.

When an additional conditioning variable C is introduced, the relationship becomes:

$$P(A\backslash B,C)P(B\backslash C)=P(A,B\backslash C) \tag{4.2}$$

Equation (4.2) extends the previous relationship by introducing an additional conditioning variable, C. It represents the joint probability of A and B given that C is observed, illustrating how further evidence can modify probabilistic dependencies within a BN.

According to the rule of conditional probability, it follows that:

$$P(A\backslash B)P(B)=P(B\backslash A)P(A). \tag{4.3}$$

Equation (4.3) follows directly from the definition of joint probability and shows that the joint probability can be expressed in either direction of the dependency. This identity provides the mathematical basis for deriving Bayes' theorem.

From these relationships, Bayes’ theorem is derived as:

$$P(B\backslash A)=\frac{P(A\backslash B)P(B)}{P(A)} \tag{4.4}$$

Equation (4.4) represents Bayes’ theorem. It enables the probability distribution of one random variable to be updated after observing evidence from another random variable, thereby providing the fundamental mechanism for probabilistic inference and belief updating in BNs.

When the additional variable C is observed, Bayes’ theorem can be expressed as:

$$P(B\backslash A,C)=\frac{P(A\backslash B,C)P(B\backslash C)}{P(A\backslash C)} \tag{4.5}$$

Equation (4.5) presents the generalised form of Bayes’ theorem with an additional conditioning variable, C. It shows how prior knowledge and newly observed evidence are combined to update probabilities in more complex dependency structures, which is the principle underlying inference in BN models.

These expressions form the foundation of probabilistic inference and belief updating in BNs.

Constructing a BN involves identifying relevant variables (nodes), specifying their possible states, defining directed dependencies between variables, and assigning conditional probability values. The relationships between parent and child nodes indicate that the probability distribution of a child node is determined by the probabilities of its parents. Each node in a BN is associated with a probability distribution, and the network is fully defined by its graphical structure and numerical parameters.

Key characteristics of BN include the ability to perform inverse reasoning, enabling conclusions to be drawn from observed outcomes to their most probable causes, the capacity to incorporate new observations without reconstructing the model, probabilistic semantics that allow reasoning with incomplete or missing data and an intuitive visual representation of complex interdependencies among variables.

The number of probability parameters required to define a BN depends directly on its structure. As the network becomes more densely connected, the number of required probabilities increases, which can lead to computational and practical challenges. The number of required parameters can be reduced by simplifying the network structure or by adopting suitable parametric probability distributions. Consequently, the selection of variables and their interconnections is a critical step in the modelling process.

A BN consists of nodes representing random variables and directed links connecting them, typically visualized using arrows [81]. Nodes receiving arrows are called child nodes, while nodes from which arrows originate are called parent nodes. For each variable A with parent nodes $B_1, \ldots, B_n$, a Conditional Probability Table $P(A \mid B_1, \ldots, B_n)$ is defined, specifying the probabilities of all possible combinations of states of A and its parents. If a variable has no parents, its probability distribution is specified as a prior distribution P(A). The entries of conditional probability tables constitute the parameters of the Bayesian model, and for each node, all combinations of parent states must be considered.

Within each column of a Conditional Probability Table (CPT), probabilities are normalized so that their sum equals one, with values constrained to the interval [0,1]. The size of a CPT increases exponentially with the number of parent nodes, which can result in computational difficulties for large and complex networks [87]. In highly complex models, conditional probability tables can become excessively large, making it difficult to justify and validate individual probability values.

5. EXPERIMENTAL DATA COLLECTION, PROCESSING AND ANALYSIS

Data were collected from multiple synchronized sources, including pre-simulation questionnaires, simulator log recordings, eye-tracking measurements, and physiological monitoring. The data collection procedure ensured temporal alignment and methodological consistency across all measurement modalities.

5.1. Data Collection

This section presents the structured data collection framework used during the experimental phase of the research. The following subsections describe each component of the data collection process in detail.

5.1.1. Pre-simulation Questionnaire

To gain insight into participants' characteristics, previous experience, knowledge, attitudes, and perceptions of autonomous ships, a questionnaire is conducted before the start of simulation exercise. The purpose of the questionnaire is to identify differences and enable classification among participants according to education level, professional experience and familiarity with autonomous technologies.

The questionnaire is structured to cover several thematic areas related to human performance and attitudes towards autonomous maritime systems. The questionnaire comprising of 25 questions divided into four sections (Appendix A). The introductory section composed of questions concern the identification of the participant group and prior experience with maritime simulators, allowing differentiation according to formal education level and practical exposure to maritime operations. Subsequent questions examine participants' familiarity with the concept of autonomous ships, their trust in technology, and their perception of the reliability of unmanned systems, as well as their attitudes towards the necessity of human supervision from shore. In this way, the questionnaire explores dimensions such as technological trust and understanding of the limits of autonomy in relation to human control, which is consistent with approaches used in previous studies [49] on the perception of autonomy and the influence of experience within the maritime industry.

Second section of the questionnaire focuses on assessing the perception of challenges that autonomous ships may encounter. Participants evaluate the importance of various factors such as system failures, cyberattacks, human factors, and communication problems using a five point Likert scale. This section examines risk perception and subjective prioritization in the context of autonomous operations, which is important for the subsequent modelling of variables related to risk perception within the BN. Third section address participants' self-assessment of performance and readiness to participate in the exercise, their ability to make decisions under pressure, and their expected level of success. These dimensions are associated with constructs such as self-confidence, stress resilience, and perceived effectiveness, which are key psychological predictors of performance in complex operational environments.

As the simulation involves the use of additional equipment for monitoring attention and physiological parameters, the questionnaire also includes questions about prior experience with such technology and possible health limitations, such as cardiovascular conditions, ongoing medical treatments, or experiences with anxiety. This section serves a dual purpose: ensuring participants' safety and identifying potential physiological factors that may influence simulation outcomes. The final part examines participants' sense of safety during the exercise and their preferences regarding the presence of an instructor or the possibility of contacting the supervisor, which allows the adjustment of supervision levels and contributes to the ethical dimension of the research.

For the group of experienced seafarers, fourth section includes questions related to sea-going experience. The purpose of these questions is to determine previous system failure experience and positions held onboard. These data make it possible to differentiate participants according to their level of professional responsibility and experience in decision-making during real operational scenarios.

The questionnaire has multiple methodological functions. It serves as an initial tool to determine participants' knowledge, attitudes, and perceptions related to autonomous systems. The results represent input variables for the BN modelling process, particularly in segments referring to the preconditions for challenges in autonomous ship operations. The questionnaire combines dichotomous (yes/no), ordinal (five point Likert scale), and open ended questions, allowing the integration of quantitative and qualitative data and enabling a multidimensional analysis of participants.

The collected data enable detailed classification of participants, identification of individual differences, and interpretation of behavior during the exercise. Incorporating the questionnaire results into the BN allows for the assessment of how initial attitudes, experience, and psychological factors influence decision-making and performance in the context of autonomous navigation.

5.1.2. Experimental Procedure and Measurement Protocol

After completing the questionnaire, participants underwent baseline physiological measurements. Blood pressure, heart rate, and body temperature were measured at rest to establish individual reference values for subsequent comparison with intra-session and post-session physiological responses.

Before commencing the simulation exercise, the eye-tracking glasses were individually calibrated for each participant by the researcher using the standard calibration procedure provided by the manufacturer. The calibration procedure ensured accurate alignment between the eye cameras and the participant's visual axis, optimizing gaze-mapping precision and minimizing measurement error. Each participant completed a separate calibration sequence to maintain consistent recording quality and data reliability across sessions (Figure 5.1)

Figure 5.1 Calibration and application of the eye-tracking system during the simulation exercise

At the predefined start of the simulation exercise, all recording systems were activated simultaneously. Continuous heart rate monitoring began at the onset of the exercise, while eye-tracking recording started concurrently with the simulator session. This synchronized activation ensured temporal consistency across data sources and enabled subsequent alignment of behavioral, visual, and physiological datasets using identical timestamps and comparable time intervals.

During the simulation, predefined system degradations affecting navigational subsystems such as Global Positioning System (GPS), radar, and ECDIS were introduced in accordance with the structured scenario design described in the Chapter 3.

In addition to automated simulator logging, key scenario events were manually recorded by the researcher in real time. The exact moment of system malfunction introduction, as well as the time at which participants verbally or behaviourally indicated recognition of the malfunction, was documented. These manually recorded timestamps served as event markers and were subsequently used to align simulator logs, gaze recordings, and physiological time series data within a unified session timeline.

Immediately after completion of the simulation exercise, post-session physiological measurements were obtained using the same protocol applied prior to the exercise. Blood pressure, heart rate, and body temperature were measured again to enable pre–post comparison of physiological indicators and assessment of potential stress-related changes induced by the simulation.

This structured recording protocol enabled the a temporally consistent and analytically robust dataset, forming the foundation for the preprocessing and analytical procedures described in the following section.

5.2. Data Analysis Procedures

After data acquisition, the recorded datasets were systematically processed to enable structured analytical evaluation. The data analysis phase comprised four main stages:

- review of the simulator session recording for operational performance assessment (event markers, manoeuvre execution, failure detection times),

- structured processing of eye-tracking recordings within the Pupil Player software environment (AOIs, gaze trajectories, fixation maps),
- computational processing of gaze data using a custom-developed Python analytical programme (derived gaze metrics and entropy-based indicators),
- processing and organization of physiological data from the wearable monitoring device (heart rate time series and baseline comparisons).

5.2.1. Review of Simulator Recording and Performance Identification

The complete simulator session recording stored on the computer was reviewed after each participant completed the exercise. This post-session analysis focused on identifying observable operational indicators, such as navigational errors, near miss situations, deviations from the planned route, and inconsistencies in speed maintenance (Figure 5.3). The criteria used to identify these operational performance indicators were established through expert judgement by the members of the NAUTiCA project team prior to the analysis. These predefined criteria ensured a consistent and reproducible evaluation of participant performance across all simulation sessions. The detailed definitions, threshold values, and classification criteria for these indicators are presented in Chapter 6, where they are formally defined as variables within the BN model.

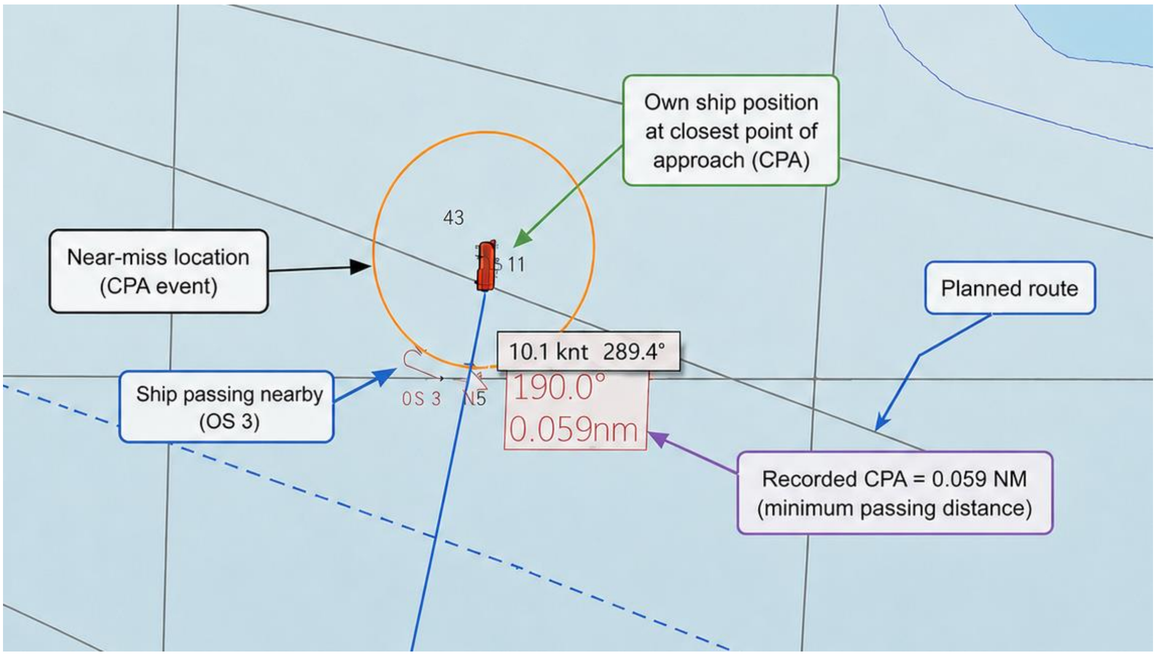

Figure 5.2 Near miss event identification during video analysis (simulator)

Particular attention was given to instances of system malfunction (GPS, radar, and ECDIS disturbances) and the participants' ability to detect and respond to these events. The objective of this stage was the systematic identification of measurable operational events, including the detection time of navigation system failures, route deviations, speed deviations from prescribed operational limits, near miss events based on the adopted CPA threshold, and requests for external assistance during the simulation. These events were subsequently synchronized with eye-tracking recordings and physiological measurements to support the integrated analysis of participant performance. Key events and reaction points were cross referenced with manually recorded timestamps established during the experimental session, ensuring consistency between observed behavior and recorded data streams.

5.2.2. Eye-Tracking Data Processing in Pupil Player

In parallel with the simulator recording review, eye-tracking recordings were processed using the Pupil Player software environment (version 3.5.1). The processing procedure followed established methodological principles for fixation and eye movement detection in eye-tracking research [88]. The scene video recorded by the outward facing camera of the eye-tracking glasses was analysed with gaze overlay enabled to visually confirm fixation accuracy and calibration stability. The initial baseline segment of each recording, characterized by simple navigational tasks without induced system failures, was used as a common reference for visual verification of gaze alignment and calibration consistency. (Figure 5.3).

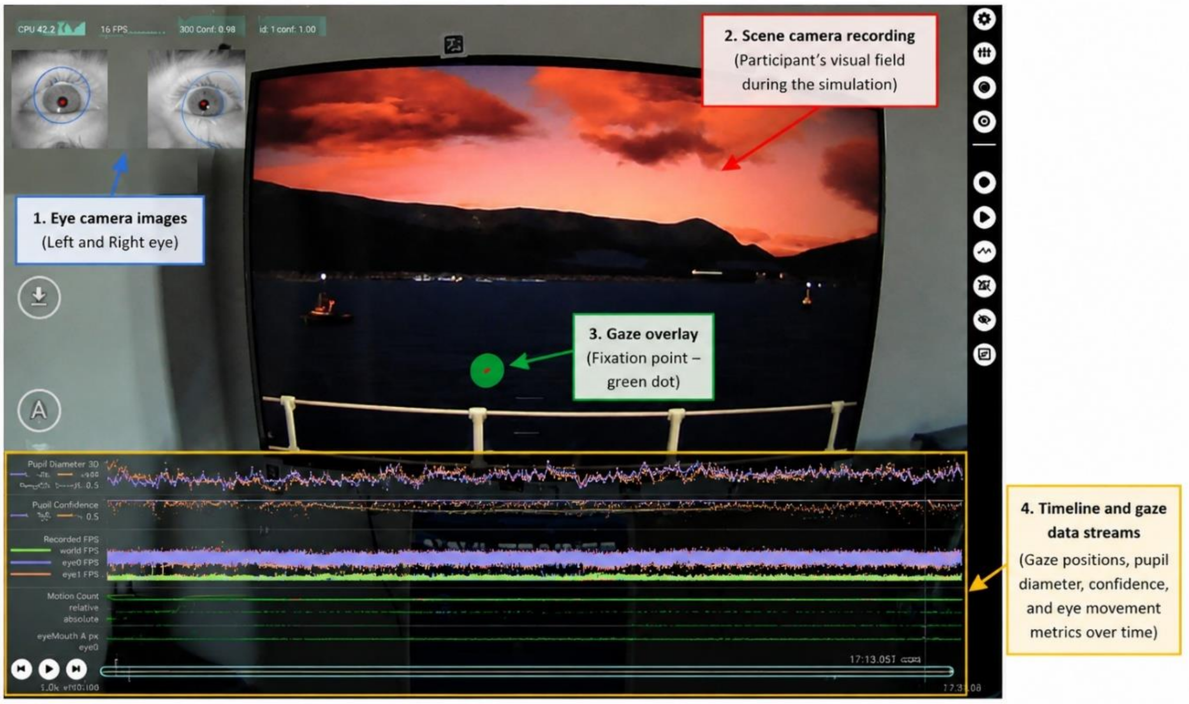

Figure 5.3 Pupil Player interface showing the scene recording, eye camera images, gaze overlay, and synchronized eye-tracking data used for visual verification and subsequent analysis

To enable structured analysis of visual attention distribution, predefined AOIs were manually inserted into each recording. These AOIs corresponded to the principal operational interfaces available to participants during the simulation exercise. The AOI configuration was applied consistently across sessions to ensure analytical comparability.

The selection of four AOIs was based on expert judgement by members of the NAUTiCA project team and reflected the principal operational interfaces required for safe navigation and decision-making within the simulation scenario. AOI boundaries were manually defined according to the physical dimensions and screen limits of each display within the simulator environment and were applied identically to all recordings, ensuring consistency and analytical comparability across all participants.

Four primary AOIs were defined:

- **Visualization Display (55-inch screen) -** the visualization AOI comprised the large 55-inch external view display, representing the out-of-window perspective from the ship's bridge. This display provided a real time visual representation of the navigational environment, including surrounding vessels, coastal features, and environmental conditions.

The visualization display did not include direct control functions. It served as a situational awareness reference interface, allowing participants to visually assess ship positioning, the relative motion of nearby traffic, and potential collision risk through visual observation.

- **Radar (24-inch screen) -** the radar AOI included the entire radar display and its associated control panel. Participants used this interface to monitor surrounding traffic and support collision avoidance, while eye-tracking analysis quantified visual attention directed to radar-based information during the simulation.
- **ECDIS (24-inch screen) -** the ECDIS AOI represented the primary electronic chart interface used for route monitoring and navigational decision-making during the simulation, while eye-tracking analysis quantified the visual attention allocated to chart-based information and route management tasks.
- **Steering System (24-inch screen) -** the Steering AOI included all elements related to direct ship control and maneuvering within the simulation environment. This interface was primarily used to control the vessel's heading and speed, while eye-tracking analysis quantified the visual attention directed to steering-related information during the simulation.

By segmenting the visual field into these four structured regions, it was possible to determine when and for how long participants directed their gaze towards specific operational interfaces during the simulation. Based on the AOI analysis, quantitative indicators of visual attention and cognitive workload were derived and then used for further processing of eye-tracking data and modelling operator performance within the BN model, as described in Chapter 6. This structured AOI configuration provided the analytical basis for subsequent quantitative analysis (Figure 5.4).

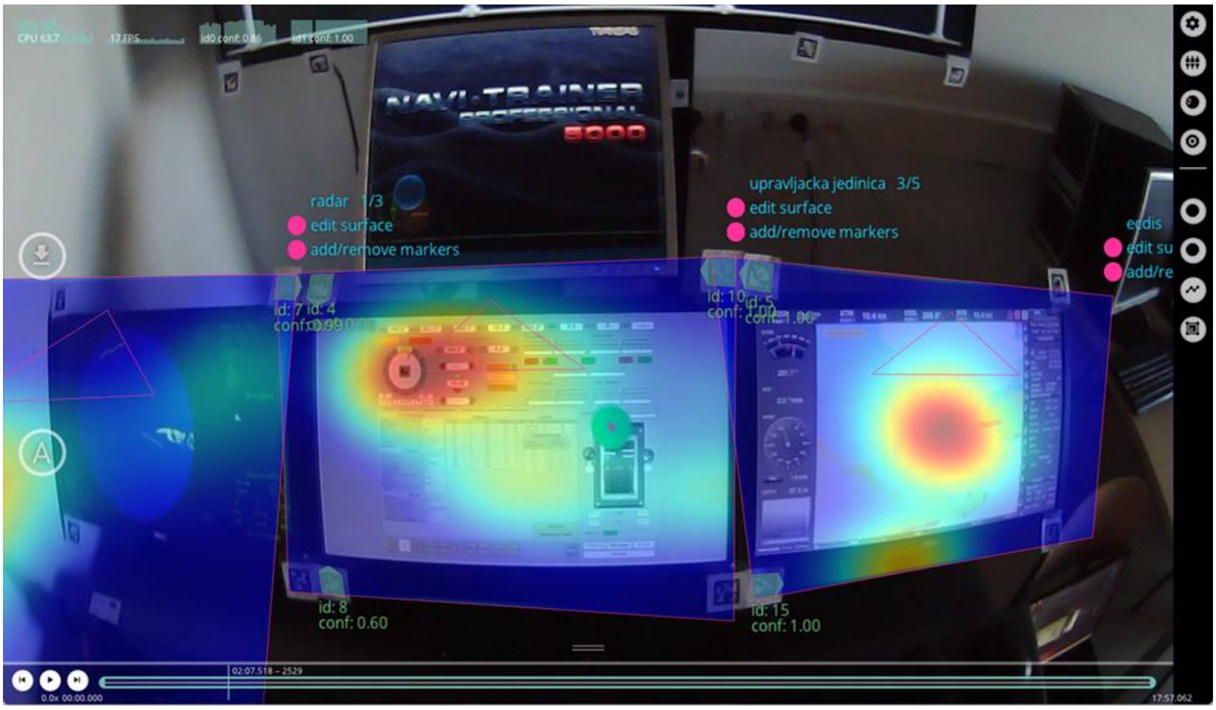

Figure 5.4 AOIs and heatmap of the participant's attention (Pupil player)

The processed eye-tracking recordings were then exported together with gaze coordinates, fixation events, timestamps, pupil diameter measurements, and AOI-related gaze information for further computational analysis in Python. These exported data served as the basis for calculating the quantitative eye-tracking metrics described in the following section.

5.2.3. Eye-Tracking Metrics Extraction and Analysis

After preprocessing the eye-tracking data, a second stage of data processing was carried out to extract metrics describing visual attention, gaze behavior, and cognitive load of participants during the simulation exercise. This stage marks the transition from structured gaze data obtained from raw Pupil Core recordings to analytically meaningful indicators of operator performance and stress response.

The processing pipeline was implemented in Python and applied to structured datasets generated for each participant. The initial output of the preprocessing stage consisted of multiple Comma-Separated Values (CSV) files containing gaze coordinates, fixation information, surface mappings, and pupil-related data. These data were reorganized into unified datasets to enable consistent analysis across participants. Key variables included timestamp information, normalized gaze coordinates (x_norm and y_norm), fixation duration, gaze movement parameters, and pupil diameter measurements. During preprocessing, gaze samples were filtered using the on surf parameter, so that only valid gaze points successfully mapped to the predefined AOIs (on surf = True) were retained for further analysis, while samples outside the defined AOIs or with unsuccessful surface mapping (on surf = False) were excluded. The same filtering procedure was systematically applied to all 65 participant recordings to ensure methodological consistency and comparability across the complete dataset. This filtering ensured that subsequent eye-tracking metrics were calculated exclusively from valid observations, thereby improving the reliability and consistency of the analysis. This structured processing of gaze data aligns with established eye-tracking methodologies, which require the transformation of raw gaze signals into interpretable behavioral metrics [89].

To enable meaningful interpretation of gaze behavior in relation to the simulated incident, the entire duration of the simulation was divided into three time intervals relative to the onset of the crisis situation: baseline, acute, and sustained. The temporal segmentation used in this dissertation was derived directly from the design of the simulation scenario described in the simulation exercise chapter. As the exercise was intentionally structured according to the principle of progressive complexity, with increasingly demanding navigational situations and deliberately introduced system disturbances, the selected observation windows correspond to distinct operational phases of the scenario rather than arbitrary time intervals. This approach enables meaningful comparison of participants' performance under different levels of cognitive and navigational workload.

The distribution of gaze across different surfaces offers fundamental insight into how participants allocate visual attention among navigational instruments. Analysing gaze frequency and time spent on each surface enables identification of dominant information sources and potential neglect of critical interfaces. Both the proportion of attention dedicated to each surface and changes in this distribution during the simulation are analysed to assess visual attention strategies under progressively changing operational conditions. Figure 5.5 presents a representative example of gaze distribution for a single participant. The dashed vertical line labelled "start of the incident" marks the beginning of the second phase of the simulation exercise, when the predefined incident scenario was introduced.

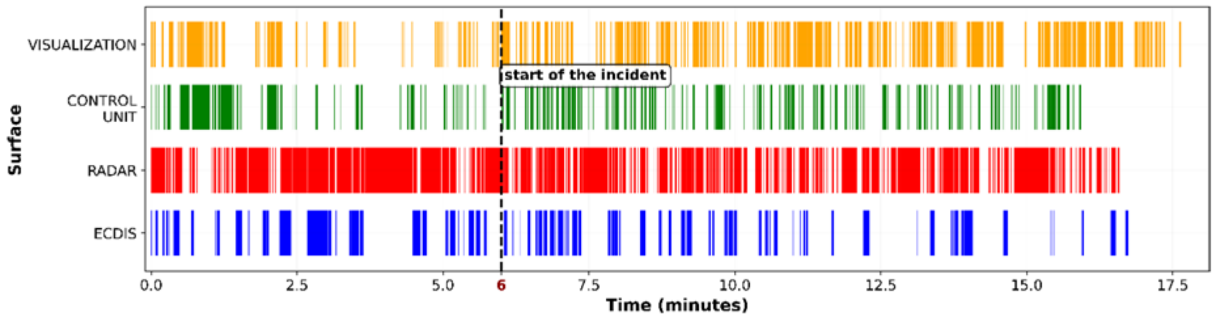

Figure 5.5 Example of gaze distribution across AOIs for a representative participant

After preprocessing, quantitative eye-tracking indicators were extracted from the processed gaze data to characterize participants' visual attention and cognitive responses during the simulation exercise. The calculated metrics included the GDI, Fixation-to-Saccade Ratio (FSR), PEI, GMS, and SGPS. All indicators were calculated separately for each participant and each predefined simulation phase (baseline, acute, and sustained), providing standardised input variables for subsequent statistical analysis and BN modelling.

The spatial distribution of gaze was visualized using heatmaps, centroids, and standard deviation ellipses for each of the 65 participants and for each predefined simulation phase. Figure 5.6 presents one representative example of the generated visualization for a single participant, whereas the same analytical procedure was systematically applied to all 65 eye-tracking recordings included in the dissertation. These graphical representations provide a visual illustration of the spatial distribution of gaze fixations and facilitate interpretation of the GDI, where more concentrated fixation patterns generally correspond to lower GDI values, while broader spatial distributions indicate higher gaze dispersion.

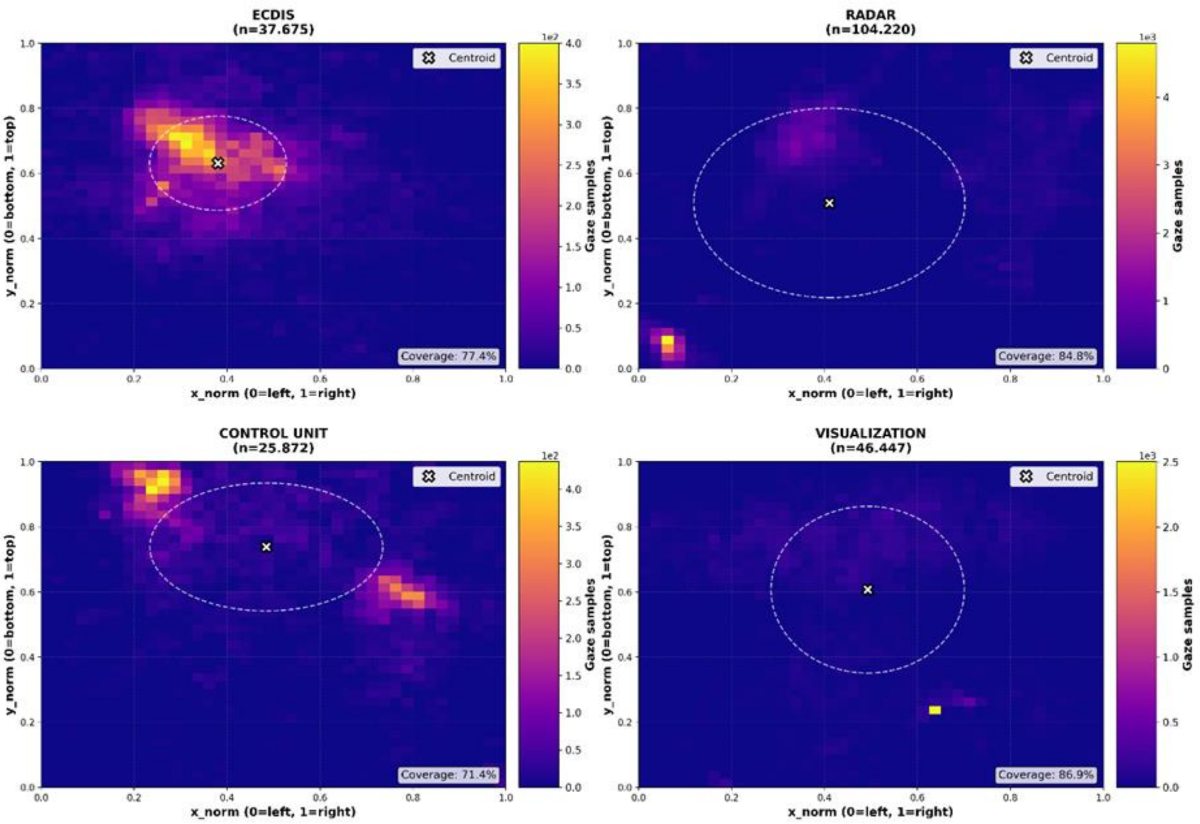

Figure 5.6 Gaze heatmap with centroid and standard deviation ellipse

The extracted eye-tracking metrics formed the basis for assessing visual attention patterns and cognitive workload throughout the simulation exercise, and for deriving the variables used in developing the BN model. The mathematical definitions, computational procedures, discretization methods, and interpretation of GDI, FSR, PEI, GMS, and SGPS are presented in detail in Chapter 6.3, where these indicators are introduced as variables of the BN model.

