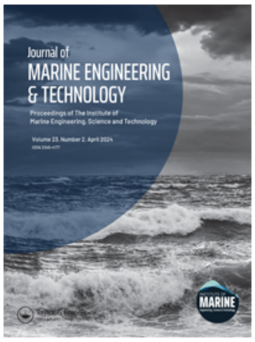

Journal of Marine Engineering & Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmar20

Human-autonomy collaboration in supervisory risk control of autonomous ships

Thomas Johansen & Ingrid Bouwer Utne

To cite this article: Thomas Johansen & Ingrid Bouwer Utne (2024) Human-autonomy collaboration in supervisory risk control of autonomous ships, Journal of Marine Engineering & Technology, 23:2, 135-153, DOI: 10.1080/20464177.2024.2319369

To link to this article: https://doi.org/10.1080/20464177.2024.2319369

JOURNAL OF MARINE ENGINEERING & TECHNOLOGY
2024, VOL. 23, NO. 2, 135–153
https://doi.org/10.1080/20464177.2024.2319369

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Human-autonomy collaboration in supervisory risk control of autonomous ships

Thomas Johansen[a,b] and Ingrid Bouwer Utne[a,b]

[a]Centre for Autonomous Marine Operations and Systems (NTNU AMOS), NTNU, Trondheim, Norway; [b]Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT

This paper presents a method for developing and testing a risk-based control system, as a first step towards including the human supervisor explicitly in the design of the system. The result is a control system with improved decision-making capabilities compared to existing control systems. The methodology presented in the paper uses the Systems Theoretic Process Analysis (STPA) to analyse the risks of an autonomous ship within its concept of operations (CONOPS), and a Human-STPA (H-STPA) is used to analyse human responsibilities and involvement. The STPA results are then used to construct a Bayesian belief network (BBN)-based risk model to assess the operational risk of the ship. This is represented as a risk cost, describing the expected cost of consequences caused by potential hazardous events. This cost is combined with fuel costs, operations costs, and the potential loss of income if new missions are not undertaken using a supervisory risk controller (SRC). The SRC is capable of making decisions about how the ship should be safely operated and notifies the human supervisor in due time when it is necessary for them to take control. The last part of the methodology presented in this paper is testing the control system using a set of verification objectives based on results from the STPA and H-STPA. A case study involving an autonomous cargo ship with a human supervisor located in a remote operation center (ROC) is included; it shows that the proposed control system can operate the ship safely in different conditions and situations. By designing the SRC to notify the human supervisor before it reaches its operational limit, the ship is able to operate in a wider range of conditions compared to when just the autonomous control system is in charge. Hence, the proposed methodology shows promising results and provides useful insights related to shared control for autonomous ships.

ARTICLE HISTORY
Received 12 March 2023
Accepted 4 February 2024

KEYWORDS
Autonomous systems; risk modelling; ship control systems; human-autonomy collaboration; systems theoretic process analysis; Bayesian belief networks

Abbreviations

AIS	Automatic Identification System
AMMS	Autonomous Machinery Management System
ANS	Autonomous Navigation System
AP	Autopilot
API	Application Programming Interface
AUV	Autonomous Underwater Vehicle
BBN	Bayesian Belief Network
CONOPS	Concept of Operations
CPT	Conditional Probability Table
DP	Dynamic Positioning
ENC	Electronic Navigational Chart
GNSS	Global Navigational Satellite System
GP	Gaussian Process
H-RIF	High-level Risk Influencing Factor
HiL	Hardware-in-the-Loop
HMI	Human Machine Interface
HSG	Hybrid Shaft Generator
I-RIF	Input Risk Influencing Factor
LNG	Liquefied Natural Gas
LoA	Level of Autonomy
Mech	Mechanical
MPC	Model Predictive Control
MRC	Minimal Risk Condition
MSO	Machinery System Operation
PID	Proportional Integral Derivative
PMS	Power Management System
PTI	Power Take In
PTO	Power Take Out
RIF	Risk Influencing Factor
ROC	Remote Operation Center
SLAM	Simultaneous Localization and Mapping
SO	Ship Operation
SRC	Supervisory Risk Controller
STL	Signal Temporal Logic
STPA	Systems Theoretic Process Analysis
UAV	Unmanned Aerial Vehicle
UCA	Unsafe Control Action
USD	United States Dollar
VHF	Very High Frequency

1. Introduction

Ship control systems have advanced from early autopilots to dynamic positioning (DP) systems, and currently they are moving towards control of autonomous ships. Autonomous ships are expected to improve general safety at sea (Wróbel et al. 2017; de Vos et al. 2021) by reducing the number of humans at risk. In general, much work has been done on identifying different risk factors and performing risk assessments of autonomous ships. Fan et al. (2020) present a framework for identifying navigational risk factors for autonomous ships. Johansen and Utne (2020) suggest using the Systems Theoretic Process Analysis (STPA) as the basis for building risk models describing autonomous ships and discuss additional methods for finding

