Chaal, Meriam; BahooToroody, Ahmad; Basnet, Sunil; Valdez Banda, Osiris; Goerlandt, Floris

Towards system-theoretic risk assessment for future ships: A framework for selecting Risk Control Options

Published in:
Ocean Engineering

DOI:
10.1016/j.oceaneng.2022.111797

Published: 01/09/2022

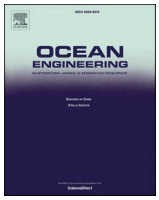
Document Version
Publisher's PDF, also known as Version of record

Ocean Engineering 259 (2022) 111797

Contents lists available at ScienceDirect

Ocean Engineering

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



Towards system-theoretic risk assessment for future ships: A framework for selecting Risk Control Options

Meriam Chaal [a,*], Ahmad Bahootoroody [a], Sunil Basnet [a], Osiris A. Valdez Banda [a], Floris Goerlandt [b]

[a] Marine and Arctic Technology Group, Department of Mechanical Engineering, Aalto University, Espoo, 11000, Finland
[b] Dalhousie University, Department of Industrial Engineering, Halifax, Nova Scotia, B3H 4R2, Canada

ARTICLE INFO

Keywords:
Risk control options
Bayesian network
STPA
Autonomous ships
Marine Formal Safety Assessment
Risk-based design

ABSTRACT

While the concept of smart shipping is expected to shape the future of the maritime industry, its safety is still a major concern. New risks might emerge when shifting from human controllers onboard, to autonomous software controllers and remote human controllers. The uncertainties associated with the emerging risks require an efficient decision-making methodology to ensure ship safety. This paper proposes a framework for selecting Risk Control Options (RCOs) of ships with higher degrees of autonomy in the context of marine risk assessment and Formal Safety Assessment (FSA). The framework uses the System Theoretic Process Analysis (STPA) for the hazard analysis and the identification of RCOs, while Bayesian Network is employed in the framework for estimating the system risk. Integrating STPA and BN offers the possibility to cover most of the steps of both risk assessment and FSA and permits the prioritization of the identified RCOs. The proposed method is applied to a concept of an autonomous seawater cooling system (SWC) as an illustrative case study. The results indicate that the RCOs including sensors health monitoring and software testing should be prioritized to reduce the risk. This is unveiled by the STPA analysis which shows the risk contribution of the associated causal scenarios.

1. Introduction

Risk assessment has long been practical for the decision-making process in different fields due to the successful experience that humans have with efficient resource allocation based on the risk assessment results (Modarres, 2016; Zhu et al., 2021). In the maritime domain, the introduction of ships with higher degrees of autonomy is expected to release the ship crew and reduce maritime operation risks (de Vos et al., 2021). Several aspects need to be covered before the introduction of ships with the highest level of autonomy. As such, the ship systems are first expected to be fitted with autonomous control capability to release the humans, still with shore-based human supervision and/or interference at least in the near future (Ramos et al., 2020). Ship systems with autonomous software controllers are considered complex systems that may introduce unpredicted hazards compared to conventional systems (Bolbot et al., 2019). Thus, assessing the risks of the future ship systems is necessary to make risk-based decisions during their development and design stages (Rolls Royce, 2016).

Maritime risk and safety are governed by the regulations and codes of the International Maritime Organization (IMO), which adopts a goal/risk-based approach to ship design. As a part of the rule-making process, IMO provides a risk-based five steps approach called the Formal Safety Assessment (FSA) framework to assess the risks of a ship design or operation and propose cost-effective options to control these risks (Breinholt et al., 2012; Skjong, 2011). In the context of FSA, quantitative risk assessment methods are preferred as they allow conducting a cost-benefit analysis to select the RCOs (IMO, 2018b). Similarly, in marine risk assessment, quantitative techniques are widely used (American Bureau of Shipping (ABS), 2020, p. 8). However, existing quantitative methods of risk assessment such as FTA, ETA, or FMEA are limited in identifying the potential accident scenarios of complex systems, detecting their causes, and effectively informing about the adequate controls (Aven, 2016). These methods cannot effectively capture the emerging hazards associated with system software and nonlinear component interactions (Sulaman et al., 2019). On the other hand, system-theoretic methods, notably STPA is designed to identify the hazards of complex and software-intensive systems (Leveson, 2011; Meng et al., 2018) but is not a part of the common methods used in the

* Corresponding author.
E-mail address: meriam.chaal@aalto.fi (M. Chaal).

https://doi.org/10.1016/j.oceaneng.2022.111797
Received 3 February 2022; Received in revised form 28 May 2022; Accepted 17 June 2022
Available online 15 July 2022

marine risk assessment techniques (American Bureau of Shipping (ABS), 2020, p. 8) or in the current FSA guidelines (IMO, 2018b). In addition, STPA is a hazard analysis technique in which, risk estimation, one of the fundamental elements of risk assessment, has not been taken into account (Bjerga et al., 2016).

Meanwhile, many researchers emphasized employing system theoretic methods in the risk assessment of autonomous ships considering their complexity and early design stages (Chaal et al., 2020; Montewka et al., 2018; Ventikos et al., 2020; Yang et al., 2020; Zhou et al., 2021). In a recent review study, Zhou et al. (2020) assessed the applicability of hazard analysis methods to autonomous ships and concluded that STPA can be considered as the most promising for the autonomous ships' design and operation. Another review study conducted by Thieme et al. (2018) argued that an adequate risk model for the Maritime Autonomous Surface Ships (MASS) should focus on assessing the control and software systems of the ship and that system-theoretic methods such as STPA and FRAM could be the most appropriate to use in this model. Moreover, Thieme et al. (2018) claimed that Bayesian Network (BN) is a suitable tool for risk modelling and estimation in the case of autonomous ships and should be considered as a part of the risk model of MASS. STPA has also been recently applied to autonomous ships in a number of scientific articles. Wróbel et al. (2019) presented a preliminary hazard identification for autonomous ships using STPA. Valdez Banda et al. (2019) proposed an STPA-based process to analyse the hazards of autonomous ships operating in the inland waterways and to derive the design options. Rokseth et al. (2019) applied STPA to define safety verification procedures for autonomous ships. Glomsrud and Xie (2019) presented a preparatory safety and security co-analysis of an unmanned ship using STPA. Yang et al. (2020) proposed a framework based on STPA to identify the unsafe transitions between different autonomy levels of marine systems in general. All the above-mentioned studies revealed the effectiveness of STPA in capturing the hazards related to the system component interaction and component unsafe behaviour in autonomous ships. They also demonstrated the use of STPA in guiding the design and managing the risks of autonomous ship operations. However, neither of these studies considered the risk quantification to estimate the level of safety to select adequate RCOs. This is since STPA does not cover the risk estimation. STPA has already been criticized for being a qualitative approach that does not consider risk quantification (Abaei et al., 2021; Ramos et al., 2020). To prioritize the control actions resulting from the STPA analysis, Gil et al. (2019) suggested evaluating the unsafe control actions identified by STPA based on the historical ship accident data. The authors highlighted that more research should be conducted to define a structured control action prioritization technique. In another recent study aiming to combine quantitative risk estimation with STPA, Bolbot et al. (2020) proposed a new safety analysis method named Combinatorial Approach for Safety Analysis (CASA) for cyber-physical systems. The method consisted of ten different steps and was designed by combining three safety analysis techniques: STPA, FTA, and ETA.

BNs, on the other hand, have been widely used to build risk models in different sectors, from process engineering (BahooToroody et al., 2019) to nuclear (DeJesus Segarra et al., 2021), aviation (Zhang and Mahadevan, 2021), civil engineering (Khalaj et al., 2020) and also the maritime field (Antão and Soares, 2019; Baksh et al., 2018). BN is one of the recommended methods for risk quantification in the FSA (IMO, 2018b) and was applied for the risk assessment and Risk Control Options selection when newly introducing substantial ship systems such as the Electronic Chart Display and Information System (ECDIS) (IMO and IMO, 2006). It was also employed to develop risk models in several important studies on maritime safety (Guedes Soares et al., 2009; Konovessis et al., 2013; Lu et al., 2019; Vassalos et al., 2010). BN has also been employed in a few studies for risk modelling of autonomous ships. Zhang et al. (2020) applied BN in integration with Event Trees to develop a probabilistic model for assessing human error in the shore-based control centres. Chang et al. (2021) applied BN in combination with FMEA and experts' opinions to estimate the risk of a set of hazards associated with the autonomous ship operation collected from the literature. Gao et al. (2021) mapped a Dynamic Fault Tree to a Dynamic Bayesian Network for assessing the reliability of unmanned vessels. Meanwhile, building a BN model is still challenging due to the lack of a clear foundation for the qualitative step of modelling especially in the context of novel technologies like autonomous shipping (Utne et al., 2020). The qualitative BN modelling in risk assessment can be compensated by integrating a qualitative method such as STPA. This integration targeting quantitative risk assessment has not yet been exploited; to the best of the authors' knowledge, only the framework presented by Utne et al. (2020) and latterly extended by Johansen and Utne (2022), highlighted the possibility of integrating the STPA analysis within a BN model where the authors aimed to use the model for the qualitative online risk assessment as a part of the autonomous ship real-time decision making. More recently Wang et al. (2022) have used the STPA control structure of an autonomous vehicle to assess the reliability of the vehicle using BN. The authors aimed to only extract the functional description from STPA without conducting the complete analysis.

This paper proposes a framework to extend the combination of STPA and BN to first, assess the risks of autonomous ship systems and second, define and select the RCOs that can mitigate the identified risks at the design phase. Doing so aims to remedy the shortcomings of both methods and exploit their strengths for a potentially system-theoretic quantitative risk assessment of future ship systems. The study also aims to provide an approach to extend the use of STPA in the risk-based IMO rule making process. The combination of STPA and BN allows the prioritization of the RCOs, which can be utilized in marine risk assessment and FSA applications when the available resources for controlling the risks are limited. The proposed framework is applied to a ship seawater cooling system as a case study. This system is substantial for the safe operation of the whole ship machinery nowadays and is expected to be onboard ships with higher degrees of autonomy. The main aim of this case study is to showcase the application of the proposed method for system-theoretic marine risk assessment, which can also be aligned for the purpose of FSA. The focus of this case study is to apply the methodology to identify the hazards that can arise if the system is automated and capable to operate without human intervention onboard. The aim is then to identify and prioritize the RCOs for risk mitigation and control.

2. The proposed framework in the context of marine risk assessment and FSA

The marine risk assessment is conducted for a certain marine system or operation to support decision-makers and bring the risks down to acceptable levels by implementing adequate controls. A marine risk assessment can be used either for risk-based design or as an alternative to demonstrate that a certain system or operation is compliant with prescriptive rules (American Bureau of Shipping (ABS), 2020, p. 1). FSA on the other hand is a part of the goal/risk-based regulatory rulemaking policy and is conducted for either regulation development or acceptance (American Bureau of Shipping (ABS), 2020, p. 1).

The main steps of a marine risk assessment are as follows (American Bureau of Shipping (ABS), 2020, p. 7).

- Risk identification: Identify the hazards, their causes, and potential consequences.
- Risk analysis: Determine the probabilities and consequences of the hazardous events.
- Risk evaluation: Use the risk analysis results and compare them with the risk acceptance criteria to make decisions.
- Risk treatment: Select one or more options of risk treatment and analyse their effectiveness. Risk treatment options may be selected

based on a balanced effort and cost of implementation against the benefits obtained.