5.2.4. Physiological Data Processing and Organization

Physiological data were retrieved from Huawei Health cloud storage after each simulation session. Heart rate measurements, recorded at 5-second intervals, were exported in CSV format and organized according to the predefined simulation phases (baseline, acute, and sustained). This procedure was applied consistently to all 65 participants, enabling direct comparison of physiological responses under different levels of navigational complexity. Figure 5.7 presents a representative example of the heart rate profile for a single participant, while the same extraction and processing procedure was applied to the complete dataset. The extracted heart rate measurements formed the basis for deriving the physiological variables incorporated into the BN model. Their mathematical definition, discretization, and interpretation are presented in detail in Section 6.2.

Figure 5.7 Representative heart rate profile recorded during the simulation exercise using the Huawei Health application

6. DEFINITION AND DESCRIPTION OF VARIABLES USED IN THE BN MODEL

This chapter presents the variables used in the BN model and provides their definitions within the context of the simulation-based experiment. For clarity, the variables are organized into four categories, reflecting different aspects of the experiment and data collection:

1. operator profile variables,
2. physiological response variables,
3. visual attention variables and
4. navigational performance variables.

Through their interactions within the BN, these variables contribute to estimating the final output variable, Operational Effectiveness, which represents a composite measure of decision-making quality and safety-related performance indicators. Table 6-1 provides an overview of all variables included in the BN model, along with their classification, role, and data source.

Table 6-1. Overview of variables used in the BN model

VARIABLE NAME	CATEGORY	ROLE IN BN	DATA SOURCE
Age	Operator Profile	Parent	Questionnaire
Maritime Education Level	Operator Profile	Parent	Questionnaire
Operational Rank	Operator Profile	Parent	Questionnaire
Operational Self-Efficacy	Operator Profile	Parent	Questionnaire
Operator Background	Operator Profile	Intermediate (child)	Derived
MASS Familiarity	Operator Profile	Parent	Questionnaire
Sea-Going Experience	Operator Profile	Parent	Questionnaire
Operator Readiness	Operator Profile	Intermediate (child)	Derived

Equipment Comfort	Operator Profile	Parent	Questionnaire
Temperature Change	Physiological Response	Parent	Physiological measurement
Blood Pressure Reactivity	Physiological Response	Parent	Physiological measurement
Heart Rate Change	Physiological Response	Parent	Physiological measurement
Physiological Response	Physiological Response	Intermediate (child)	Derived
Physiological Stress Level	Physiological Response	Intermediate (child)	Derived
Cognitive Overload	Physiological and Cognitive Response	Intermediate (child)	Derived
Gaze Dispersion Index (GDI)	Visual Attention	Parent	Eye-tracking
Pupil Entropy Index (PEI)	Visual Attention	Parent	Eye-tracking
Gaze Movement Score (GMS)	Visual Attention	Parent	Eye-tracking
Stress-induced Gaze Pattern Score (SGPS)	Visual Attention	Intermediate (child)	Derived
Visual Attention Quality	Visual Attention	Intermediate (child)	Derived
Radar Error	Navigational Performance	Parent	Simulation
GPS Error	Navigational Performance	Parent	Simulation
Response to Navigation System Failures	Navigational Performance	Intermediate (child)	Derived

Navigation Complexity	Navigational Performance	Parent	Simulation
Navigation Control	Navigational Performance	Intermediate (child)	Derived
Voyage Execution	Navigational Performance	Intermediate (child)	Derived
Route Adherence	Navigational Performance	Parent	Simulation
Arrival Time to Assigned Position	Navigational Performance	Parent	Simulation
Near Miss	Navigational Performance	Parent	Simulation
Exercise Performance	Navigational Performance	Intermediate (child)	Derived
Operational Effectiveness	Output	Final Output	Derived

For variables that could not be discretized directly from empirical measurements, state definitions and discretization thresholds were determined through a structured expert elicitation process conducted within the NAUTiCA project. Expert judgement was not limited to defining discretization thresholds but also informed several stages of BN development, including the conceptual definition of latent variables, specification of variable states, construction of conditional probability tables, and verification of the logical consistency of relationships between variables. The expert panel comprised faculty members and researchers with expertise in maritime navigation, ship command, simulator-based training, human factors, and physiological assessment. Final state definitions and classification criteria were established through expert consensus to ensure methodological consistency, practical interpretability, and coherence throughout the BN model.

6.1. Operator Profile Variables

Operator profile variables describe the background characteristics of participants in the simulation exercise. These variables provide contextual information about the participants and enable analysis of potential performance differences related to experience, education level, and familiarity with maritime technologies.

6.1.1. Age

The Age variable represents the chronological age of participants at the time of the simulation-based exercise and is classified as an operator profile variable. As a fundamental demographic characteristic, age is frequently considered in human factors research because of its potential influence on cognitive processing, decision-making, situational awareness, and adaptability to complex operational environments.

Age-related differences in cognitive functioning, particularly in processing speed and the ability to respond to time-critical situations, have been extensively documented in cognitive research. According to Salthouse, ageing is associated with gradual changes in information processing efficiency, which may affect performance in tasks requiring rapid perception, interpretation, and response under pressure [90]. In maritime operations, where decision-making often occurs in dynamic and uncertain environments, such differences may influence operators' ability to manage complex navigational situations.

Studies in maritime human factors and accident analysis emphasise the importance of operator characteristics in evaluating safety and performance outcomes. Analyses of collisions and groundings indicate that human factors, including individual differences among seafarers, play a significant role in accident causation and Operational Effectiveness [91]. Although age alone does not determine performance, it is an important contextual variable that may be associated with experience, risk perception, and behavioral patterns during navigation tasks.

This categorization reflects the heterogeneous structure of the sample, which comprises maritime high school students, university students, and experienced seafarers. The selected thresholds allow a meaningful distinction between early-career individuals, participants in the transitional phase of professional development, and more experienced operators.

Based on data collected from 65 participants, the Age variable was discretized into three categories, young (≤25 years), middle (26–35 years), and older (>35 years) resulting in the following distribution:

- young: 33 participants (51%),
- middle: 16 participants (25%),
- older: 16 participants (24%).

The slight adjustment to the probability of the "older" category was made to account for rounding and to ensure that the displayed probability distribution sums to 100% within the BN.

Although Age does not have incoming dependencies, it may influence several intermediate variables within the model, particularly those related to operator characteristics, cognitive workload, and decision-making performance. Thus, Age contributes indirectly to the overall assessment of operator effectiveness by providing additional context about participants' backgrounds and potential behavioral tendencies during the simulation exercise.

In the BN model implemented in GeNIe, the Age variable is defined as an independent (root) variable, meaning it has no parent nodes and is not influenced by other variables in the model. The probability distribution is derived directly from empirical data collected during the simulation exercise and is entered into the model as a prior probability distribution.

6.1.2. Maritime Education Level

The variable Maritime Education Level represents the highest level of formal maritime education attained by each participant. It reflects each participant's educational background in the maritime field and includes both secondary maritime education and higher academic qualifications. Each participant was assigned to a single category corresponding to their most advanced level of education, thus avoiding overlap between educational stages.

To ensure a consistent and analytically meaningful representation, the variable was defined by prioritizing the highest completed or ongoing level of maritime education. Participants with higher education qualifications were classified according to their academic level, while those without university-level education were categorized according to completed maritime secondary education. This hierarchical approach is methodologically appropriate for BN modelling, as it provides mutually exclusive states and reduces ambiguity in interpreting operator characteristics.

The Maritime Education Level variable, based on a sample of 65 participants, shows:

- high school: 39 participants (60%),
- undergraduate: 4 participants (6%),
- graduate: 22 participants (34%).

This categorization preserves the distinction between basic and advanced formal education while keeping the number of states manageable for probabilistic modelling. This distribution reflects the heterogeneous composition of the sample, which includes participants at various stages of maritime education and professional development.

Formal maritime education is widely recognized as a key factor in developing navigational performance, situational awareness, and decision-making capabilities, particularly in increasingly complex and technology-driven maritime environments. Recent studies emphasise that the transition to MASS requires both traditional seafaring knowledge and advanced theoretical understanding, as well as system-oriented competencies developed through formal education [92].

Within the BN structure, Maritime Education Level is treated as an external root variable, meaning that it has no parent nodes and that its values are determined directly from the empirical dataset.

6.1.3. Operational Rank

The variable Operational Rank represents the highest seaman's rank of each participant. It indicates the participant's position within the shipboard hierarchy and, consequently, their level of involvement in navigational tasks, decision-making, and safety management. In maritime operations, rank is not merely a formal title but a direct indicator of authority, accountability, and exposure to complex operational situations.

Shipboard organization follows a well-defined hierarchy, with responsibilities increasing at higher ranks. Officers are involved in navigation, watchkeeping, and operational decision-making, while the captain holds full command responsibility, including final authority over navigation, safety, and crisis response. In contrast, individuals without Operational Rank, such as students, typically lack direct experience in real maritime environments and have not faced decision-making under operational constraints. Therefore, the Operational Rank variable provides a meaningful distinction between participants with different levels of practical experience and responsibility.

From the perspective of human performance, differences in rank can be interpreted using established models of cognitive behavior. According to Rasmussen’s framework, human performance is categorized into skill-based, rule-based, and knowledge-based levels, with the dominant mode depending on the individual’s expertise and familiarity with the task environment [93]. Highly experienced operators typically rely on skill-based and rule-based behavior, characterized by rapid pattern recognition and efficient execution of routine actions. In contrast, less experienced individuals are more likely to operate at the knowledge-based level, where decision-making is slower, more analytical, and involves greater cognitive demand. Thus, Operational Rank serves as an indirect indicator of the dominant cognitive strategies used during navigation and problem-solving.

In addition to cognitive processing, Operational Rank is closely related to situational awareness, which is widely recognized as a critical component of safe navigation. Situational awareness involves perceiving elements in maritime environment, understanding their meaning, and projecting their future status, all of which are essential for effective decision-making in dynamic maritime environments [94]. Higher-ranking and more experienced personnel typically demonstrate more developed situational awareness, enabling them to anticipate potential hazards and respond proactively to changing conditions. Lower levels of experience are often associated with a reduced ability to integrate multiple sources of information and maintain a comprehensive understanding of the operational context.

The Operational Rank Distribution variable, based on a sample of 65 participants, shows:

- student category: 34 participants (52%),
- officer: 23 participants (35%),
- captain: 8 participants (13%).

This distribution includes both inexperienced and highly experienced individuals, enabling the analysis to capture a broad range of behavioral patterns and perceptual differences related to maritime expertise.

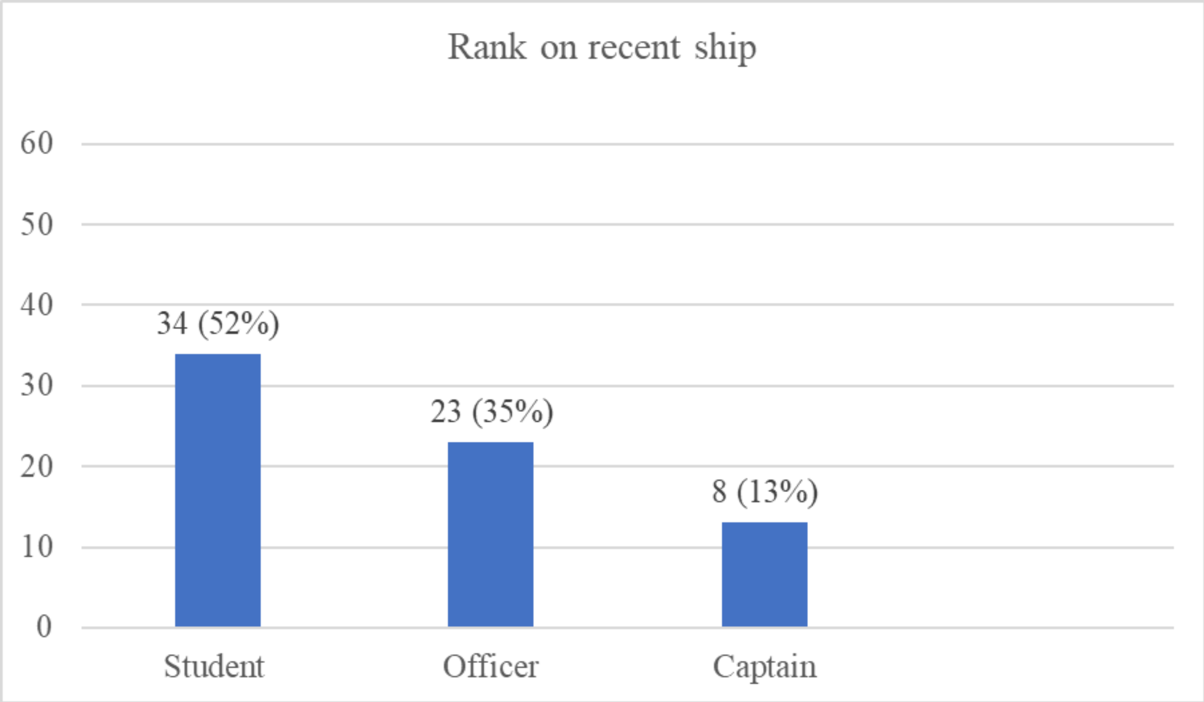

Figure 6.1 Distribution of participant rank on the most recent ship (N = 65)

By incorporating Operational Rank into the model, it is possible to account for the effect of professional hierarchy on cognitive and perceptual aspects of autonomous ship operations. This enables a more nuanced understanding of how different levels of experience and responsibility influence decision-making processes and responses to complex navigational scenarios, ultimately enhancing the interpretability and robustness of the BN model.

Within the BN, Operational Rank is defined as an independent variable with no parent nodes. Its values are taken directly from the survey data collected before the simulation exercise and represent a fixed characteristic of each participant.

6.1.4. Operational Self-Efficacy

The variable Operational Self-Efficacy reflects participants' self-assessment of their ability to cope successfully with challenges during the simulator exercise. Before the simulation began, participants were asked to rate their confidence in managing the operational demands and potential difficulties that might arise during the exercise.

Self-efficacy refers to an individual's belief in their capability to organize and execute the actions required to manage prospective situations. These beliefs influence how people approach difficult tasks, the effort they invest, and their persistence when facing obstacles [95].

Individuals with higher perceived self-efficacy tend to approach complex situations with greater confidence and are more likely to engage actively in problem-solving and decision-making.

In maritime simulator exercise, perceived self-efficacy may influence how participants respond to unexpected events, system failures, and increasing task complexity. Participants who believe they can handle operational challenges may demonstrate more proactive behavior and maintain better situational awareness, whereas lower confidence may result in hesitation or increased reliance on external assistance.

The variable was derived from the survey question: "Do you believe you will successfully cope with the challenges during the exercise?" Participants rated their response on a five point Likert scale, where 1 indicated no confidence at all and five indicated complete confidence.

For BN modelling, the variable Operational Self-Efficacy was discretized into three categories:

- low confidence – responses 1–2,
- moderate confidence – response 3,
- high confidence – responses 4–5.

This discretization represents different levels of perceived operational capability and enables the variable to be incorporated into the BN model as an indicator of participants' initial confidence before the simulation exercise.

Based on responses from 65 participants, the distribution of Operational Self-Efficacy levels on shows that most participants reported a relatively high level of confidence before the simulation exercise. Specifically, 38 participants (58%) selected level 4 and 8 participants (12%) selected level 5 on the Likert scale, resulting in a combined 70% of participants classified as having high confidence. A smaller proportion of participants, 18 (28%), reported a moderate level of confidence (level 3), while only one participant (2%) indicated low confidence (level 1), and no responses were recorded for level 2. This distribution suggests that the sample was generally characterized by a high level of perceived readiness to engage in the simulator exercise.

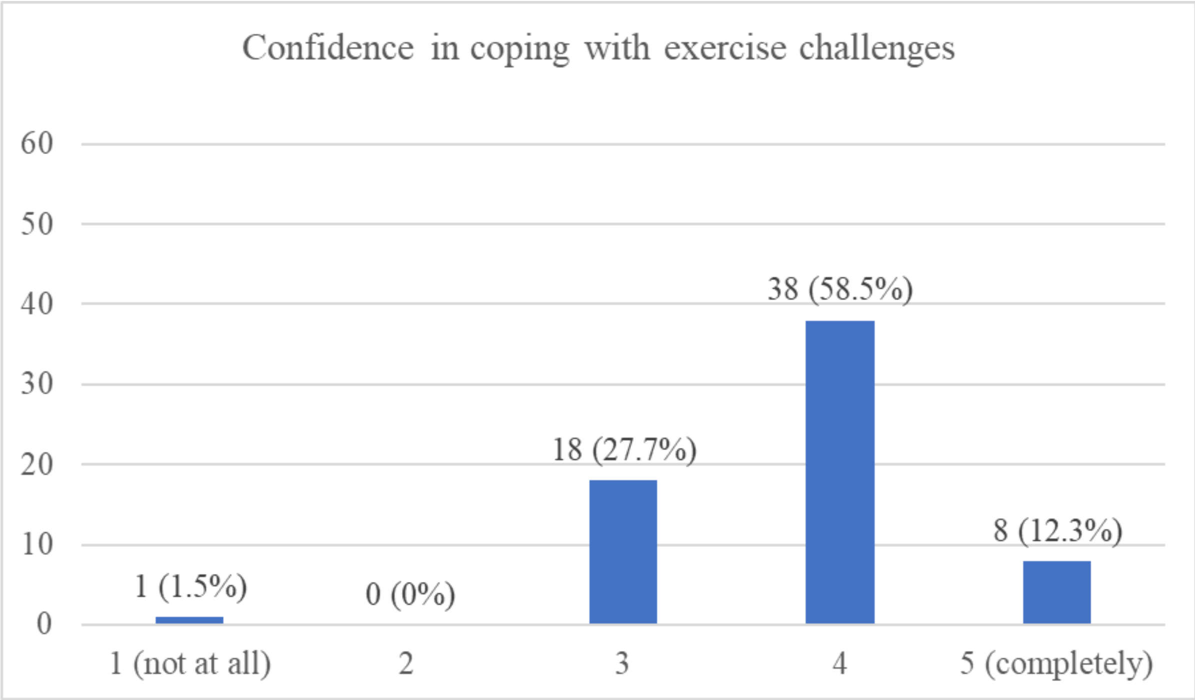

Figure 6.2 Distribution of confidence in coping with exercise challenges

The Operational Self-Efficacy variable, based on a sample of 65 participants, shows:

- low confidence: 1 participant (2%),
- moderate confidence:18 participants (28%),
- high confidence: 46 participants (70%).

These values defined the prior probability distribution for the Operational Self-Efficacy variable in the BN model.

Within the BN, this variable is defined as an independent (root) variable with no parent nodes in the model.

6.1.5. Operator Background

The variable Operator Background represents a composite measure of the participant's baseline professional profile before engagement in the simulator-based experiment. It integrates key demographic, educational, professional, and perceptual characteristics to capture the overall level of preparedness derived from accumulated knowledge, experience, and self-perceived capability. Unlike performance-based variables, which reflect behavior during the simulation, this variable provides a structured representation of the operator’s pre-existing conditions that may influence decision-making processes in autonomous ship operations.

In the BN model, Operator Background is defined as an intermediate (derived) node with four parent variables:

- Age,
- Maritime Education Level,
- Operational Rank,
- Operational Self-Efficacy.

These variables were selected to capture complementary dimensions of the operator profile. Age is included as an indicator of general maturity and potential exposure to maritime environments, although its influence is considered secondary. Maritime Education Level reflects formal knowledge and theoretical understanding of maritime systems. Operational Rank represents the participant’s professional role and responsibility level on board, and is treated as the most informative indicator of real world performance. Operational Self-Efficacy captures the participant’s subjective perception of their ability to perform navigational tasks under demanding conditions.

The variable is defined using three mutually exclusive states: limited, intermediate, and advanced. This classification allows a clear distinction between lower, moderate, and higher levels of professional background, while maintaining compatibility with probabilistic modelling requirements. The discretization does not imply strict deterministic categorization but rather reflects dominant tendencies expressed through conditional probabilities. The Operator Background variable, derived from the combination of three parent variables, was distributed as follows:

- limited (41%),
- intermediate (35%),
- advanced (24%).

The CPT for Operator Background was constructed using an expert-based weighting approach .A hierarchical influence structure was defined, with Operational Rank assigned the strongest influence (approximately 50%), followed by Maritime Education Level (30%), while Operational Self-Efficacy and Age were treated as secondary modifiers (10% each). This structure reflects the assumption that professional responsibility and role within ship operations provide the most reliable indicator of performance, while formal education enhances knowledge-based capacity. Self-efficacy is included as a psychological factor influencing confidence and decision-making tendencies, but not as a primary determinant of actual performance. Age contributes marginally as a proxy for general experience and cognitive maturity.

To ensure accuracy, the CPT was validated using consistency checks and scenario-based evaluation. These tests confirmed that increases in rank, education, or combined favourable conditions systematically shift the output distribution from limited to advanced, while unfavourable combinations have the opposite effect. The resulting structure preserves interpretability, maintains probabilistic realism, and avoids deterministic behavior.

The modelling approach for this variable is consistent with established practices in BN construction based on expert judgement, particularly in complex domains where empirical data are incomplete [96].

6.1.6. MASS Familiarity

The variable MASS Familiarity represents participants' level of familiarity with the concept of MASS. It was derived from a questionnaire in which participants assessed their knowledge of MASS using a five point Likert scale, where 1 indicates no familiarity with the concept and five indicates excellent familiarity.

Understanding familiarity with autonomous ship concepts is important because the development and implementation of MASS represent a significant technological and operational change in maritime transport. The introduction of such systems requires seafarers and future operators to understand new operational principles, automation systems, and the evolving roles of human operators in ship management and navigation. Findings from the study conducted by Kondo *et al.* indicate that the emergence of autonomous shipping will require adjustments in maritime education and training, as well as a broader understanding of autonomous ship operations among maritime professionals [97].

The categories were defined by grouping the original Likert scale responses into broader levels to ensure sufficient representation of observations within each category and to facilitate probabilistic modelling.

Based on the collected survey data (N = 65), the distribution of responses (Figure 6.3) that participants generally have a low to moderate level of familiarity with the concept of MASS.

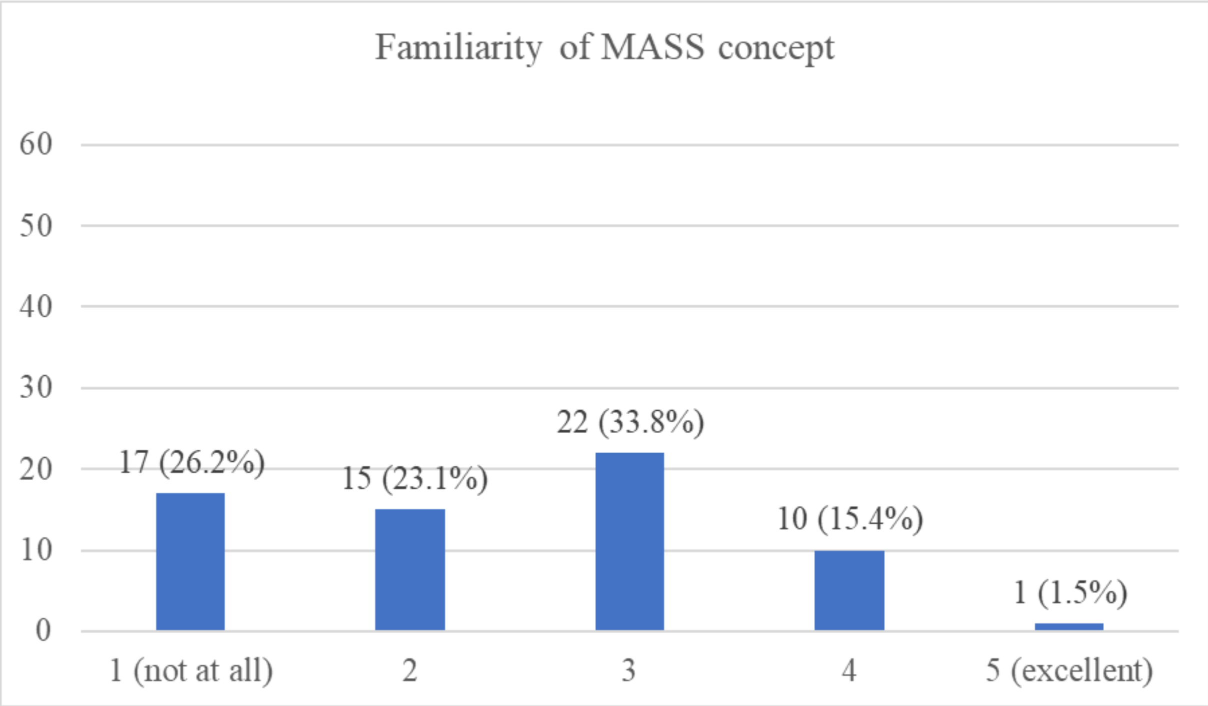

Figure 6.3 Distribution of participant familiarity with the MASS concept (N = 65)

Based on data collected from 65 participants, the MASS Familiarity variable was discretized into three categories, low familiarity (1 and 2 responses), moderate familiarity (3 response), and high familiarity (4 and 5 responses) resulting in the following distribution:

- low familiarity: 32 participants (49%),
- moderate familiarity: 22 participants (34%),
- high familiarity: 11 participants (17%).

The results indicate that nearly half of the participants have limited familiarity with the concept of MASS, while a relatively small proportion demonstrates a high level of knowledge. This reflects the still emerging nature of autonomous shipping technologies and the limited exposure of maritime professionals to MASS in practical contexts.

This variable is treated as an independent (root) variable, representing participants' prior knowledge before the simulation exercise, and therefore does not have parent nodes within the BN structure.

6.1.7. Sea-Going Experience

The variable Sea-Going Experience refers to the participant's prior practical experience gained through service on board ships. Experience acquired during real maritime operations is a key factor in developing navigational performance, as it enables seafarers to become familiar with bridge procedures, navigational equipment, and the operational conditions encountered during ship navigation. Practical exposure to ship operations supports the gradual development of judgement and the ability to interpret complex navigational situations.

Findings from the study conducted by Porathe, Prison, and Man in maritime human factors highlight that operational experience is important for developing situational awareness and for operators' ability to understand and interpret navigational information in demanding maritime environments [98]. Seafarers with practical Sea-going Experience are generally better prepared to recognize potential risks and respond appropriately to operational challenges.

In this dissertation, the variable is derived from the questionnaire in which participants reported the total time they had spent sailing on board ships.

The Sea-going Experience variable, based on a sample of 65 participants, shows:

- no sea-going experience (0 years):33 participants (52%),
- limited sea-going experience (1–10 years): 15 participants (23%),
- extensive sea-going experience (>10 years): 16 participants (25%).

This distribution reflects the heterogeneous composition of the sample, which includes both participants without real world maritime experience and experienced seafarers with extended sea service.

This classification was adopted to ensure a clear distinction between participants with no practical exposure to ship operations, those with basic to moderate operational experience, and those with substantial professional experience at sea.

The inclusion of a separate category for participants with no sea-going experience is methodologically important, as this group constitutes a qualitatively distinct population in terms of operational familiarity, decision-making strategies, and reliance on theoretical knowledge. In contrast, participants with practical experience are expected to exhibit more developed mental models of ship operations and a greater ability to manage complex navigational situations.

In the BN model, Sea-Going Experience is treated as an independent (root) variable with no parent nodes, and its probability distribution is derived directly from the empirical data. As a background variable, it serves as an important contextual factor influencing higher-level constructs such as decision-making performance, Navigation Control, and overall Operational Effectiveness.

6.1.8. Operator Readiness

The variable Operator Readiness represents a composite measure of an individual's preparedness to perform decision-making tasks effectively in the context of MASS operations. Rather than reflecting a single observable characteristic, this variable integrates multiple dimensions of operator capability into a unified probabilistic construct. It captures the combined effects of practical experience, professional background, and system-specific knowledge, providing a higher-level representation of operational competence within the BN model.

Within the model, Operator Readiness is defined as a child node with three parent variables:

- Sea-Going Experience,
- Operator Background,
- MASS Familiarity.

These variables were selected to reflect the most relevant determinants of operator performance in autonomous maritime environments. Sea-Going Experience represents direct exposure to real navigational conditions and operational responsibilities, and is treated as the dominant contributor to readiness. Operator Background, derived from demographic and professional characteristics, provides a structured representation of baseline competence. MASS Familiarity introduces a domain-specific dimension, reflecting the operator's understanding of autonomous systems and their operational principles. This selection is consistent with recent research on operators of autonomous ships, which emphasises that effective performance in remote control centres depends on a combination of practical experience, professional performance, and system-specific knowledge [99].

Following the expert-based state definition methodology described at the beginning of this chapter, the variable was classified into three mutually exclusive states: low, moderate, and high. This categorization ensures both interpretability and compatibility with the overall BN structure. The CPT was constructed using a weighting approach, in which Sea-Going Experience and Operator Background jointly determine the primary direction of the outcome, while MASS Familiarity acts as a secondary modifier that refines the distinction between intermediate and high readiness levels.

The probabilistic relationships were designed to satisfy monotonic consistency, so that improvements in the parent variables lead to a gradual increase in the likelihood of higher readiness states. For example, combinations characterized by extensive Sea-going Experience, advanced Operator Background, and high MASS Familiarity consistently produce a dominant high readiness outcome (approximately 75%–80%), while profiles with no experience and limited background are associated with a dominant low readiness state (typically exceeding 70%). Intermediate configurations, such as limited experience combined with an intermediate background and moderate familiarity, result in a balanced distribution with moderate readiness as the most probable state (around 50%), reflecting realistic operational uncertainty.