CONTACT Thomas Johansen tjoha@ntnu.no

more data. Chaal et al. (2020) propose using the STPA to model the ship control structure in order to describe the system functionalities. Valdez Banda et al. (2019) use the STPA for the hazard analysis of autonomous passenger ferries. The STPA is also used in Wróbel et al. (2018) to develop a model for analysing safety and providing recommendations for designing autonomous vessels. However, none of these papers use the results of the risk analyses to control autonomous ships. Other works have proposed using risk models to predict the loss of AUVs (Brito and Griffiths 2016; Loh, Brito, Bose, Xu, Nikolova et al. 2020; Loh, Brito, Bose, Xu and Tenekedjiev 2020) or to manage uncertainty in AUV missions (Brito 2016). Risk is also included in multiple papers discussing collision avoidance (Hu et al. 2017; Lyu and Yin 2019; Wang et al. 2019; Woo and Kim 2020; Gil 2021; Li et al. 2021), but at a more general level.

Even with the continuous development and improvement of ship control systems, it is expected that humans will remain important in the safe and efficient operation of autonomous systems (Ramos et al. 2020b, 2020a). Therefore, an important issue when developing autonomous ships is designing control systems that support the safe transition between autonomous and human control.

Ramos et al. (2020b) present a method for analysing cooperation between humans and autonomous ships called the Human-System Interaction (H-SIA) method. The method is used in a case study to analyse a collision scenario. Ramos et al. (2020a) present a generic approach for analysing failures in the interaction between the system and humans and demonstrate this approach by analysing an autonomous ship. Hogenboom et al. (2021) discuss how the available time affects risk when humans must take over control in DP operations. Parhizkar et al. (2020) propose a risk management framework for DP operations to provide decision support to human supervisors and test the framework in a case study on DP drilling operations. Wu et al. (2022) summarise and review techniques for analysing human and organisational factors related to maritime accidents, and they provide ideas for further development, with a focus on humans. All these papers discuss important aspects of human-system cooperation for ships, but they do not discuss how to include humans as a part of the control system or specify the responsibilities of a human supervisor in shared control scenarios.

Huang et al. (2020) present a collision avoidance system that is focused on human-machine interaction. The collision avoidance system is designed such that the decision-making process is easy to follow and interactive for human supervisors. However, the control system is limited to only considering collision avoidance and is not a more high-level control system. Liu et al. (2022) discuss multiple issues and challenges related to human-machine cooperation with autonomous ships. They also discuss unsolved problems that should be tackled as part of further development and therefore provide ideas for further work. Rødseth et al. (2021) propose an operational envelope that includes sharing control responsibilities between humans and the control system. They show how this can be done in a general way to account for most geographical areas and operations, but they do not demonstrate how this information can explicitly be used to design the control system.

Porathe (2021) discusses how to design the autonomous control system to provide better decision support for human supervisors of autonomous ships. The paper suggests having a copy of the control system running in a remote operation center (ROC) such that data are readily available to human supervisors. However, the paper lacks a description of the actual control system and how the human supervisor should be included. Dittmann et al. (2021) describe how to design a control system complying with international regulations on watch-keeping with a remote control center as part of the control system. They discuss how to design the system to share information with human supervisors and how to transfer control between the system and human supervisors. A control structure is suggested but how the different parts function is not specified.

Table 1. Summary of key aspects of the proposed control system compared to existing control systems.

	Proposed control system	Existing control systems
Main features/tasks	High-level risk-based decision making. Optimum control of autonomous ships.	Control of specific functions and subsystems. Optimising energy consumption.
Integration with humans	Controller designed to notify human supervisor in case of emergencies.	Human supervisor/operator assumed to constantly monitor control system in case of emergencies.
Possible application areas	Control of autonomous ships. Decision support system for human operators and supervisors.	Control of autonomous ships. Decision support system for human operators and supervisors.
Limitations and challenges addressed	Including risk and safety in optimum control of ships. Inclusion of human supervisors.	Automation of ship control systems. Optimum control of ship subsystems.

Utne et al. (2020) propose using risk models in the control system, i.e. a supervisory risk controller (SRC), to improve the decision-making capabilities and intelligence of the system. Thieme et al. (2021) describe how to use risk analysis methods to design control systems and propose four areas where this can be implemented. Johansen and Utne (2022) propose a control system using Bayesian belief network (BBN)-based risk models and show how this implementation can contribute to high-level decisions, such as selecting the optimal machinery and control mode to ensure the safe and efficient operation of an autonomous ship. Similar control systems have been proposed for autonomous underwater vehicles (AUVs) performing under-ice mapping (Bremnes et al. 2019, 2020). Yang and Utne (2022) present a set of criteria for an online risk model for autonomous marine systems and discuss potential methods for building the model. All these works show how risk modelling can be used to improve control systems, but they lack the perspective of shared control and the inclusion of the human supervisor and his/her responsibilities in the system and operations for different levels of autonomy.