On the other hand, IMO describes the FSA steps as follows (IMO,2018b).

- Problem definition: the ship system and its functions are represented with a generic model description.
- Hazard identification: the safety specialists investigate accidents and incidents data and examine the system model to identify the possible hazards leading to the accidents.
- Risk analysis: the hazards' causes and consequences are analysed, followed by a risk estimation and evaluation.
- Risk Control Options: the possible safety barriers, or Risk Control Measures (RCM) are defined and ranked by their potential to reduce the risks. Accordingly, Risk Control Option (RCO) is a set of RCMs that can mitigate the risks when implemented together.
- Cost-Benefit assessment: the cost of implementing each RCO is estimated, and the RCOs are ranked by both total cost and risk reduction potential.

While the purposes of marine risk assessment and FSA are different, the main steps of these two processes remain the same and are approximately the same for the risk assessment process as defined by the Society of Risk Analysis (SRA). Correspondingly, and in line with these processes, Fig. 1 summarizes the steps of the proposed methodology. As a risk assessment can be conducted using different techniques or combinations of techniques (American Bureau of Shipping (ABS), 2020), the proposed methodology can represent a technique of system-theoretic risk assessment for different marine applications. In the first part, the hazard analysis is conducted using STPA. The outcomes of STPA are then used to model the BN and conduct the risk analysis and RCOs steps. The outcomes of STPA are the losses, system-level hazards, the control structure, the UCAs, the causal scenarios, and the safety constraints or requirements (Leveson and Thomas, 2018). All these outcomes, except the control structure, are then applied in the qualitative risk analysis using BN. The control structure was used only for the STPA analysis. The qualitative risk analysis part is then followed by quantitative risk analysis, mainly by filling the Conditional Probability Tables (CPTs) of the BN using the available technical data and/or expert judgment data. Applying STPA in conjunction with BN covers almost all the steps of the marine risk assessment and FSA. However, the Cost-Benefit analysis step (with a dashed outline in Fig. 1) is not under the scope of the proposed methodology. Nevertheless, this step was partly covered when using the BN model to examine the RCOs effect on the likelihood of untoward events. Similarly, the uncertainty analysis is not covered by the proposed steps of the methodology. In addition, the risk evaluation step, which initially compares the identified risks against risk-acceptance criteria to treat only high-level risks, is modified in this framework. The system-theoretic aspect of the method suggests identifying and attempting the treatment of all the risks without an initial prioritization. This is because a causal scenario initially viewed with a minor risk contribution can cause different UCAs due to the interactions between the system components. Thus, the same causal scenario can have a higher risk contribution after the completion of the analysis. Therefore, the evaluation and prioritization in the proposed framework come after the risk treatment step, when all the risks are considered and the corresponding RCOs are defined. In this step, if the resources are limited, then the comparison and prioritization can support the decision-making.

A thorough explanation of the proposed framework steps is provided

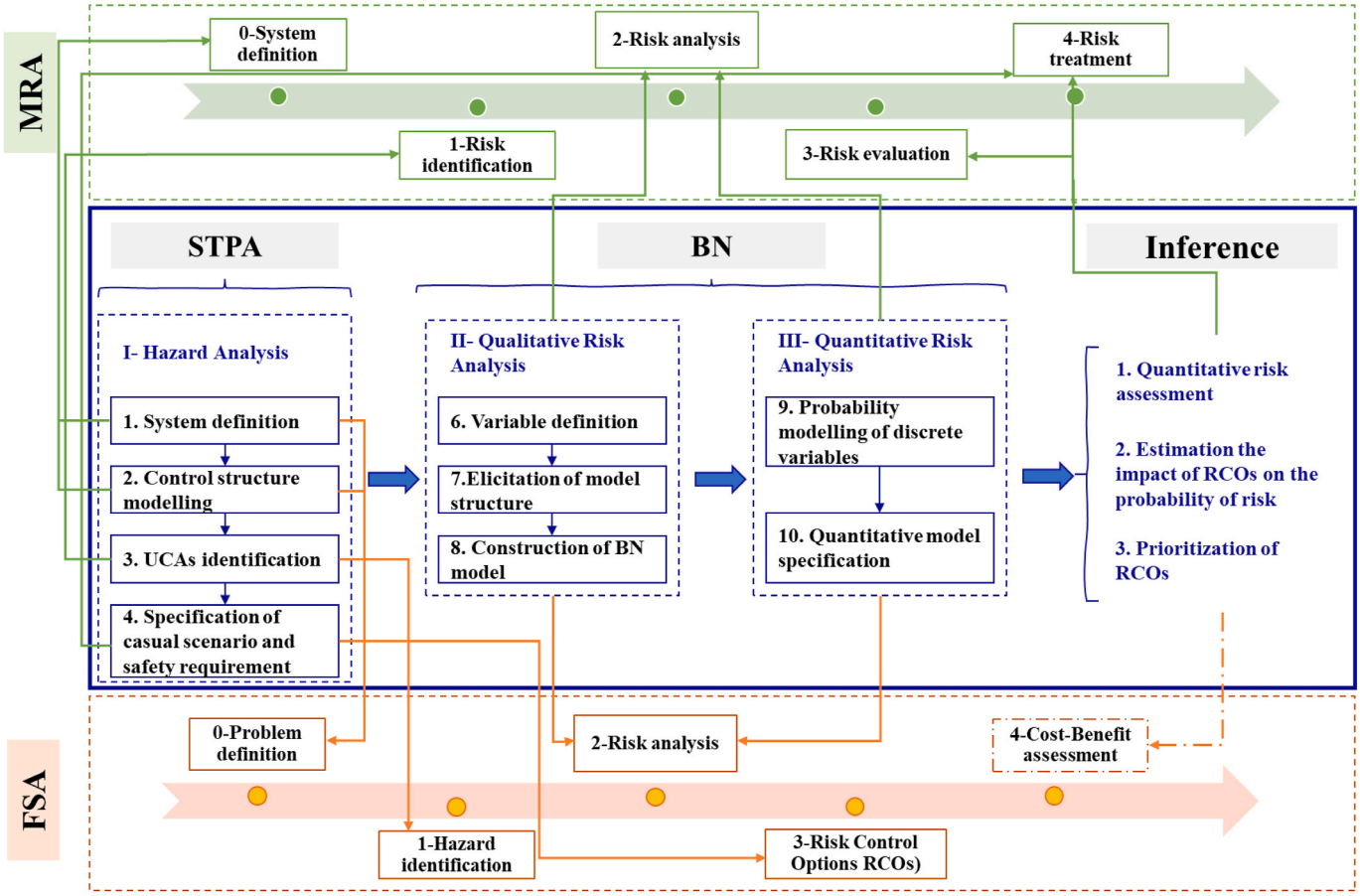


Fig. 1. The proposed STPA-BN framework.

in the following sections.

2.1. Hazard analysis; application of STPA

The first step of the proposed methodology (Fig. 1) is the hazard analysis. Through the application of STPA the losses, the system-level hazards, the Unsafe Control Actions (UCAs) (actions that can lead to the hazards), as well as the scenarios that might cause the UCAs to occur, can be explicitly identified. Given the identified losses, hazards, and UCAs, STPA can also provide the safety constraints to secure the safe operation and prevent specified scenarios. There are four steps in STPA as follows. A detailed discussion on hazard identification using STPA with a wide range of engineering applications is presented by (Leveson, 2011; Leveson and Thomas, 2018):

a. Define the system and the purpose of the analysis

In this step, the system and its boundaries need to be defined together with its function and the purpose of the analysis. The scope of the whole analysis is affected by the system definition and boundaries, which specify what is included in the system (components, sub-systems, context) and what is part of the environment. The purpose of the analysis must be defined in this step as well. Generally, the purpose of the analysis includes the identification of the losses and the system-level hazards considering the previously defined system boundaries. The losses involve value to the stakeholders and can be considered as the consequences of hazards, such as loss of human lives, loss of mission, damage to the property, the environment etc. Upon specifying the losses and system-level hazards, a safety constraint (aimed to prevent the hazard) needs to be determined for each system-level hazard.

b. Model the control structure of the system

In this step, a functional model of the system is developed, called Control Structure. Depending on the extent of the system, the control structure can include several layers to cover the sub-systems and their components. A control structure is an ensemble of feedback control loops that captures the control relationships and interactions between different system components. The controllers and controlled processes in each control loop are identified from the available system information. The system information is also used to define the control actions provided by the controllers to the controlled processes and the variables that describe the controlled processes' state. Depending on the controlled processes state, the controllers should provide adequate actions to enforce the safety constraints.

c. Identify Unsafe Control Actions

In this step, the control structure is analysed to identify how control actions could lead to the hazards and losses defined in the first step. A control action can be unsafe if it violates the safety constraints. Four different ways make a control action unsafe; a control action that, if provided, causes hazard, a control action that, if not provided, causes hazard, a control action that, if provided too late or too early, causes hazard, a control action that, if applied for too long or too short, causes hazard.

d. Identify causal scenarios and safety requirements

This step identifies the reasons behind the occurrence of unsafe control actions and the reasons why safe control actions might not be appropriately executed, leading to a hazard or a loss. The causal scenarios are then used to define the mitigations and create the safety requirements of the system. The causal scenarios can be failures related to the controller, inadequate control algorithm, missing or unsafe control input, inadequate process model, failures of the controlled process, etc.

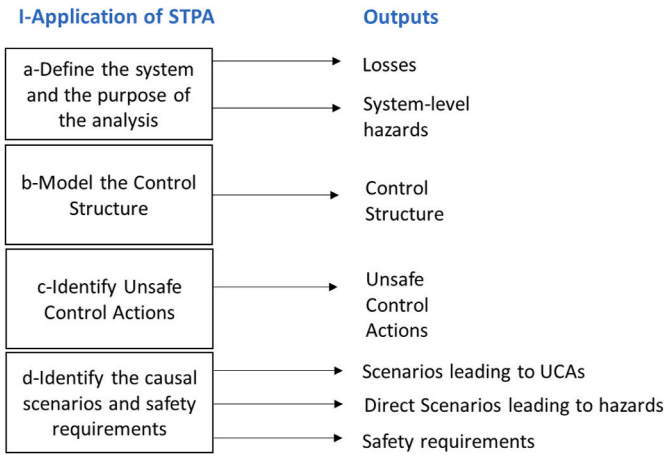


Fig. 2. Outputs of the application of STPA.

A summary of the outputs of the four steps of STPA is given in Fig. 2.

It should be noted that the first step of the STPA has a major role in framing the whole analysis. In the proposed system-theoretic risk assessment framework specifically, the first step affects the generation of the results that satisfy the aim of the analysis whether it is a marine risk assessment or an FSA. In the case of FSA, the generic ship model should be defined at the first step of STPA when defining the system.