The Operator Readiness variable, derived from the combination of three parent variables, was distributed as follows:

- low readiness (61%),
- moderate readiness (31%),
- high readiness (8%).

These values are not predefined but are derived from the BN structure, reflecting the combined effect of the prior distributions of the parent variables and the specified conditional probability relationships within the CPT. This distribution reflects the sample structure, which includes a substantial proportion of less experienced participants, and is consistent with expectations for a heterogeneous population in a simulator-based training environment.

The validity of the CPT was verified through scenario-based sensitivity testing in the GeNIe software environment. Extreme cases and balanced configurations were evaluated to ensure logical and stable behavior of the variable. The results confirmed that Operator Readiness responds consistently to changes in input conditions, demonstrating both robustness and interpretability of the modelled relationships

6.1.9. Equipment Comfort

The Equipment Comfort variable reflects participants' subjective assessment of the comfort or discomfort associated with wearing the experimental equipment during the simulation exercise. In the exercise, participants wore eye-tracking glasses to monitor visual attention and a smartwatch to monitor physiological indicators during the simulation. Findings from the study conducted by Niehorster, *et al.*, on using eye-tracking systems and other measurement devices, experimental conditions and equipment may influence the naturalness of participant behavior during task performance [100]. Although such devices are designed for use in everyday and research settings, the presence of additional equipment can cause discomfort or reduce the naturalness of behavior during task performance.

Data for this variable were collected using a survey before the simulation, in which participants were asked whether they expected to feel discomfort due to the equipment they would wear during the exercise. This approach provided insight into participants’ initial perceptions before exposure to the experimental task.

The empirical distribution of responses, based on a sample of 65 participants, indicates that most of them did not anticipate discomfort when using the experimental equipment.

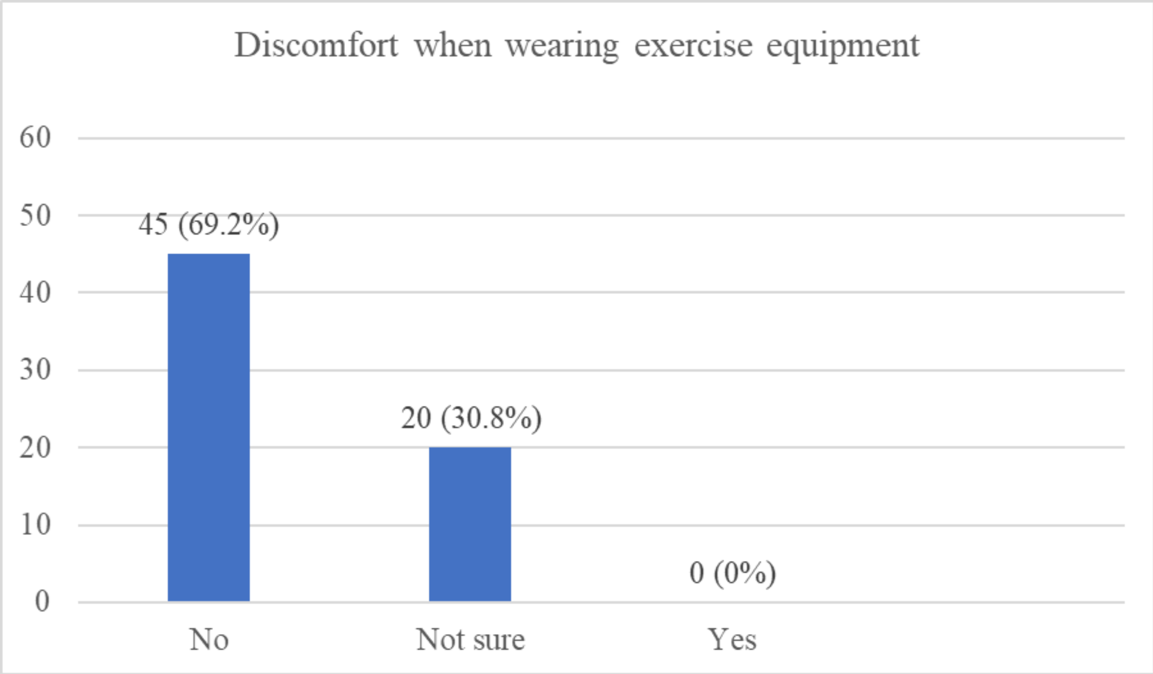

Figure 6.4 Distribution of responses on discomfort when wearing experimental equipment (N = 65)

For the purposes of the BN, the variable was defined with three possible states:

- comfortable: 45 participants (69%),
- not sure: 20 participants (30%),
- uncomfortable: 0 participants (1%).

Although no observations were recorded for the uncomfortable state, a small non-zero probability (1%) was assigned to this category in the model. This adjustment ensured numerical stability and proper functioning of the BN in the GeNIe environment, where zero probability states can cause computational limitations. The final probability distribution used in the model was therefore 69% for comfortable, 30% for not sure, and 1% for uncomfortable, without affecting the overall interpretation of the observed data.

In the BN, this variable was modelled as an independent input variable without parent nodes, as it represents an individual characteristic present before the simulation begins. Within the model, it serves as a contextual factor that may contribute to the interpretation of participant behavior, particularly through its potential influence on cognitive load and physiological stress during task execution.

6.2. Physiological Response Variables

The Physiological Response Variables indicate participants' physiological reactions during the simulation exercise. These variables were used to capture changes associated with cognitive workload and stress during navigation tasks in the simulation.

6.2.1. Temperature Change

Temperature Change represents the difference between body temperature measured immediately before and after completion of the simulation exercise. This variable was calculated as the difference between post temperature and pre-temperature, and reflects short-term physiological variation occurring during exposure to the simulated navigational scenario.

All measurements were conducted under controlled indoor laboratory conditions to minimize environmental influence. Given the relatively short interval between measurements, the observed variation is assumed to represent acute physiological adjustment rather than long regulatory processes.

Acute psychological stress has been associated with mild elevations in core body temperature resulting from sympathetic nervous system activation. This phenomenon, commonly described as stress-induced hyperthermia, reflects autonomic activation rather than infection related fever mechanisms [101]. Clinical observations indicate that psychological stress alone may contribute to measurable increases in body temperature in humans, a condition often referred to as psychogenic fever [102].

In the simulator exercise, participants experienced increasing cognitive load, system degradations, and decision-making pressure. These operational conditions may activate autonomic regulatory mechanisms and lead to subtle physiological responses.

The variable was calculated as the difference between post-simulation and pre-simulation body temperature, as shown in the following expression:

$$\Delta Temperature_{Change} = T_{post} - T_{pre} \tag{6.1}$$

where T_{post} denotes the body temperature measured immediately after the simulation exercise, and T_{pre} denotes the body temperature measured immediately before the simulation.

Based on a sample of 65 participants, discretization into three states was defined according to the magnitude of the observed temperature change:

- decrease: 8 participants (12%),
- no Change: 28 participants (28%),
- increase: 39 participants (60%).

A threshold of 0.1°C was adopted as it represents the minimum measurement resolution of the recording device, allowing physiologically meaningful temperature changes to be distinguished from negligible measurement variability.

In the BN model, Temperature Change was implemented as a root node. This variable does not depend on any other component of the network.

6.2.2. Blood Pressure Reactivity

Blood pressure was measured immediately before and after the simulator exercise using a sociometric wrist-based device. Both systolic blood pressure (SBP) and diastolic blood pressure (DBP) were recorded in millimeters of mercury (mmHg). Measurements were taken under consistent conditions for all participants to ensure data comparability.

The analysis focused on the change between pre- and post-simulation values rather than absolute blood pressure levels. For each participant, the difference in systolic and diastolic pressure (ΔSBP and ΔDBP) was calculated as:

$$\Delta SBP = SBP_{after} - SBP_{before} \tag{6.2}$$

$$\Delta DBP = DBP_{after} - DBP_{before} \tag{6.3}$$

This approach enabled assessment of the cardiovascular response to the cognitive and operational demands of the autonomous ship simulation task.

Absolute blood pressure values may vary considerably between individuals due to physiological and lifestyle factors. By analysing change values instead of raw measurements, the influence of baseline differences was reduced, allowing clearer interpretation of stress-related responses within a heterogeneous participant sample.

Cardiovascular reactivity is widely recognized as an indicator of acute psychophysiological stress response. Experimental research has also demonstrated that cognitively demanding tasks can significantly influence systolic and DBP levels during computer-based work [103].

Based on the calculated changes in systolic and diastolic blood pressure (ΔSBP and ΔDBP), each participant was classified according to the direction and magnitude of blood pressure change.

For classification, a threshold-based approach was applied. Changes within the range of ±10 mmHg were considered stable, while values exceeding this range were classified as either increase or decrease, depending on the direction of change.

Based on a sample of 65 participants, the distribution of participants across the defined categories was as follows:

- decrease: 16 participants (25 %),
- stable: 42 participants (64%),
- increase:7 participants (11%).

These results indicate that the majority of participants did not exhibit significant changes in blood pressure following the simulation task, suggesting a moderate overall physiological response. However, a notable proportion of participants demonstrated either increased or decreased blood pressure, reflecting individual variability in cardiovascular reactivity under cognitive and operational stress conditions.

The observed distribution was used to define the prior probabilities of the Blood Pressure Reactivity variable in the BN model implemented in GeNIe.

In the BN model, Blood Pressure Reactivity was implemented as a root node. This variable does not depend on any other component of the network.

6.2.3. Heart Rate Change

Heart rate was continuously recorded throughout the navigational exercise using a wrist worn device. Measurements were collected in real time during task execution, enabling monitoring of physiological responses under conditions involving cognitive workload, time pressure, and system degradations introduced during the scenario.

For analytical purposes, the representative heart rate value for each participant was calculated as the mean heart rate over the entire simulation period. This value was defined as:

$$HR_{mean} = \frac{1}{N}\sum_{I=1}^{N} HR_{\mathrm{i}} \tag{6.4}$$

HR_{mean} denotes the mean heart rate value for a given participant during the simulation, HR_i represents each individual heart rate measurement recorded at 5-second intervals, and N is the total number of heart rate measurements collected for that participant over the full duration of the exercise. In practical terms, this calculation involved summing all recorded heart rate values for a participant and dividing the result by the total number of recorded measurements. This approach provides a stable indicator of overall cardiovascular activation and reduces the influence of short-term fluctuations that may occur during individual events within the exercise.

Heart rate is widely recognized as a sensitive marker of autonomic nervous system activation in response to psychological stress and cognitive demands [104]. Cardiovascular responses to controlled stress exposure have been extensively examined in experimental research and are commonly interpreted as indicators of task-related physiological strain [105].

Within of this dissertation, heart rate was incorporated into the BN as an indicator of physiological activation. The defined categories indicate relative levels of physiological activation based on the distribution of mean heart rate values observed in the dataset. The classification thresholds were defined based on the distribution of the observed mean heart rate values. Specifically, the low category includes values less than or equal to 88 bpm, the moderate category includes values between 88 and 101 bpm, while the high category includes values greater than 101 bpm.

These categories are defined relative to the study sample and do not represent absolute clinical thresholds; rather, they reflect differences in physiological response among participants under the same experimental conditions. Based on data collected from 65 participants, the Heart Rate Change variable was discretized into three categories, low, moderate and high resulting in the following distribution:

- low: 19 participants (29%),
- moderate: 26 participants (40%),

- high: 20 participants (31%).

In the BN, this variable is defined as an external input variable with no parent nodes. The values are taken directly from the processed experimental data and entered into the model as prior probabilities.

6.2.4. Physiological Response

Physiological Response acts as an aggregated indicator of the participant's immediate physiological reaction to the simulator-based navigational task. This variable was introduced to combine the main measurable bodily responses associated with task-related stress and activation into a single intermediate node, thereby reducing model complexity while preserving the most relevant physiological information. Rather than interpreting heart rate, blood pressure, and temperature separately at higher levels of the BN, this variable summarizes their combined effect in a structured and probabilistically consistent form. In the current model, Physiological Response is defined as a child node with three parent variables:

- Temperature Change,
- Blood Pressure Reactivity,
- Heart Rate Change.

Following the expert-based state definition methodology described at the beginning of this chapter, the variable Physiological Response was classified into three mutually exclusive states. This categorization ensures both interpretability and compatibility with the overall BN structure:

- low (32%),
- moderate (41%),
- high (27%).

Blood Pressure Reactivity was treated as the dominant parent, Heart Rate Change as the secondary contributor, and Temperature Change as a minor corrective factor. The relative influence of the parent variables was approximated as 60% for blood pressure, 30% for heart rate, and 10% for temperature. This weighting structure was intentionally selected so that the CPT would be driven primarily by cardiovascular reactivity, while still allowing heart rate to meaningfully modify the response and temperature to provide only limited adjustment. In practical terms, blood pressure determined the general direction of the resulting state, heart rate shifted the result upward or downward within that direction, and temperature acted only as a final fine-tuning input.

The three parent variables were incorporated into the model based on empirically observed state distributions from the participant sample.

The variable's behavior was verified through scenario-based tests in the GeNIe environment, confirming that the node responded logically across a range of input conditions. The model demonstrated stable transitions between states and maintained the intended hierarchical influence of the parent variables, with blood pressure as the primary driver of the response, heart rate providing secondary modulation, and temperature acting as a minor corrective factor.

In the baseline configuration, the resulting distribution of the Physiological Response variable indicates that the moderate state is the most prevalent outcome. This aligns with the expected physiological profile of participants exposed to controlled but cognitively demanding simulator conditions, where measurable activation is present but does not typically reach extreme levels.

Within the BN, this variable functions as an intermediate integrative node, translating multiple lower-level physiological indicators into a unified representation of acute physiological activation in response to operational stress.

6.2.5. Physiological Stress Level

The variable Physiological Stress Level is a derived variable within the BN model, intended to capture the overall operator stress during the simulator-based navigation task. Rather than reflecting a single measured parameter, this variable integrates multiple sources of influence, combining physiological activation, operational performance, and contextual factors. It acts as an intermediate node that translates lower-level indicators into a unified representation of stress experienced under realistic maritime conditions. Within the BN, Physiological Stress Level is modelled as a child node with four parent variables:

- Physiological Response,
- Response to Navigation System Failures,
- Sea-Going Experience, and

- Equipment Comfort.

The CPT was constructed using a structured expert elicitation approach to determine the relative importance of each contributing factor. As described earlier in this chapter, the expert group comparatively assessed the influence of each parent variable on the target node, and the final weighting scheme represents a consensus-based evaluation of their relative operational importance within the BN model. Physiological Response and Response to Navigation System Failures were assigned the highest influence (35% each), as they directly represent the operator's physiological activation and ability to manage critical system disruptions. Sea-Going Experience was assigned a moderate influence (20%), reflecting its role as a mitigating factor that can reduce stress through operational familiarity and accumulated experience. Equipment Comfort was assigned a lower influence (10%), representing a secondary source of strain that may influence, but is not expected to dominate, the overall physiological stress response.

The Physiological Stress Level, derived from the combination of four parent variables, was distributed as follows:

- moderate stress (45%),
- high stress (55%).

Categorization was chosen to balance interpretability and model complexity, allowing clear differentiation between stress levels while maintaining a manageable conditional probability structure.

The CPT was structured to ensure a consistent and monotonic relationship between input conditions and the resulting stress level. Combinations with low physiological activation, effective response to failures, high experience, and comfortable operating conditions were assigned a high probability of low stress. Conversely, combinations characterized by elevated physiological response, poor handling of system failures, lack of experience, and uncomfortable equipment conditions were mapped to High Stress. This approach ensured that the model avoids abrupt or inconsistent behavior and instead reflects a realistic progression of stress under varying operational conditions [106].

Particular attention was given to maintaining logical consistency across the CPT, ensuring that increases in physiological activation and degradation in operational performance consistently led to higher stress levels. At the same time, the moderating role of experience ensures that, even under adverse conditions, more experienced operators are less likely to experience extreme stress than less experienced participants. Similarly, equipment-related discomfort contributes to stress escalation but does not override the influence of the primary factors.

The behavior of the variable was further verified through scenario-based testing within the GeNIe environment. Under ideal conditions – characterized by low physiological activation, effective response to failures, high experience, and comfortable equipment – the model converged to a dominant low stress state. In contrast, adverse conditions resulted in a dominant High Stress outcome, while intermediate scenarios produced a balanced distribution centred around moderate stress. These results confirm that the variable behaves in accordance with expected domain logic and that the weighting structure produces stable and interpretable outcomes.

6.2.6. Cognitive Overload

The variable Cognitive Overload is a derived construct within the BN model. Unlike directly measured variables, this construct integrates multiple underlying influences and provides a unified representation of the operator's internal state during task execution. In the context of MASS operations, Cognitive Overload is particularly relevant as it directly affects situational awareness, decision-making accuracy, and the ability to respond effectively to system disturbances [107].

Within the BN, Cognitive Overload is modelled as a child node with four parent variables:

- VAQ,
- Physiological Stress Level,
- Navigation Complexity,
- Equipment Comfort.

The CPT was constructed based on expert judgement to represent the relative importance of each contributing factor. VAQ was assigned the highest influence (45%), as it directly reflects the operator's ability to maintain focused and efficient visual processing under dynamic conditions. Physiological Stress Level was assigned a substantial influence (35%), capturing the effect of physiological activation on cognitive functioning. In contrast, Navigation

Complexity (10%) and Equipment Comfort (10%) were treated as secondary, corrective factors that modulate the overall cognitive load without dominating the outcome.

The Cognitive Overload, derived from the combination of four parent variables, was distributed as follows:

- low (20%),
- moderate (36%),
- high (44%).

This enables a clear and interpretable distinction between different degrees of cognitive strain while maintaining a manageable level of complexity for probabilistic modelling. The states were intentionally limited to three categories to balance model sensitivity and robustness, ensuring that the CPTs remain tractable and logically consistent.

The structure of the CPT was calibrated to ensure monotonic and interpretable relationships between inputs and outputs. Lower levels of VAQ and higher levels of physiological stress consistently increase the probability of the high Cognitive Overload state. Conversely, favourable conditions, such as high attention quality and low stress, shift the distribution towards the low state. Navigation Complexity and Equipment Comfort introduce additional adjustments, but their influence remains intentionally limited so as not to override the primary cognitive and physiological drivers.

The validity of the CPT was verified through scenario-based sensitivity testing in the GeNIe software environment. Extreme cases and balanced configurations were evaluated to ensure logical and stable behavior of the variable. The results confirmed that the variable Cognitive Overload responds consistently to changes in input conditions, demonstrating the robustness and interpretability of the modelled relationships.

6.3. Visual Attention Variables

Visual attention variables are a key component of behavioral analysis in this dissertation, providing objective insight into how operators interact with the navigational environment during the simulation. These variables are derived from eye-tracking data and capture aspects of visual processing, including spatial distribution, temporal dynamics, and cognitive load.

Together, they enable a detailed assessment of attention allocation and support the identification of patterns associated with effective and ineffective navigation performance.

6.3.1. Gaze Dispersion Index (GDI)

GDI measures the spatial dispersion of gaze within a specific AOI and a defined temporal window. It is calculated as the radial standard deviation of normalized horizontal and vertical gaze coordinates, using all valid gaze samples recorded on the observed surface:

$$GDI = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{6.5}$$

Where σ_x^2 and σ_y^2 denote the variance of normalized gaze coordinates in the horizontal and vertical directions$\left(x_{norm,}y_{norm} \in [0,1]\right)$.

The GDI is calculated using all gaze samples within the observed surface, without filtering for fixation events. If the number of available gaze samples is insufficient, the GDI value is considered unavailable. The calculation is performed separately for each surface (radar, ECDIS, controls, visualization) and for each temporal phase (baseline, acute, sustained).

Higher GDI values indicate greater spatial dispersion of gaze, suggesting visual search behavior, surface scanning, or difficulty in identifying relevant information. Lower values indicate focused monitoring and narrowly directed attention.

From an interpretative perspective, GDI reflects the extent of visual exploration behavior. Lower values indicate concentrated gaze patterns and stable attention directed towards specific elements, while higher values indicate greater spatial dispersion, suggesting active visual scanning or difficulty in locating relevant information. In complex operational environments, increased dispersion is often associated with higher cognitive demand and adaptive search behavior, particularly under conditions of uncertainty or system degradation [108].

Interpretation of GDI (in the context of the present scenario) includes four descriptive levels. These levels are:

- very low GDI (< 0.0002) indicates highly focused gaze behavior, with the participant consistently monitoring a single point on the surface. In inexperienced participants, this may reflect stress-induced “tunnel vision”, whereas in experienced operators it may represent expected behavior during routine monitoring.

- low GDI (0.0002–0.0008) indicates focused gaze with limited spatial dispersion. The participant maintains attention within a confined area, which may reflect targeted information processing or the initial phase of visual search.
- medium GDI (0.0008–0.002) indicates moderate visual exploration, where the participant scans the surface and monitors relevant elements without excessive dispersion. This is considered a functional viewing pattern that supports situational awareness.
- high GDI (> 0.02) indicates highly dispersed gaze behavior, with the participant actively searching for information across the entire surface. In inexperienced participants, this often reflects disorientation, confusion, or early signs of panic, while in experienced operators it may indicate broad environmental scanning.

For the purposes of constructing the composite SGPS index, the inverse normalized form of GDI was used:

$$GDIinv = 1 - GDInorm \tag{6.6}$$

This transformation allows higher values to represent more focused and stable gaze behavior, aligning the directionality of the variable with other components of the composite index.

Although GDI was initially interpreted using four descriptive categories (very low, low, medium, and high) to provide a detailed behavioral understanding of gaze dispersion. Based on a sample of 65 participants, discretization into three states was defined:

- low:8 participants (13%),
- medium: 10 participants (15%),
- high: 47 participants (72%).

This reduction was implemented to improve model stability and reduce sparsity in the conditional probability tables.

In this aggregation, the very low and low categories were merged into the low state, as they represent gaze behavior characterized by limited spatial dispersion, focused attention, or functionally controlled visual exploration. The medium state corresponds to the medium GDI category, reflecting increased spatial dispersion and broader visual scanning behavior. The high state represents the most pronounced gaze dispersion, typically associated with intensive visual search, uncertainty, or elevated cognitive load.

The dominance of the high category reflects the demanding nature of the simulation environment, which included system degradations, complex navigation conditions, and participants with varying levels of experience.

In the BN, GDI is defined as an independent root variable without parent nodes. Its values are taken directly from the eye-tracking data collected during the simulation experiment and serve as empirical inputs to the model.

6.3.2. Pupil Entropy Index (PEI)

The PEI index measures cognitive load based on the variability of pupil diameter. It is calculated as the Shannon entropy of the distribution of pupil diameter values within a defined time window, using discretization into 50 histogram bins.

$$PEI = -\sum_{i=1}^{n} p(x_i) \cdot log_2 p(x_i) \tag{6.7}$$

Where:

- n — the number of histogram bins (set to 50),
- $p(x_i)$ — the probability that pupil diameter values fall into the i-th bin.

The result is expressed in bits (logarithmic base 2). PEI is calculated over defined time windows (baseline, acute, sustained) using all pupil diameter values within the observed interval. Higher PEI values indicate greater variability in pupil diameter, which is associated with increased activation of the sympathetic nervous system, reflecting higher cognitive load, mental effort, and emotional arousal. Lower PEI values indicate a more stable pupil diameter, associated with parasympathetic dominance, reflecting a more relaxed state, lower workload, or potentially fatigue.

Low PEI values indicate a relaxed state, characterized by stable pupil diameter and low variability. This is typically expected during baseline conditions or routine task execution. Moderate PEI values indicate a functional level of cognitive engagement, where the participant is attentive and actively processing information without being overloaded. High PEI values indicate increased cognitive load or stress, reflected in greater variability of pupil diameter. In less experienced participants, such values may also indicate overload, anxiety, or difficulty managing task demands.

Although entropy-based measures have been used to assess attentional and cognitive processes, their absolute values are highly dependent on experimental conditions and task characteristics [109], [110]. Therefore, PEI values should be interpreted relative to the specific dataset and experimental context rather than using fixed universal thresholds.

For BN modelling, the continuous PEI variable values were discretized into three categories using a data-driven approach based on the empirical distribution of the dataset. The low category included values ≤ 4.45, the moderate category included values between 4.45 and 4.81, and the high category included values > 4.81.

Based on a sample of 65 participants, discretization into three states was:

- low: 20 participants (31%),
- moderate: 25 participants (38%),
- high: 20 participants (31%).

This distribution reflects a balanced representation of cognitive load levels, with a slight predominance of moderate values, suggesting that most participants operated under functional but cognitively demanding conditions during the simulation.

Within the BN, the PEI variable is treated as a root node, with values derived directly from the processed eye-tracking data.

6.3.3. Gaze Movement Score (GMS)

GMS quantitatively measures the frequency of intentional gaze shifts over time, reflecting the dynamic aspect of visual behavior during task execution. While the GDI captures the spatial distribution of fixations, GMS describes the temporal dynamics of visual attention, indicating how often a participant redirects their gaze between different AOI within the operational environment.

The metric was developed to capture active visual search behavior in a complex and dynamic simulator scenario. During navigation tasks, especially under conditions of system degradation and increased cognitive demand, operators continuously shift their gaze between navigational displays, external visual cues, and control interfaces. The frequency of these gaze shifts reflects the intensity of environmental scanning, situational awareness updating, and adaptive responses to operational challenges. Previous research has shown that eye movements are systematically guided between relevant locations during visual search tasks [111]. Increased eye movement activity has also been associated with higher levels of mental workload, supporting the interpretation of gaze dynamics as an indicator of cognitive demand [112].

The calculation of GMS is based on detecting gaze shifts with amplitudes exceeding a predefined threshold, thus excluding micro-movements and drift during fixations. The metric is defined as:

$$GMS = \frac{Number\ of\ gaze\ shifts\ with\ amplitude\ >\ threshold}{Duration\ of\ the\ time\ window\ (s)} \tag{6.8}$$

The amplitude threshold is set at 0.005 in normalized screen coordinates, which corresponds to approximately 0.5% of the screen width or height. The denominator represents the actual duration of the observed time window in seconds. The analysis was conducted within defined temporal segments of the simulation (baseline, acute, and sustained phases), with a minimum time window duration of 10 seconds required to ensure the reliability of the calculated values.

The threshold of 0.005 was selected during the optimization of the eye-tracking processing algorithm developed for this dissertation. Several candidate values were evaluated, and the selected threshold provided an appropriate balance between excluding micro-movements within fixations and preserving intentional gaze shifts between distinct AOIs. This threshold excludes physiological noise while preserving meaningful gaze transitions related to attentional reallocation. A higher threshold (e.g., 0.01) would detect only larger gaze shifts, resulting in lower GMS values, while a lower threshold (e.g., 0.001) would include micro-movements, increasing sensitivity to noise and reducing specificity. The selected value balances sensitivity and robustness, ensuring consistent comparison across participants. A higher threshold (e.g., 0.01) would detect only larger gaze shifts, resulting in lower GMS values, while a lower threshold (e.g., 0.001) would include micro-movements, increasing sensitivity to noise and reducing specificity. The selected value balances sensitivity and robustness, ensuring consistent comparison across participants.

From a behavioral perspective, GMS reflects the intensity of visual scanning. Lower values indicate stable gaze behavior, characterized by prolonged fixation on specific elements, typically associated with routine monitoring and efficient information processing. Moderate values indicate functional visual scanning, where the participant actively monitors multiple sources of information in a controlled and organized manner. Higher values indicate frequent gaze redirection, suggesting intensified information seeking, increased task demand, or potential difficulty in maintaining stable attention.

For interpretative purposes in this dissertation, GMS values were analysed using behaviorally meaningful ranges. Values below 0.5 shifts per second indicate stable and focused viewing. Values between 0.5 and 1.0 shifts per second indicate moderate and functional scanning behavior. Values above 1.0 shifts per second indicate frequent scanning behavior, which may be associated with increased cognitive demand. Values above 1.5 shifts per second indicate highly intensive scanning, potentially reflecting elevated stress, disorganized visual behavior, or Cognitive Overload.

Changes in GMS across simulation phases offer additional insight into behavioral adaptation. A sudden increase in GMS relative to baseline (for example, from 0.3 to 1.2 shifts per second) indicates stress-induced visual scanning, where the participant moves from stable monitoring to rapid gaze reallocation in response to unexpected events. Sustained high GMS values during later phases may indicate prolonged cognitive strain or difficulty adapting to operational demands. Conversely, relatively low GMS values during high-demand phases, particularly when combined with elevated physiological indicators, may reflect focused and controlled attention rather than disorganized scanning.

Compared to other visual metrics used in this dissertation, a high GMS value often co-occurs with a high GDI, reflecting both increased spatial dispersion and frequent gaze relocation. However, GMS provides complementary information by quantifying the rate of gaze transitions, thereby capturing the temporal dimension of visual behavior. When interpreted together with the PEI, GMS enables differentiation between cognitive states. High GMS combined with high PEI may indicate stress accompanied by active scanning, whereas low GMS combined with high PEI may reflect focused but cognitively demanding processing.

For BN modelling, based on a sample of 65 participants the continuous GMS values were discretized into three categories:

- low: 28 participants (44%),
- medium: 17 participants (24%),
- high: 20 participants (32%).

This distribution indicates that most participants exhibited stable or moderately dynamic visual behavior, while a substantial proportion demonstrated elevated scanning activity, particularly during more demanding phases of the simulation scenario.

In the BN, the GMS variable is defined as a root node, with no parent nodes in the model. Its values are determined directly from empirical data collected during the simulation exercise, in which gaze movement frequency was calculated for each participant. The observed data were used to establish the prior probability distribution of the variable and serve as direct input to the model.