In general, previous works on control systems for autonomous ships focus either on the control system or human control. A limited number of papers discuss collaboration but without discussing how to design the control system to support and interact with the human supervisor. Safe and efficient collaboration between the human supervisor and the autonomous system is decisive for safe operation. Hence, the objective of this paper is to present a methodology for designing and testing a risk-based control system, focusing on both the autonomous control system and the human supervisor. The control system is designed to notify the human supervisor to provide them with time to react and make alternative plans when necessary. The proposed control system is tested in a case study involving an autonomous coastal cargo ship. This paper is the first attempt to include both the autonomous control system and the human supervisor in the SRC to ensure safe ship operation. An overview of key differences between the proposed control system and existing control systems is shown in Table 1.

2. Background

2.1. Level of autonomy

The level of autonomy (LoA) is used to describe the functionality of autonomous systems and how they are related to the human

Table 2. Levels of autonomy, adopted from Utne et al. (2017).

LoA	Type	Description
1	Automatic operation / Remote control	The system operates automatically with a remote human operator. The human operator has full control of the system. The system can have pre-programmed functions implemented.
2	Management by consent	The control system can make recommendations about specific parts of the operation. The human operator still controls the operation. The system can perform many tasks independently, if they are approved by the human operator.
3	Management by exception	The system automatically executes the mission plan and has the ability to make small changes when the available time is too short for human intervention. The human supervisor can take control of the system or change the plan. The human supervisor is notified by the system when it is necessary to take over or update the plan.
4	Highly autonomous operation	The system automatically plans and executes the operation. The system can change and alter the plan during operation. Humans can be informed about the operation, but the system operates independently.

operator/ supervisor. In this paper, four LoAs are used; they are based on Utne et al. (2017) and shown in Table 2.

Level one describes an automated system in which the human operator has full control of the system. The system is dependent on human supervisors who monitor and control the system. The human operator and the system can be located in different places. In level two, the system has more automation, but it still needs a human operator to make decisions about how it should operate. At level three, the system can follow a plan. If the operation deviates from the plan, the system can suggest changes to the plan, but the human supervisor must accept these changes. If the operation goes according to plan, the human supervisor is 'out of the loop.' At level four, the system operates without human control. Humans can be informed about the progress of the system, but the system is operating independently. The human supervisor has limited or no ability to take control of the ship, but they may provide input to the system. It is important to note that a system may switch between different LoAs in operation, i.e. high and low LoAs, and the system may also include sub-systems operating at different LoAs at the same time.

This paper focuses on an autonomous ship operating at LoA 3. The ship can follow preplanned routes, choose which preplanned route to follow, and change the speed, machinery mode, and control mode. To make bigger changes to the plan, such as deviating from the preplanned routes due to weather conditions, the human supervisor must assess the situation and agree to the new route proposed by the control system. To support the decision-making abilities of the human supervisor, the control system should be designed to provide enough time and information for human intervention. If the human supervisor need to react, the controller should still maintain the ship in a safe condition by for example maintaining its current position using DP. Collision avoidance is considered outside the scope of this work due to the complexity of building a control module for handling this. It is also assumed that collision avoidance would function otuside the control system proposed in this paper due to the criticality of such decisions and the time available to avoid collisions with other ships.

2.2. Human-autonomy collaboration

The ship considered in this paper is an unmanned cargo ship operating along the Norwegian coast. The ship has no crew aboard but is connected to a remote operation center (ROC). In the ROC, a human supervisor has access to the same information and data as the control system on the ship, but he/she also has the ability to remotely take control of the ship. The human supervisor, however, is not monitoring the ship during normal operation. Only after a notification will the human supervisor take control of the ship, and therefore they need some amount of time to obtain a sufficient awareness of the situation and react appropriately.

There are three main types of notifications sent from the control system to the human supervisor. First, the control system sends a notification when it is unable to maintain the safe operation of the ship or when it determines that it is likely to lose control in the near future. The control system is designed to go into a 'minimal risk condition' mode if it determines that it is unsafe to continue and it also notifies the supervisor. To exit this mode, the human supervisor has to take remote control of the ship or indicate that the control system can continue to operate. Second, the control system will notify the human supervisor of potential problems that he or she can contribute to avoiding or mitigating. The final type of notification is sent when the control system loses control, and it is impossible to avoid an accident. In these cases, the human supervisor's role is to start coordinating rescue operations to limit negative consequences and salvage the ship.

Control systems for autonomous ships are designed to reduce the need for human control while still operating in a safe and efficient manner. However, humans are still expected to be involved in operating the ship, especially when the situation exceeds the autonomous capabilities of the ship. Humans will then function more as supervisors who monitor the ship and assist when necessary rather than as operators or crews onboard responsible for the daily operation of the ship.

Since the autonomous ship in this paper is operating at either LoA 3 or LoA 1, the human supervisor receives these three types of notifications only. These notifications are mainly caused by failures or conditions that exceed the operational limits or safety constraints of the control system. In any case, the amount of time (Hogenboom et al. 2021) and information available to the human supervisor are important for a successful intervention. If the amount of time is too short or information is missing, there is less of a chance for the human supervisor to successfully take control and handle the situation. Providing a detailed analysis of human reaction times, human reliability, risk-based decision support for the supervisor, and human-machine interaction is, however, outside the scope of this paper and should be the subject of future work.