2.2. Qualitative risk analysis; modelling the Bayesian Network

Herein a BN is developed using the outputs of STPA in the previous part of the methodology as given in Fig. 2. The STPA outputs used for the BN model are the losses (L), the system-level hazards (H), the Unsafe Control Actions (UCA), causal scenarios leading to UCAs (Sc), the direct causal scenarios leading to hazards (DSc) and the Safety Requirements. BN is a method based on the subjective (Bayesian) theory of probability and can combine statistical data with experts' opinions to estimate the risk (Kelly and Smith, 2009, 2011). It can quantify different notions of uncertainty; aleatoric (randomness) and epistemic (lack of knowledge) (BahooToroody et al., 2020; Goerlandt and Montewka, 2015). A BN is a Directed Acyclic Graph (DAG) that links different variables to represent their dependencies, including the cause-effect relationship (Barber, 2010; Neapolitan, 2004). It then constitutes a model for reasoning and answering various queries about a system (Nielsen and Jensen, 2009). The graph is a set of nodes and arcs, where each node represents a variable, and each arc represents a conditional dependency between the interlinked variables. The BN calculates the joint probability distribution of the variables using equation (1) (Barber, 2010):

$$P(U) = \prod_{i=1}^{n} P(X_i | pa(X_i)) \tag{1}$$

where $P(U)$ is the joint probability distribution, n is the number of variables, and $pa(X_i)$ is the parent set of variables for x_i.

The BN is modelled based on the cause-effect relationship between the STPA outputs. The outputs of STPA have a causal relationship, where losses are caused by system-level hazards, system-level hazards are caused by unsafe control actions and unsafe control actions are caused by causal scenarios (Antoine, 2013; Puisa et al., 2019). Fig. 3 shows how the STPA outputs are mapped into the BN model. In the BN model, the losses are accounted to play a pivotal role of consequences, linking to the causing system-level hazard as the top node of the network. Similarly, the system-level hazard node is linked to its parent nodes of UCAs and direct causal scenarios, and finally, the UCAs are linked to their respective causal scenarios. The safety requirements represent the Risk Control Measures (RCM) in the FSA framework and thus are added to the

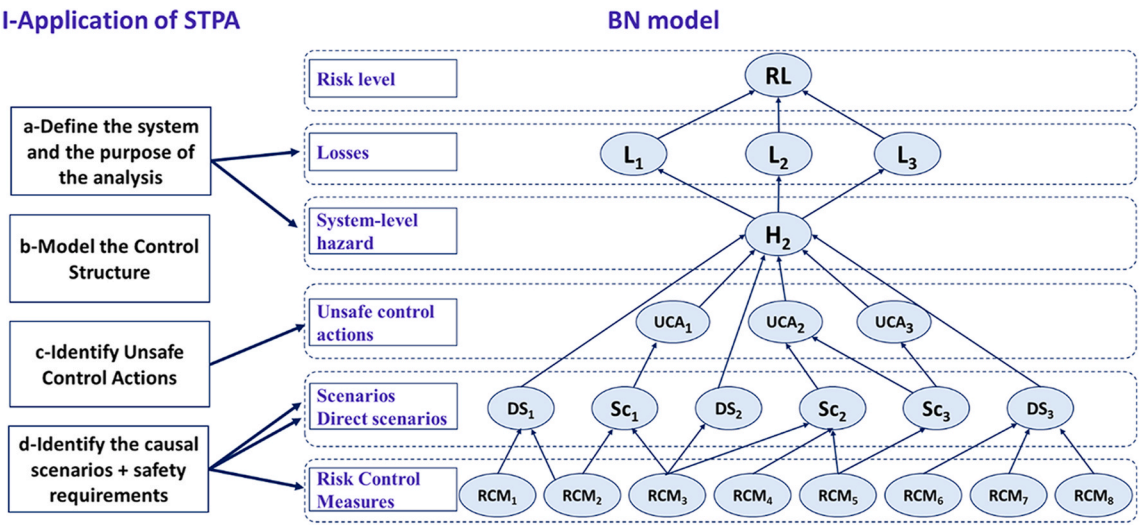


Fig. 3. Mapping STPA outputs into a BN model.

bottom level of the BN to prevent the occurrence of causal scenarios. As the losses are considered the final consequences in the STPA method, the probability of loss can define the risk level when the loss is estimated. The analysis aims to prevent the losses by design, using the safety requirements named RCMs in this study.

2.3. Quantitative risk analysis; filling the CPTs of the BN

This step aims to define each node's states and provide quantitative information by filling the CPTs. The reasoning process can then be achieved by propagating the probabilistic evidence in any direction in the model (top-down, down-top). The probabilistic evidence is propagated through the model to determine the probability of occurrence of each event (remarked by the nodes in BN) and, ultimately, the probability of risk levels. The states assigned to the different nodes of the BN are given in Table 1. The states of "high, medium and low" in risk level represent different levels of risk given the scope of the analysis and vary with the probability of occurrence of the associated hazard.

Regarding the loss node, the "yes" state means that the system will experience that loss. For the nodes of hazard, UCA and causal scenario, the state of "uncontrolled" and "occurred" represents the occurrence of a hazard, a UCA, or a causal scenario. For the nodes of RCM, the state of "implemented" means that the RCM is implemented/fulfilled in the system.

Table 1
States of the BN nodes.

Node	States
Risk level	High, Medium, Low
Loss	Yes, No
Hazard	Controlled, Uncontrolled
UCA	Denied, Occurred
Causal scenario (DSc, Sc)	Controlled, Occurred
Risk Control Measure (RCM)	Implemented, Not implemented

To fill the CPTs, the Boolean Logic with OR gates between UCAs, causal scenarios and the hazard was adopted. This is because a loss can be caused by one or more hazards, a hazard by one or more UCAs or direct causal scenarios, and a UCA by one or more causal scenarios. As for the CPTs of the risk level node, both expert judgment and operational data are considered to simulate the different levels of the risk. The data needed to fill the CPTs consists of two categories. The first category includes the probabilities of occurrence of the causal scenarios, which can be collected from historical data (literature and/or system manufacturers), test data, or experts' opinions. Depending on the nature of the causal scenario, the data type may differ, e.g., the probability of failure of a certain component or the probability of error of a certain controller. In case the data contains a failure rate, the exponential probability distribution is recommended by (Pui et al., 2017) to account for the randomness of the failure data. The exponential cumulative failure probability is given by (Leoni et al., 2021):

$$F(t) = \int_0^t f(t)dt = 1 - e^{-\lambda t} \tag{2}$$

where λ is the failure rate in time t. Measuring the changes of variables in both sides of Equation (2) with respect to the change of time results in Equation (3) as the exponential probability density function representing the reliability of a component or a controller (Pui et al., 2017):

$$f(t) = \lambda e^{-\lambda t} \tag{3}$$

Accordingly, the reliability of the system and the probability of failure can be achieved by equations (4) and (5), respectively (Pui et al., 2017):

$$R(t) = \int_t^\infty f(t)dt = e^{-\lambda t} \tag{4}$$

$$P(t) = 1 - e^{-\lambda t} \tag{5}$$

Depending on the analysis aim, the operation time (t) can be one year

or five years, as commonly used in FSA (IMO, 2018b).

The second category of the data covers the risk reduction potential of the RCMs. It represents the capability of the RCM to prevent causal scenarios.

As for the CPTs of risk level and RCM nodes, expert opinion and operation data are assigned to simulate different levels of the risk incorporated in the network as well as the impact of RCM on them.

3. Illustrative case study

An illustrative case study of a critical system in a ship machinery plant is offered to demonstrate how the developed methodology can be applied. The case study is analysing the system assuming that it operates fully independent of human intervention while remotely monitored by shore-based human controllers. This can be considered a remotely operated ship system at autonomy degree 3 as defined by IMO (2018a). Given this assumption, and to reflect better the capabilities of the proposed method, the seawater cooling system (SWC) has been selected since any neglected surge condition of SWC will lead to ship blackout, no matter how high the autonomy degree of the ship would be. SWC's role is to dissipate the heat transferred to the freshwater (intermediate cooling fluid) when producing the energy. Thus, it avoids damaging the machinery system components due to excessive heat. Therefore, daily inspections are currently required. While the MASS Concepts are still in the pre-study phase, SWC is expected to be onboard such future ships. According to the consulted industry experts, the SWC will likely remain a critical system from both operational and safety points of view even if MASS will run on different energy sources (Hydrogen, Ammonia, Batteries …). The SWC for such ships will supposedly be highly automated and capable of operation without human intervention while fitted with an additional remote operation mode. Considering the IMO definition of ship autonomy level 3, SWC should be capable of performing its function while shore-based human controllers can intervene only if required. The case study aims to showcase how the proposed methodology can define (based on the STPA study) and prioritize (based on BN calculations) the RCOs to mitigate the risks arising with the assumed mode of operation. To this end, GeNie software is used as a tool for modelling the BN of the proposed method (BayesFusion, 2020). A detailed discussion on the application of each step of the developed methodology in the case study is given in the ensuing sections.

3.1. System description

A SWC is illustrated in Fig. 4 depicting how the Fresh Water (FW) and the Sea Water (SW) cooling systems function together to dissipate the heat.

Fig. 5 shows a complete drawing of the standard ship SWC system considered in this study. This system is assumed to operate independent of onboard-human intervention and will be monitored by the shore-based crew. The components of the system are illustrated using the technical documentation of a passenger ship operating in the Mediterranean Sea. The SWC in Fig. 5 comprises seawater inlet and outlet valves, coolers' valves, coolers, seawater pumps, seawater pressure sensors, and seawater temperature sensors. The pumps circulate the seawater in an open circuit from the seawater inlet valves to the coolers and then, overboard. The heat is transferred from the fresh cooling water (intermediate cooling fluid) to the seawater through the main coolers' plates. The hot seawater is then discharged overboard. This system has three redundant pumps and two redundant coolers that can be isolated or put in service by switching their corresponding valves. The Differential Pressure (DP) sensor indicates the clogging condition of the cooler (if the cooler is dirty). The water pressure sensor and temperature sensor

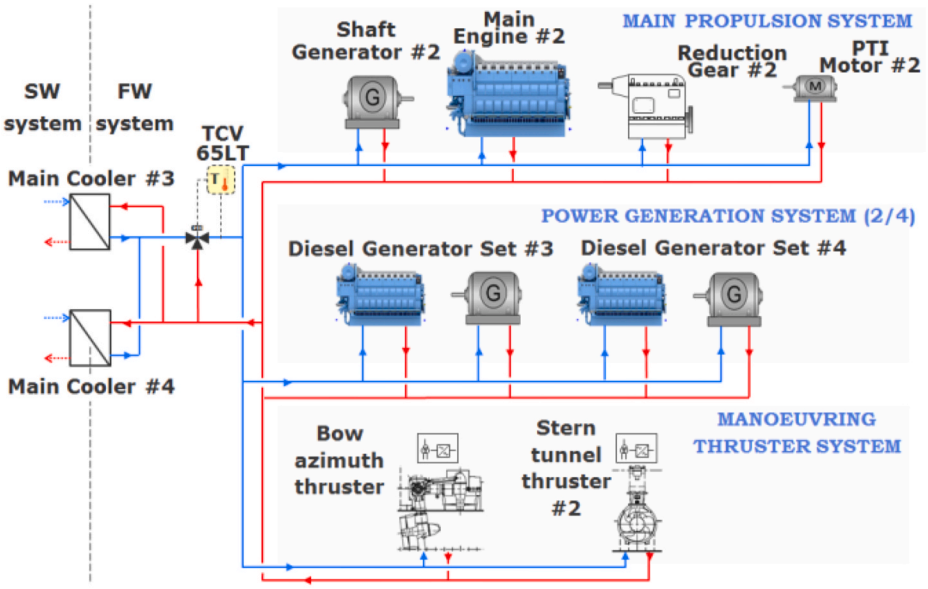


Fig. 4. Illustration of FW and SW cooling systems (Zymaris et al., 2016).