6.3.4. Stress-induced Gaze Pattern Score (SGPS)

SGPS is a composite indicator designed to capture the combined effects of visual attention behavior and physiological response during the simulator-based navigation task. This variable integrates multiple dimensions of operator performance, enabling a unified representation of stress level and cognitive load under dynamically changing operational conditions. Unlike individual indicators considered separately, SGPS offers a holistic measure of operator state by combining behavioral and physiological signals within a single framework.

The SGPS index is calculated in several steps. First, all input components (PEI, GDI, and GMS) are normalized to a common scale to ensure comparability between variables measured in different units. Each component is transformed to the range [0, 1] using the min–max normalization method:

$$X_{norm} = \frac{X - X_{min}}{(X_{max} - X_{\mathrm{min}})} \tag{6.9}$$

As a lower GDI indicates focus and potential stress, the inverse GDI from formula 6.6 is defined as follows.

After normalization, the SGPS index is calculated as a weighted sum of the normalized input variables. The weighting scheme reflects the relative importance of each component in representing operator stress and cognitive load. The weights were established through the structured expert elicitation process described earlier in this chapter, during which the expert group comparatively assessed the contribution of each component to the SGPS index. The selected weighting scheme was not intended to represent a universally optimal configuration, but rather a consensus-based evaluation of the relative importance of the physiological and gaze-based indicators within the methodological framework proposed in this dissertation. Consequently, the physiological component (PEI) was assigned the highest weight (0.4), reflecting its direct representation of autonomic activation associated with mental workload, whereas the gaze-based indicators GDI and GMS were each assigned a weight of 0.3, reflecting their complementary contribution to the assessment of visual attention behavior. The SGPS index is calculated as follows:

$$SGPS = 0.4 \cdot PEI_{norm} + 0.3 \cdot GDI_{inv} + 0.3 \cdot GMS_{norm} \tag{6.10}$$

This formulation ensures that physiological activation has the greatest influence on the final stress score, while visual behavior modulates the overall result.

The mean SGPS value was approximately 0.634, with a standard deviation of 0.050. To preserve the central tendency of the observed data while differentiating between lower and higher stress responses, classification thresholds were defined using a mean ± 0.5 standard deviation approach. Values below approximately 0.609 were classified as low, values between 0.609 and 0.660 as moderate, and values above 0.660 as high. Applying these thresholds resulted in the following distribution:

- low: 18 participants (28%),
- moderate: 24 participants (37%),
- high: 23 participants (34%).

This data-driven discretization ensures that the SGPS categories reflect the actual variability observed in the experimental dataset, rather than relying on arbitrary or externally defined thresholds.

Within the BN, SGPS is defined as a derived variable based on three parent variables: GMS, GDI and PEI.

6.3.5. Visual Attention Quality (VAQ)

The variable VAQ represents an interpretative assessment of the effectiveness of visual attention during the simulator-based navigation task. It reflects how efficiently the operator processes visual information from multiple sources, including radar, ECDIS, and the external visual environment, under dynamically changing operational conditions.

Unlike directly measured eye-tracking variables, VAQ is not derived from a single raw indicator; it is defined as a higher-level interpretative construct based on the composite variable SGPS. Introducing VAQ as a separate node creates a conceptual abstraction layer between raw eye-tracking measurements and higher-level cognitive and operational variables, improving overall interpretability while preserving the information contained in the original data.

Visual attention is a crucial aspect of cognitive processing in complex and dynamic operational environments, as it directly affects perception, situation assessment, and decision-making. Physiological and behavioral indicators from eye-tracking provide reliable insights into attentional processes in real time. Pupil dynamics, in particular, are sensitive markers of attentional engagement and cognitive processing load, allowing high temporal resolution analysis of attention allocation [113]. Increased system complexity and higher levels of automation place greater demands on attentional resources, which can significantly impact operator performance and decision-making quality [114].

The VAQ variable, derived from the SGPS variable using the CPT, was distributed as follows:

- low (30%)
- moderate (35%)
- high (35%)

In the BN, this variable is defined as an intermediate node with SGPS as its single parent. The CPT was designed to preserve the dominant SGPS state while assigning smaller probabilities to adjacent states, introducing a softened probabilistic transition that reflects the inherent uncertainty of human visual attention processes. This structure improves model robustness while maintaining a close conceptual relationship between SGPS and VAQ.

The behavior of the VAQ node was verified through evidence-based consistency tests performed in GeNIe software. The results confirmed that the node behaves as intended, with the corresponding state becoming dominant in each test scenario while preserving the softened probabilistic transition defined by the CPT.

6.4. Navigational Performance Variables

The navigational performance variables are objective indicators derived from simulator logs, used to assess participants' behavior during the simulation exercise. They reflect key aspects of navigational performance and safety, including vessel proximity, speed control, recognition of system faults, and the ability to maintain the planned navigational track.

6.4.1. Radar Error

The Radar Error variable represents the time required for a participant to recognize and identify a radar system failure during a simulation scenario. Radar is navigation instrument, enabling target detection, assessment of relative vessel motion, and maintenance of safe distances between vessels, especially in conditions of increased traffic density and limited maneuvering space. Timely detection of radar system malfunctions is therefore an important aspect of situational awareness and safe ship control.

During the simulation exercise, the radar was intentionally degraded at a certain point to simulate a technical failure in the navigation system. Participants were not informed in advance of the exact moment of failure, allowing observation of their ability to recognize changes in system operation during the navigation activities. The fault detection time was measured from the moment of the radar malfunction to the moment the participant responded to the problem, either verbally or through a navigational action indicating that the fault had been recognized.

The ability to recognize a navigation radar malfunction in a timely manner is considered an important element of risk management in navigation. Delays in detecting such problems can lead to misjudgment of the traffic situation and increase the risk of unsafe passages between vessels or other navigational errors. For this reason, the detection time of a radar fault is a relevant indicator of the monitoring of navigation systems and the ability of participants to recognize the degradation of key navigational instruments [115].

For BN modelling, the Radar Failure Detection Time variable was discretized into three states: < 2 min, representing participants who detected the radar failure within two minutes; > 2 min, representing participants who detected the failure after more than two minutes; and Not detected, representing participants who did not detect the failure during the simulation.

Based on the collected data from 65 participants, the distribution across the defined states was:

- < 2 min: 15 participants (23%)
- 2 min: 17 participants (26%)
- not detected: 33 participants (51%)

The high proportion of participants in the not detected category (51%) suggests limited monitoring effectiveness and situational awareness under the simulation conditions. This finding may be partly explained by the composition of the sample, which consisted predominantly of students with relatively limited practical experience in recognizing system failures.

In the BN structure, the Radar Error variable is defined as an input (root) variable, meaning there are no parent variables. Its values are determined directly based on the observed behavior of participants during the simulation, according to the time required to recognize a radar system fault after its occurrence.

6.4.2. Global Positioning System GPS Error

The GPS Error variable describes participants' ability to promptly recognize the loss of GPS signal during a simulation exercise. In the simulation scenario, the loss of GPS signal was deliberately simulated to assess participants' situational awareness and their ability to notice changes in the availability of navigation information.

In navigation practice, Global Navigation Satellite Systems (GNSS) are among the primary sources of positional data, but their availability can be limited by various technical problems, signal degradation, or interference with signal reception [116]. The literature also highlights that interference and interruptions of GNSS signals, including intentional interference, can pose a significant safety risk to maritime navigation [117]. In such situations, the navigator must promptly recognize the loss of signal and rely on other available navigation sources, such as radar data, visual landmarks, or electronic navigation charts.

For BN modelling, the GPS Signal Loss Detection Time variable was discretized into three states: < 30 s, representing participants who detected the GPS signal loss within 30 seconds; 30–60 s, representing participants who detected the signal loss between 30 and 60 seconds; and > 60 s, representing participants who detected the signal loss after more than 60 seconds.

Based on the collected data from 65 participants, the distribution across the defined states was:

- < 30 s: 24 participants (37%)
- 30–60 s: 24 participants (37%)
- >60 s: 17 participants (26%).

This categorization allows quantification of participants' reaction speed and integration of the results into a BN model for assessing operational efficiency during autonomous ship control.

Within the BN, this variable is defined as an independent variable, with no parent nodes in the model. The values of the variable are determined directly from the data collected during the simulation experiment, where the detection time of GPS signal loss was measured for each participant during the navigation exercise. The empirical data obtained are used as input values for the variable in the BN.

6.4.3. Response to Navigation System Failures

The variable Response to Navigation System Failures measures the operator's effectiveness in identifying and reacting to critical degradations in key navigational systems during the simulator-based exercise. It reflects the participant's ability to detect and respond appropriately to malfunctions in radar and GPS, which were deliberately introduced during the scenario to replicate realistic operational disturbances encountered in maritime navigation.

Within the model, Response to Navigation System Failures is defined as a child node with two parent variables:

- Radar Error,
- GPS Error.

These indicators capture the time required for participants to recognize system anomalies, a critical component of safe navigation and operational decision-making. The ability to detect and correctly interpret such failures plays a fundamental role in preventing operational breakdowns and ensuring navigational safety, particularly in complex maritime environments where delayed or missed detection may result in serious consequences [118].

According to expert judgement, the parent variables were assigned unequal importance in the construction of the CPT. Radar Error Detection Time was considered the dominant factor, accounting for approximately 60% of the overall influence, while GPS Error Detection Time contributed approximately 40%. This weighting reflects the greater importance of radar monitoring for maintaining situational awareness and collision avoidance in maritime operations.

For modelling purposes, the variable was discretized into three qualitative states. Good, moderate, and poor, representing different levels of response effectiveness. The CPT was constructed using expert reasoning, guided by the principle that rapid detection of both system failures results in a high probability of effective response, whereas delayed detection or failure to detect significantly reduces response quality.

Special emphasis was placed on radar failure detection, which was treated as a dominant factor within this subnetwork due to its critical role in maintaining situational awareness and collision avoidance. Consequently, scenarios in which radar failure was not detected are associated with a high probability of the poor state, regardless of GPS performance. This modelling approach ensures that the CPT structure reflects realistic operational priorities and the relative importance of different navigational systems.

The Response to Navigation System Failures variable, derived from the combination of two parent variables, was distributed as follows:

- good (27%),
- moderate (30%),
- poor (43%).

Within the BN this variable is modelled as a child node of Radar Error and GPS Error. It directly depends on these two parent variables and represents their combined effect on operator performance in failure situations.

6.4.4. Navigation Complexity

The variable Navigation Complexity represents the overall operational difficulty of the navigational environment in which the participant undertakes the simulator-based task. It reflects the combined influence of several situational factors, including traffic density, spatial constraints of the navigational area, proximity to coastal structures, and the requirement to monitor and interpret multiple sources of navigational information simultaneously, such as radar, ECDIS, and visual inputs. The introduction of system disturbances during the simulation increases task demands and operational uncertainty.

In simulator-based studies, task complexity is consistently linked to increased mental workload, especially in environments that require continuous monitoring, processing information from multiple sources, and making time-critical decisions. As concurrent demands rise, operators must divide their attention across competing tasks, resulting in reduced performance and greater cognitive strain. Empirical research using eye-tracking and simulator-based navigation tasks demonstrates that higher task complexity is evident in altered visual attention patterns and increased indicators of cognitive load [119].

Following the expert-based definition methodology described at the beginning of this chapter, the prior probability distribution of this variable was defined based on the characteristics of the simulation scenario. The resulting prior distribution was:

- low (25%)
- moderate (50%)
- high (25%)

reflecting a balanced operational context in which moderately complex situations are most frequent, while still accounting for both simpler and more demanding conditions.

Navigation Complexity is defined as a root (contextual) variable without parent nodes.

6.4.5. Navigation Control

The variable Navigation Control represents a central performance construct within the BN model, capturing the operator’s ability to maintain effective and stable control of the vessel during the simulator-based navigation task. Unlike outcome-oriented variables, this construct reflects the process-level operational capability of the participant, integrating cognitive, behavioral, and situational influences into a unified probabilistic representation. In the context of MASS, Navigation Control is particularly critical, as operators must supervise complex systems, respond to unexpected failures, and maintain situational awareness under dynamically changing conditions [120].

Within the model, Response to Navigation Control is defined as a child node with four parent variables:

- VAQ,
- Response to Navigation System Failures,
- Cognitive Overload,
- Navigation Complexity.

The three primary variables (VAQ, Response to Navigation System Failures, and Cognitive Overload) were assigned equal weights because the expert group considered them to represent complementary and equally important dimensions of Navigation Control. VAQ reflects the operator's ability to acquire and process navigational information, Response to Navigation System Failures represents the capability to recognize and manage critical system degradations, and Cognitive Overload captures the operator's internal cognitive capacity to maintain effective performance under demanding conditions. Navigation Complexity was assigned a lower weight (10%) because it acts as a contextual factor that influences Navigation Control indirectly by increasing task demands, rather than directly reflecting the operator's performance.

Within the model, Navigation Control is defined as a three-state variable. The prior probability distribution of the variable is determined by the BN structure and the expert-defined CPT, resulting in the following state distribution:

- low (35%),
- moderate (33%),
- high (32%).

This definition allows a clear distinction between inadequate, acceptable, and effective levels of control, while maintaining a balance between interpretability and computational tractability of the CPT. The three-level structure is consistent with common practice in human performance modelling, where excessive granularity can reduce model robustness without significantly improving explanatory power.

The probabilistic structure of the CPT follows logically consistent pattern. High levels of VAQ and effective responses to navigation system failures increase the likelihood of achieving high Navigation Control. In contrast, elevated Cognitive Overload significantly reduces this likelihood. Navigation Complexity was defined as a secondary factor. Its role is to slightly modify the outcome, but not to determine it. Final result is primarily driven by the three dominant variables. Navigation Complexity can influence the probability distribution, but cannot override their combined effect. For example, even under favourable conditions (high attention quality, effective response, and low overload), increased complexity results only in a moderate reduction in the probability of high Navigation Control, rather than a complete shift towards lower performance states. Conversely, in adverse conditions, low complexity does not substantially compensate for poor operator performance.

The validity of the CPT was verified through scenario-based sensitivity testing in the GeNIe software environment. Extreme cases and balanced configurations were evaluated to ensure logical and stable behavior of the variable. The results confirmed that the variable Navigation Control responds consistently to changes in input conditions, demonstrating the robustness and interpretability of the modelled relationships.

6.4.6. Voyage Execution

The variable Voyage Execution represents an integrated measure of the operator's overall effectiveness in performing the simulator-based navigation task. Unlike process-oriented variables that describe specific aspects of behavior, such as visual attention or physiological response, this variable captures the final quality of task execution, reflecting how successfully the participant managed the navigation scenario under dynamically changing conditions. In MASS operations, Voyage Execution is particularly important as it encapsulates the operator's ability to maintain control, adapt to system disturbances, and achieve operational objectives within a constrained, safety critical environment. This aligns with simulator-based maritime research linking ship handling, navigation performance, and decision-making under stressful conditions [121].

Within the model, Voyage Execution is defined as a child node with three parent variables:

- Sea-Going Experience,
- Navigation Control,

- Navigation Complexity.

These variables were selected to represent the key dimensions influencing execution performance like Operator Background, real time operational capability, and environmental or task-related difficulty. CPT was constructed using a structured expert-based approach, with predefined relative weights assigned to each parent variable to reflect their influence on the outcome. Navigation Control was assigned the dominant influence (60%), as it directly captures the operator's ability to maintain stable vessel control and respond effectively to navigational challenges. Sea-Going Experience was assigned a secondary influence (25%), reflecting the role of accumulated knowledge and practical familiarity in supporting decision-making and execution. Navigation Complexity was assigned a corrective influence (15%), representing the impact of external task demands and environmental difficulty, while intentionally limiting its ability to independently determine the outcome.

In the BN, Voyage Execution is defined as a three-state variable:

- low (48%),
- moderate (33%),
- high (19%).

This discretization enables a clear and interpretable distinction between unsuccessful, partially successful, and successful execution outcomes, while maintaining a manageable level of model complexity. The classification reflects overall performance quality, with the low state corresponding to inadequate task execution with significant deficiencies, the moderate state indicating acceptable but suboptimal performance, and the high state representing effective and stable task completion.

The CPT was constructed to ensure logical consistency with this weighting structure. To ensure the robustness and validity of the CPT, the variable was tested using multiple scenario-based configurations of input variables. These tests confirmed that the model behaves consistently with the predefined influence structure: Navigation Control is the primary determinant of execution quality, experience provides a stabilizing effect, and complexity introduces controlled variability. This validation approach supports the internal consistency of the model and enhances the interpretability of the Voyage Execution variable within the broader Bayesian framework.

6.4.7. Route Adherence

The variable Route Adherence indicates whether the participant-maintained navigation along the route during the simulation exercise. In the experimental exercise, participants were required to navigate the ship from the starting position to the assigned destination by following a predefined route displayed on the navigation system.

Keeping the ship on the intended route is an important indicator of navigational performance, particularly in coastal navigation, where traffic density, navigational constraints, and limited maneuvering space require continuous monitoring of the ship's position relative to the planned track. Effective route monitoring demonstrates the navigator's situational awareness and ability to interpret available navigational information correctly and apply appropriate course adjustments during navigation. Research on ship track-keeping and route monitoring shows that maintaining a ship on the planned navigation track is a fundamental element of safe navigation and operational control during navigational tasks [122].

In this dissertation, Route Adherence was assessed by analysing the vessel's track recorded during the simulator exercise. Participants were evaluated on whether they maintained navigation along the planned route during the simulation.

Based on the collected data from 65 participants, the distribution across the defined states was:

- route followed: 34 participants (52%),
- route not followed: 31 participants (48%).

Within the BN model, Route Adherence is defined as an external variable without parent nodes.

6.4.8. Arrival Time to Assigned Position

The variable Arrival Time to Assigned Position is a performance indicator in the simulator exercise. The primary objective of the navigation task was for participants to safely navigate the ship to a predefined assigned position within the simulated environment. Therefore, successful arrival at the assigned position, as well as the time taken to reach it, provides a significant indicator of overall navigational performance.

During the simulation exercise, participants were required to manage ship navigation while simultaneously dealing with gradually introduced equipment failures and operational challenges. Under these conditions, the ability to complete the voyage and reach the assigned position reflects several important competencies, including situational awareness, decision-making under uncertainty, and the ability to maintain safe ship control despite system degradations. Findings from the study by Gould *et al.*, based on maritime simulator experiments, show that performance during simulated navigation tasks can serve as an important indicator of operator effectiveness and the ability to manage complex navigational situations [123]. A similar simulator-based study by Múller-Plath *et al.* [124], analysing the interaction between navigators and automated navigation systems, also uses task completion and operational performance as key measures of navigational capability in controlled experimental environment.

For each participant, the total time required to reach the predefined assigned position was measured from the beginning of the exercise until the vessel arrived at the designated location. Based on the recorded time on the collected data from 65 participants, the distribution across the defined states was:

- < 18 min: 19 participants (29%),
- 18–22 min: 37 participants (57%),
- > 22 min: 6 participants (9%),
- Not arrived:3 participants (5%).

These categories allow the model to distinguish between efficient task completion, delayed arrival, and unsuccessful completion of the navigation task. The inclusion of the Not arrived category is particularly important because it captures situations in which participants were unable to reach the assigned position during the simulation exercise.

These results indicate that most participants were able to complete the navigational task within the expected time frame, although some required additional time, reflecting differences in performance under simulated operational conditions. The cases in which participants failed to reach the assigned position represent instances of reduced Operational Effectiveness and highlight the challenges associated with managing complex navigation scenarios and system failures.

Within the BN model, Arrival Time to Assigned Position is treated as an observed performance variable derived directly from simulator data and represents one of the key indicators used to assess participants' operational performance during the experiment.

6.4.9. Near Miss

The Near Miss variable is an objective safety indicator of participants' navigation performance during the autonomous ship control simulation exercise. In maritime safety practice, a near miss is defined as a sequence of events or conditions that could have resulted in loss, but did not, due to a break in the chain of events, as described in IMO guidance on near miss reporting [125]. Such potential losses may include human injury, environmental damage, or negative operational or economic consequences.

In the BN model, the Near Miss variable is treated as an observed performance variable derived directly from simulator data and serves as a key indicator for assessing participants' navigational safety and decision-making performance under increased operational complexity.

This general definition does not prescribe a universal numerical threshold for minimum Closest Point of Approach (min CPA), as safe passing distances depend on the navigational context, including traffic density, environmental conditions, vessel characteristics, and maneuvering constraints. In maritime risk analysis, CPA-based indicators are widely used to assess collision risk and safety margins between ships [126].

For this dissertation, a Near Miss event was operationally defined as a critical approach to another ship or object in which the minimum passage distance between the own ship and the other object minimum CPA was equal to or less than 0.05 NM. This distance, approximately 92.6 meters, was adopted as the critical safety distance for the defined simulation scenario, which was characterized by increased traffic density and limited maneuvering space.

Such events are not considered random but are interpreted as indicators of a reduced safety margin, potentially untimely reactions, incorrect assessment of the traffic situation, or insufficient situational awareness. Data on the minimum CPA were obtained from the simulator log (radar/AIS). The analysis was conducted at the level of a single encounter, with each encounter meeting the defined threshold recorded as a single Near Miss event.

Based on the recorded time on the collected data from 65 participants, the distribution across the defined states was:

- no near miss: 9 participants (14%),
- one near miss: 26 participants (40.0%),

- two or more near miss: 30 participants (46%).

Within the BN model, the Near Miss variable is treated as an observed performance variable derived directly from simulator data and represents a key indicator used to assess participants' navigational safety and decision-making performance under conditions of increased operational complexity.

6.4.10. Exercise Performance

The variable Exercise Performance represents an aggregated measure of participant success in the simulator-based navigation task. It is designed to capture the overall outcome of the exercise by integrating safety-related events, procedural compliance, and task completion efficiency into a single variable. Unlike process-oriented variables that describe how the task was performed, this variable reflects what was ultimately achieved during the simulation, making it a outcome indicator within the BN model.

Within the model, Response to Navigation System Failures is defined as a child node with three parent variables:

- Near Miss,
- Route Adherence,
- Arrival Time to the Assigned Position.

The CPT was constructed using a structured expert-based weighting approach, combined with logical consistency rules derived from the operational characteristics of the exercise. A dominant influence was assigned to Near Miss events (60%), reflecting their direct relationship with collision risk and critical decision-making failures. Route Adherence and Arrival Time were assigned equal secondary influence (20% each), as they represent important but less critical aspects of performance.

These variables were selected to represent the most relevant and observable aspects of performance in a controlled maritime simulation environment Near Miss events provide a direct proxy for navigational safety, Route Adherence reflects compliance with planned operational procedures, and arrival time captures the efficiency and temporal accuracy of task execution, which is consistent with established approaches to human factor-based performance assessment in maritime safety [127].

Following the expert-based definition methodology described at the beginning of this chapter, the prior probability distribution of this variable was defined based on the characteristics of the simulation scenario. The resulting prior distribution was:

- good (16%),
- moderate (28%),
- low (56%).

Specific logical constraints were incorporated into the CPT to ensure realistic model behavior. In particular, the occurrence of multiple Near Miss events systematically drives the variable towards the low-performance state, regardless of other conditions. Similarly, failure to reach the assigned position (not arrived) is treated as a near-failure condition, resulting in a dominant probability of low-performance. These rules ensure that the model prioritizes safety and task

Within the BN, Exercise Performance serves as an intermediate outcome variable linking lower-level behavioral indicators to higher-level constructs such as Operational Effectiveness. By consolidating multiple aspects of task performance into a single node, it reduces model complexity while preserving interpretability, thus supporting analytical clarity and practical applicability in assessing operator performance in autonomous maritime operations.

6.5. Operational Effectiveness (Output Variable)

The variable Operational Effectiveness represents the final outcome construct within the BN model and is intended to capture the overall quality of participant performance in the simulator-based navigation exercise. Unlike intermediate variables that describe specific aspects of operator state or behavior during the task, this variable provides a higher-level assessment of how successfully the participant managed the scenario as a whole. In the context of MASS, this type of outcome variable is particularly relevant because operator performance is no longer reflected only through isolated technical actions, but through the ability to maintain control, make appropriate decisions, and ensure safe task execution under conditions of uncertainty and system disturbance. This understanding aligns with research performed by Lynch, *et al.*, which emphasises that decision-making in MASS operations must be viewed as a multidimensional human–machine performance process rather than a single isolated action [128].

In this dissertation, Operational Effectiveness is defined as a composite outcome of decision-making quality and safety-related performance indicator. Its role is to summarize whether the participant was able to respond adequately to navigational challenges, keep operational stability, and complete the assigned task without major degradation in performance. As a composite variable within the BN, Operational Effectiveness integrates the combined probabilistic influence of its parent variables through the corresponding CPT. Rather than relying on predefined weighting coefficients, the relative contribution of each parent variable is determined by the BN structure and conditional probability relationships. The influence of individual parent variables on Operational Effectiveness is quantitatively evaluated through the sensitivity analysis presented in the model validation chapter. For this reason, this variable is positioned as the final outcome node of the model, where the combined effects of lower-level cognitive, behavioral, and navigational influences are integrated into a single interpretable representation of overall performance.

Following the expert-based definition methodology described at the beginning of this chapter, the prior probability distribution of this variable was defined based on the characteristics of the simulation scenario. The resulting prior distribution was:

- poor (43%),
- moderate (35%),
- high (22%).

This three-state structure was chosen to preserve interpretability while maintaining a manageable level of model complexity. The poor state corresponds to unsatisfactory operational performance, indicating that the participant showed considerable difficulty in maintaining effective task execution, preserving decision quality, or ensuring sufficient safety margins during the simulation. The moderate state reflects partially satisfactory performance, where the scenario was managed with some limitations, inconsistencies, or reduced operational stability. The high state represents effective and stable performance, characterized by appropriate decisions, maintained control, and successful execution of the navigational task under the given experimental conditions.

This suggests that the simulation scenario was sufficiently demanding to result in a predominance of less favourable overall outcomes, consistent with the presence of dynamic navigational demands, failure-related disturbances, and varying levels of participant preparedness. However, the proportions of the moderate and high states indicate that a significant part of the sample was still able to maintain acceptable or effective operational performance despite these constraints.

Operational Effectiveness should be considered the most comprehensive performance indicator within the BN structure. It does not refer to a single isolated response, but to the overall operational outcome of the participant's decision-making quality and safety-related behavior throughout the entire exercise. In this sense, it provides the clearest probabilistic representation of whether the participant was able to perform the assigned MASS-related task effectively under simulated operational pressure.

7. BAYESIAN NETWORK MODEL FOR MASS OPERATOR ASSESSMENT

This chapter introduces the BN model developed for assessing operator performance in MASS operations. It outlines the overall structure of the model and the relationships between the key variables influencing Operational Effectiveness.

7.1. Qualitative Model Structure

The qualitative structure of the BN model describes the relationships between key factors influencing operator performance in MASS operations. The model is designed to capture interactions among human, physiological, perceptual, and operational variables, providing a structured representation of decision-making processes under dynamic and uncertain conditions. In this context, the BN approach enables the representation of conditional dependencies and uncertainty within a complex system of interacting factors [129].

The model is organized into interconnected levels, reflecting different aspects of operator behavior and system interaction. At the initial level, the structure includes variables describing the operator's background and preparedness, such as Maritime Education Level, Operational Rank, Operational Self-Efficacy, MASS Familiarity, and Sea-going Experience. These variables define the baseline conditions under which the operator enters the simulator-based navigation task and form the basis for higher-level readiness-related constructs.

The second level of the model incorporates physiological and perceptual variables, reflecting the internal state of the operator during task execution. Physiological indicators, including heart rate, blood pressure reactivity, and temperature change, are combined into a broader representation of physiological stress. Visual attention indicators derived from eye-tracking data contribute to the description of attentional behavior and cognitive processing. These variables are not treated as isolated measures but as functionally connected elements that jointly influence the operator's ability to maintain effective situational control under demanding operational conditions.

At the operational level, the model includes variables related to Navigation Complexity, system failure response, and directly observed navigational performance. These influences are further integrated through intermediate variables such as Cognitive Overload, Navigation Control, Voyage Execution, and Exercise Performance. Qualitatively, these variables represent the core functional layer of the model, as they connect lower-level inputs with higher-level performance consequences.

A central role within the model is assigned to intermediate variables that act as integration points between multiple sources of influence. Cognitive Overload represents the combined effect of perceptual, physiological, and contextual pressures on the operator, while Navigation Control reflects the ability to maintain effective vessel control during the exercise. Voyage Execution and Exercise Performance capture broader aspects of task completion and overall behavioral effectiveness. Through these intermediate variables, the model represents both direct and indirect pathways leading to successful or degraded operator performance.

All causal pathways ultimately converge on the final outcome variable, Operational Effectiveness, which represents the overall assessment of operator performance in the context of MASS operations. This variable integrates multiple dimensions of performance, safety, and decision-making quality and serves as the final output of the BN model.

The model was intentionally structured to reflect realistic cause–effect relationships expected in simulator-based MASS operations. Directional links between variables were established using expert knowledge, the conceptual design of the simulation exercise, and evidence from the literature describing relationships among human, technical, and operational factors. This hierarchical organization enables the model to capture the complexity of operator performance without reducing it to a single isolated indicator.

The qualitative structure of the BN model was developed through an iterative process that combined evidence from the literature, expert knowledge, and the conceptual design of the simulation exercise. The initial network structure was established based on identified relationships among human, technical, and operational factors. The proposed dependencies and hierarchical organization were then refined through expert consultation and consistency analysis. This process produced a structured qualitative model representing the conditional relationships among the variables included in the BN (Figure 7.1).