To reduce the likelihood of hazardous events, the control system has the option to enter a minimal risk condition (MRC) mode when the risk becomes too high. ISO (2020) defines the MRC as 'a condition to which a user or an automated driving system may bring a vehicle after performing the minimal risk manoeuvre in order to reduce the risk of a crash when a given trip/voyage cannot be completed.' For the autonomous ship in this paper, it is very difficult to eliminate all risk, but the risk can still be reduced to a level that is as low as reasonably practicable (ALARP). For further information on the definition of ALARP, please see HSE (2001).

3. Methodology

The proposed methodology extends and further develops the work in Utne et al. (2020), Johansen and Utne (2022), and Johansen et al. (2023) by adding more advanced functionalities to the controller, such as the ability to select different routes to follow, and adding a specific MRC mode that the ship can enter when the risk becomes too high. Furthermore, interaction with the human supervisor is considered; it is not included in the above-mentioned studies. Specifically, the SRC in this paper is a high-level controller that can manage the ship control system. The SRC makes decisions, such as selecting the control mode for the navigation system and selecting how the machinery system should be operated. The methodology proposed for developing the SRC in this paper is a five-step process:

- Perform an STPA of the ship and a fault tree analysis (FTA) of critical sub-systems.
- Extend the STPA with a Human-STPA (H-STPA).
- Develop an online risk model and assign inputs for the different nodes, such as sensor measurements and data from electronic navigational charts (ENCs).
- Set up the SRC and integrate it with the rest of the control system, including the motion and machinery controllers.
- Verify the control system in scenarios based on the STPA and H-STPA.

The STPA is used to get a good overview of unsafe control actions related to the ship and control system within its concept of operations (CONOPS). This forms the basis for building the risk model, deciding what data need to be extracted from ENCs to use in the control system, and setting up the SRC. To ensure a safe interaction between the human supervisor and the autonomous system, an H-STPA is performed. The results from this analysis are also used when setting up the control system to enable the human supervisor to interact with the control system in a safe and efficient manner. Then, the system is tested in different scenarios, which are formulated based on the STPA and H-STPA results, to verify that it functions as intended. The testing should include both easy and challenging scenarios.

3.1. Extended STPA and fault tree analysis

The STPA is based on Leveson (2011), and the extended STPA proposed in Johansen and Utne (2022) includes consequences as part of the analysis. In the traditional STPA, losses are defined as a starting point, which to some extent indicates the consequences. When developing control systems, however, it is necessary to include consequences in more detail. Hence, 'losses' are here called hazardous events, and consequences are explicitly described. This is also in line with the bow-tie model (Rausand and Haugen 2020). The STPA starts by describing the ship and the CONOPS, including the machinery, propulsion, and control system. The CONOPS should provide information about the intended routes and/or area of operation, potential cargo aboard the ship, schedule, and limitations concerning when and where the ship can sail.

The STPA then defines hazardous events that, under certain conditions, can cause negative consequences for the ship. The rest of the analysis follows the normal STPA process by identifying system-level hazards, unsafe control actions (UCAs), loss scenarios, and causes.

Critical systems related to power, propulsion, and navigation sensors that emerge from the STPA are then analysed using a qualitative FTA. The reason for this is that such systems are monitored, and the FTA provides information about whether the ship still has the necessary redundancy to continue or if it should notify the human supervisor about the situation to obtain assistance and enter the MRC.

3.2. Human-STPA

A Human-STPA is used to identify causal factors that affect the human supervisor's ability to intervene. This is done using an STPA by modelling the human supervisor as a human controller, as proposed by France (2017). Each possible action from the human supervisor is a control action that can be analysed. The focus in this step is on how the control system should be designed to make it as safe and efficient as possible for a human supervisor to take control of the ship and to make decisions.

The analysis uses the same control structure as the regular STPA but it focuses on the human supervisor instead of the SRC. The rest of the analysis follows the same approach and considers the same hazardous events and system-level hazards. As with the SRC, the human supervisor has a set of available control actions that is analysed to identify UCAs and specify scenarios in which these UCAs may occur.

3.3. Building the online risk model

The STPA results are used as the basis for building the BBN risk model. A detailed description of this process can be found in Utne et al. (2020). The BBN is made into an online risk model by connecting input risk influencing factors (RIFs), i.e. by connecting parent nodes to the control system, and deciding when to update nodes with new information. This includes describing the data required from electronic navigational charts (ENCs), which are required for the path planning and safe navigation of the ship.

An important source of data for the online risk model is ENCs. These contain navigational information about the area, such as the water depth, land, and navigational marks. However, the charts contain so much data that these data need to be processed to be useful in both the online risk model and the rest of the control system. The ENC module used in this paper is based on the work of Blindheim and Johansen (2022). The ENCs provide necessary information about the area around the ship so that the SRC can include this information in its decision-making process. The module is based on SeaCharts, an open-source Python package for displaying and manipulating charts. The module uses FGDB 10.0 charts with 2D data concerning the relevant areas. These data are processed and filtered to avoid giving irrelevant data to the control system. The relevant data are stored in shapefiles for different water depths and land. This makes it easier to find the water depth and the distance to points where the ship can ground.