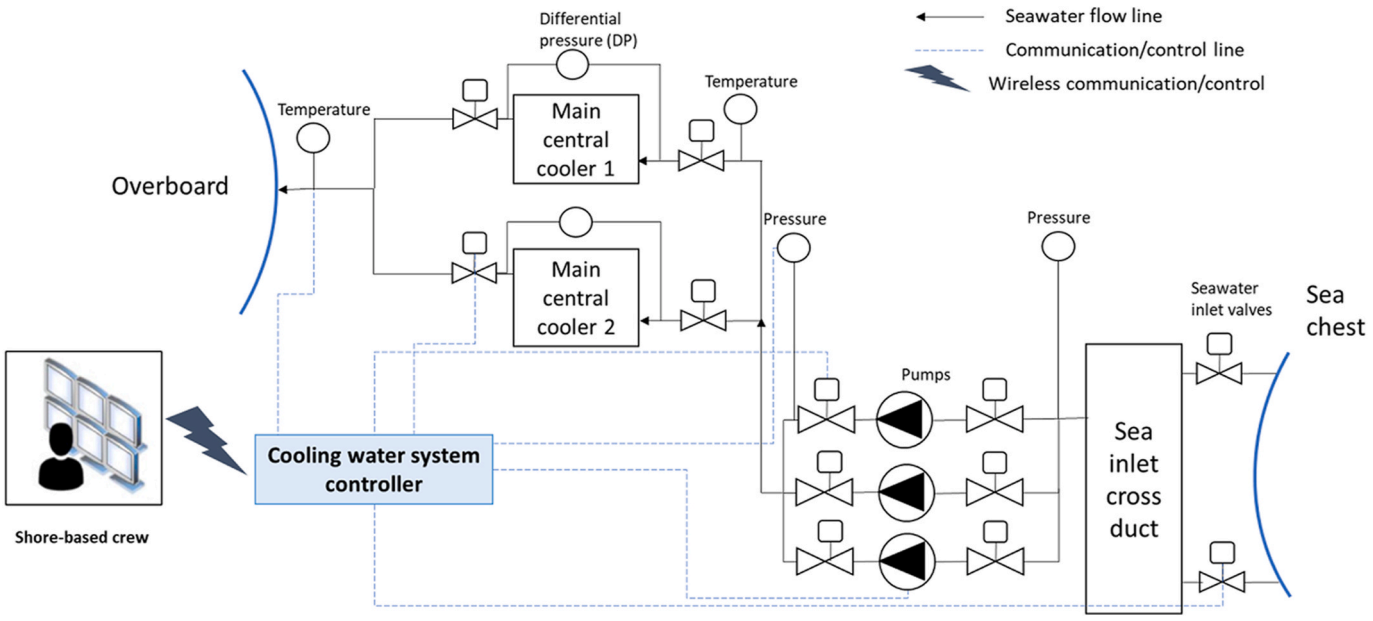


Fig. 5. Seawater cooling system (SWC) diagram of the reference ship.

indicate the current pressure and temperature of the water in the system. For the autonomous operation compared to the traditional process, the software controller of the system is responsible for the operation, alarm, and activation of all the system components. The controller receives the information from the system components and sensors. Instead of the onboard ship crew, a shore-based crew will monitor the SWC system and be able to take over the software controller when needed.

4. Results

4.1. Results of step I: Application of STPA to the seawater cooling system

a The system-level losses, hazards, and safety constraints

The SWC system as defined in the previous section is the subject of this STPA analysis. The operational context is defined as "operation during seagoing from Port A to Port B". Therefore, the analysis focuses on the prevention of the losses related to the SWC system and its function, which are three system-level losses as follows:

- Loss: (L1) non-severe damage to the seawater cooling system, (L2) loss of cooling mission, (L3) loss of cooling mission with severe damage to the SWC system.

Accordingly, the system-level hazard and system-level safety constraints were identified. Only one system-level hazard leading to the losses was identified.

- System-level hazard: (H) The seawater flow decreases below the safe limit.
- System-level safety constraint: The seawater flow shall not decrease below the defined safe limit.

The loss of cooling mission means that the system is no longer dissipating the heat coming from the machinery systems. A hazardous condition that can lead to this loss is (H), which reflects an insufficient seawater flow in the system and, consequently, inadequate heat dissipation. The safety constraint is then defined following the STPA procedure. This constraint is meant to prevent the hazard and should be respected when modelling the control structure of the system so that the flow is monitored and controlled to be consistently above the safe limit. In this type of system, the water pressure and temperature indicate that there is sufficient water flow.

b The control structure

As in this study, the system is supposed to operate without onboard human intervention, the control actions given by the humans onboard should be added to the control structure under the software controller. For this purpose, the control hierarchy and the responsibilities of the controllers were identified following Part 3 of the framework proposed by Chaal et al. (2020). In the framework, it is proposed that ship machinery operators should answer a questionnaire to identify the control structure elements. The answers to the questionnaire in the case of the ship SWC system are presented in the Appendix. Fig. 6 presents the resulting control structure of the seawater cooling system. In the conventional seawater cooling system, the crew onboard open and close the seawater inlet valves when the filters are clogged. The crew onboard are also responsible for switching between the coolers when required. In the autonomous operation, as shown in Fig. 6, the cooling water system controller is responsible for these actions instead of the crew. There are three levels of hierarchy, the shore-based crew is at the top of the hierarchy, the cooling water system controller is at the intermediate level, and the physical system processes at the lowest level. The shore-based crew as a controller of the system software controller can set the mode of operation to autonomous or remote. Under the autonomous mode, the system operates autonomously without interaction with the shore-based crew. Under the remote mode, the shore-based crew can take over the system controller to change setpoints or other actions. The arrows down represent the control actions given by the cooling water system controller to the cooler valves to switch between the cooler in service and the standby one. The controller also sends start/stop and close/open control actions to the pumps and the seawater inlet valves. The arrow up represents the feedback that the cooling water controller receives about the status of the controlled processes. The controller receives feedback through the corresponding sensors, such as the pressure and temperature sensors or valve position sensors. The water temperature and pressure indicate if the system functions properly; therefore, they should be maintained at safe levels by adequate control. The controller sends feedback about the system status with the generated alarms to the shore-based crew, who monitors the system operation.

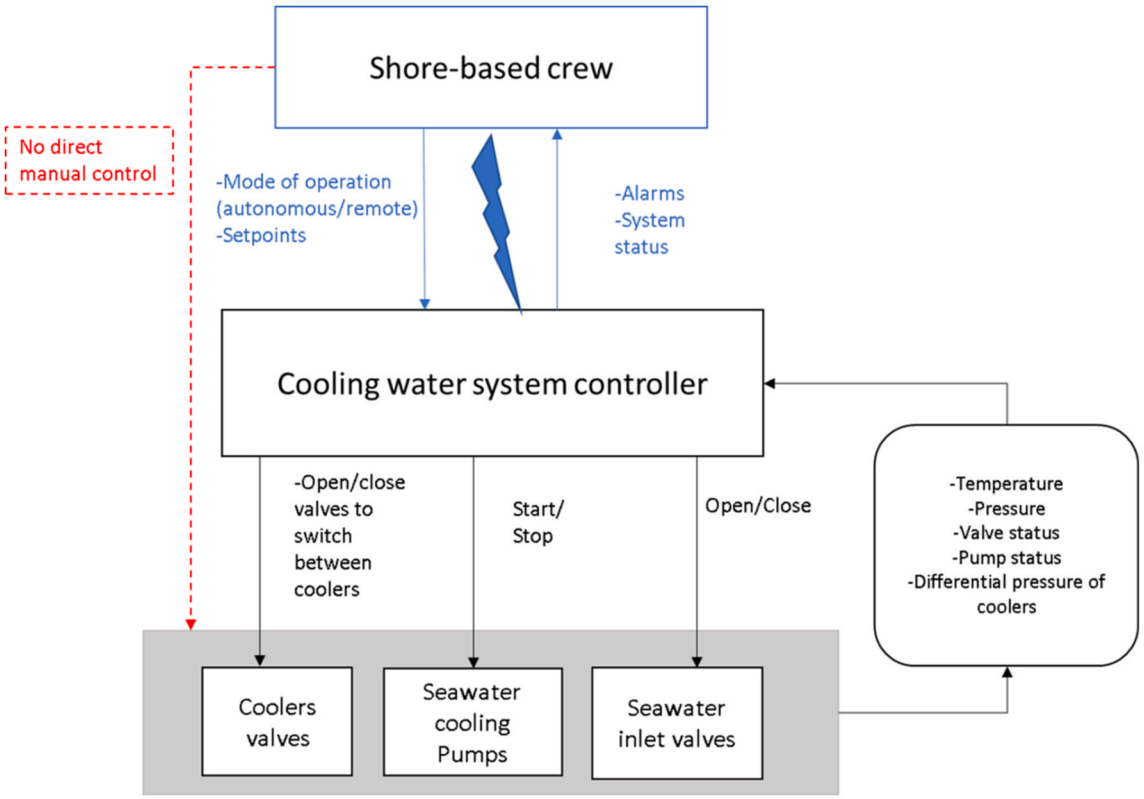


Fig. 6. Control structure of an autonomous seawater cooling system.

c The Unsafe Control Actions

The UCAs analysed in this case study are related to the control actions provided by the SWC controller that operates autonomously. Table 2 presents the seven UCAs that were identified during the analysis. Given this table, one point to be noticed is that the control action to open/close the cooler valves is critical because it is also required in case a cooler is defective (UCA-2). This was previously detected and handled by the crew onboard. In addition, assigning the autonomous control of seawater valves to the software controller might cause unsafe control actions such as UCA-3, which is critical during operation. Another point to notice from Table 2 is about the UCAs that are mainly of the type *"provided causing a hazard"* and *"not provided causing a hazard"*. These two guidewords were sufficient to demonstrate the application of the method in this case because most of the control actions sent to the physical components of the system, such as valves or pumps, are discrete control actions (not continuous in time).

Table 2
The identified Unsafe Control Actions.

Controller - Cooling water system controller		
Control action	Not providing causes hazard	Providing causes hazard
Open cooler valves	**UCA-1:** Controller does not open the standby cooler valves when the differential pressure is high. **UCA-2:** Controller does not open the standby cooler valves when there is a water leakage due to a defective main cooler.	Not Applicable (NA)
Close cooler valves	NA	**UCA-3:** Controller closes the cooler valves during normal operation.
Open seawater inlet valve	**UCA-4:** Controller does not open the standby seawater inlet valve when the inlet filter is clogged.	NA
Close seawater inlet valve	NA	**UCA-5:** Controller closes the seawater inlet valve during normal operation.
Start pump	**UCA-6:** Controller does not start the standby pump when the pump in service fails.	NA
stop pump	NA	**UCA-7:** Controller stops the pump during normal operation.

d The causal scenarios and safety requirements

There were nine identified causal scenarios belonging to two classes of scenarios. Table 3 shows the causal scenarios leading to UCA1 as an example. The scenarios are causing the occurrence of the UCAs, and the

Table 3
Identified causal scenarios for UCA1.

UCA	Scenarios related to UCA	Direct scenarios leading to hazard (not to UCA)
UCA1: Controller does not open the standby cooler valves when the differential pressure is high	**Sc1:** Controller does not receive feedback about the high DP due to a break in the feedback line	**DSc1:** Controller sends "open" command to the cooler valves, but the command is not executed due to valve failure
	Sc2: Controller receives wrong feedback about the DP due to a DP sensor problem	**DSc2:** Controller sends "open" command to the valve, but the valve does not receive it due to a break in the control line
	Sc3: Controller misinterprets the correct feedback about the DP due to a design flaw or change in the algorithm	
	Sc4: The controller hardware fails	

Table 4
The causal scenario types.