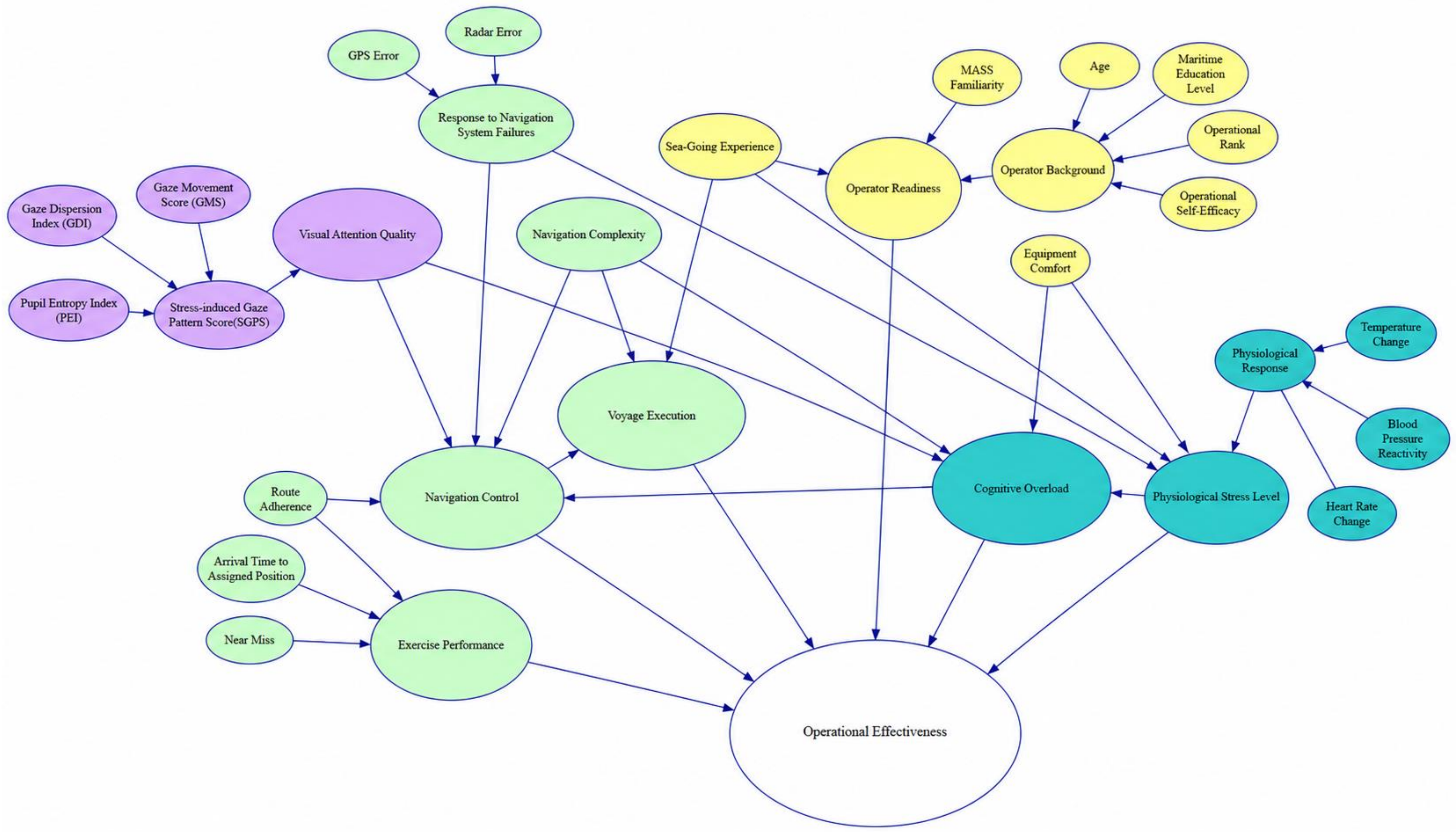

Figure 7.1 Qualitative structure of the BN model for MASS operator assessment (GeNIe-author)

The variable Perceived Primary MASS Risk, although initially considered during the conceptual design phase, was not included in the final model structure. This decision was based on its limited contribution to the model's explanatory power and the lack of a clear, consistent influence on key performance-related variables. The variable reflects subjective perceptions rather than observable behavior or measurable performance, making it less suitable for a model focused on Operational Effectiveness. Its exclusion therefore contributes to a clearer, more consistent, and more interpretable structure.

7.2. Quantitative Model Specification

The quantitative specification of the BN model defines how the relationships in the qualitative structure are represented using probabilities. While the qualitative model shows how variables are connected, the quantitative model assigns numerical values that determine the strength and behavior of these relationships. This step is the core computational component of the model, enabling the propagation of uncertainty and the calculation of outcome probabilities. BN are typically defined by a combination of a graphical structure and a set of conditional probability distributions that govern the relationships between variables.

Most variables in the model are defined as discrete nodes with a limited number of states, usually grouped into three categories (e.g., low, moderate, high). This discretization was intentionally applied to balance model interpretability and computational efficiency. Some variables are defined using specific state structures based on the nature of the observed data, such as categorical outcomes or time-based classifications. This approach ensures that the model accurately reflects real operational conditions while remaining easy to interpret and computationally manageable.

The relationships between variables are quantified using CPTs, which define the probability distribution of each child node given all possible combinations of its parent nodes. For each node, the CPT provides a complete set of probabilistic rules governing how input conditions are transformed into output states. In this model, CPTs were constructed using a structured expert-based approach, supported by empirical distributions obtained from the simulator-based experiment.

For variables directly derived from collected data, such as Near Miss occurrence, Arrival Time at the assigned position, and detection of navigation system failures, prior probability distributions were defined based on observed frequencies. This ensured that the model reflects realistic behavioral patterns observed during the exercise. These empirical inputs form the foundation of the quantitative model and provide a data-driven basis for further probabilistic reasoning.

For intermediate variables, such as Cognitive Overload, Navigation Control, Voyage Execution, and Exercise Performance, the CPTs were defined using a weighted influence logic. In this approach, parent variables were assigned relative importance based on their conceptual relevance and expected impact on the child variable. The resulting probability distributions were constructed to ensure logical consistency, so that favourable input conditions increase the probability of favourable outcomes, while adverse conditions shift the distribution towards lower performance states.

Special attention was given to maintaining monotonic behavior within CPTs, ensuring that the model responds predictably to changes in input variables. This ensures that improvements in key contributing factors, such as VAQ or System Failure Response, systematically lead to increased probabilities in higher-level performance variables. Conversely, deteriorations in these inputs result in a consistent shift towards lower outcome states. This property is essential for preserving interpretability and avoiding contradictory model behavior, and is commonly emphasised in BN parameterization approaches [130].

The final outcome variable, Operational Effectiveness, integrates the influence of several higher-level performance indicators, including Navigation Control, Voyage Execution, Exercise Performance, and Operator Readiness, each representing a complementary aspect of operator performance. Navigation Control reflects the participant's ability to maintain safe and accurate navigation under varying operational conditions, while Voyage Execution assesses the successful completion of the assigned navigational task according to the planned route and operational objectives. Exercise Performance reflects the overall quality of task execution by integrating safety-related events, procedural compliance, and operational success. Operator Readiness indicates the participant's preparedness to respond effectively to operational challenges. By combining these variables within a single CPT, the Operational Effectiveness node offers a comprehensive assessment of overall operator performance, capturing the combined influence of navigational, behavioral, and operational factors rather than relying on a single performance indicator. The quantitative definition of this variable ensures that it reflects not only isolated performance aspects but also their combined effect within a complex operational context.

The model was implemented in the GeNIe software environment, where all CPTs were defined and verified. The implementation allowed systematic testing of model behavior under different input conditions, ensuring that the probabilistic relationships are internally consistent and aligned with the intended logic of the model. During this process, multiple validation checks were performed to confirm that the model produces realistic and interpretable outputs across a wide range of scenarios.

Overall, the quantitative model specification transforms the conceptual framework into a functional probabilistic system. By combining empirical data with structured expert knowledge, the model is capable of representing complex interactions between variables and generating meaningful predictions of operator performance in MASS operations.

The complete quantitative specification of the model, including the defined probability distributions for all variables, is presented in Figure 7.2.

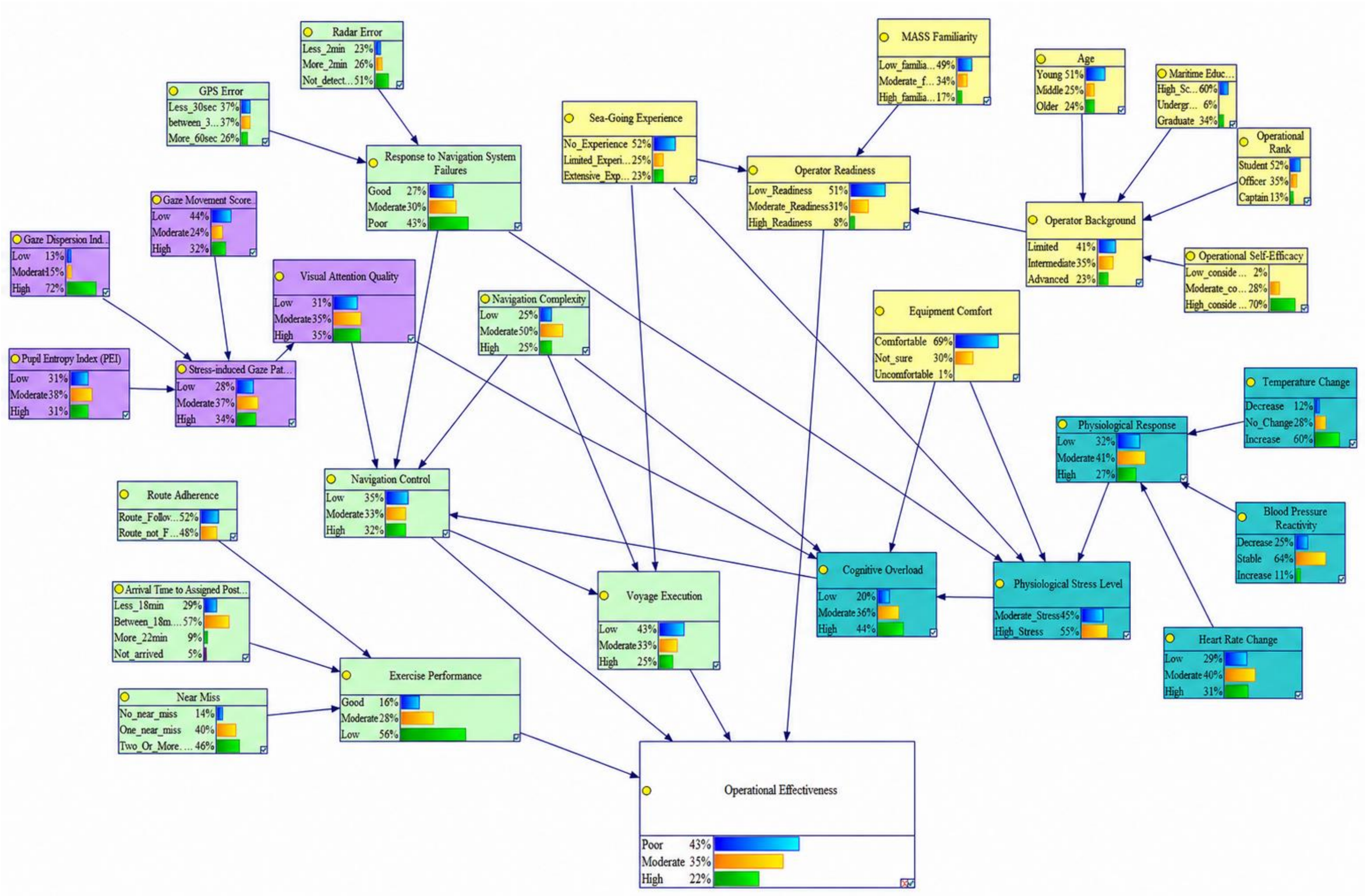

Figure 7.2 Quantitative specification of the BN model with prior probability distributions (GeNIe-author)

7.3. Model Validation

Following the qualitative and quantitative specification of the BN model presented in Sections 7.1 and 7.2, the final stage of model development focuses on validating its internal consistency and robustness. In this dissertation, validation is carried out through sensitivity analysis, which provides a systematic approach to examining how the model behaves under controlled variation of inputs and parameters. This analysis is widely used in BN validation [131]. The purpose of this section is not to interpret the model's outcomes, but to assess whether the structure and parameterization produce coherent, stable, and logically consistent responses.

Figure 7.3 shows the final BN configuration used for validation. The intensity of the red shading reflects the relative dominance of particular states under the current model configuration, while lighter tones indicate more balanced probability distributions. Although this visual cue is not used as a formal analytical measure, it provides an intuitive indication of how probability mass is distributed across the network. The observed pattern suggests that the propagation of probabilities follows the expected direction from lower-level variables to intermediate constructs and finally to the outcome node. From a validation perspective, this indicates that the model behaves consistently and that the relationships between variables are logically structured.

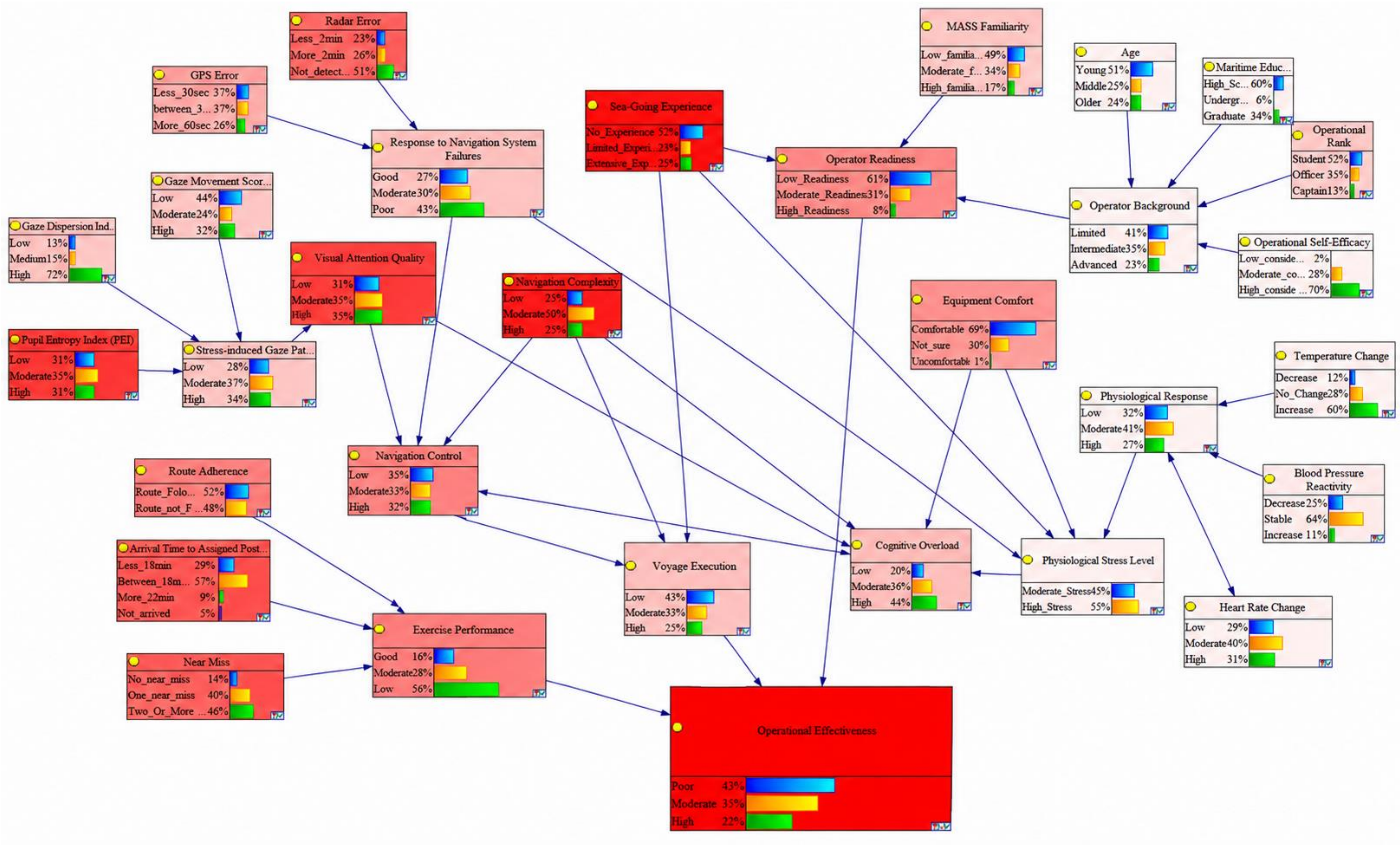

Figure 7.3 Final BN model used for sensitivity analysis (GeNIe-author)

To complement the visual inspection of the BN structure, sensitivity analysis was first applied to the Cognitive Overload variable. As one of the principal intermediate variables within the BN, Cognitive Overload represents the central mechanism through which physiological, visual attention, technical, and operational factors interact before influencing higher-level performance variables. Because the objective was to identify the factors contributing to a low level of Cognitive Overload, the sensitivity ranking was based on the Cognitive Overload = low state, while the corresponding influence on the high state was also calculated to provide a balanced assessment of parameter importance. Table 7-1 presents the five most influential variables affecting Cognitive Overload, expressed as approximate ranges of probability change (ΔP). The ranking confirms that the node is primarily influenced by variables directly associated with cognitive processes, supporting the logical consistency of the BN structure and the validity of the probabilistic relationships within this intermediate layer of the model.

Table 7-1. Combined sensitivity ranking of the five most influential variables affecting Cognitive Overload

Rank	Variable	ΔP (Poor)	ΔP (High)	Average Influence (ΔP)
1	Visual Attention Quality (VAQ)	0.010	0.013	0.0115
2	Equipment Comfort	0.008	0.010	0.0090
3	Physiological Stress Level	0.006	0.009	0.0075
4	Pupil Entropy Index (PEI)	0.005	0.007	0.0060
5	Navigation Complexity	0.005	0.006	0.005

Following the validation of the Cognitive Overload node, the same sensitivity analysis procedure was applied to Operational Effectiveness, which represents the final output variable of the proposed BN model. Unlike Cognitive Overload, which reflects an intermediate cognitive mechanism, Operational Effectiveness integrates the combined influence of all preceding variables and therefore provides an overall assessment of operator performance. Table 7-2 ranks the ten variables with the greatest influence on Operational Effectiveness, expressed as approximate probability change ranges (ΔP) representing the change in outcome probability in response to variations in input variables derived from the tornado diagrams. The table serves as a consistency check, confirming that the model is driven by variables logically positioned within the network rather than by peripheral or unintended elements. This verification is particularly important in models integrating multiple levels of abstraction, as it ensures that the flow of influence aligns with the intended design of the BN.

Table 7-2. Combined sensitivity ranking of the ten most influential variables affecting Operational Effectiveness

Rank	Variable	ΔP (Poor)	ΔP (High)	Average Influence (ΔP)
1	Navigation Control	0.015	0.005	0.0100
2	Exercise Performance	0.013	0.004	0.0085
3	Near Miss	0.010	0.007	0.0085
4	Voyage Execution	0.012	0.004	0.0080
5	Sea-Going Experience	0.008	0.006	0.0070
6	Operator Readiness	0.007	0.003	0.0050
7	Visual Attention Quality (VAQ)	0.006	0.004	0.0050
8	Arrival Time to Assigned Position	0.004	0.005	0.0045
9	Physiological Stress Level	0.004	0.003	0.0035
10	Cognitive Overload	0.003	0.003	0.0030

The ΔP values presented in Table 7-1 were obtained directly from the Sensitivity to Findings analysis implemented in the GeNIe software. Each ΔP value represents the change in the predicted probability of the selected Operational Effectiveness state resulting from variations in an individual input variable propagated throughout the BN. Although Navigation Control achieved the highest ΔP value, the differences among the first-ranked variables are relatively small, indicating that they represent a group of comparably influential factors rather than distinctly separated levels of influence.

The behavioral properties of the model are further examined using a tornado diagram representing sensitivity to results a commonly used graphical tool for sensitivity analysis and ranking variable influence [132]. Figure 7.4 shows how variations in influential variable states affect the probability range of the selected target outcome, Operational Effectiveness (poor). The diagram shows that the strongest influence comes from variables related to Navigation Control and Exercise Performance, which are positioned close to the outcome node. This suggests that the model responds primarily to the most relevant parts of the network. Meanwhile, the overall structure remains stable, with no unexpected changes in the pattern of influence. The diagram confirms that the model behaves consistently and that its structure is logically sound.

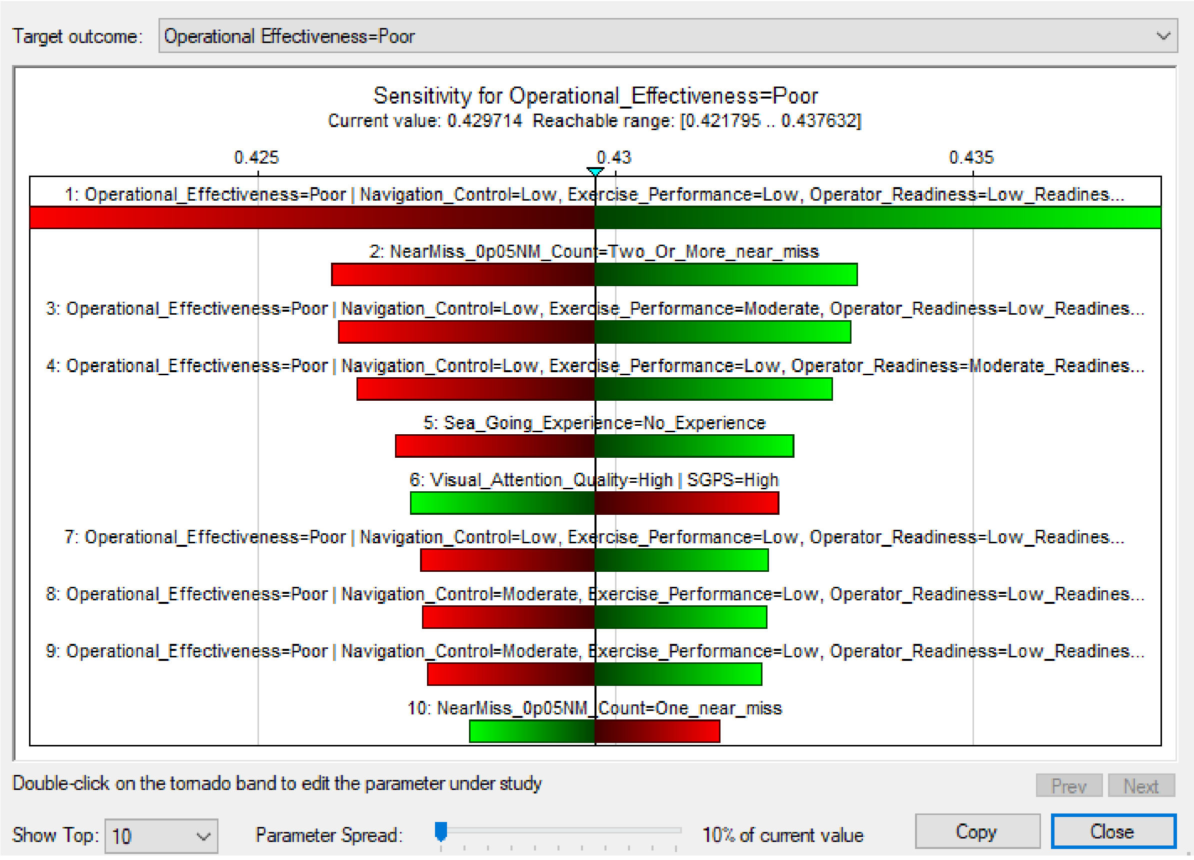

Figure 7.4 Tornado diagram showing sensitivity of Operational Effectiveness (Poor state) to variations in influential variable states (GeNIe-author)

Local sensitivity to results alone is not sufficient to demonstrate robustness regarding uncertainty in parameter values. Therefore, an additional sensitivity-to-parameters analysis was conducted Figure 7.5 shows the tornado diagram generated under a ±20% parameter perturbation, representing a deliberate variation of CPT values. This step evaluates whether the model maintains stable behavior when its numerical specification is subject to moderate uncertainty. Such testing is particularly relevant in BN models that incorporate expert-informed parameterization, as it provides a means of assessing whether the observed behavior depends on precise numerical calibration.

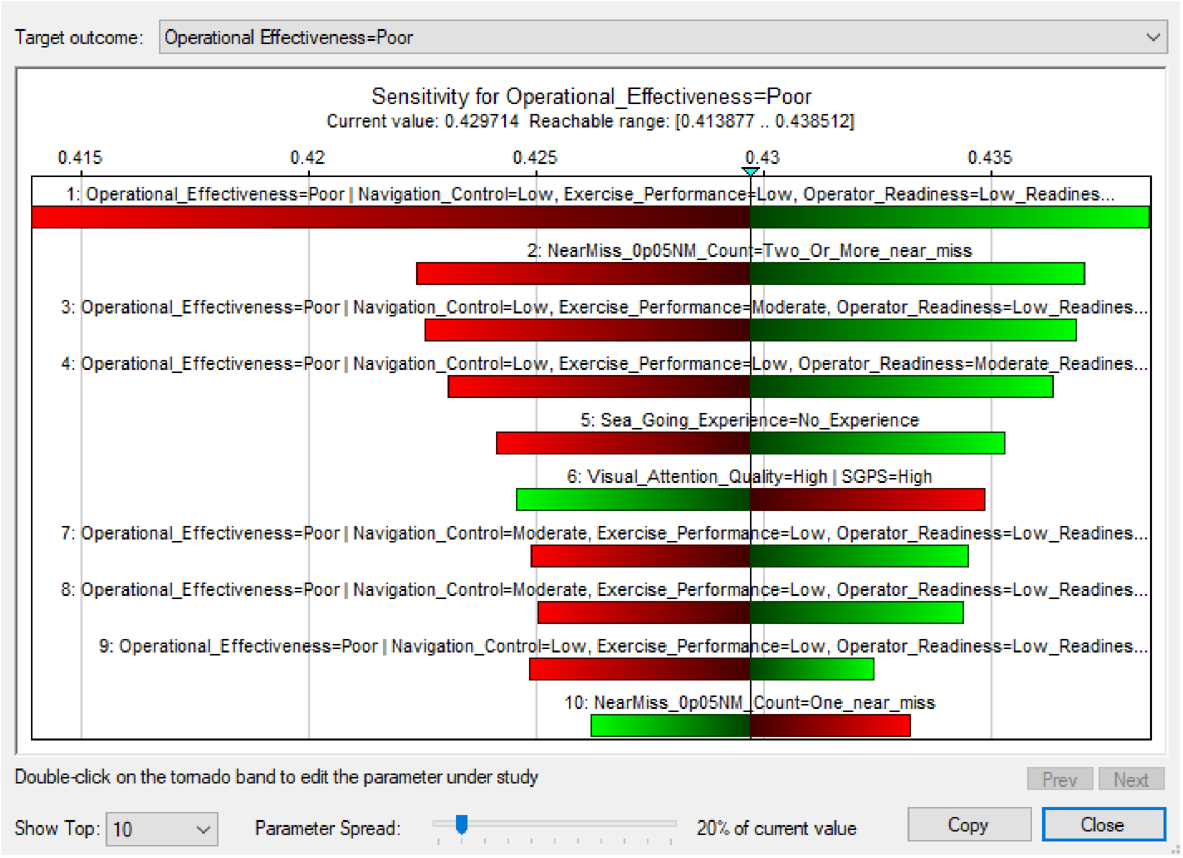

Figure 7.5 Tornado diagram showing sensitivity of Operational Effectiveness (Poor state) under ±20% parameter perturbation (GeNIe-author)

The comparison between the baseline sensitivity analysis and the ±20% perturbation scenario enables a direct assessment of model robustness. From a validation perspective, the preservation of a consistent influence structure under increased parameter variation indicates that the model is not sensitive to specific probability values. Instead, its behavior is governed primarily by the topology of the network and the logical dependencies defined between variables. This property is essential for ensuring that the model can be used as a reliable analytical framework, particularly in complex domains such as MASS operations, where uncertainty and variability are characteristic of the system.

The sensitivity analysis performed in this section confirms that the BN model exhibits coherent and stable behavior under both variation of input states and controlled perturbation of parameters. The combination of visual inspection, quantitative ranking, and graphical sensitivity assessment provides a comprehensive validation approach, demonstrating that the model is internally consistent and structurally robust. This establishes sufficient confidence to support the use of the model in the subsequent analysis of results, where its outputs will be interpreted in the context of decision-making processes in autonomous maritime operations.

8. RESULTS AND DISCUSSION

Building on the development and validation of the BN model presented in Chapter 7, this chapter presents the results obtained by integrating simulation-based experimental data with probabilistic modelling, followed by a discussion of their implications for operator performance in autonomous ship operations. The results combine behavioural performance indicators, physiological responses, eye-tracking measurements, and BN-based inferences to provide a comprehensive evaluation of operator performance. The discussion interprets these findings by examining the interactions between human, technical, and operational factors identified through the BN model.

8.1. Results

The results are based on data collected during a simulation-based experiment involving 65 participants with different levels of maritime education and professional experience. The experimental sample comprised maritime high school students, university maritime students, certified deck officers and experienced captains, providing a broad representation of potential future and current operators involved in autonomous ship operations. This chapter presents the results derived from the experimental data and the BN model.

Figure 8.1 illustrates the composition of the participant sample. Officers constituted the largest participant group (36%), followed by maritime high school students (26%) and faculty students (26%), while experienced captains accounted for the remaining 12% of the sample.