The ENC module is set up based on the required data from the online risk model and the SRC. The data are then used to describe how much open water is around the ship, how much room the ship has to maneuver, and other relevant information to improve the decision-making process of the SRC.

3.4. Setting up the supervisory risk controller (SRC)

The SRC combines the risk cost, fuel cost, operation cost, and a penalty cost for the potential future loss of income and delays:

$$Cost(d) = R(d) + F(d) + O(d) + L(d). \tag{1}$$

The expected risk cost, $R(d)$, is taken directly from the risk model. The expected fuel cost, $F(d)$, is derived for the remaining route. The operation cost, $O(d)$, describes the additional operation costs (not fuel). The potential future loss, $L(d)$, represents the extra time used because the ship is not sailing full speed all the time and potentially misses deadlines because it is not able to follow the planned schedule. Notifications to the human supervisor are included based on the

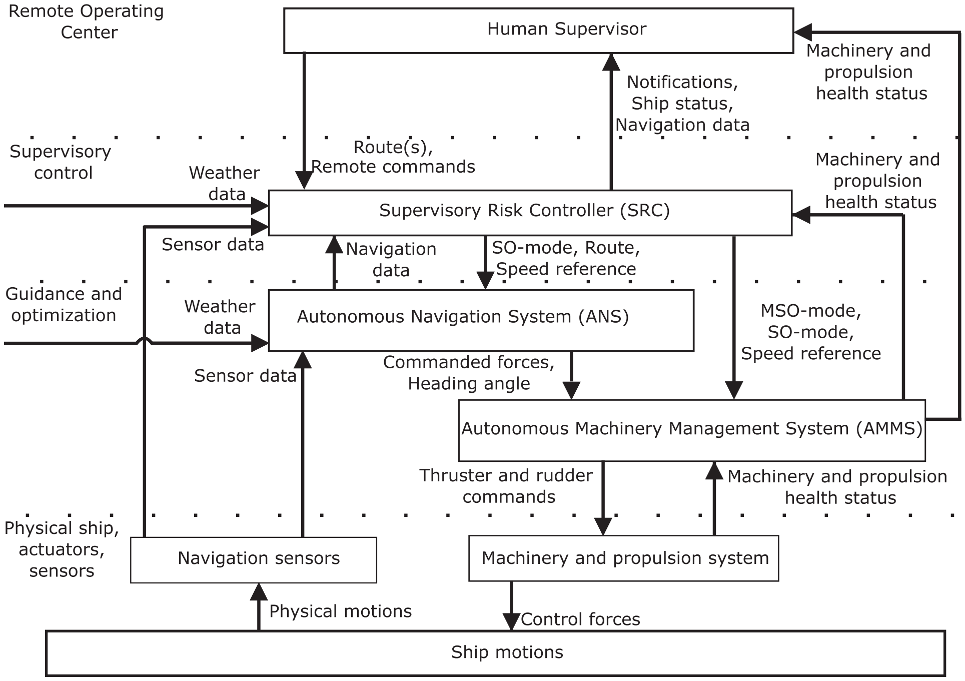
Remote Operating Center
Human Supervisor
Notifications, Ship status, Navigation data
Machinery and propulsion health status
Supervisory control
Weather data
Route(s), Remote commands
Machinery and propulsion health status
Supervisory Risk Controller (SRC)
Sensor data
Navigation data
SO-mode, Route, Speed reference
Guidance and optimization
Weather data
Autonomous Navigation System (ANS)
MSO-mode, SO-mode, Speed reference
Sensor data
Commanded forces, Heading angle
Autonomous Machinery Management System (AMMS)
Thruster and rudder commands
Machinery and propulsion health status
Physical ship, actuators, sensors
Navigation sensors
Machinery and propulsion system
Physical motions
Control forces
Ship motions

Figure 1. Control structure (adapted from Johansen and Utne (2022) and Johansen et al. (2023)).

results from the H-STPA. The SRC is also implemented with a route checker to see if the ship is able to follow the route. If not, the SRC can either switch to an alternative route or notify the human supervisor. Changing the route is possible if an alternative route is provided and the ship has not passed the point where the alternative route starts.

3.5. Verification of the control system

The fifth step is to verify that the control system works as intended, focusing on its functional behaviour, performance (Pedersen et al. 2022), and safety to ensure that it is ready for further use. This is done by simulating the ship in different scenarios within the CONOPS using verification objectives identified from the STPA and H-STPA. Verification objectives are formulated using a modified version of the method proposed in Rokseth et al. (2018).

In Rokseth et al. (2018), causal scenarios are used to specify safety constraints that, if violated, can lead to UCAs. Safety constraints are also used to derive verification objectives. Objectives are then processed to verify that the proposed control system can operate as intended without violating the safety constraints.

The verification objectives are processed in this study by simulating the ship in various scenarios to see if the objectives are satisfied. This is done by setting up a set of simulations and checking that all objectives are satisfied. An alternative method involves using an automated testing framework, as proposed in Torben et al. (2022).