UCAs	Causal Scenario numbers	Causal Scenario types
UCA1	Sc1, Sc2, Sc3, Sc4, DSc1, DSc2	Software error, Sensor failure, Valve failure, controller hardware failure, communication line failure
UCA2	Sc5, DSc1, DSc2	Software error, valve failure, communication line failure
UCA3	Sc4, Sc6	Software error, controller hardware failure
UCA4	Sc1, Sc2, Sc3, Sc4, DSc1, DSc2	Sensor failure, controller hardware failure, Software error, communication line failure
UCA5	Sc4, Sc6	Software error, controller hardware failure
UCA6	Sc1, Sc2, Sc4, DSc2, DSc3	Software error, controller hardware failure, Pump failure, Communication line failure
UCA7	Sc4, Sc6	Software error, controller hardware failure

direct scenarios are driving the improper execution or non-execution of safe control actions.

Table 4 presents a summary of the identified causal scenarios for all the UCAs. The scenarios can also be described by their type as communication line failure, sensor failure, software error, controller hardware failure, or valve failure. Most of the identified causal scenarios are software errors due to the SWC controller type. This controller might send unsafe control actions to the controlled processes because of the wrong process algorithm or process model.

After identifying the causal scenarios, the possible RCMs that can prevent the scenarios are formulated with the help of the experts involved in the study. Table 5 presents an example of the RCMs to prevent the causal scenarios related to UCA1. As shown in Table 5, a causal scenario can be prevented by more than one RCM. In some cases, an RCM can prevent more than one causal scenario. In the formulated RCMs, the role of the shore-based crew was considered. The crew can act using a redundant control module, such as in the case of RCM9 in Table 5.

4.2. Step II and step III: BN model

The structured results of STPA were used to build the BN model presented in Fig. 7. The hazard (H – The seawater flow decreases below the safe limit) may lead to the three losses (L1 – non-severe damage to the seawater cooling system, L2 – loss of cooling mission and L3 – loss of cooling mission with severe damage to the seawater cooling system), which have been considered for the prediction of risk level associated with the system. On the other hand, the incorporated hazard node in the network is linked to the nodes of the seven identified UCAs and the three direct scenarios (DSc1, DSc2, DSc3). As presented in section 2.3, the first category of the data used in this study includes the failure rates of the physical components in the system and the software error rate. This data is collected from reference sources in the relevant literature, as depicted in Table 6. The data was fed into the BN model following the Boolean logic described in Section 2.3.

The data was used to fill the CPTs of the BN model and determine the probability of the UCAs, the hazard, the losses, and the risk level node. In the BN model shown in Fig. 7, the state of the RCM is set to "not implemented" for all the RCMs, which gives the risk picture before implementing the RCMs. The estimated risk profile highlights that the considered hazard will result in low, medium, and high risk with probabilities of 0.62, 0.17, and 0.21, respectively. Besides, the probability of the losses at this state is equal to 0.36, 0.27, and 0.18 for non-severe damage, loss of cooling mission, and severe damage to the system, respectively, which can be considered relatively high compared to the results of the traditional risk assessment methods.

To infer how different nodes' values affect the target node (system-level hazard), a sensitivity analysis was conducted using the GenIe software feature. The sensitivity analysis highlights the effect of each node on the risk level node (please see Fig. 8). As can be seen, the causal scenarios (except for the second scenario) appear in dark red, meaning

Table 5
The risk control measures for UCA1.

UCA	Causal scenarios	Safety requirements
UCA1: Controller does not open the standby cooler valves when the differential pressure is high	**Sc1:** Controller does not receive feedback about the high DP due to a break in the feedback line	RCM1: The cooling water system wirings shall be checked and maintained at adequate intervals to limit their failures. RCM2: The system shall generate an alarm.
	Sc2: Controller receives wrong feedback about the DP due to a DP sensor problem	RCM3: Two redundant DP sensors shall be installed. RCM4: The installed sensor shall have a low failure rate. RCM5: The condition of the sensors shall be monitored to anticipate their maintenance and prevent their failure. RCM6: The cooler shall be fitted with intelligent health monitoring features using different sensors (pressure, temperature) input.
	Sc3: Controller misinterprets the correct feedback about the DP due to a design flaw or change in the algorithm	RCM7: The software functionalities must be intensively tested during the design. RCM8: The controller algorithm functionalities shall be tested for the possible errors at adequate operation intervals and generate an alarm. RCM9: The shore-based crew shall have a redundant control module of the system
	Sc4: Failure of the controller hardware	RCM2: The controller shall generate an alarm. RCM9: The shore-based crew shall have a redundant control module of the system

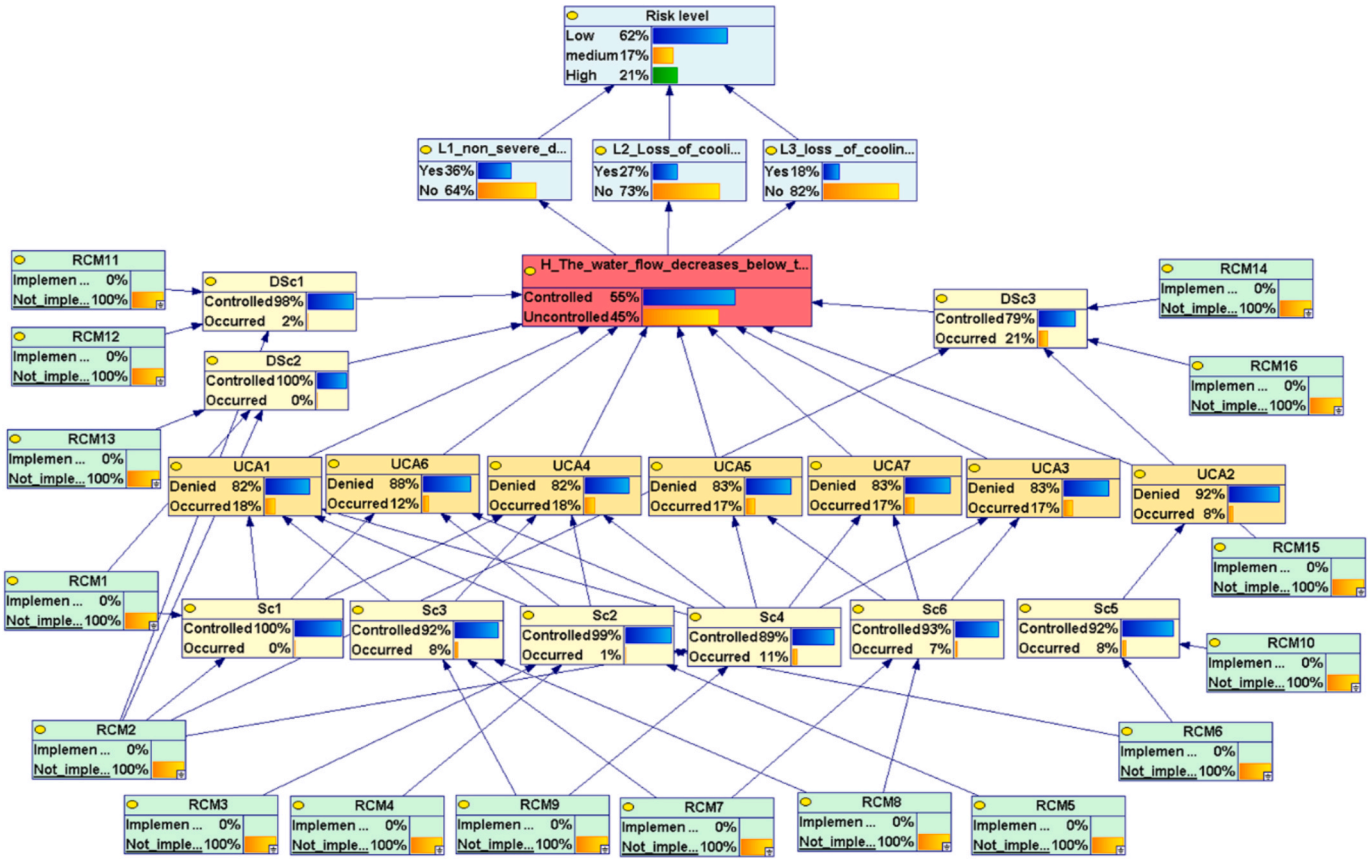


Fig. 7. The constructed BN model for the analysis of the seawater cooling system.

Table 6
Data sources.

Type of scenario	Rate (/h)	Source	Annual rate (/8760h)	Scenario probability
Communication line failure rate	2,50 × 10−8	Chai et al. (2016)	0,0002190	0,0002190
Pressure sensor failure rate	6,30 × 10-7	Schüller et al. (1997)	0,0055188	0,0055036
Error rate for software function	1,00 × 10-5	SINTEF (2006)	0,0876000	0,0838727
Controller hardware failure rate	1,50 × 10-5	SINTEF (2006)	0,1314000	0,1231330
Valve failure rate	1,80 × 10-6	Schüller et al. (1997)	0,0157680	0,0156443
Water pump failure rate	3,02 × 10-5	(OREDA companies, 2015)	0,2645520	0,2324503

they all have a crucial effect on the given hazard. This quality of interactions can be explained through the OR gate logic assigned in the network, where any of the scenarios leads to the connected UCAs and consequently to the hazard.

It should be noted that the analysis was conducted without setting any specific state for the RCMs to identify which RCMs have a more critical impact on the safety of the system. Correspondingly, different red colour shades are illustrated for the RCMs nodes, with RCM2, RCM7, RCM8, and RCM9 having the most significant contribution to system safety. It can be noticed from Table 5 that these critical RCMs are related to the system software safety, which is a critical component in the autonomous operation as it relies on the software controller. In addition, these RCMs prevent more than one scenario, which also explains their high impact on system safety. The same sensitivity analysis results show that the safety requirements RCM11, RCM12, RCM13, RCM1, RCM3, RCM4, and RCM5 have less impact on the system's safety.