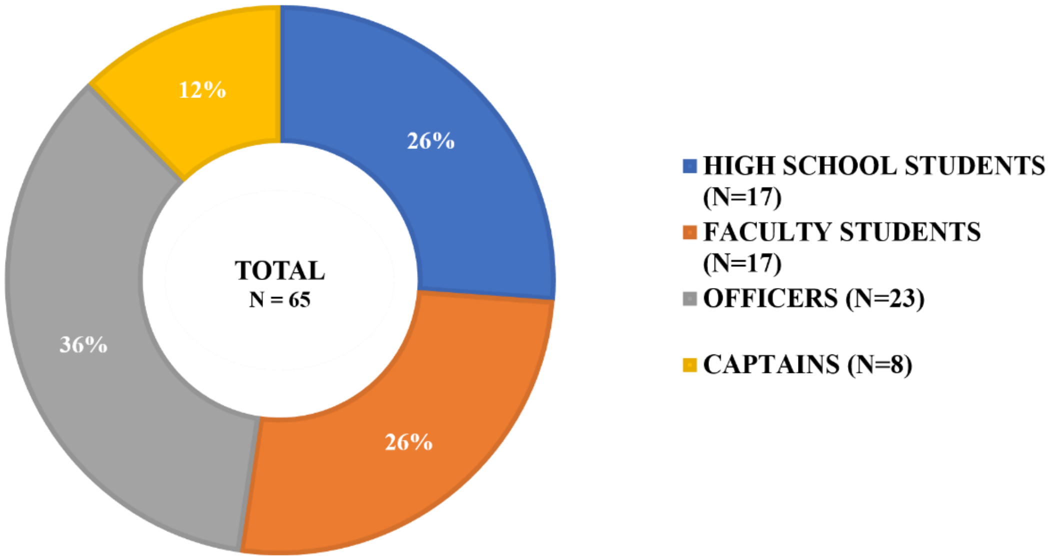

Figure 8.1 Distribution of participants according to educational and professional background (N = 65)

GPS failure detection performance varied across the four participant groups, revealing differences in how quickly participants recognized the introduced navigation system failure (Figure 8.2). Although all three detection categories were represented in every group, the overall distribution indicates that participants with professional sea-going experience generally achieved a higher proportion of rapid detections, whereas delayed detections were more common among the student groups.

The proportion of participants detecting the GPS failure within the shortest time interval was highest among officers and captains, while faculty students showed the lowest proportion of rapid detections. Conversely, delayed detection after more than 60 seconds occurred predominantly among the two student groups, whereas this category represented only a small proportion of responses among experienced seafarers. Intermediate detection times were observed across all participant groups, indicating that a substantial proportion of participants recognized the system failure within one minute of its occurrence.

The distribution demonstrates clear differences in GPS failure detection performance between participant groups and suggests differences associated with maritime education and operational experience on the timely recognition of navigation system failures.

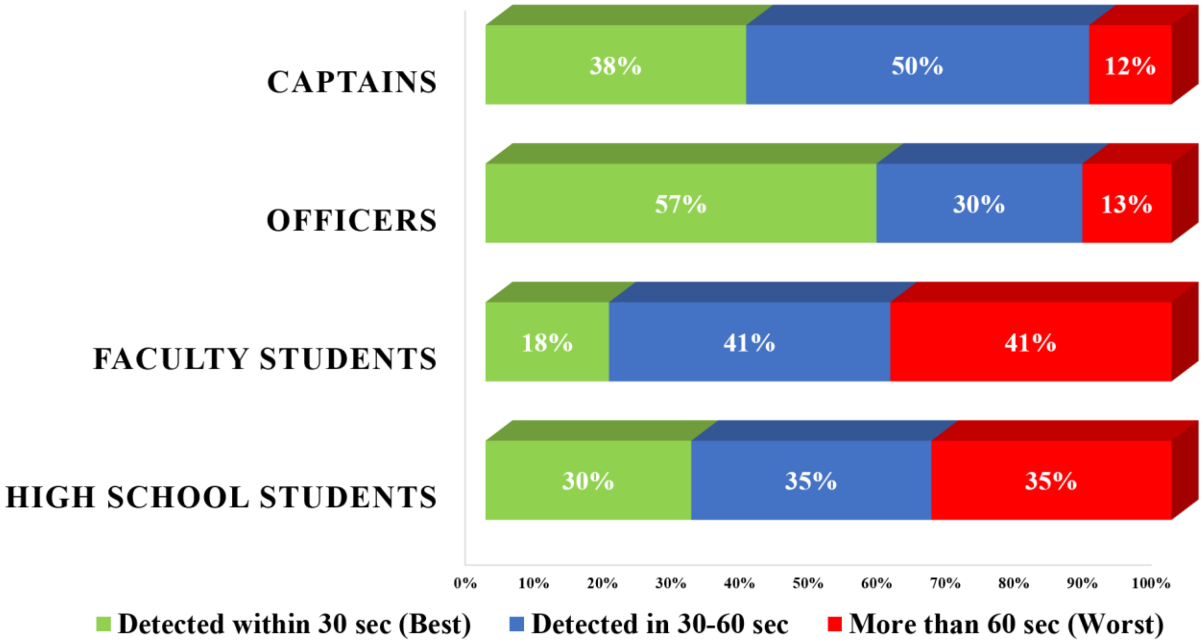

Figure 8.2 Distribution of GPS failure detection times according to participant group

Radar failure detection was assessed by measuring participants' ability to recognize the introduced radar failure during the simulation exercise (Figure 8.3). Detection performance was classified into three categories: detected within two minutes, detected more than two minutes, and not detected.

The results show clear differences in radar failure detection performance between participant groups. Early detection was more frequently observed among experienced seafarers, with captains showing the highest proportion of failures identified within the first two minutes. Officers also demonstrated comparatively strong detection performance, whereas both student groups recorded substantially lower proportions of early detections. In contrast, undetected radar failures were considerably more common among students, particularly faculty students, for whom this category represented the dominant outcome. Although delayed detection (more than two minutes) was observed in all participant groups, its distribution was more even than that of the early detection and undetected categories.

The results indicate that the ability to recognize radar failures differed considerably between participant groups, with experienced seafarers generally showing more favourable detection performance than participants with limited operational experience. These results complement the GPS failure detection results and provide additional insight into participants' responses to different types of navigation system failure.

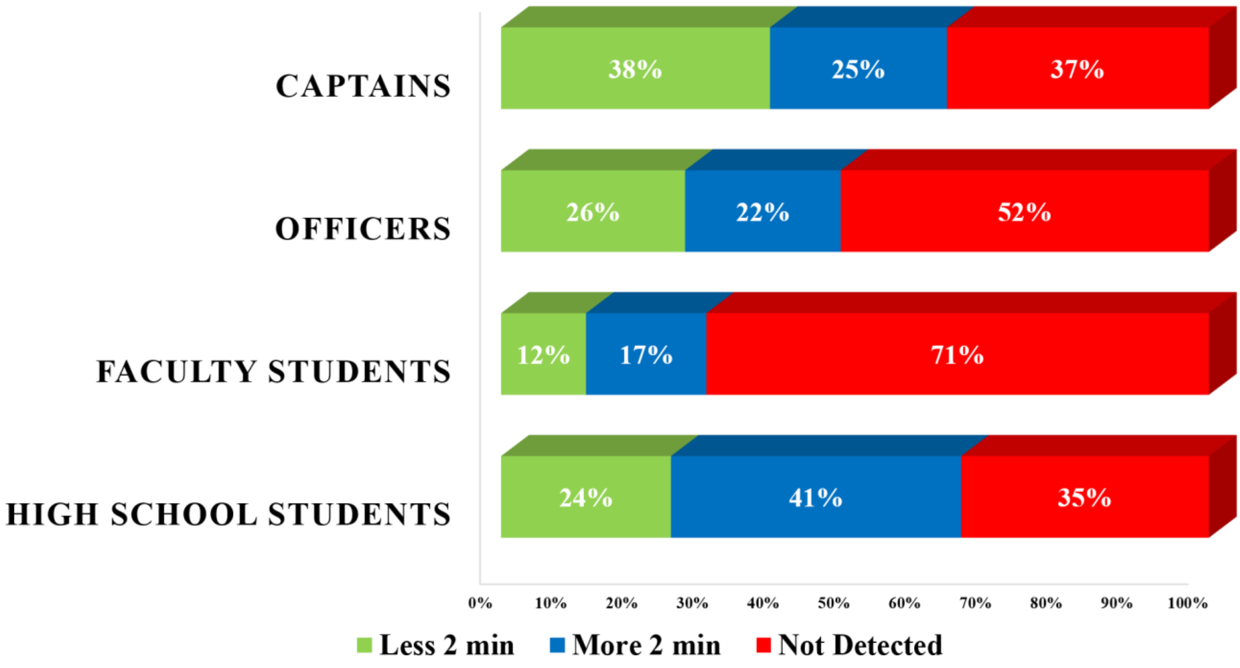

Figure 8.3 Distribution of radar failure detection outcomes according to participant group

Route Adherence was evaluated by determining whether participants maintained the predefined route throughout the simulation exercise (Figure 8.4).

The results indicate clear differences in Route Adherence among the participant groups. Maritime high school students recorded the highest proportion of successful Route Adherence, followed by faculty students, whereas officers and captains showed lower proportions of maintaining the predefined route. Conversely, deviations from the planned route occurred more frequently among the more experienced participant groups, with captains showing the highest proportion of route deviations. Despite these differences, both outcomes were observed across all participant categories, indicating variability in navigational behavior irrespective of educational background or professional experience.

The results demonstrate that Route Adherence differed among the four participant groups and provide an additional indicator of navigational performance within the simulation exercise. When considered together with the GPS and radar failure detection results, Route Adherence offers a broader description of participants' operational behavior during the simulated navigation task.

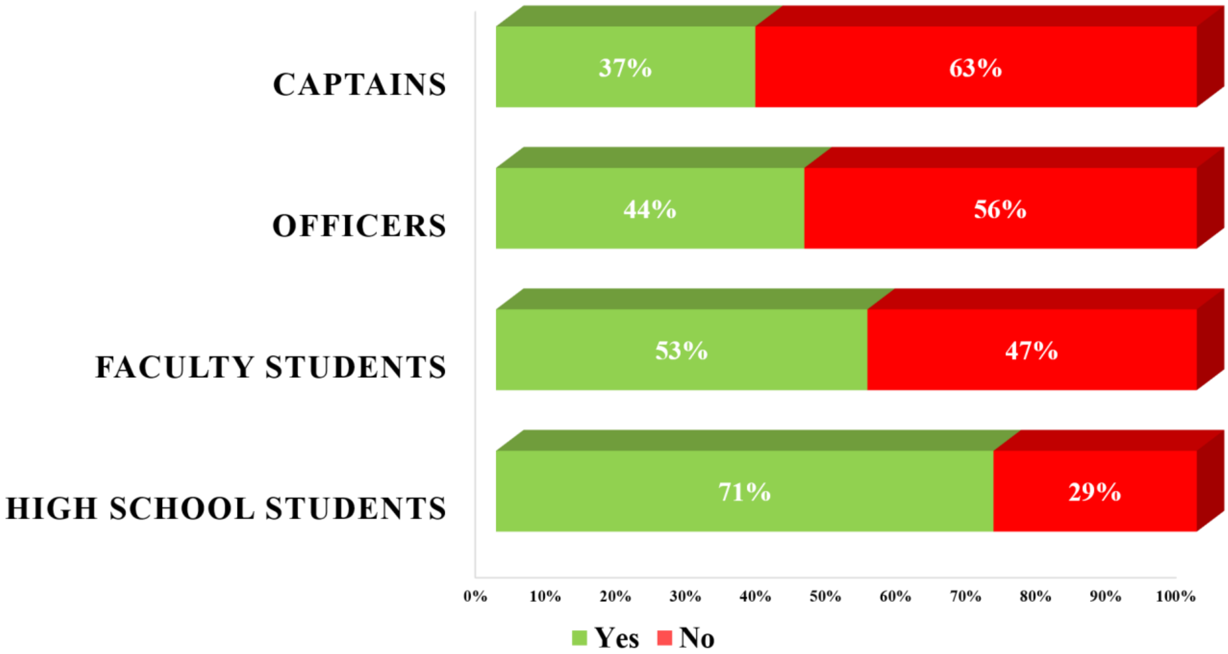

Figure 8.4 Distribution of Route Adherence according to participant group

Near miss events were recorded throughout the simulation exercise as an indicator of navigational safety performance (Figure 8.5). Each participant was classified according to the number of near miss events observed during the exercise, resulting in three outcome categories: 0, 1, and 2 near miss events.

The distribution of near miss events differed among the participant groups, indicating variability in operational safety performance during the simulation. Participants with professional sea-going experience completed a higher proportion of simulations without any near miss events, whereas this outcome was observed less frequently among the student groups. Across all four groups, however, 1 or 2 near miss events were the most common outcomes, demonstrating that safety critical situations occurred throughout the sample regardless of participants' educational or professional background.

The highest proportion of simulations completed without a near miss event was observed among officers, followed by captains and faculty students, while no participants from the maritime high school group completed the exercise without experiencing at least one near miss event. Conversely, two near miss events occurred more frequently among the student groups and captains, whereas officers showed the lowest proportion in this category. The results indicate measurable differences in navigational safety performance between participant groups and provide an additional operational indicator for the overall assessment of performance during the simulation exercise.

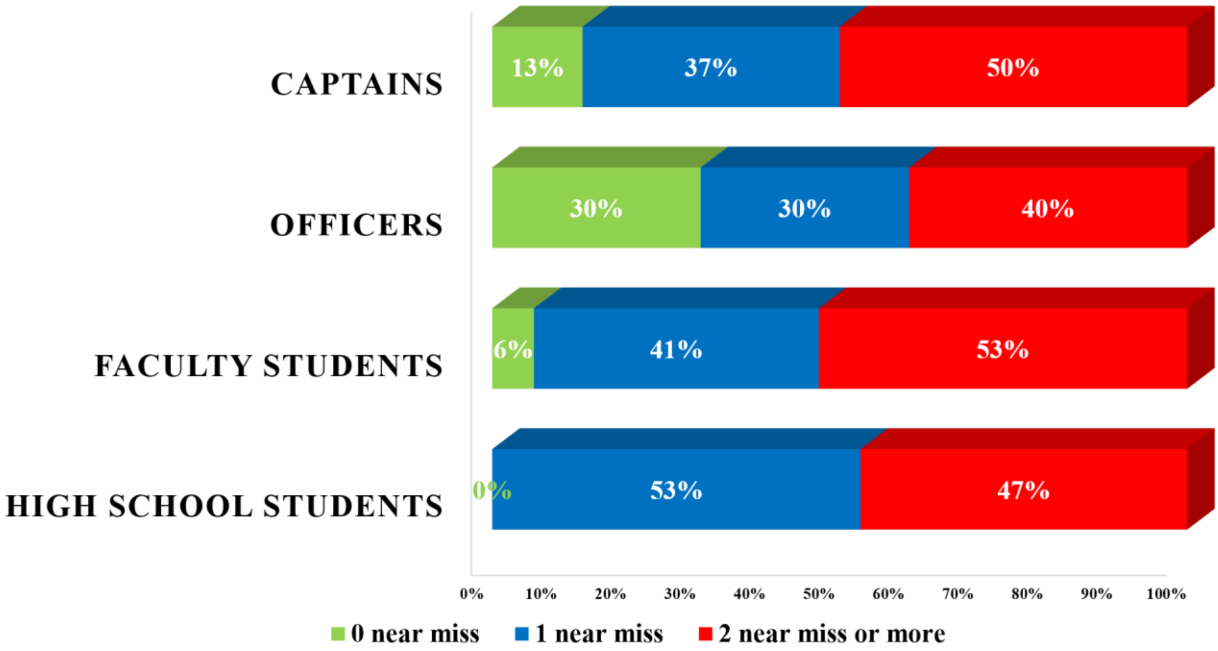

Figure 8.5 Distribution of near miss events according to participant group

Arrival time was evaluated as an additional indicator of navigational performance by recording the time required for participants to reach the predefined destination during the simulation exercise (Figure 8.6).

The distribution of arrival times differed among the participant groups, indicating variability in task completion during the simulation. Most participants completed the exercise within the expected 18 to 22 minutes interval, which was the dominant outcome for officers, faculty students, and captains. In contrast, maritime high school students most frequently completed the simulation in less than 18 minutes, and no participants from this group required more than 22 minutes to reach the assigned destination.

Longer completion times and failure to reach the assigned destination were observed only in a relatively small proportion of participants. Completion times exceeding 22 minutes occurred primarily among captains, while the proportion of participants who did not reach the destination remained relatively low across all groups. The results indicate that, despite differences in arrival time distributions, the majority of participants successfully completed the simulation within the expected operational timeframe.

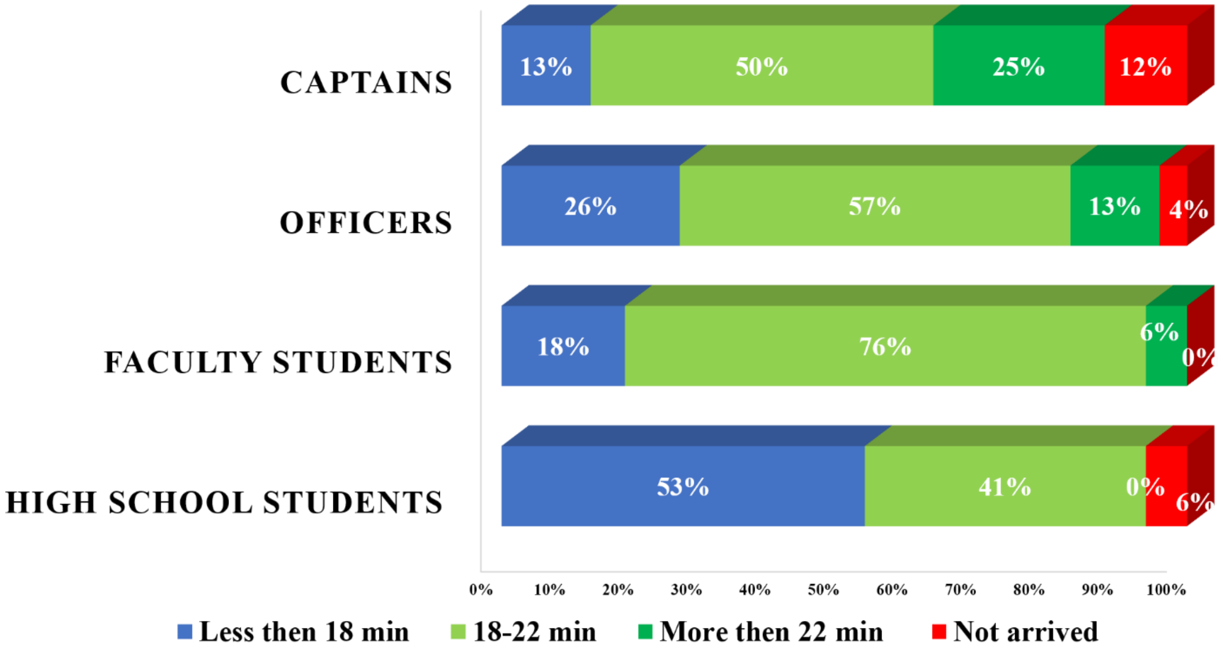

Figure 8.6 Distribution of arrival times according to participant group

The distribution of operator-related variables reflects differences in participants' preparedness and professional backgrounds within the experimental sample. The Operator Readiness variable was predominantly characterized by low and moderate readiness, with only a small proportion of participants achieving a high level of readiness. This distribution corresponds to the heterogeneous composition of the participant sample, which included individuals with different educational backgrounds and varying levels of operational experience. Similarly, Sea-going Experience ranged from participants without practical sea-going experience to certified officers and experienced captains, providing a broad basis for evaluating operator performance across different levels of professional com.

Physiological measurements showed that participants responded differently to the demands of the simulation exercise. Heart rate and blood pressure recordings reflected individual variations in physiological activation during task execution, indicating differences in responses to increasing navigational complexity, navigation system failures, and time-critical decision-making. Taken together, these measurements confirm that the simulation environment elicited measurable physiological responses suitable for inclusion in the proposed performance evaluation framework.

The cognitive-related variables also exhibited substantial variability across the participant sample. Cognitive Overload was distributed across all predefined states, suggesting that participants experienced different levels of mental workload throughout the simulation. Likewise, VAQ, derived from eye-tracking measurements, varied in terms of visual scanning behavior, fixation characteristics, and information acquisition while participants monitored navigational instruments and the surrounding traffic environment. These observations indicate differences in visual attention strategies and situational monitoring during task execution under simulated operational conditions.

The operator-related, physiological, and cognitive variables complement the navigational performance indicators by providing additional insight into participants' behavioral and cognitive responses during the simulation exercise. Their integration within the BN enables a multidimensional assessment of operator performance by simultaneously considering behavioral, physiological, and cognitive aspects of decision-making. The BN integrates all previously described navigational, physiological, and cognitive variables into a unified probabilistic model. Within this model, variables such as Navigation Control, Exercise Performance, Near Miss, and Voyage Execution represent key components that contribute to the overall assessment of operator performance.

To further illustrate the inferential capabilities of the developed BN model, three participant profiles were selected from the experimental dataset. These profiles represent the participant with the highest Operational Effectiveness outcome, the participant whose result was closest to the central tendency of the sample, and the participant with the lowest Operational Effectiveness outcome. Table 8-1 presents three representative participant profiles selected from the experimental dataset. The variables included in the table were chosen to provide a concise overview of participants' demographic, professional, and operational characteristics, enabling comparison of typical operator profiles rather than highlighting the most influential variables identified in the BN sensitivity analyses.

Table 8-1. Characteristics and BN-based performance comparison of selected participant profiles

Characteristic / Variable	High-performance profile	Representative performance profile	Low-performance profile
Age (years)	41	44	18
Participant group	Captain	Officer	Maritime high school student
Sea-going Experience	15 years	17 years	No experience
Education	Graduate maritime education	Graduate maritime education	Maritime High school (in progress)
Visual Attention Quality	High	Moderate	Low
Operator Readiness	High	Low	Low
Cognitive Overload	Moderate	High	High
Navigation Control	High	Moderate	Low
Response to Navigation System Failures	Good	Moderate	Moderate
Voyage Execution	High	High	Low
Exercise Performance	Good	Moderate	Low
Operational Effectiveness	High	Moderate	Poor

As shown in Table 8-1. the high-performance profile is an experienced captain with 15 years of sea-going experience and a graduate maritime education. This participant achieved favourable values across most variables included in the BN model, including high VAQ, high Operator Readiness, high Navigation Control, and high Voyage Execution, while experiencing only moderate Cognitive Overload. Together, these characteristics resulted in a high level of Exercise Performance and, consequently, the highest Operational Effectiveness among the three representative profiles.

The representative performance profile corresponds to an experienced officer with 17 years of sea-going experience and a graduate maritime education. Although this participant demonstrated high Voyage Execution and moderate Navigation Control, the overall assessment was influenced by lower Operator Readiness, moderate VAQ, and high Cognitive Overload. As a result, the BN model estimated moderate Exercise Performance and an overall moderate level of Operational Effectiveness.

The low-performance profile represents a maritime high school student with no sea-going experience. This participant exhibited low VAQ, low Operator Readiness, low Navigation Control, and low Voyage Execution, together with high Cognitive Overload. These conditions resulted in low Exercise Performance and the lowest Operational Effectiveness among the analysed representative profiles.

Comparison of the three representative profiles shows that Operational Effectiveness is determined by the combined influence of multiple variables rather than by a single performance indicator. Profiles characterized by favourable levels of VAQ, Operator Readiness, Navigation Control, and Voyage Execution consistently achieved higher Exercise Performance and higher Operational Effectiveness. Conversely, lower values of these variables, particularly when accompanied by high Cognitive Overload, were associated with reduced Exercise Performance and lower Operational Effectiveness. These results illustrate the capability of the developed BN model to evaluate operator performance by simultaneously considering interactions among human, physiological, technical, and operational factors.

Figure 8.7 presents clear distinction between the three representative participant profiles across all assessed performance indicators. The radar chart confirms that the developed BN model can distinguish between different levels of operator performance through the integrated assessment of multiple performance indicators. The figure presents the posterior probability (%) of the favourable state for each performance indicator obtained through BN inference, enabling direct comparison of the three selected participant profiles.

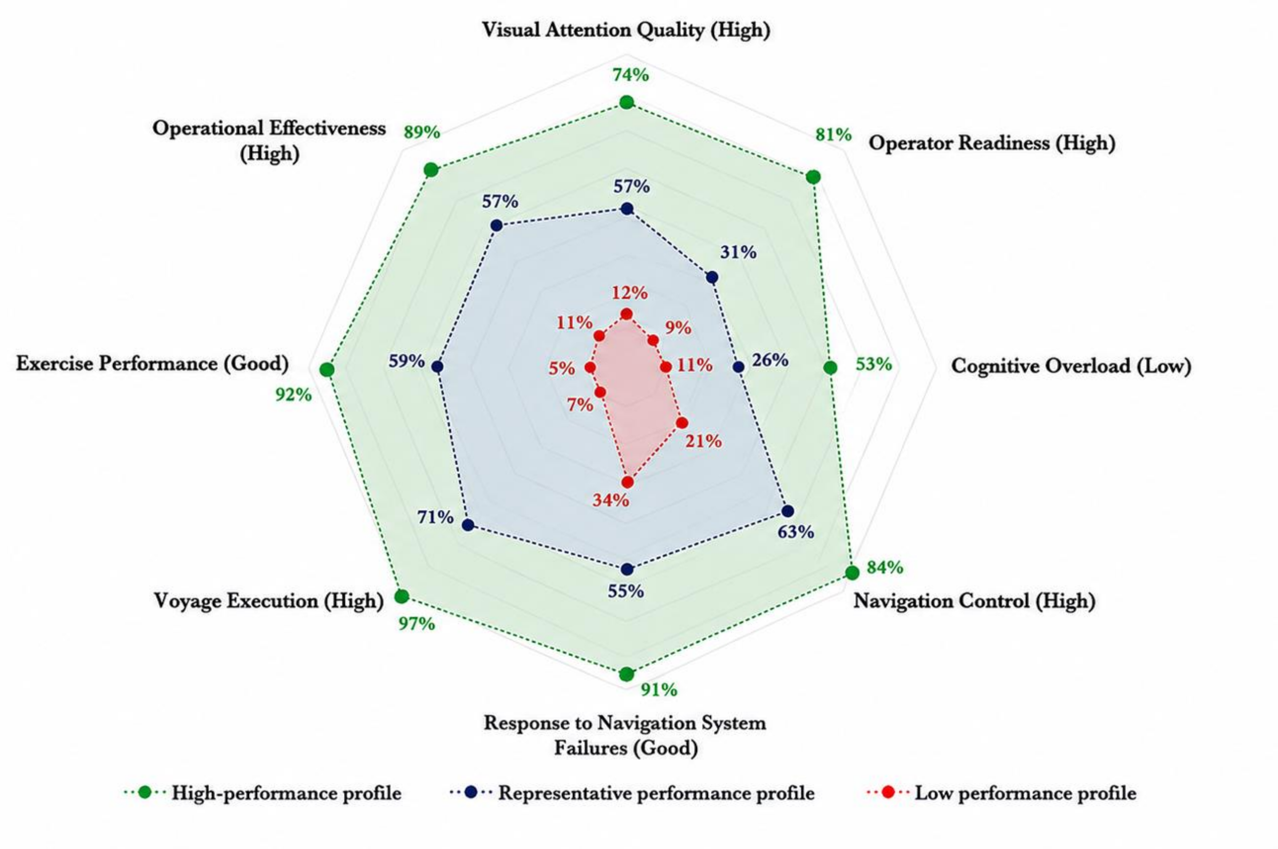

Figure 8.7 Comparison of selected participant profiles based on posterior probabilities of favourable performance states

Taken together, the results show that participant performance in the simulation exercise was characterized by substantial variability across navigational, physiological, cognitive, and operational indicators. Differences between participant groups were consistently reflected in several complementary measures, including navigation system failure detection, route adherence, near miss occurrence, arrival time, physiological responses, and cognitive performance.

The BN successfully integrated these heterogeneous data sources into a unified probabilistic framework capable of distinguishing between different levels of operator performance. The selected participant profiles further illustrate the model's ability to represent diverse combinations of behavioral, cognitive, physiological, and operational characteristics that contribute to overall Operational Effectiveness.

8.2. Discussion

The results presented in the previous section show that operator performance in MASS operations cannot be attributed to a single dominant factor, but instead arises from the interaction of multiple human, technical, cognitive, and environmental variables. Variability observed in Route Adherence, Arrival Time, Near Miss occurrences, System Failure detection, Physiological Response, and visual attention patterns suggests that performance degradation is systematic and depends on interacting influences rather than random error.

This interpretation is further supported by the sensitivity analysis presented in Chapter 7, which provides additional insight into the relative influence of the variables within the developed BN model. The analysis identified Navigation Control, Exercise Performance, Near Miss, and Voyage Execution as the variables exerting the greatest influence on Operational Effectiveness. Rather than representing isolated performance indicators, these variables capture the cumulative effects of human, physiological, visual attention, technical and operational factors defined in the qualitative and quantitative model specification (Tables 7-1. and 7-2.) and confirmed through the model validation process. Their dominant influence therefore demonstrates that overall operator performance in MASS operations arises from the interaction of multiple lower-level variables rather than from any single factor. This probabilistic interpretation provides the framework for understanding the experimental findings discussed in the following sections.