4. Case study: autonomous coastal cargo ship

The purpose of the case study is to test the methodology and assess if the SRC will make reasonable decisions compared to a conventional ship. The role of the human supervisor is to make the overall plan for the SRC to follow. Furthermore, notifications from the SRC are received in the ROC, where the human supervisor is located; they have a communication link to the ship and the ability to assess the situation and intervene if necessary. This is the first step towards the design and implementation of human-in-the-loop control systems for autonomous ships.

4.1. Step 1: extended STPA with FTA

The ship in the case study is 80 m long and 15 m wide. It is equipped with a liquid natural gas (LNG)-powered main engine (ME), two diesel generators (DGs), and a hybrid shaft generator (HSG) for power production. The HSG can be used as a generator to obtain electrical power from the ME or as an electric engine powered by the DGs. The propulsion and steering system consists of a main propeller, two electric tunnel thrusters, and steering machinery controlling the rudder. The control system consists of an autonomous navigation system (ANS), an autonomous machinery management system (AMMS), and the SRC. The ANS handles the navigation and motion control, the AMMS controls the power production and propulsion, and the SRC makes high-level decisions for the rest of the control system to carry out. The full STPA control structure of the autonomous ship is shown in Figure 1. Collision avoidance is considered outside the scope of this work and therefore not included in the STPA control structure.

The ANS has a DP controller, an autopilot (AP) controller, and an observer for data processing. The DP controller is used for low-speed maneuvering and for station-keeping, while the AP controller is used to control the ship at higher speeds. The DP controller provides the required surge, sway, and yaw forces to control the position, heading, and speed of the ship. The AP controller is a line-of-sight (LoS) guidance controller that provides a heading reference based on the route and current ship position. The observer is used to process and check the data coming from navigation sensors, such as the GNSS. The data coming from these sensors must be filtered to remove noise and checked to confirm that these data are valid and do not contain measurement errors.

The AMMS consists of a power management system (PMS), thrust allocation (TA), a speed controller, and a rudder controller. The PMS is responsible for power production. Thrust allocation is used to convert the force commands from the DP controller into individual thrust commands for the propulsion system. The speed controller is used to control the load on the main propeller according to the speed reference, and the rudder controller converts the heading angle to a rudder angle for the steering machinery.

The SRC consists of the BBN risk model, ENC module, fuel consumption estimator, and controller. The SRC can select two different ship operation modes (SO-modes): DP-mode and AP-mode. When the ship is operated in AP-mode, the ANS uses the LoS controller to send a heading reference to the rudder controller in the AMMS. The speed reference is sent directly to the speed controller in the AMMS. The speed controller outputs a load percentage for the main propeller to maintain the desired speed, and the rudder controller provides a rudder angle to maintain the necessary heading angle. In AP-mode, the main propeller provides forward thrust, and the rudder controls the heading.

When operating in DP-mode, the DP controller calculates the force necessary to follow the desired route or maintain a certain position when it is used for station-keeping. The general force demand is mapped to individual thruster commands in the TA. In DP-mode, the main propeller provides thrust (surge), and the two tunnel thrusters control the sway and yaw. Since each degree of freedom (DOF) can be controlled directly, the DP-mode provides more accurate control of the ship than the AP-mode, but only at low speeds at which the tunnel thrusters are still efficient.

There are three different machinery system operation modes (MSO-modes), namely power take out (PTO), power take in (PTI), and mechanical (Mech). In PTO, the ME drives the main propeller. The HSG is used as a generator to produce electricity. In PTI, the two DGs produce electricity, and the HSG is used as an electric engine to power the main propeller. Mech uses the ME to power the main propeller and DGs to produce electricity.

The SRC is designed to manage the ANS and AMMS by setting the MSO-mode, SO-mode, and speed reference. It also has the option to enter an MRC, notify the human supervisor when necessary, and switch to an alternative route. The selection of modes and the speed reference is done using an optimisation algorithm that calculates the cost of operating the ship and selects the set with the lowest total cost. The MRC is entered when the risk cost becomes too high for the ship to continue sailing or when the ship loses redundancy in critical systems. When this happens, the ship will begin station-keeping and use the DP-controller to maintain its current position. The MRC is not included explicitly in the risk model since model updates are paused when this condition is triggered and remain paused until the situation is assessed by the human supervisor. Route changes are not directly linked to the risk model; instead, they are based on how much the ship drifts and deviates from its course in different weather conditions.

The STPA in this paper is based on a workshop with 12 participants that focused on risk analysis, ship control systems, and the verification of control systems. The experts have 5–30 years of experience in both academia and industry. The workshop was conducted in three sessions. The first two sessions were used to identify different UCAs, which were discussed and analysed in the third session. The sessions focused on the ship's machinery system and grounding and collision, but they also considered how selecting the wrong SO-mode could lead to hazardous events. The control structure, shown in Figure 1, includes the SRC and control responsibilities, as described above, in addition to the AMMS and the ANS. As the SRC is a novel functionality, the results from the workshops have been used as a basis for the analysis in this paper, but with some modifications to account for the changed control structure and control responsibilities due to the SRC. The STPA considers two hazardous events:

- HE1: The ship collides/allides with a ship/obstacle.
- HE2: The ship grounds or has contact with the seafloor.