The developed BN model in this study can prioritize the RCMs using sensitivity analysis (see Table 7). As a result, it is expected that a set of requirements including (RCM2)&(RCM7)&(RCM8)&(RCM9) can substantially improve the safety of the system if it is implemented. As a Risk

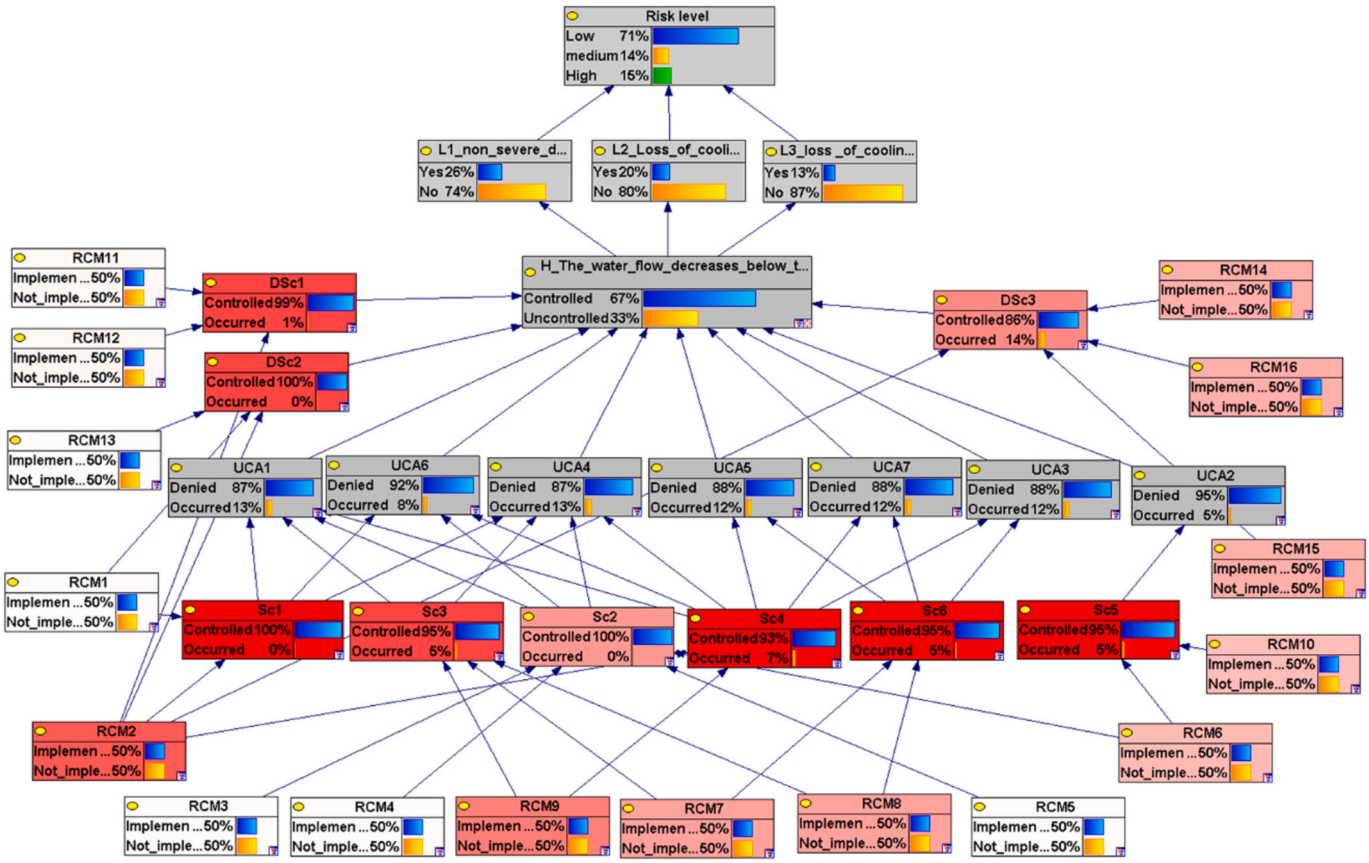


Fig. 8. Sensitivity analysis to identify the effect of the BN nodes on the risk level.

Table 7
Ranking of the risk control measures (RCMs).

Rank	Risk Control Measures
1	RCM1, RCM3, RCM4, RCM5, RCM11, RCM12, RCM13
2	RCM6, RCM10, RCM14, RCM15, RCM16
3	RCM2, RCM7, RCM8, RCM9

Control Option (RCO) is a set of RCMs, two sets of RCOs were defined; RCM2, RCM7, RCM8, RCM9 assigned to RCO1 and RCM11, RCM12, RCM13, R14 given to RCO2. The updating property of BN was applied to compare the posterior probability of the top node, assuming that different RCOs were implemented. Table 8 presents the results of this comparison where RCO1 outperforms RCO2 in reducing the probability of hazard as well as increasing the possibility of operation at a minimum possible risk level. As remarked in Table 8, the probability of hazard is expected to be reduced by 15.2% when applying RCO1 compared to 2,56% when implementing RCO2. The probability that the risk associated with the system will be at a low level increased by 11.7%, establishing RCO1. In comparison, this probability increases only by 1.5% given that RCO2 is hired, highlighting that RCO1 can be more efficient considering the safety level of the given system. This striking difference in the probability of risk level must draw the attention of designers, policymakers, and asset managers to the importance of the prioritization of RCOs and their influence on the overall risk profile.

To certify the system operates at a low-risk level, backward propagation through the developed risk-based STPA-BN model is used to estimate the critical probability values of the root nodes. The risk level was set at "low risk", and the backward propagation gave the estimations outlined in Table 9. The same approach is also practical if the system's strategy accepts more risks (medium or high level). As shown, the probability of the losses (L1, L2 and L3) had increased dramatically when the higher level of risk was assigned to the network. The probability of all nodes has been updated given that the system was exposed to a medium and high level of risks. This posterior probability of nodes can be accounted as the safe operational limit since experiencing a high level of risk through the system would not necessarily require all Scenarios, UCAs, and hazards to occur; rather, exceeding the predicted threshold assigned for each node will result in operating at the high level of risk.

The present study can also specify the UCAs and scenarios that have remarkable deviation through different levels of risk. This deviation informs the designers and other decision-makers about the UCAs and scenarios that can potentially lead to severe damage to the system. As shown in Fig. 9, DSc2, Sc1, and Sc2 have minimum deviation as the system is experiencing a higher level of risk, and on the contrary, DSc3, UCA4, and UCA1 have the maximum deviation.

Table 8
Impact of the implemented RCOs.

Node		Prior probability	Updated probability given RCO1	Updated probability given RCO2
Hazard		0.33	0.18	0.31
Risk level	Low	0,0520	0,0597	0,0527
	Medium	0.11	0.06	0.01
	High	0.15	0.08	0.14

Table 9
Critical probability values of nodes.

Node (state)	Prior probability	100% low risk level	100% medium risk level	100% high risk level
		Posterior probability	Posterior probability	Posterior probability
L1 (Yes)	0.2624	0.0663	0.6215	0.8611
L2 (Yes)	0.1968	0.0167	0.4661	0.7998
L3 (Yes)	0.1312	0.0002	0.2428	0.6448
H (Uncontrolled)	0.3280	0.1040	0.8010	0.9559
UCA1 (Occurred)	0.1278	0.0405	0.3122	0.3726
UCA2 (Occurred)	0.0509	0.0161	0.1244	0.1484
UCA3 (Occurred)	0.1167	0.0370	0.2849	0.3400
UCA4 (Occurred)	0.1278	0.0405	0.3122	0.3726
UCA5 (Occurred)	0.1167	0.0370	0.2849	0.3400
UCA6 (Occurred)	0.0790	0.0250	0.1931	0.2304
UCA7 (Occurred)	0.1167	0.0370	0.2849	0.3400
DSc1 (Occurred)	0.0110	0.0035	0.0269	0.0321
DSc2 (Occurred)	0.0001	4.4169e-005	0.0003	0.0004
DSc3 (Occurred)	0.1412	0.04480	0.3448	0.4114
Sc1 (Occurred)	0.0009	0.0002	0.0022	0.0027
Sc2 (Occurred)	0.0037	0.0011	0.0090	0.0108
Sc3 (Occurred)	0.0533	0.0169	0.1301	0.1553
Sc4 (Occurred)	0.0748	0.0237	0.1826	0.2179
Sc5 (Occurred)	0.0509	0.0161	0.1244	0.1484
Sc6 (Occurred)	0.0452	0.0143	0.1105	0.1319

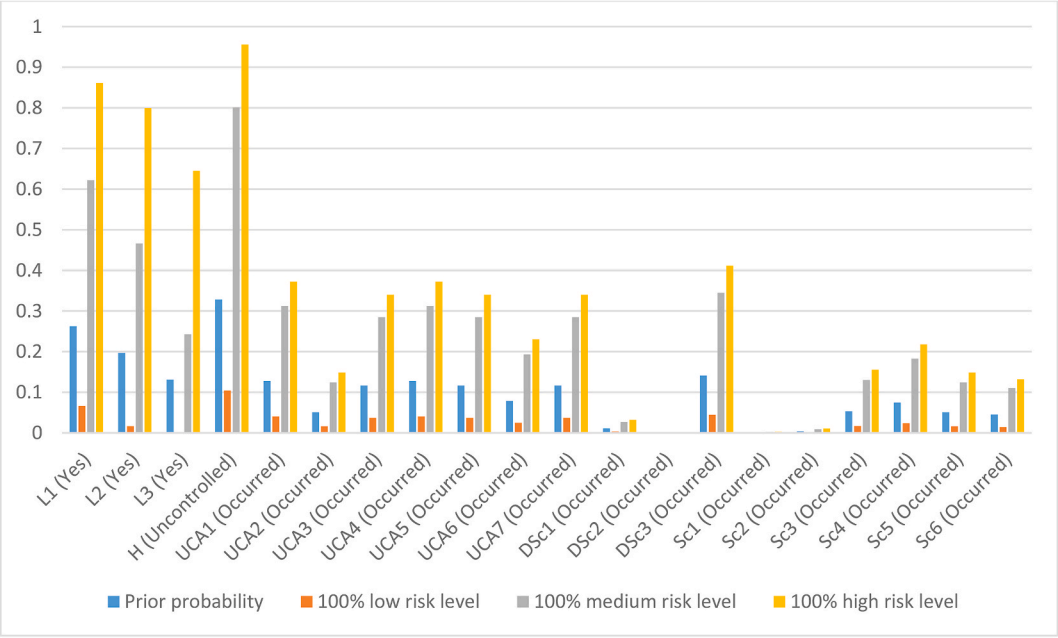


Fig. 9. Critical probability values of nodes.

5. Discussion

The methodology proposed in this paper uses the STPA analysis results in developing a BN model to assess the risks of a ship with higher degrees of autonomy and offers an efficient RCOs selection process. The proposed combination of STPA and BN is an approach towards a system-theoretic framework for marine risk assessment applications and FSA for new maritime technologies such as ship systems without onboard human intervention. Although autonomous ship concepts are nowadays at the development phases, using STPA to analyse the safety of ship systems with autonomous and remote-control capabilities can support the design of safer future ships. Transferring all the system control actions to the system software controller gives rise to new hazardous scenarios that should be identified and mitigated with adequate RCOs. STPA demonstrated the outstanding capability to identify the inadequate controls that might lead to the identified hazards of the automated SWC system. It also allowed identifying the causal scenarios that are related to the software control problems and the failure of physical components in the system. With STPA, it is possible to analyse the system as a whole and identify all the hazards leading to the losses and their respective causes. This is done without an initial prioritization based on accident data as followed in traditional hazard analysis methods. The proposed approach is proactive and aims at avoiding the hazardous conditions and their possible causes identified during the analysis by implementing the effective RCOs. In STPA the safety requirements are defined in parallel during the analysis to mitigate the identified causal scenarios, which makes it an appropriate method to generate the RCOs. An RCO can suggest a change in the control algorithm, a backup by the human remote controller, or other different measures. In any case, the RCO can be less/more or equally effective than other RCOs in mitigating the system risk. Therefore, the prioritization of the RCOs in this approach came after the STPA analysis when setting the desired risk

level and selecting the RCOs that prevent the causal scenarios with the highest contribution to risk. This was possible by combining STPA with BN, which helped explore the causal scenarios' impact on the risk picture. The BN development after STPA analysis was simple and systematic, due to the traceability and cause-effect relationship of the STPA outcomes, which was also highlighted by Utne et al. (2020). This represents one of the strengths of the developed BN model compared with other BN applications for the risk-based design of autonomous ships. For instance, Chang et al. (2021) identified the nodes of the BN and their interdependencies directly from literature and experts' opinions, which is a common approach in modelling the qualitative part of BN in the maritime field applications.