System disturbances, particularly radar and GPS errors, played a critical role in shaping performance outcomes. Participants who did not detect these disturbances promptly showed reduced navigation stability and a higher likelihood of unsafe situations. Within the BN model, delayed or ineffective responses to navigation system failures decrease the probability of maintaining high Navigation Control, while timely detection acts as a stabilizing factor even under adverse conditions. This finding is consistent with the sensitivity analysis, which identified Navigation Control as one of the principal determinants of Operational Effectiveness. Because Navigation Control reflects the combined influence of visual attention, cognitive state, operator response, and navigational conditions, reduced ability to recognize and respond to system failures propagates through the BN model and ultimately decreases the probability of achieving high Operational Effectiveness. These results underscore continuous system monitoring and early failure recognition as core competencies in MASS operations.

Physiological and eye-tracking data provide further insight into the mechanisms underlying performance variability. Increased cognitive load was associated with reduced stability of visual attention and less effective monitoring of navigational information. Within the BN model, higher Cognitive Overload contributes to decreased VAQ, which negatively affects Navigation Control and decision-making accuracy. This confirms that cognitive limitations are a key constraint in MASS operations, especially when operators must respond to system degradation and complex traffic situations simultaneously.

Near Miss events are a key safety indicator in the analysis. Their occurrence was not random, but linked to specific combinations of degraded attention, delayed system response, and reduced control. The BN model shows that low VAQ combined with delayed detection of system failures increases the probability of near miss events, indicating a breakdown in situational awareness and decision-making processes. This confirms that safety critical events in MASS operations arise from interacting deficiencies rather than isolated errors.

An important finding of the dissertation concerns the distinction between apparent and actual performance quality. The results show that successful route completion alone does not necessarily indicate effective or safe navigation performance. In several cases, participants were able to maintain or complete the planned route, while other variables, including VAQ, Cognitive Overload, and Response to Navigation System Failures, indicated degraded performance conditions. Within the BN model, such cases are not interpreted as fully successful performance. Even when Route Adherence is achieved, the presence of low attention quality, delayed failure detection, or high cognitive load reduces the probability of overall performance quality. This indicates that some apparently successful outcomes may represent coincidental or unstable success rather than reliable operator performance. The model therefore distinguishes between truly effective performance and superficially successful outcomes, preventing isolated indicators such as route completion from being interpreted as evidence of safe and competent operation.

The analysis in this dissertation, based on a sample of 65 participants, yields several key insights. First, performance degradation is not caused by isolated failures but by the interaction of multiple factors, particularly reduced situational awareness, delayed response to system failures, and increased Cognitive Overload. Second, successful task execution, such as maintaining or completing a predefined route, does not necessarily reflect effective or safe operator performance. Third, traditional indicators of competence, such as experience and Operator Readiness, have a conditional influence on performance and may lose part of their positive effect under system failure and elevated cognitive load.

Considering the results obtained, it is expected that the operation of MASS will rely on a team-based approach, closely resembling traditional shipboard organization. This conclusion is supported by observed interactions between key variables within the model, where performance was influenced not by a single factor but by the combined effects of visual attention, physiological stress, system failure response, and navigational complexity. In this context, the captain's role remains central, as the individual with the highest level of responsibility for overall vessel safety and decision-making. The results indicate that a single operator is unlikely to manage all system inputs and operational demands effectively in a remote control environment, particularly under increased cognitive load and system disturbances. The impact of Cognitive Overload and delayed responses to navigation system failures further supports the need for a coordinated team structure, with responsibilities distributed among multiple operators. The increased demands associated with continuous monitoring of digital interfaces suggest that effective working periods may be shorter than in onboard operations, requiring more frequent operator rotation to maintain performance and safety. These results confirm that, despite increasing automation, safe and efficient MASS operation will continue to depend on the interaction between a responsible captain (main operator) and a well-coordinated team.

Within the analysed sample, age-related performance patterns were also observed. Older participants tended to experience more difficulties in visual information processing and situational visualization, as reflected in reduced VAQ and less stable gaze behavior. In contrast, younger participants generally demonstrated better visual tracking and attention stability but more frequently encountered difficulties in navigation-related tasks, including route management, decision-making under traffic conditions, and response to system failures. Within the BN model, these differences can be interpreted as variations in the balance between cognitive processing and operational experience. Reduced VAQ in older participants increases the probability of degraded Navigation Control, while in younger participants, lower levels of experience and operational judgement contribute to less effective decision-making despite relatively stable attention patterns.

The results indicate that neither age group demonstrates consistently superior performance. Instead, different limitations predominate depending on the operator profile. The results therefore suggest that optimal performance is achieved not at the extremes of age or experience, but in profiles that combine sufficient operational experience with preserved cognitive and attentional capacity. Within the analysed sample, the BN model indicated that the most favourable overall performance was associated with the profile of a mid-aged shipmaster, suggesting that a balance between professional experience, operational judgement, and maintained cognitive flexibility provides the most stable performance under complex MASS-related conditions.

The results confirm the hypothesis that the capability of ship operators can be determined through simulator-based training and BN modelling, enabling the identification of key combinations of factors that influence decision-making and human error in autonomous ship operations.

Results also show that operator performance is not driven by a single factor, but by the interaction of multiple variables. In particular, reduced attention, delayed responses to system failures, increased Cognitive Overload, and weakened Navigation Control act together to shape performance, directly affecting safety critical outcomes and overall Operational Effectiveness.

The dissertation further identifies key technical, cognitive, and problem-solving factors that influence safe and effective performance in MASS operations. Variables such as Navigation Control, VAQ, Cognitive Overload, Response to Navigation System Failures, Exercise Performance, Near Miss events, and Voyage Execution were shown to play a central role in shaping performance outcomes. This dissertation provides an empirically grounded and scientifically robust model for assessing operator capability in autonomous maritime environments. By integrating simulation-based data, physiological measurements, eye-tracking indicators, and Bayesian probabilistic modelling, the dissertation advances existing approaches to operator performance assessment and provides a foundation for evidence-based training, objective evaluation methods, and decision-support systems for future MASS operations. The comparison confirms that Operational Effectiveness results from the combined influence of multiple interconnected cognitive, behavioral, and operational factors, rather than a single performance indicator.

8.3. Limitations and Directions for Future Research

The following limitations should be considered when interpreting the results of this dissertation:

- **Sample size and participant samples**. The research involved 65 participants, which is sufficient to identify consistent relationships within the dataset but limits broader statistical generalisation. All participants were recruited from the Croatian maritime education and training environment, which may reduce the applicability of the results to other geographical, cultural, and operational contexts.
- **Simulation-based research environment**. Although high fidelity maritime simulators provide a realistic and controlled platform for studying operator behavior in MASS operations, they cannot fully replicate the complexity of real world autonomous maritime environments. Long-term supervisory control, interaction with remote control centres, system-level uncertainties, and the operational consequences of decision-making are only partially represented.
- **Eye-tracking equipment constraints.** The Pupil Core eye-tracking system requires adequate ambient lighting to accurately detect pupil movements and diameter changes. Reduced daylight and low-visibility scenarios occasionally affected the quality and completeness of eye-tracking recordings, potentially influencing some derived visual attention metrics. To minimize this limitation, eye-tracking data were cross-validated with behavioral and physiological indicators.
- **Scenario-specific experimental design**. The simulation exercise was based on a predefined set of navigational tasks, environmental conditions, and system failures. Consequently, the results primarily reflect operator behavior within the selected MASS scenarios and may not fully represent the diversity of operational situations encountered in future autonomous ship operations.
- **BN model assumptions**. Several variables were discretized using expert judgement to ensure compatibility with the BN structure. Although this approach follows established

modelling practice, different discretization strategies or alternative expert assessments could lead to slightly different probability distributions and inference results.

- **Cross-sectional assessment approach.** The research evaluated participant performance during a single experimental session. As a result, long-term learning effects, adaptation processes, skill retention, and performance development over time were not examined.

Despite these limitations, the consistency of the observed relationships within the analysed sample suggests that the identified patterns reflect meaningful characteristics of operator behavior in complex navigation scenarios. The results therefore provide a valuable foundation for further refinement. Accordingly, future research should focus on the following directions:

- **Larger and more diverse participant populations**. Future studies should increase the sample size and include experienced seafarers from different countries, shipping sectors, and training backgrounds. This would improve the generalisability of the model and enable validation across different operational and cultural contexts.
- **Longitudinal performance assessment.** Repeated assessments over time would allow investigation of learning effects, performance development, adaptation processes, and long-term retention of the skills required for MASS operations.
- **Further development of the BN model**. Future work could incorporate larger datasets, the use of Dynamic Bayesian Networks to represent temporal dependencies, Noisy-MAX functions for more efficient probability modelling, and comparisons between expert-based and data-driven discretization methods.
- **Integration of real operational data.** The proposed model should be validated using data collected from autonomous and semi-autonomous ship operations, remote control centres, and international simulator facilities to improve its robustness and practical relevance.
- **Multimodal human performance assessment.** Additional physiological and behavioral data sources, such as wearable sensors, communication analysis, EEG, and other cognitive workload indicators, could be integrated to improve the predictive capability of the BN model.
- **Practical implementation.** Future research should investigate the application of the proposed model in maritime training institutions, shipping companies, and remote

control centres to assess operator performance and support operator training for autonomous ship operations.

9. CONCLUSION

This doctoral dissertation addresses the increasing need for a structured and empirically grounded approach to evaluating human performance in MASS operations. The increasing automation in maritime transport is changing the role of seafarers, shifting responsibilities from direct ship handling to system supervision, monitoring, and decision-making under uncertainty. In this context, traditional methods of training and performance assessment are no longer sufficient, creating a need for new data-driven evaluation models.

The research framework is based on a systematic literature review conducted by the author, providing a comprehensive overview of existing approaches to MASS operations, training, and human factors. The review identifies a lack of experimental and data-driven studies integrating simulator-based performance indicators with probabilistic modelling, which directly motivated the research presented in this dissertation.

The primary objective of this research was to develop and validate a model for assessing the performance of ship operators in the control of autonomous ships using simulation-based experiments and BN modelling. To achieve this, a comprehensive methodological model was established, integrating high fidelity simulation, behavioral performance indicators, physiological measurements, and eye-tracking data into a unified analytical approach. The experimental setup included a maritime simulator, eye-tracking glasses, and wearable physiological monitoring devices, enabling the collection of objective and synchronized data on operator behavior and physiological responses. Developing the proposed model successfully met the dissertation’s second research objective.

The simulation exercise was specifically designed to include dynamic operational conditions, such as navigation system failures and increased traffic complexity, requiring participants to continuously adapt their decisions. This approach ensured a realistic representation of MASS operational challenges and enabled the observation of performance under conditions of stress and uncertainty. The first research objective was successfully achieved by developing and implementing a simulation-based navigational exercise.

The results indicate that operator performance in MASS operations is determined not by a single factor, but by the interaction of multiple variables, including Cognitive Overload, Visual

Attention, Response to System Failures, and Navigation Control. The simulation-based exercise captured realistic behavioral patterns under controlled yet complex navigational conditions, demonstrating that performance degradation occurs through combinations of reduced situational awareness, delayed reactions, and increased cognitive load.

A key finding is that successful task execution does not necessarily imply effective decision-making. The analysis showed that participants were sometimes able to maintain the predefined navigation route despite degraded cognitive and operational conditions. The BN model captures these discrepancies by integrating multiple performance indicators, preventing isolated outcomes from being interpreted as evidence of performance. This represents a significant shift in the evaluation of operator performance, emphasising the importance of analysing both the decision-making process and the final outcome.

The developed BN model offers a structured probabilistic framework for analysing the relationships between human element, technical, and environmental factors that influence performance in autonomous maritime operations. By modelling dependencies between variables such as VAQ, Cognitive Overload, Navigation Control, and Response to Navigation System Failures, the model enables descriptive and predictive analysis of operator performance. Sensitivity analysis confirmed the model's internal consistency and identified key variables with the greatest influence on Operational Effectiveness. This demonstrates the suitability of BN as a robust tool for modelling uncertainty and supporting decision-making in autonomous maritime operations.

The third research objective was achieved through the development and validation of a BN model that describes the relationships among human, technical, and operational factors that influence operator performance in MASS operations. The validated model enables analysis of their interdependencies and quantifies their influence on Operational Effectiveness under different operational conditions. By integrating empirical simulation data with probabilistic reasoning, the proposed BN model provides a structured framework for evaluating operator performance and supporting evidence-based decision-making in autonomous ship operations.

The results show that traditional indicators of competence, such as experience and Operator Readiness, have a conditional impact on performance. While these factors contribute positively under normal conditions, their effect is diminished in situations involving system failures and increased cognitive demand. This finding suggests that operator performance in MASS operations cannot be adequately assessed using traditional indicators alone. Effective assessment requires consideration of cognitive, behavioral, and system interaction factors, as well as experience and Operator Readiness.

Within the analysed sample, further insights were identified regarding operator profiles. The results indicate that optimal performance is associated with a balance between professional experience and cognitive adaptability, rather than maximum experience alone. Age-related patterns showed that older participants experienced more difficulties in visual processing and attention stability, while younger participants had limitations in navigation-related decision-making. These results suggest that effective performance in MASS operations requires a combination of experience and preserved cognitive flexibility.

The results in this chapter support the dissertation hypothesis. The results show that simulator derived performance indicators, physiological measurements, eye-tracking variables, and BN modelling can identify combinations of factors influencing decision-making quality, human error, and overall Operational Effectiveness in autonomous ship operations. The observed relationships between Cognitive Overload, VAQ, Navigation Control, system failure response, and performance outcomes confirm the suitability of the proposed approach for assessing operator capability in MASS environments. The proposed model not only identifies key variables affecting performance but also explains how their interactions shape operational outcomes.

The results of the sensitivity analysis and Bayesian inference further support the conclusions of this research. Navigation Control, Exercise Performance, Near Miss, and Voyage Execution were identified as the variables exerting the greatest influence on Operational Effectiveness. In contrast, reduced VAQ, increased Cognitive Overload, and delayed responses to navigation system failures were associated with reduced Operational Effectiveness and a higher probability of operator error. These results confirm the proposed research hypothesis and demonstrate that the developed BN model enables the identification of the key factors, and their interactions, that influence the quality of decision-making, the likelihood of operator error, and overall operator performance in MASS operations.

The scientific contribution of this dissertation is reflected in several key aspects. First, the research provides an empirically grounded model that integrates behavioral, physiological, and technical data, addressing a significant gap in the existing literature. Second, it establishes a methodological framework that combines simulation-based experimentation with BN modelling for performance assessment. Third, it contributes to a deeper understanding of human factors in autonomous maritime environments, particularly in relation to cognitive workload and visual attention. The research offers a foundation for the development of evidence-based training and assessment systems tailored to the requirements of MASS operations.

In addition to its scientific contribution, the proposed model has significant practical implications. The model can be used in maritime education and training institutions as an objective assessment tool, supporting the evaluation of Operator Readiness and the identification of critical competencies. It can also be integrated into simulation environments and remote control centres to provide real time or post-Exercise Performance analysis. More broadly, such applications can support the safe implementation of autonomous maritime technologies by enabling data-driven decision-making and reducing the risk of human error.

Despite its contributions, the dissertation acknowledges certain limitations related to sample size, geographical scope, simulation conditions, and technical constraints of data collection. These limitations do not diminish the validity of the results but define the boundaries within which the results should be interpreted and highlight directions for future research.

This dissertation provides a comprehensive and scientifically robust model for evaluating operator performance in MASS operations. By integrating simulation-based data with Bayesian probabilistic modelling, the research advances current approaches to human performance assessment and contributes to the safe and effective transition towards autonomous shipping.

The results show that in MASS operations, a correct outcome does not necessarily reflect correct decision-making, redefining how operator performance should be evaluated. This research moves beyond traditional assessment approaches by introducing an empirically grounded, data-driven model that captures the complex interaction of human, technical, and cognitive factors in autonomous ship operations. The proposed model is a step towards the development of objective, evidence-based training and decision-support systems necessary for the safe implementation of autonomous maritime technologies.

10. REFERENCES

[1] "IMO Autonomous Shipping," https://www.imo.org/en/mediacentre/hottopics/pages/autonomous-shipping.aspx. Accessed: Sep. 09, 2025. [Online]. Available: https://www.imo.org/en/mediacentre/hottopics/pages/autonomous-shipping. aspx

[2] A. N. Pilatis, D. N. Pagonis, M. Serris, S. Peppa, and G. Kaltsas, "A Statistical Analysis of Ship Accidents (1990–2020) Focusing on Collision, Grounding, Hull Failure, and Resulting Hull Damage," *J. Mar. Sci. Eng.*, vol. 12, no. 1, Jan. 2024, doi: 10.3390/jmse12010122.

[3] Ø. J. Rødseth, L. A. L. Wennersberg, and H. Nordahl, "Improving safety of interactions between conventional and autonomous ships," *Ocean Engineering*, vol. 284, p. 115206, Sep. 2023, doi: 10.1016/J.OCEANENG.2023.115206.

[4] H. Fan, H. Jia, X. He, and J. Lyu, "Navigating uncertainty: A dynamic Bayesian network-based risk assessment framework for maritime trade routes," *Reliab. Eng. Syst. Saf.*, vol. 250, p. 110311, Oct. 2024, doi: 10.1016/J.RESS.2024.110311.

[5] X. Li and K. F. Yuen, "A human-centred review on maritime autonomous surfaces ships: impacts, responses, and future directions," *Transp. Rev.*, vol. 44, no. 4, pp. 791–810, 2024, doi: 10.1080/01441647.2024.2325453.

[6] K. M. Lynch, V. A. Banks, A. P. J. Roberts, S. Radcliffe, and K. L. Plant, "Maritime autonomous surface ships: can we learn from unmanned aerial vehicle incidents using the perceptual cycle model?," *Ergonomics*, vol. 66, no. 6, pp. 772–790, 2023, doi: 10.1080/00140139.2022.2126896.

[7] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," in *Journal of clinical epidemiology*, Oct. 2009, pp. e1–e34. doi: 10.1016/j.jclinepi.2009.06.006.

[8] T. Meštrović, I. Pavić, M. Maljković, and A. Androjna, "Challenges for the Education and Training of Seafarers in the Context of Autonomous Shipping: Bibliometric

Analysis and Systematic Literature Review," *Applied Sciences*, vol. 14, no. 8, p. 3173, Apr. 2024, doi: 10.3390/app14083173.

[9] "IMO, Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS), 3rd June 2021." Accessed: Sep. 21, 2025. [Online]. Available: https://wwwcdn.imo.org/localresources/en/MediaCentre/PressBriefings/Documents/MSC.1-Circ.1638%20-%20Outcome%20Of%20The%20Regulatory%20Scoping%20ExerciseFor%20The%20Use%20Of%20Maritime%20Autonomous%20Surface%20Ships...%20(Secretariat).pdf

[10] I. Vio and M. Brdar, "Maritime Autonomous Surface Ships-International and National Legal Framework," *Pomorski zbornik*, vol. 62, pp. 141–155, May 2022, Accessed: Oct. 29, 2025. [Online]. Available: https://hrcak.srce.hr/file/408033

[11] M. Yoshida, E. Shimizu, M. Sugomori, and A. Umeda, "Regulatory requirements on the competence of remote operator in maritime autonomous surface ship: Situation awareness, ship sense and goal-based gap analysis," *Applied Sciences (Switzerland)*, vol. 10, no. 23, pp. 1–27, Dec. 2020, doi: 10.3390/app10238751.

[12] C. Hult, G. Praetorius, and C. Sandberg, "On the future of maritime transport – discussing terminology and timeframes," *TransNav*, vol. 13, no. 2, pp. 269–273, Jun. 2019, doi: 10.12716/1001.13.02.01.

[13] A. J. Fenton and I. Chapsos, "Ships without crews: IMO and UK responses to cybersecurity, technology, law and regulation of maritime autonomous surface ships (MASS)," *Front. Comput. Sci.*, vol. 5, 2023, doi: 10.3389/fcomp.2023.1151188.

[14] International Maritime Organization, "Current Status of MASS Technological Development Project and Regulatory Framework in Japan," London,UK, Mar. 2026.

[15] International Maritime Organization, "Development of a Goal-Based Instrument for Maritime Autonomous Surface Ships (MASS): Report of the Working Group (MSC 111/WP.9)," London, UK, May 2026.

[16] International Maritime Organization, “International Code of Safety for Maritime Autonomous Surface Ships (MASS Code) (Resolution MSC.595(111)),” London,UK, May 2026.

[17] G. R. Emad and S. Ghosh, “Identifying essential skills and competencies towards building a training framework for future operators of autonomous ships: a qualitative study,” *WMU Journal of Maritime Affairs*, vol. 22, no. 4, pp. 427–445, Dec. 2023, doi: 10.1007/s13437-023-00310-9.

[18] A. Stefani and L. Apicella, “A new educational model for Marine 4.0 technologies,” in *Proceedings of the International Ship Control Systems Symposium*, Institute of Marine Engineering Science & Technology, 2022. doi: 10.24868/10724.

[19] T. Fonseca, K. Lagdami, J.-U. Schröder-Hinrichs, K. Lagdamia, and J.-U. Schroder-Hinrichsa, “The 20th Comemorative Annual General Assembly Proceedings of the International Association of Maritime Universities,” IAMU) Conferenc e. [Online]. Available: https://www.researchgate.net/publication/337947996

[20] B. Belev, A. Penev, Đ. Mohović, and A. Perić Hadžić, “Autonomous ships in maritime education model course 7.01,” *Pomorstvo*, vol. 35, no. 2, pp. 388–394, Dec. 2021, doi: 10.31217/p.35.2.20.

[21] C. Campos Toresano, A. Mujal Colilles, and M. Castells Sanabra, “The next step on the maritime education and training in the era of autonomous shipping,” in *Maritime Transport - 9th International Conference on Maritime Transport (Maritime Transport IX): Barcelona,27-29June,2022*, Universitat Politècnica de Catalunya. Iniciativa Digital Politècnica. doi: 10.5821/mt.11004.

[22] E. Demirel, “The Prospective Posture of Maritime Education in Support of Smart Shipping,” *Journal of Marine Technology and Environment*, vol. 2022, no. 1, pp. 26–39, Apr. 2022, doi: 10.53464/JMTE.01.2022.04.

[23] A. Sharma and T.-E. Kim, “Exploring technical and non-technical competencies of navigators for autonomous shipping,” *Maritime Policy and Management*, vol. 49, no. 6, pp. 831–849, 2022, doi: 10.1080/03088839.2021.1914874.

[24] T. eun Kim and S. Mallam, “A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations,” *WMU Journal of Maritime Affairs*, vol. 19, no. 2, pp. 163–181, Jun. 2020, doi: 10.1007/s13437-020-00203-1.

[25] G. Reza Emad, S. Narayanan, and A. Kataria, “On the road to Autonomous Maritime Transport: A conceptual framework to meet training needs for future ship operations,” 2022. doi: 10.54941/ahfe1002500.

[26] K. Bogusławski, M. Gil, J. Nasur, and K. Wróbel, “Implications of autonomous shipping for maritime education and training: the cadet’s perspective,” *Maritime Economics & Logistics*, vol. 24, no. 2, pp. 327–343, Jun. 2022, doi: 10.1057/s41278-022-00217-x.

[27] S. C. Mallam, S. Nazir, and A. Sharma, “The human element in future Maritime Operations – perceived impact of autonomous shipping,” *Ergonomics*, vol. 63, no. 3, pp. 334–345, Mar. 2020, doi: 10.1080/00140139.2019.1659995.

[28] J. Kim, “A Fundamental Study of the Sustainable Key Competencies for Remote Operators of Maritime Autonomous Surface Ships,” *Sustainability*, vol. 16, no. 12, p. 4875, Jun. 2024, doi: 10.3390/su16124875.

[29] A. L. Aalberg, “Pride and mistrust? The association between maritime bridge crew officers’ professional commitment and trust in autonomy,” *WMU Journal of Maritime Affairs*, vol. 23, no. 4, pp. 551–574, Dec. 2024, doi: 10.1007/s13437-024-00329-6.

[30] M. Belabyad *et al.*, “The human element in autonomous shipping: a study on skills and competency requirements,” *WMU Journal of Maritime Affairs*, vol. 24, no. 4, pp. 575–605, Dec. 2025, doi: 10.1007/s13437-025-00366-9.

[31] Y. Li, Z. Duan, and Z. Liu, “Study on Risk-based Operators’ Competence Assessment of Maritime Autonomous Surface Ships,” in *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, IEEE, Jul. 2019, pp. 1412–1417. doi: 10.1109/ICTIS.2019.8883599.

[32] H. Hwang, T. Hwang, and I.-H. Youn, “Effect of Onboard Training for Improvement of Navigation Skill under the Simulated Navigation Environment for Maritime

Autonomous Surface Ship Operation Training," *Applied Sciences*, vol. 12, no. 18, p. 9300, Sep. 2022, doi: 10.3390/app12189300.

[33] J. Kim, J. Park, and K. Cho, "Continuous Autonomous Ship Learning Framework for Human Policies on Simulation," *Applied Sciences*, vol. 12, no. 3, p. 1631, Feb. 2022, doi: 10.3390/app12031631.

[34] T. Kuntasa and T.-C. Lirn, "A conceptual model of autonomous ship remote operators' competency," *Journal of Navigation*, vol. 76, no. 6, pp. 653–674, Nov. 2023, doi: 10.1017/S0373463324000055.

[35] M. Lutzhoft, A. Hynnekleiv, J. V Earthy, and E. S. Petersen, "Human-centred maritime autonomy - An ethnography of the future," *J. Phys. Conf. Ser.*, vol. 1357, no. 1, p. 012032, Oct. 2019, doi: 10.1088/1742-6596/1357/1/012032.

[36] M. Belabyad, C. Kontovas, R. Pyne, and C.-H. Chang, "Skills and competencies for operating maritime autonomous surface ships (MASS): a systematic review and bibliometric analysis," *Maritime Policy & Management*, vol. 53, no. 1, pp. 1–26, Jan. 2026, doi: 10.1080/03088839.2025.2475177.

[37] A. L. Aalberg, S. M. Holen, T. Kongsvik, and A. M. Wahl, "Does it do the same as we would? How trust in automated shipboard systems relates to seafarers' professional identity," *Saf. Sci.*, vol. 172, p. 106426, Apr. 2024, doi: 10.1016/j.ssci.2024.106426.

[38] R. Song, E. Papadimitriou, R. R. Negenborn, and P. van Gelder, "Safety and efficiency of human-MASS interactions: towards an integrated framework," *Journal of Marine Engineering & Technology*, vol. 24, no. 3, pp. 159–178, May 2025, doi: 10.1080/20464177.2024.2414959.

[39] M. Minami, M. Kobayashi, K. Hikida, and K. Kokubun, "Development of the Comprehensive Simulation System for Autonomous Ships," in *Journal of Physics: Conference Series*, Institute of Physics, 2022. doi: 10.1088/1742-6596/2311/1/012012.

[40] S. Ahvenjärvi, J. Lahtinen, M. Löytökorpi, and M. M. Marva, "Istlab – new way of utilizing a simulator system in testing & demonstration of intelligent shipping technology and training of future maritime professionals," *TransNav*, vol. 15, no. 3, pp. 569–574, 2021, doi: 10.12716/1001.15.03.09.

[41] M. Martelli, A. Virdis, A. Gotta, P. Cassara, and M. DI Summa, “An Outlook on the Future Marine Traffic Management System for Autonomous Ships,” *IEEE Access*, vol. 9, pp. 157316–157328, 2021, doi: 10.1109/ACCESS.2021.3130741.

[42] N. Smirnov and S. Tomforde, “Navigation Support for an Autonomous Ferry Using Deep Reinforcement Learning in Simulated Maritime Environments,” in *Proceedings - 2022 IEEE International Conference on Cognitive and Computational Aspects of Situation Management, CogSIMA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 142–149. doi: 10.1109/CogSIMA54611.2022.9830689.

[43] V. Ronca *et al.*, “Neurophysiological Assessment of An Innovative Maritime Safety System in Terms of Ship Operators’ Mental Workload, Stress, and Attention in the Full Mission Bridge Simulator,” *Brain Sci.*, vol. 13, no. 9, Sep. 2023, doi: 10.3390/brainsci13091319.

[44] Southampton Solent University, “Warsash Maritime Autonomous Surface Ships (MASS) Research Centre.” Accessed: Oct. 15, 2025. [Online]. Available: https://www.solent.ac.uk/research/research-areas/environment-and-engineering/warsash-maritime-autonomous-surface-ships-mass-research-centre

[45] M. G. Lokuketagoda, T. Miwa, S. Jayasinghe, and P. Dev Ranmuthugala, “Training engineers for remotely operated ships of the future,” 2018. Accessed: Oct. 31, 2025. [Online]. Available: https://figshare.utas.edu.au/articles/conference_contribution/Training_engineers_for_remotely_operated_ships_of_the_future/23099006?file=40827965

[46] G. Reza Emad and A. Kataria, “Challenges of simulation training for future engineering seafarers - A qualitative case study,” in *Human Factors in Transportation*, AHFE International, 2022. doi: 10.54941/ahfe1002501.

[47] Damitha Sandaruwan, Nihal Kodikara, Chamath Keppitiyagama, and Rexy Rosa, “A Six Degrees of Freedom Ship Simulation System for Maritime Education,” *The International Journal on Advances in ICT for Emerging Regions 2010 03 (02) :34 - 47*, Mar. 2010.

[48] K. Wróbel, M. Gil, Y. Huang, and R. Wawruch, “The Vagueness of COLREG versus Collision Avoidance Techniques—A Discussion on the Current State and Future

Challenges Concerning the Operation of Autonomous Ships," Dec. 01, 2022, *MDPI*. doi: 10.3390/su142416516.

[49] J. Chan, D. Golightly, R. Norman, and K. Pazouki, "Perception of Autonomy and the Role of Experience within the Maritime Industry," *J. Mar. Sci. Eng.*, vol. 11, no. 2, Feb. 2023, doi: 10.3390/jmse11020258.