Three system-level hazards can lead to these hazardous events:

- H1: The ship violates the minimum distance of separation to a ship/obstacle.
- H2: The ship violates the minimum distance of separation to shore.
- H3: The ship sails in too-shallow water.

The next step in the STPA is identifying UCAs. In this case study, UCAs are used to identify scenarios that should be checked during the verification process. Three types of UCAs are used in the case study: not providing a control action, providing an unsafe control action, or providing a control action at the wrong time (too late/early). The STPA also includes a fourth type of UCA, i.e. a signal lasts too long or stops too soon. However, since all signals considered in this case study are discrete, this is not relevant here. To build the BBN risk model, the relevant control actions are setting the MSO-mode, SO-mode, and speed reference, since these are decisions made by the SRC. In this work, changing the route is considered relevant for verification purposes but not for building the risk model since this decision is not made based on the risk cost. Instead, this decision is based on how much space the ship needs to maneuver with different wind and current conditions. Entering the MRC is also a different type of control action since this action is triggered when the ship is unable to continue sailing and continues until the human supervisor has assessed the situation. However, these actions are still important to consider when verifying the resulting control system.

Table 3 shows 13 different UCAs: four for selecting the MSO-mode, four for selecting the SO-mode, one for setting the speed reference, two for changing which route to follow, and two for entering the MRC. These UCAs are grouped together into six more general UCAs, as shown in Table 4. This makes the analysis easier to follow since it limits the number of UCAs describing the same type of situation.

Setting the speed reference is not explicitly included in the list of UCAs since it will impact the other UCAs as an RIF. UCA-1 and UCA-3 focus on failures that cause the machinery and propulsion system to be unable to function as intended. UCA-2 is related to the maximum power available in each mode, depending on the machinery parts used, and the ability to predict how much power the ship needs in different situations. UCA-4 is related to the ability to control the ship with respect to the SO-mode, speed reference, and conditions around the ship. Based on the six different UCAs shown in Table 4, the scenarios shown in Table 5 are specified.

The final part of the extended STPA is analysing the consequences of the hazardous events. This is necessary in order to be able to quantify the input data used for the optimisation of the control system. The consequences are first divided into either damage to the ship, damage to other ships/objects/structures, and harm to humans. Based on IMO (2018), these consequences are either severe, significant, minor, or nonexistent. Severe damage to the autonomous ship means that the ship is unable to continue without assistance and that it needs extensive repairs.

Significant damage means that the ship can get back to shore without assistance but will need extensive repairs before it can sail again. Minor damage must be repaired during the next planned maintenance period, but the ship can still sail with the damage that has been sustained. Severe damage to other objects/structures means that it needs immediate extensive repairs. Significant damage requires bigger repairs but is not as time critical. Minor damage should be repaired during the next planned maintenance period. Fatalities or serious injuries to humans are considered severe consequences.

Table 3. UCAs identified in the STPA.

Control action	Type	Context
MSO-mode	Providing a control action	Selecting an MSO-mode using failed machinery
MSO-mode	Providing a control action	Selecting an MSO-mode that produces a max. power that is too low
MSO-mode	Not providing a control action	Not selecting a different MSO-mode when machinery parts fail
MSO-mode	Not providing a control action	Not selecting a different MSO-mode when the current MSO-mode produces too little power
SO-mode	Providing a control action	Selecting an SO-mode using failed propulsion parts
SO-mode	Providing a control action	Selecting an SO-mode unable to control the ship with the current speed reference
SO-mode	Not providing a control action	Not selecting a different SO-mode when propulsion parts fail
SO-mode	Not providing a control action	Not selecting a different SO-mode when the speed reference is too high or low for the current mode
Speed reference	Providing a control action	Setting the speed reference too high when the ship has limited space to maneuver
Route selection	Not providing a control action	Not changing to an alternative route when the ship is unable to follow the initial route
Route selection	Providing a control action too late	Changing to an alternative route after passing the point where it was possible to change the route
MRC	Providing a control action	Entering the MRC when the ship is unable to maintain the current position due to propulsion failures
MRC	Providing a control action too late	Not entering the MRC when the traffic or conditions become too difficult for the ship to continue

Table 4. UCAs for the case study.

UCA	Description	Hazard(s)
UCA-1	The SRC changes to an MSO-mode that depends on failed parts of the machinery system.	H1, H2, H3
UCA-2	The SRC changes to an MSO-mode that is unable to produce the necessary power.	H1, H2, H3
UCA-3	The SRC changes to an SO-mode that depends on failed parts of the machinery system.	H1, H2, H3
UCA-4	The SRC changes to an SO-mode that is unable to maintain sufficient control of the ship.	H1, H2, H3
UCA-5	The SRC fails to change to an alternative route when the ship is unable to follow the original route.	H1, H2, H3
UCA-6	The SRC fails to enter the MRC when the situation makes it necessary.	H1, H2, H3

Table 5. Scenarios that could lead to UCAs.