Another strength of the proposed framework is the advantage BN offers to map the interdependencies between the nodes and update the whole model whenever new data or evidence is available. This is one of the strengths of the BN compared to FTA. The same strength was highlighted by Bolbot et al. (2020) in their application of STPA in conjunction with FTA to ship scrubber systems as part of the new safety analysis method CASA. The authors concluded that BN would potentially replace FTA in future research (Bolbot et al., 2020). Once the CPTs are filled with the data, the developed BN supports the reasoning about the hazard analysis results, which is the aim of FSA in the marine risk-based design approach to estimate the uncertainties about the undesired events and provide insights that can be useful and meaningful for the decision making. The results of FSA have been usually communicated to the stakeholders to help them make risk-informed decisions. With the utilization of BN, it was possible to achieve this aim by first feeding the data into the model using probabilities to estimate the uncertainty about the likelihood of the causal scenarios identified with STPA. In addition, with the sensitivity analysis, it was possible to identify the input uncertainties that dominate the uncertainties about the output (the risk level) and compare the effectiveness of different RCOs in reducing these uncertainties. The developed BN model can also suggest the prioritization of recognized RCMs using backward and forward propagations. Given that the network can update features in light of new evidence, different combinations of RCMs can be deemed to be implemented through forward propagation. The model can estimate the updated probability of a hazard and its associated risk level. On the other hand, using backward propagation, it is possible to set the risk level at the minimum level, and BN can determine the critical probability values of the root nodes. Either way, risk managers, maintenance engineers, and policymakers can exploit the proposed framework to specify the sequence of RCMs to be implemented to ensure the system is operating at a minimum possible risk level. The backward propagation also offers safety engineers the possibility to identify the threshold for the UCAs and causal scenarios to respect a predefined safety level.

Applying the method to the autonomous SWC system showed that many causal scenarios were related to software errors. This is due to the nature of the autonomous ship systems where the control is shifting from human operator onboard to software controllers. Furthermore, the application demonstrated the importance of testing the software functionalities, which was one of the effective RCMs in reducing the risk of the system. Moreover, some of the causal scenarios, such as the sensor failure, appeared to be critical to the safety of the software-controlled ship system. For the mitigation of some unsafe scenarios, the health of the sensors appeared to be more effective than the full redundancy of other physical components of the system. This is due to the STPA analysis results which unveiled the importance of the sensors providing feedback to the software controller and their real contribution to the risk of the system. With the STPA analysis, it was possible to map the interactions of the sensor and the software components of the analysed system, which was also reflected in the relatively high probability of the hazard node. In addition to the initial aim of the analysis, the application of the proposed methodology can reveal when the shore-based control operators are needed. This was identified, for example, when the SWC system controller was the cause of the UCA. Such information can support the system designers in defining the role of humans in the autonomous ship systems operation and when their supervisory role should switch to control role. In the case study, the experts involved in the analysis suggested that a redundant control module should be available as a means of controlling the SWC system when the main controller has an error.

Although the proposed framework has been effective in meeting the objectives of this study, it showed some limitations. Finding the data to feed the model was challenging. The failure rates data that was employed in this study is conventional system data. However, using the data of the conventional ship system at this point is a realistic approach to support the decisions about the reliability and control of the future ship systems. In addition, an autonomous ship concept will require a cooling system that is most probably going to be similar to the current systems but with a different control setup. On the other hand, simulation data of the SWC system components can be also used when available to tackle the issue of data availability. It is also worth mentioning that the techniques combined in the framework do not cover all the steps of a complete marine risk assessment or FSA. Nevertheless, the proposed approach combines the strengths of STPA and BNs, which are widely considered good methods for hazard analysis and risk analysis. The uncertainty analysis for example, which is an important part of the risk assessment, was not covered in the proposed framework. Such analysis usually considers the assessment of the strength of evidence. Some published research work proposed generic approaches for the assessment of these uncertainties (Aven, 2013; Goerlandt and Reniers, 2016). Other studies suggested approaches that are tailored to STPA and BN (Lu et al., 2022; Wróbel et al., 2018).A similar approach can be considered to assess the strength of evidence when applying the method proposed in this work. Additionally, the cost of implementing the RCOs was not estimated and thus the CBA was not complete. Therefore, a potential future research can focus on investigating the possible integration of the RCM cost of implementation to the BN model since the cost analysis was beyond the scope of this work. This would then cover the cost-benefit analysis, thus provide a holistic ranking of the RCOs. In future research, applying the framework to a generic autonomous ship model for FSA can also be considered when adequate data becomes available. In this case, the unavailability of the SWC system due to an accident can be examined. Another future research topic can focus on analysing systems with multiple hazards to explore the effectiveness of the method with multiple interrelated hazards. This was not the case in the analysis of the SWC system because only one hazard was identified.

6. Conclusion

This paper presents an integration of STPA and BN through a risk-based model to identify and specify the priority order of RCOs. The proposed framework is applicable for the purpose of marine risk assessment and FSA. A case study of SWC in the machinery plant of a ship is considered to verify the applicability of the proposed framework. The system was assumed to operate without onboard-human intervention while monitored by the shore-based crew. The first part of the proposed methodology starts with the application of STPA steps. The outcomes of STPA then constitute the basis for the qualitative modelling of the BN in the second part. Lastly, the CPTs of the developed BN model are filled with the data to enable the reasoning process and extract useful information to make risk-informed design decisions. The application of the framework was systematic and straightforward due to the structured outcomes of STPA, which has been considered as a support for modelling the BN.

The methodology allowed estimating the probabilities of different nodes incorporated in the BN model and predicting the probability of different risk levels specified for the system. In addition, the proposed framework allowed the identification of the safety-critical aspects of the analysed system and the effective RCMs to control the risks of these aspects. The intensive testing of the software controller functionalities

and the sensor health monitoring techniques appeared to be substantial for reducing the risks associated with the autonomous SWC system.

To ensure the system operates at a minimum possible risk level (low), backward analysis is carried out in the BN model to determine the critical probability values of the root nodes. This STPA-BN model can also suggest the prioritization of recognized RCOs using backward and forward propagations that can be exploited by ship designers, safety engineers, and policymakers to analyse any target systems of the autonomous ship.

CRediT authorship contribution statement

Meriam Chaal: Conceptualization, Methodology, Model development, Data curation, Writing – original draft. **Ahmad Bahootoroody:** Advisory, Validation, Writing – original draft, Visualization. **Sunil Basnet:** Contributing to the case study, Visualization, Writing – review & editing. **Osiris A. Valdez Banda:** Supervision, Writing – review & editing. **Floris Goerlandt:** Conceptualization, Writing – review & editing.

Declaration of competing interest

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

Acknowledgement

The study presented in this article started under the Sea for Value (S4V) research program and was culminated during the Enablers and Concepts for Automated Maritime Solutions (ECAMARIS) program. The S4V and ECAMARIS programs are partially funded by Business Finland. The contributions of the last author are supported by the Natural Sciences and Engineering Research Council of Canada, through the Canada Research Chairs Program.

The authors want to express their gratitude to the Helsinki Shipyard project engineer and the two marine engineers who participated in the brainstorming sessions to conclude this work.

Appendix

Control the seawater cooling system				
Element type	Main question	Answer to the main question	Additional question	Answer to additional question
A-Feedback	1-What information does the operator need for controlling the seawater cooling system	- The actual seawater temperature. - The actual seawater pressure. - The differential water pressure at the coolers. - The pumps status (running/stopped). - The valves status (open/closed). - The active alarms.	2-From where is the information provided?	- The system components sensors. - The interface of the software controller.
B-Control actions	3-What are the actions taken as output of the specified function?	- Set the mode of operation (automatic/manual). - Set the setpoint of alarms. - Set the setpoint for standby pumps operation. - Start/stop the pumps. - Open/Close seawater inlet valves. - Open/close valves to switch between coolers.	4-To which function, sub-function or component is the output given?	- The system software controller. - To the system components (pumps, valves).
C-Controlled process	5-To which function, sub-function, or component is the output given?	Given in the previous answer	NA	
D-Other inputs to, outputs from components	6-What other information can influence the operator actions	No more answers	NA	

References

Abaei, M.M., Hekkenberg, R., BahooToroody, A., 2021. A multinomial process tree for reliability assessment of machinery in autonomous ships. Reliab. Eng. Syst. Saf. 210, 107484 https://doi.org/10.1016/j.ress.2021.107484.

American Bureau of Shipping (ABS), 2020. Guidance Notes on Risk Assessment Applications for the Marine and Offshore Industries.

Antão, P., Soares, C.G., 2019. Analysis of the influence of human errors on the occurrence of coastal ship accidents in different wave conditions using Bayesian Belief Networks. Accid. Anal. Prev. 133, 105262 https://doi.org/10.1016/j.aap.2019.105262.

Antoine, B., 2013. Systems Theoretic Hazard Analysis (STPA) Applied to the Risk Review of Complex Systems : an Example from the Medical Device Industry (Thesis). Massachusetts Institute of Technology.

Aven, T., 2013. Practical implications of the new risk perspectives. Reliab. Eng. Syst. Saf. 115, 136–145. https://doi.org/10.1016/j.ress.2013.02.020.

Aven, T., 2016. Risk assessment and risk management: review of recent advances on their foundation. Eur. J. Oper. Res. 253, 1–13. https://doi.org/10.1016/j.ejor.2015.12.023.

BahooToroody, A., Abaei, M.M., Arzaghi, E., BahooToroody, F., De Carlo, F., Abbassi, R., 2019. Multi-level optimization of maintenance plan for natural gas system exposed to deterioration process. J. Hazard Mater. 362, 412–423. https://doi.org/10.1016/j.jhazmat.2018.09.044.

BahooToroody, A., De Carlo, F., Paltrinieri, N., Tucci, M., Van Gelder, P.H.A.J.M., 2020. Bayesian Regression Based Condition Monitoring Approach for Effective Reliability Prediction of Random Processes in Autonomous Energy Supply Operation, 201. Reliability Engineering & System Safety, 106966. https://doi.org/10.1016/j.ress.2020.106966.

Baksh, A.-A., Abbassi, R., Garaniya, V., Khan, F., 2018. Marine transportation risk assessment using Bayesian Network: application to Arctic waters. Ocean Eng. 159, 422–436. https://doi.org/10.1016/j.oceaneng.2018.04.024.

Barber, D., 2010. Bayesian Reasoning and Machine Learning.

BayesFusion, 2020. GeNie.

Bjerga, T., Aven, T., Zio, E., 2016. Uncertainty treatment in risk analysis of complex systems: the cases of STAMP and FRAM. Reliab. Eng. Syst. Saf. 156, 203–209. https://doi.org/10.1016/j.ress.2016.08.004.

Bolbot, V., Theotokatos, G., Bujorianu, L.M., Boulougouris, E., Vassalos, D., 2019. Vulnerabilities and safety assurance methods in Cyber-Physical Systems: a comprehensive review. Reliab. Eng. Syst. Saf. 182, 179–193. https://doi.org/10.1016/j.ress.2018.09.004.

Bolbot, V., Theotokatos, G., Boulougouris, E., Psarros, G., Hamann, R., 2020. A novel method for safety analysis of cyber-physical systems—application to a ship exhaust gas scrubber system. Saf. Now. 6, 26. https://doi.org/10.3390/safety6020026.