[50] A. Kennard, P. Zhang, and S. Rajagopal, "Technology and training: How will deck officers transition to operating autonomous and remote-controlled vessels?," *Mar. Policy*, vol. 146, Dec. 2022, doi: 10.1016/j.marpol.2022.105326.

[51] S. Jo, E. D'agostini, and J. Kang, "From seafarers to e-farers: Maritime cadets' perceptions towards seafaring jobs in the industry 4.0," *Sustainability (Switzerland)*, vol. 12, no. 19, pp. 1–18, Oct. 2020, doi: 10.3390/su12198077.

[52] J. Nasur and K. Bogusławski, "Awareness and attitude of maritime students towards the introduction of autonomous merchant ships – preliminary results," *TransNav*, vol. 14, no. 4, pp. 859–863, 2020, doi: 10.12716/1001.14.04.10.

[53] E. Hannaford and E. Van Hassel, "Risks and benefits of crew reduction and/or removal with increased automation on the ship operator: A licensed deck officer's perspective," *Applied Sciences (Switzerland)*, vol. 11, no. 8, Apr. 2021, doi: 10.3390/app11083569.

[54] P. Trucco, E. Cagno, F. Ruggeri, and O. Grande, "A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation," *Reliab. Eng. Syst. Saf.*, vol. 93, no. 6, pp. 845–856, 2008, doi: 10.1016/j.ress.2007.03.035.

[55] S. D. Kristensen, R. Skjetne, and I. B. Utne, "Comparison of methods for using risk information in path planning for autonomous surface vessels," in *Journal of Physics: Conference Series*, Institute of Physics, 2024. doi: 10.1088/1742-6596/2867/1/012039.

[56] J. D. Palbar Misas, R. Hopcraft, K. Tam, and K. Jones, "Future of maritime autonomy: cybersecurity, trust and mariner's situational awareness," *Journal of Marine Engineering and Technology*, vol. 23, no. 3, pp. 224–235, 2024, doi: 10.1080/20464177.2024.2330176.

[57] P. Li, Y. Wang, and Z. Yang, "Risk assessment of maritime autonomous surface ships collisions using an FTA-FBN model," *Ocean Engineering*, vol. 309, Oct. 2024, doi: 10.1016/j.oceaneng.2024.118444.

[58] M. Aydin, E. Akyuz, O. Turan, and O. Arslan, "Validation of risk analysis for ship collision in narrow waters by using fuzzy Bayesian networks approach," *Ocean Engineering*, vol. 231, Jul. 2021, doi: 10.1016/j.oceaneng.2021.108973.

[59] C. Guo, S. Haugen, and I. B. Utne, "Risk assessment of collisions of an autonomous passenger ferry," *Proc. Inst. Mech. Eng. O J. Risk Reliab.*, vol. 237, no. 2, pp. 425–435, Apr. 2023, doi: 10.1177/1748006X211050714.

[60] X. Zhao, H. Yuan, and Q. Yu, "Autonomous vessels in the yangtze river: A study on the maritime accidents using data-driven bayesian networks," *Sustainability (Switzerland)*, vol. 13, no. 17, Sep. 2021, doi: 10.3390/su13179985.

[61] M. Zhang, D. Zhang, H. Yao, and K. Zhang, "A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration," *Saf. Sci.*, vol. 130, Oct. 2020, doi: 10.1016/j.ssci.2020.104838.

[62] C. Fan, V. Bolbot, J. Montewka, and D. Zhang, "Advanced Bayesian study on inland navigational risk of remotely controlled autonomous ship," *Accid. Anal. Prev.*, vol. 203, Aug. 2024, doi: 10.1016/j.aap.2024.107619.

[63] S. I. Sezer, S. Il Ahn, E. Akyuz, R. E. Kurt, and P. Gardoni, "A hybrid human reliability analysis approach for a remotely-controlled maritime autonomous surface ship (MASS- degree 3) operation," *Applied Ocean Research*, vol. 147, Jun. 2024, doi: 10.1016/j.apor.2024.103966.

[64] C.-H. Chang, C. Kontovas, Q. Yu, and Z. Yang, "Risk assessment of the operations of maritime autonomous surface ships," *Reliab. Eng. Syst. Saf.*, vol. 207, p. 107324, Mar. 2021, doi: 10.1016/j.ress.2020.107324.

[65] P. Weber, G. Medina-Oliva, C. Simon, and B. Iung, "Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas," *Eng. Appl. Artif. Intell.*, vol. 25, no. 4, pp. 671–682, 2012, doi: 10.1016/j.engappai.2010.06.002.

[66] R. P. de Oliveira, G. Carim Junior, B. Pereira, D. Hunter, J. Drummond, and M. Andre, "Systematic Literature Review on the Fidelity of Maritime Simulator Training," Nov. 01, 2022, *MDPI*. doi: 10.3390/educsci12110817.

[67] I. Mraković and R. Bošnjak, "Comparative Analyses of Manoeuvring Patterns in Real and Virtual AtoN Environment," *TransNav*, vol. 17, no. 3, pp. 617–623, Sep. 2023, doi: 10.12716/1001.17.03.13.

[68] I. Toman, P. Vidan, M. Barić, and L. Grbić, "Realism of TRANSAS NTPRO 5000 Radar Simulation in Search and Rescue Training Exercises," *Transactions on Maritime Science*, vol. 9, no. 1, pp. 99–105, Apr. 2020, doi: 10.7225/toms.v09.n01.009.

[69] M. Baldauf, M. Kitada, R. Mehdi, and D. Dalaklis, "E-Navigation, Digitalization and Unmanned Ships: Challenges for Future Maritime Education and Training," in *INTED2018 Proceedings*, IATED, Mar. 2018, pp. 9525–9530. doi: 10.21125/inted.2018.2374.

[70] Y. Liu, X. Hou, O. Sourina, D. Konovessis, and G. Krishnan, "Human Factor Study for Maritime Simulator-Based Assessment of Cadets," ASME 2016 35th International Conference on Ocean, Offshore and Arctic Engineering, 2016. doi: 10.1115/OMAE2016-54772.

[71] Yara Knowledge group, "Yara Birkeland." Accessed: Sep. 24, 2025. [Online]. Available: https://www.yara.com/news-and-media/media-library/press-kits/yara-birkeland-press-kit

[72] O. Atik and O. Arslan, "Use of Eye Tracking for Assessment of Electronic Navigation Competency in Maritime Training," *J. Eye Mov. Res.*, vol. 12, no. 3, pp. 1–15, 2019, doi: 10.16910/jemr.12.3.2.

[73] R. Mao, G. Li, H. P. Hildre, and H. Zhang, "Analysis and Evaluation of Eye Behavior for Marine Operation Training - A Pilot Study," *J. Eye Mov. Res.*, vol. 12, no. 3, pp. 1–14, 2019, doi: 10.16910/jemr.12.3.6.

[74] W. Zhang, Y.-N. Zhou, Y. Zhou, and J.-G. Wang, "Validation of the watch-type HUAWEI WATCH D oscillometric wrist blood pressure monitor in adult Chinese,"

Blood Press. Monit., vol. 27, no. 5, pp. 353–356, Oct. 2022, doi: 10.1097/MBP.0000000000000608.

[75] Imotions, “Pupil Core glasses.” Accessed: Oct. 01, 2025. [Online]. Available: https://imotions.com/products/hardware/pupil-core/?srsltid=AfmBOoo_l0Cfns9DOtdRFKvfdiyUN6UFR4JfGNhIpvgBIeMjWYxEI8aV

[76] “Huawei watch D2.” Accessed: Apr. 23, 2026. [Online]. Available: https://consumer.huawei.com/levant/wearables/watch-d2/specs/

[77] I. Petrović and S. Vujičić, “Use of Eye-Tracking Technology to Determine Differences Between Perceptual and Actual Navigational Performance,” *J. Mar. Sci. Eng.*, vol. 13, no. 2, Feb. 2025, doi: 10.3390/jmse13020247.

[78] International Chamber of Shipping, Ed., *BRIDGE PROCEDURES GUIDE*, Sixth Edition. Marisec Publications, 2022. [Online]. Available: https://www.ics-shipping.org/publications/bridge-procedures-guide-sixth-edition

[79] Hydrographic Institute of the Republic of Croatia, “Chart no.47,HHI.” Accessed: Oct. 02, 2025. [Online]. Available: https://www.hhi.hr/vijesti/karta-lucki-plan-br-47-novo-izdanje

[80] J. Pearl, *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000.

[81] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. New York: Springer, 2007.

[82] U. B. Kjærulff and A. L. Madsen, *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. New York: Springer, 2008.

[83] M. Rausand, *Risk Assessment: Theory, Methods, and Applications*. Hoboken, NJ: John Wiley & Sons, 2011.

[84] E. Charniak, “Bayesian networks without tears,” *AI Mag.*, vol. 12, no. 4, pp. 50–63, Dec. 1991, doi: httsp:doi.org/10.1609/aimag.v12i4.918.

[85] L. Vojković, “MODEL OF RISK ASSESSMENT OF MARITIME ACCIDENTS ON PASSENGER VESSELS IN NON-LINEAR COASTAL NAVIGATION,” University of Rijeka, Faculty of Maritime Studies, 2019. Accessed: Feb. 09, 2026. [Online]. Available: https://urn.nsk.hr/urn:nbn:hr:187:270669

[86] S. Vujičić, “EVACUATION EFFICIENCY ASSESSMENT MODEL ON CRUISE SHIPS,” Rijeka, 2018. Accessed: Feb. 09, 2026. [Online]. Available: https://urn.nsk.hr/urn:nbn:hr:187:520652

[87] A. G. Eleye-Datubo, A. Wall, A. Saajedi, and J. Wang, “Enabling a Powerful Marine and Offshore Decision-Support Solution Through Bayesian Network Technique,” *Risk Analysis*, vol. 26, no. 3, pp. 695–721, Jun. 2006, doi: 10.1111/j.1539-6924.2006.00775.x.

[88] M. Nyström and K. Holmqvist, “An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data,” *Behav. Res. Methods*, vol. 42, no. 1, pp. 188–204, Feb. 2010, doi: 10.3758/BRM.42.1.188.

[89] A. T. Duchowski, *Eye Tracking Methodology*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-57883-5.

[90] T. A. Salthouse, “The processing-speed theory of adult age differences in cognition.,” *Psychol. Rev.*, vol. 103, no. 3, pp. 403–428, 1996, doi: 10.1037/0033-295X.103.3.403.

[91] C. Chauvin, S. Lardjane, G. Morel, J.-P. Clostermann, and B. Langard, “Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS,” *Accid. Anal. Prev.*, vol. 59, pp. 26–37, Oct. 2013, doi: 10.1016/j.aap.2013.05.006.

[92] E. Veitch, O. A. Alsos, T. Cheng, K. Senderud, and I. B. Utne, “Human factor influences on supervisory control of remotely operated and autonomous vessels,” *Ocean Engineering*, vol. 299, p. 117257, May 2024, doi: 10.1016/j.oceaneng.2024.117257.

[93] J. Rasmussen, “Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, no. 3, pp. 257–266, May 1983, doi: 10.1109/TSMC.1983.6313160.

[94] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, Mar. 1995, doi: 10.1518/001872095779049543.

[95] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change.," *Psychol. Rev.*, vol. 84, no. 2, pp. 191–215, 1977, doi: 10.1037/0033-295X.84.2.191.

[96] K. L. Hassall *et al.*, "Facilitating the elicitation of beliefs for use in Bayesian Belief modelling," *Environmental Modelling & Software*, vol. 122, p. 104539, Dec. 2019, doi: 10.1016/j.envsoft.2019.104539.

[97] T. Kondo, S. Fujimoto, M. Rooks, T. Konishi, and N. Yasuta, "A Study on the Implications of Maritime Autonomous Surface Ships on Maritime Education and Training," *TransNav*, vol. 19, no. 4, pp. 1179–1185, Dec. 2025, doi: 10.12716/1001.19.04.15.

[98] T. Porathe, J. Prison, and Y. Man, "Situation Awareness in Remote Control Centres for Unmanned Ships," in *Human Factors in Ship Design & Operation*, RINA, Feb. 2014, pp. 105–114. doi: 10.3940/rina.hf.2014.12.

[99] M. Bachari-Lafteh and A. Harati-Mokhtari, "Operator's skills and knowledge requirement in autonomous ships control centre," *Journal of International Maritime Safety, Environmental Affairs, and Shipping*, vol. 5, no. 2, pp. 74–83, Apr. 2021, doi: 10.1080/25725084.2021.1949842.

[100] D. C. Niehorster, T. H. W. Cornelissen, K. Holmqvist, I. T. C. Hooge, and R. S. Hessels, "What to expect from your remote eye-tracker when participants are unrestrained," *Behav. Res. Methods*, vol. 50, no. 1, pp. 213–227, Feb. 2018, doi: 10.3758/s13428-017-0863-0.

[101] C. H. Vinkers *et al.*, "Stress-induced hyperthermia and infection-induced fever: Two of a kind?," *Physiol. Behav.*, vol. 98, no. 1–2, pp. 37–43, Aug. 2009, doi: 10.1016/j.physbeh.2009.04.004.

[102] T. Oka, "Psychogenic fever: how psychological stress affects body temperature in the clinical population," *Temperature*, vol. 2, no. 3, pp. 368–378, Jul. 2015, doi: 10.1080/23328940.2015.1056907.

[103] N. Hjortskov, D. Rissen, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Sogaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *Eur. J. Appl. Physiol.*, vol. 92, no. 1–2, pp. 84–89, Jun. 2004, doi: 10.1007/s00421-004-1055-z.

[104] Y. Chida and A. Steptoe, "Greater Cardiovascular Responses to Laboratory Mental Stress Are Associated With Poor Subsequent Cardiovascular Risk Status," *Hypertension*, vol. 55, no. 4, pp. 1026–1032, Apr. 2010, doi: 10.1161/HYPERTENSIONAHA.109.146621.

[105] A. C. Phillips, A. T. Ginty, and B. M. Hughes, "The other side of the coin: Blunted cardiovascular and cortisol reactivity are associated with negative health outcomes," *International Journal of Psychophysiology*, vol. 90, no. 1, pp. 1–7, Oct. 2013, doi: 10.1016/j.ijpsycho.2013.02.002.

[106] G. Robert J. Hockey, "Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework," *Biol. Psychol.*, vol. 45, no. 1–3, pp. 73–93, Mar. 1997, doi: 10.1016/S0301-0511(96)05223-4.

[107] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, Jan. 2015, doi: 10.1080/00140139.2014.956151.

[108] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *Int. J. Ind. Ergon.*, vol. 24, no. 6, pp. 631–645, Oct. 1999, doi: 10.1016/S0169-8141(98)00068-7.

[109] V. Skaramagkas *et al.*, "Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 260–277, 2023, doi: 10.1109/RBME.2021.3066072.

[110] K. Krejtz, T. Szmidt, A. T. Duchowski, and I. Krejtz, "Entropy-based statistical analysis of eye movement transitions," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, New York, NY, USA: ACM, Mar. 2014, pp. 159–166. doi: 10.1145/2578153.2578176.

[111] B. C. Motter and E. J. Belky, “The guidance of eye movements during active visual search,” *Vision Res.*, vol. 38, no. 12, pp. 1805–1815, Jun. 1998, doi: 10.1016/S0042-6989(97)00349-0.

[112] J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, and J. R. Brannan, “Eye movement indices of mental workload,” *Acta Psychol. (Amst).*, vol. 75, no. 1, pp. 75–89, Oct. 1990, doi: 10.1016/0001-6918(90)90067-P.

[113] S. M. Wierda, H. van Rijn, N. A. Taatgen, and S. Martens, “Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8456–8460, May 2012, doi: 10.1073/pnas.1201858109.

[114] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, May 2000, doi: 10.1109/3468.844354.

[115] G. Rutkowski, “Practical aspects of using radar in maritime navigation with analysis of its potential errors and limitations,” *Scientific Journal of Gdynia Maritime University*, no. 113, pp. 17–28, Mar. 2020, doi: 10.26408/113.02.

[116] M. Ding, W. Chen, and W. Ding, “Performance analysis of a normal GNSS receiver model under different types of jamming signals,” *Measurement*, vol. 214, p. 112786, Jun. 2023, doi: 10.1016/j.measurement.2023.112786.

[117] A. Grant, P. Williams, N. Ward, and S. Basker, “GPS Jamming and the Impact on Maritime Navigation,” *Journal of Navigation*, vol. 62, no. 2, pp. 173–187, Apr. 2009, doi: 10.1017/S0373463308005213.

[118] C. Hetherington, R. Flin, and K. Mearns, “Safety in shipping: The human element,” *J. Safety Res.*, vol. 37, no. 4, pp. 401–411, Jan. 2006, doi: 10.1016/j.jsr.2006.04.007.

[119] Y. Ma, Q. Liu, L. Yang, and M. He, “Recognition of marine navigators’ workload based on eye movement features using bridge simulation,” *Saf. Sci.*, vol. 178, p. 106607, Oct. 2024, doi: 10.1016/j.ssci.2024.106607.

[120] J. P. Chan, R. Norman, K. Pazouki, and D. Golightly, “Autonomous maritime operations and the influence of situational awareness within maritime navigation,” *WMU Journal of Maritime Affairs*, vol. 21, no. 2, pp. 121–140, Jun. 2022, doi: 10.1007/s13437-022-00264-4.

[121] H. Xue, J.-F. Røds, Ø. Haugseggen, A. J. Christensen, B.-M. Batalden, and O. T. Gudmestad, “A study on the effects of rapid training method related to ship handling in decision-making skills under stressful situations,” *Journal of Navigation*, vol. 77, no. 3, pp. 395–414, May 2024, doi: 10.1017/S0373463324000213.

[122] A. S. Kamis, A. F. A. Fuad, A. Ashaari, C. W. M. Noor, and S. A. Ali, “Development of WOP Mathematical Model for Optimum Track-Keeping. A Ship Simulation Study Using VLCC, Focusing on Hard Over Rudder Turning Circle with Three Stages of Validation Analysis,” *Polish Maritime Research*, vol. 28, no. 3, pp. 156–174, Sep. 2021, doi: 10.2478/pomr-2021-0043.

[123] K. S. Gould, B. K. Røed, E.-R. Saus, V. F. Koefoed, R. S. Bridger, and B. E. Moen, “Effects of navigation method on workload and performance in simulated high-speed ship navigation,” *Appl. Ergon.*, vol. 40, no. 1, pp. 103–114, Jan. 2009, doi: 10.1016/j.apergo.2008.01.001.

[124] G. Múller-Plath *et al.*, “How Does Maritime Situation Awareness Depend on Navigation Automation and Mental Workload? A Sea Simulator Experiment,” *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 17, no. 4, pp. 963–969, 2023, doi: 10.12716/1001.17.04.23.

[125] INTERNATIONAL MARITIME ORGANIZATION, “GUIDANCE ON NEAR-MISS REPORTING MSC-MEPC.7-Circ.7.” Accessed: Apr. 09, 2026. [Online]. Available: https://www.register-iri.com/wp-content/uploads/MSC-MEPC.7-Circ.7.pdf

[126] J. Montewka, F. Goerlandt, and P. Kujala, “Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents,” *Ocean Engineering*, vol. 40, pp. 50–61, Feb. 2012, doi: 10.1016/j.oceaneng.2011.12.006.

[127] H. P. Berg, “Human Factors and Safety Culture in Maritime Safety (revised),” *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 7, no. 3, pp. 343–352, 2013, doi: 10.12716/1001.07.03.04.

[128] K. M. Lynch, V. A. Banks, A. P. J. Roberts, S. Radcliffe, and K. L. Plant, “What factors may influence decision-making in the operation of Maritime autonomous surface ships? A systematic review,” *Theor. Issues Ergon. Sci.*, vol. 25, no. 1, pp. 98–142, Jan. 2024, doi: 10.1080/1463922X.2022.2152900.

[129] Heckerman David, “A Tutorial on Learning with Bayesian Networks (MSR-TR-95-06),” 1995. Accessed: Apr. 20, 2026. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-85066-3_3

[130] Y. Zhou, N. Fenton, and C. Zhu, “An empirical study of Bayesian network parameter learning with monotonic influence constraints,” *Decis. Support Syst.*, vol. 87, pp. 69–79, Jul. 2016, doi: 10.1016/j.dss.2016.05.001.

[131] E. Castillo, J. M. Gutierrez, and A. S. Hadi, “Sensitivity analysis in discrete Bayesian networks,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 27, no. 4, pp. 412–423, Jul. 1997, doi: 10.1109/3468.594909.

[132] T. G. Eschenbach, “Spiderplots versus Tornado Diagrams for Sensitivity Analysis,” *Interfaces (Providence).*, vol. 22, no. 6, pp. 40–46, Dec. 1992, doi: 10.1287/inte.22.6.40.

LIST OF FIGURES AND DIAGRAMS

LIST OF TABLES

LIST OF ABBREVIATIONS

AIS	Automatic Identification System
ANS	Autonomous Navigation Systems
AOI	Area of Interest
AtoN	Aids to Navigation
BN	Bayesian Network
COLREG	International Convention for the Prevention of Collisions at Sea
CPA	Closest Point of Approach
CPT	Conditional Probability Table
CREAM	Cognitive Reliability and Error Analysis Method
CSV	Comma-Separated Values
DAG	Directed Acyclic Graph
DBP	Diastolic Blood Pressure
EBL	Electronic Bearing Line
EBP	Experience-Building Phase
ECDIS	Electronic Chart Display and Information System
ER	Evidential Reasoning
FMEA	Failure Mode and Effects Analysis
FTA-FBN	Fault Tree and Fuzzy Bayesian Network
FSR	Fixation-to-Saccade Ratio
GeNIe	Graphical Network Interface
GDI	Gaze Dispersion Index
GMS	Gaze Movement Score
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
HMIs	HMIs Human–Machine Interfaces
HR	Heart Rate
HRV	Heart Rate Variability
IALA	International Association of Marine Aids to Navigation and Lighthouse Authorities

ICS	International Chamber of Shipping
IMO	International Maritime Organization
IMO MASS Code	IMO International Code of Safety for MASS
MASS	Maritime Autonomous Surface Ships
min CPA	minimum Closest Point of Approach
MSC	Maritime safety Committee
NAUTiCA	Navigational Autonomous Training and Competence Advancement
OSV	Offshore Support Vessel
PEI	Pupil Entropy Index
PI	Parallel Indexing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RBN	Recursive Bayesian Networks
ROC	Remote Operation Centres
ROT	Rate of Turn
6DoF	Six Degrees of Freedom
SBP	Systolic Blood Pressure
SGPS	Stress-induced Gaze Pattern Score
SOLAS	International Convention for the Safety of Life at Sea
STCW	International Convention on Standards of Training, Certification and Watchkeeping for Seafarers
TCPA	Time to Closest Point of Approach
TEU	Twenty-foot Equivalent Unit
VAQ	Visual Attention Quality
VRM	Variable Range Marker

LIST OF SYMBOLS

Δ	Change in a measured variable
ΔBP	Change in blood pressure
ΔHR	Change in heart rate
ΔP(X)	Change in probability of variable X
ΔP_{High}	Change in probability of the “High” outcome
ΔP_{Poor}	Change in probability of the “Poor” outcome
GDI_{Max}	Maximum observed value of GDI
GDI_{Min}	Minimum observed value of GDI
H	Shannon entropy used to calculate PEI
i	Index of histogram bin
n	Total number of histogram bins
p (x_i)	Probability of value occurrence in the i-th bin
r	Radial dispersion of gaze
t	Time or observation window
x,y	Normalized gaze coordinates
X_i	Value of pupil diameter in the i-th bin
log_2	Logarithm with base 2
σ	Standard deviation
σ_x^2	Variance of gaze in horizontal direction
σ_y^2	Variance of gaze in vertical direction

APPENDICES

Appendices include additional materials and documentation that readers can optionally refer to for a deeper understanding of the work or the process that led to it. For example, this may include research materials (e.g., survey questions, stimuli, examples of standardised test questions, databases), supplementary data analyses, and other relevant supporting materials.

Appendix A: Survey Questionnaire Used in Data Collection

Pre-simulation Questionnaire

Dear Participant,

The aim of the study is to gain a better understanding of participants' knowledge, experience, attitudes, and perceptions regarding MASS, as well as their readiness to operate autonomous navigation systems.

This questionnaire is anonymous. The collected data will be used exclusively for scientific purposes within research conducted at the Faculty of Maritime Studies, including the preparation of a doctoral dissertation in the field of autonomous navigation. Participation is entirely voluntary, and you may withdraw from the study at any time without providing a reason and without any consequences.

Please answer the questions by selecting the provided options or by giving additional information where required.

Thank you in advance for your time and participation in this research.

For simplicity, the masculine grammatical form is used in this questionnaire. All expressions apply equally to people of all genders.

For pupils, students, and seafarers – total of 26 questions

1. Which participant group do you belong to?

☐ Maritime high school pupil

☐ Maritime faculty student

☐ Experienced seafarer (with sea-going experience)

2. Have you previously participated in simulations on a maritime simulator?

☐ Yes

☐ No

3. How familiar are you with the concept of Maritime Autonomous Surface Ships (MASS)?(1 = not familiar at all, 5 = excellent knowledge)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

4. Are you familiar with any example of an autonomous ship?

☐ Yes, please specify (optional): ___________________

☐ No

☐ Not sure

5. Do you believe that an autonomous ship can operate reliably without a crew on board?

☐ No, not at all

☐ Yes, partially / in limited situations

☐ Yes, completely

☐ Not sure

6. Do you think autonomous ships should be under continuous human supervision from the shore?

☐ Yes, at all times

☐ Yes, but only in complex or high risk situations

☐ No, the technology is reliable enough to operate without supervision

☐ Not sure

7. Please assess the importance of the following challenges for autonomous ships:

(1 = least significant challenge, 5 = most significant challenge)

System failures ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Cyber attacks ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Human factor in supervision ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Poor communication ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Other (please specify): ___________________ ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

8. Do you believe you will successfully cope with the challenges during the exercise?

(1 = not at all, 5 = completely)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

9. How prepared do you feel to participate in this exercise?

(1 = unprepared, 5 = fully prepared)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

10. How do you expect to react if a system failure or an unclear situation occurs during the exercise?

☐ I will remain calm and try to resolve the problem

☐ I will request assistance as soon as I feel uncertain

☐ I am not sure; it depends on the situation

11. How do you assess your ability to make decisions under pressure?

(1 = very poor, 5 = excellent)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

12. Have you previously used devices such as eye-tracking glasses or smartwatches in education or research?

☐ Yes

☐ No

13. Do you feel uncomfortable about the equipment you will be wearing during the exercise (e.g., glasses, watch)?

☐ Yes

☐ No

☐ I am not sure, this is my first time using such equipment

14. Have you ever experienced cardiovascular issues

(e.g., irregular heartbeat, tachycardia, chest pain, etc.)?

☐ Yes

☐ No

If yes, have you consulted a medical professional regarding this?

☐ Yes

☐ No

15. Are you currently taking any medication or therapy

(e.g., for heart conditions, epilepsy, psychological difficulties)?

☐ Yes

☐ No

If yes, please specify which and why (optional): ________________________

16. Have you ever experienced panic attacks, severe anxiety, loss of control, or similar symptoms?

☐ Yes, several times

☐ Yes, once

☐ No

☐ Not sure

17. Do you feel safe being alone in the simulator bridge room?

☐ Yes

☐ No

☐ Not sure

☐ I would prefer someone to be in the room with me

☐ I would like to know where the instructor is located so I can call them if needed

For experienced seafarers only (if you selected "experienced seafarer" in Question 1)

18. Please state your age and gender:

----

19. Please indicate your education level:

- Maritime high school ☐ Yes ☐ No

☐ Undergraduate degree (3 years, BSc)

☐ Graduate degree (5 years, MSc)

☐ PhD

☐ Other

20. How much total sea-going experience do you have?

_________years

21. What was your rank on your most recent ship?

☐ Master

☐ Officer

☐ Cadet

☐ Other: ___________

22. What type of vessel are you currently sailing on?

----

23. In which country is the company you are currently employed by registered?

----

24. Have you ever experienced a serious failure of one of the key ship systems in practice?

☐ Yes

☐ No

25. Which radar system manufacturers have you worked with on board (e.g. FURUNO, ANSCHÜTZ, SPERRY)? Please list the systems you are confident using.

____

26. Which ECDIS system manufacturers have you worked with on board (e.g. FURUNO, ANSCHÜTZ, SPERRY)? Please list the systems you are confident using.

____

Thank you for your participation!

BIOGRAPHY

Toni Meštrović is a doctoral candidate in Maritime Technologies at the Faculty of Maritime Studies, University of Split, Croatia, where he works as a research and teaching assistant in the Nautical Department. He obtained his Master's degree from the Faculty of Maritime Studies and has completed all courses for the doctoral programme. His research interests include MASS, simulation-based assessment of seafarer performance, human–machine interaction, and risk analysis in autonomous and remote ship operations. His doctoral research focuses on developing a simulation-based experimental model for evaluating operator performance in MASS operations using Bayesian Networks. He is actively involved in research projects, including the NAuTiCA project.

LIST OF PUBLISHED WORKS (Titles)

- Impact of the IHO Resolutions and Standards for ENCs on the Use of ECDIS
- A Comparative Analysis of Piracy in Somalia and Eastern Africa: Decadal Shifts and Emerging Trends
- Ship Manoeuvring in Shallow and Narrow Waters: Predictive Methods and Model Development Review
- Challenges for the Education and Training of Seafarers in the Context of Autonomous Shipping: Bibliometric Analysis and Systematic Literature Review
- The Influence of Shallow Waters on the Manoeuvring of Large Ships
- How Important Is Training in Marine Firefighting Equipment – SCABA?
- Challenges for Seafarers' Education and Training in the Context of Autonomous Ships Development