Breinholt, C., Ehrke, K.-C., Papanikolaou, A., Sames, P.C., Skjong, R., Strang, T., Vassalos, D., Witolla, T., 2012. SAFEDOR–The implementation of risk-based ship design and approval. Procedia - Social Behav. Sci. Transp. Res. Arena 48, 753–764. https://doi.org/10.1016/j.sbspro.2012.06.1053, 2012.

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

Chai, M., Reddy, B.D., Sobrayen, L., Panda, S.K., Die, W., Xiaoqing, C., 2016. Improvement in efficiency and reliability for diesel- electric propulsion based marine vessels using genetic algorithm. In: 2016 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific). https://doi.org/10.1109/ITEC-AP.2016.7512944.

Chang, C.-H., Kontovas, C., Yu, Q., Yang, Z., 2021. Risk assessment of the operations of maritime autonomous surface ships. Reliab. Eng. Syst. Saf. 207, 107324 https://doi.org/10.1016/j.ress.2020.107324.

ORDEA companies, 2015. Offshore Reliability Data Handbook OREDA. Trondheim, Norway.

de Vos, J., Hekkenberg, R.G., Valdez Banda, O.A., 2021. The impact of autonomous ships on safety at Sea – a statistical analysis. Reliab. Eng. Syst. Saf. 210, 107558 https://doi.org/10.1016/j.ress.2021.107558.

DeJesus Segarra, J., Bensi, M., Modarres, M., 2021. A bayesian network approach for modeling dependent seismic failures in a nuclear power plant probabilistic risk assessment. Reliab. Eng. Syst. Saf. 213, 107678 https://doi.org/10.1016/j.ress.2021.107678.

Gao, C., Guo, Y., Zhong, M., Liang, X., Wang, H., Yi, H., 2021. Reliability analysis based on dynamic Bayesian networks: a case study of an unmanned surface vessel. Ocean Eng. 240, 109970 https://doi.org/10.1016/j.oceaneng.2021.109970.

Gil, M., Wróbel, K., Montewka, J., 2019. Toward a method evaluating control actions in STPA-based model of ship-ship collision avoidance process. J. Offshore Mech. Arctic Eng. 141, 051105 https://doi.org/10.1115/1.4042387.

Glomsrud, J.A., Xie, J., 2019. A Structured STPA Safety and Security Co-analysis Framework for Autonomous Ships 8.

Goerlandt, F., Montewka, J., 2015. A framework for risk analysis of maritime transportation systems: a case study for oil spill from tankers in a ship–ship collision. Saf. Sci. 76, 42–66. https://doi.org/10.1016/j.ssci.2015.02.009.

Goerlandt, F., Reniers, G., 2016. On the assessment of uncertainty in risk diagrams. Saf. Sci. 84, 67–77. https://doi.org/10.1016/j.ssci.2015.12.001.

Guedes Soares, C., Jasionowski, A., Jensen, J., McGeorge, D., Papanikolaou, A., Poylio, E., Sames, P., Juhl, J., Vassalos, D., 2009. Risk-based Ship Design – Methods, Tools and Applications.

IMO, 2018a. MSC 99/5/6: REGULATORY SCOPING EXERCISE for the USE of MARITIME AUTONOMOUS SURFACE SHIPS (MASS) Considerations on Definitions for Levels and Concepts of Autonomy. London, UK.

IMO, 2018b. MSC-MEPC.2/Circ.12/Rev.2 : REVISED GUIDELINES for FORMAL SAFETY ASSESSMENT (FSA) for USE IN the IMO RULE-MAKING PROCESS.

IMO, 2006. MSC 81/24/5 - Any Other Business, FSA Study on ECDIS/ENCs, Submitted by Denmark and Norway. In: IMO, ed. MSC 81/24/5. London.

Johansen, T., Utne, I.B., 2022. Supervisory risk control of autonomous surface ships. Ocean Eng. 251, 111045 https://doi.org/10.1016/j.oceaneng.2022.111045.

Kelly, D.L., Smith, C.L., 2009. Bayesian inference in probabilistic risk assessment—the current state of the art. Reliab. Eng. Syst. Saf. 94, 628–643. https://doi.org/10.1016/j.ress.2008.07.002.

Kelly, D., Smith, C., 2011. Bayesian Inference for Probabilistic Risk Assessment: A Practitioner's Guidebook. Springer Science & Business Media.

Khalaj, S., BahooToroody, F., Mahdi Abaei, M., BahooToroody, A., De Carlo, F., Abbassi, R., 2020. A methodology for uncertainty analysis of landslides triggered by an earthquake. Comput. Geotech. 117, 103262 https://doi.org/10.1016/j.compgeo.2019.103262.

Konovessis, D., Cai, W., Vassalos, D., 2013. Development of Bayesian network models for risk-based ship design. J. Mar. Sci. Appl. 12, 140–151. https://doi.org/10.1007/s11804-013-1179-9.

Leoni, L., BahooToroody, F., Khalaj, S., Carlo, F.D., BahooToroody, A., Abaei, M.M., 2021. Bayesian estimation for reliability engineering: addressing the influence of prior choice. IJERPH 18, 3349. https://doi.org/10.3390/ijerph18073349.

Leveson, N., 2011. Engineering a Safer World, <systems Thinking Applied to Safety.

Leveson, N., Thomas, J.P., 2018. STPA HANDBOOK.

Lu, L., Goerlandt, F., Valdez Banda, O.A., Kujala, P., Höglund, A., Arneborg, L., 2019. A Bayesian Network risk model for assessing oil spill recovery effectiveness in the ice-covered Northern Baltic Sea. Mar. Pollut. Bull. 139, 440–458. https://doi.org/10.1016/j.marpolbul.2018.12.018.

Lu, L., Goerlandt, F., Banda, O.A.V., Kujala, P., 2022. Developing fuzzy logic strength of evidence index and application in Bayesian networks for system risk management. Expert Syst. Appl. 192, 116374 https://doi.org/10.1016/j.eswa.2021.116374.

Meng, X., Chen, G., Shi, J., Zhu, G., Zhu, Y., 2018. STAMP-based analysis of deepwater well control safety. J. Loss Prev. Process. Ind. 55, 41–52. https://doi.org/10.1016/j.jlp.2018.05.019.

Modarres, M., 2016. Risk Analysis in Engineering: Techniques, Tools, and Trends. CRC press.

Montewka, J., Wróbel, K., Heikkila, E., Valdez-Banda, O., Goerlandt, F., Haugen, S., 2018. Challenges, solution proposals and research directions in safety and risk assessment of autonomous shipping, 12. Los Angeles.

Neapolitan, R.E., 2004. Learning Bayesian Networks. Pearson Prentice Hall, Upper Saddle River, NJ.

Nielsen, T.D., Jensen, F.V., 2009. Bayesian Networks and Decision Graphs. Springer Science & Business Media.

Pui, G., Bhandari, J., Arzaghi, E., Abbassi, R., Garaniya, V., 2017. Risk-based maintenance of offshore managed pressure drilling (MPD) operation. J. Petrol. Sci. Eng. 159, 513–521. https://doi.org/10.1016/j.petrol.2017.09.066.

Puisa, R., Bolbot, V., Ihle, I., 2019. Development of functional safety requirements for DP- driven servicing of wind turbines. In: Presented at the European STAMP Workshop & Conference, 2019 Finland.

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

Rokseth, B., Haugen, O.I., Utne, I.B., 2019. Safety verification for autonomous ships. MATEC Web Conf. 273 https://doi.org/10.1051/matecconf/201927302002, 02002.

Rolls Royce, 2016. Remote and Autonomous Ships the Next Steps.

Schüller, J.C.H., Brinkman, J.L., Van Gestel, P.J., van Otterloo, R.W., 1997. Methods for Determining and Processing Probabilities: Red Book. Committee for the Prevention of Disasters, The Hague, Netherlands.

SINTEF, 2006. Reliability Data for Safety Instrumented Systems PDS Data Handbook. Trondheim, Norway.

Skjong, R., 2011. Formal Safety Assessments and Risk Based Design in the Maritime Industry.

Sulaman, S.M., Beer, A., Felderer, M., Höst, M., 2019. Comparison of the FMEA and STPA safety analysis methods–a case study. Software Qual. J. 27, 349–387. https://doi.org/10.1007/s11219-017-9396-0.

Thieme, C.A., Utne, I.B., Haugen, S., 2018. Assessing ship risk model applicability to marine autonomous surface ships. Ocean Eng. 165, 140–154. https://doi.org/10.1016/j.oceaneng.2018.07.040.

Utne, I.B., Rokseth, B., Sørensen, A.J., Vinnem, J.E., 2020. Towards supervisory risk control of autonomous ships. Reliab. Eng. Syst. Saf. 196, 106757 https://doi.org/10.1016/j.ress.2019.106757.

Valdez Banda, O.A., Kannos, S., Goerlandt, F., van Gelder, P.H.A.J.M., Bergström, M., Kujala, P., 2019. A systemic hazard analysis and management process for the concept design phase of an autonomous vessel. Reliab. Eng. Syst. Saf. 191 https://doi.org/10.1016/j.ress.2019.106584.

Vassalos, D., Spyrou, K., Themelis, N., Mermiris, G., 2010. Risk-based Design for Fire Safety – A Generic Framework.

Ventikos, N.P., Chmurski, A., Louzis, K., 2020. A systems-based application for autonomous vessels safety: hazard identification as a function of increasing autonomy levels. Saf. Sci. 131, 104919 https://doi.org/10.1016/j.ssci.2020.104919.

Wang, F., Araújo, D.F., Li, Y.-F., 2022. Reliability assessment of autonomous vehicles based on the safety control structure. In: Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. https://doi.org/10.1177/1748006X211069705, 1748006X2110697.

Wróbel, K., Montewka, J., Kujala, P., 2018. Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels. Reliab. Eng. Syst. Saf. 178, 209–224. https://doi.org/10.1016/j.ress.2018.05.019.

Wróbel, K., Krata, P., Montewka, J., 2019. Preliminary Results of a System-Theoretic Assessment of Maritime Autonomous Surface Ships' Safety. ResearchGate.

Yang, X., Utne, I.B., Sandøy, S.S., Ramos, M.A., Rokseth, B., 2020. A systems-theoretic approach to hazard identification of marine systems with dynamic autonomy. Ocean Eng. 217, 107930 https://doi.org/10.1016/j.oceaneng.2020.107930.

Zhang, X., Mahadevan, S., 2021. Bayesian network modeling of accident investigation reports for aviation safety assessment. Reliab. Eng. Syst. Saf. 209, 107371 https://doi.org/10.1016/j.ress.2020.107371.

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

Zhou, X.-Y., Liu, Z.-J., Wang, F.-W., Wu, Z.-L., Cui, R.-D., 2020. Towards applicability evaluation of hazard analysis methods for autonomous ships. Ocean Eng. 214, 107773 https://doi.org/10.1016/j.oceaneng.2020.107773.

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

Zhu, T., Haugen, S., Liu, Y., 2021. Risk information in decision-making: definitions, requirements and various functions. J. Loss Prev. Process. Ind. 72, 104572 https://doi.org/10.1016/j.jlp.2021.104572.

Zymaris, A.S., Alnes, Ø.Å., Knut Erik, K., NikolaosM, P.K., 2016. Towards a Model-Based Condition Assessment of Complex Marine Machinery Systems Using Systems Engineering - PDF Free Download. Presented at the European Conference of the PHM Society, Bilbao, Spain.