Scenario	Description	UCA
Sc-1	The SRC selects PTO as the MSO-mode when a fault in the ME results in a loss of power.	UCA-1
Sc-2	The SRC selects PTO as the MSO-mode when a fault in the HSG results in a loss of electric power.	UCA-1
Sc-3	The SRC selects Mech as the MSO-mode when a fault in the ME results in a loss of propulsion power.	UCA-1
Sc-4	The SRC selects Mech as the MSO-mode when a fault with t he DGs results in a loss of electric power.	UCA-1
Sc-5	The SRC selects PTI as the MSO-mode when a fault in the HSG results in a loss of propulsion power.	UCA-1
Sc-6	The SRC selects PTI as the MSO-mode when a fault with the DGs results in a loss of power.	UCA-1
Sc-7	The SRC selects PTO as the MSO-mode when the load on the main propulsion system is higher than the power the ME can produce when it is also powering the HSG.	UCA-1
Sc-8	The SRC selects PTI as the MSO-mode when the total load on the machinery is higher than the power the DGs can produce.	UCA-2
Sc-9	The SRC selects AP as the SO-mode when a fault in the steering machinery results in a loss of steering for the ship.	UCA-3
Sc-10	The SRC selects AP as the SO-mode when a fault with the main propeller results in a loss of propulsion for the ship.	UCA-3
Sc-11	The SRC selects DP as the SO-mode when a fault with the main propeller results in a loss of propulsion for the ship.	UCA-3
Sc-12	The SRC selects DP as the SO-mode when a fault with the tunnel thrusters results in a loss of steering for the ship.	UCA-3
Sc-13	The SRC selects AP as the SO-mode when the speed is too low for the rudder to control the ship.	UCA-4
Sc-14	The SRC selects AP as the SO-mode when the ship is maneuvering in very tight areas where the AP-controller is unable to provide sufficient control.	UCA-4
Sc-15	The SRC selects DP as the SO-mode when the speed is too high for the tunnel thrusters to produce the necessary thrust to maneuver the ship.	UCA-4
Sc-16	The SRC fails to change the route, because the control system underestimates the current conditions.	UCA-5
Sc-17	The SRC fails to enter the MRC while it can still do so safely because the current conditions are underestimated.	UCA-6
Sc-18	The SRC enters the MRC when it is unable to maintain its position due to a failure with the tunnel thrusters.	UCA-6

Less serious injuries are considered significant consequences, and insignificant injuries such as scratches and bruises are minor consequences.

The qualitative fault tree analysis focuses on three critical subsystems identified in the STPA as being especially important for operating the ship:

- The machinery system;
- The propulsion system;
- The navigation sensors and communication system.

These sub-systems are analysed in more detail to identify when the ship is unable to continue sailing because one of these systems fails or because redundancy is lost so that the control system can notify the human supervisor and make alternative plans. A fault tree analysis, even though it is a qualitative analysis, provides information about what components are necessary to operate the ship in the different SO- and MSO-modes. The same fault trees are used to identify situations in which the ship loses redundancy in the same systems.

The information from the fault trees is used to construct the BBN so that specific components can be monitored in more detail. Each sub-system fault tree is represented by a node in the BBN to monitor the status of each sub-system. These components receive input from nodes in the BBN that describe the individual components.

Figure 2 shows that the machinery system can fail in two ways. If the ME, DG1, and DG2 fail, the ship loses power. It will also lose

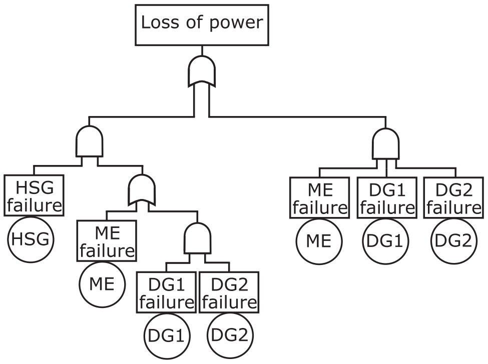

Figure 2. Fault tree showing a loss of power production for the autonomous ship.

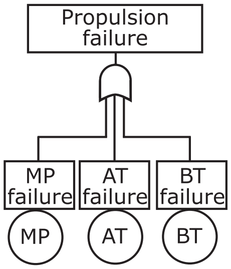

Figure 3. Fault tree showing a loss of propulsion for the autonomous ship.

power if the HSG fails and either the ME or both DGs fail. If the HSG, ME, or both DGs fail, the ship loses redundancy in the power production system. It will, however, still have the necessary power for propulsion and navigation.

The propulsion system is analysed in Figure 3. The propulsion system is considered to have failed if the MP or either of the two tunnel thrusters fails. The MP is critical since it provides forward thrust in both DP-mode and AP-mode. The tunnel thrusters are considered critical since they are necessary to control the ship if it enters the MRC. If the steering machinery fails, the ship can only operate in DP-mode and therefore loses redundancy.

Figure 4 shows the fault tree for the navigation and communication system. This system consists of the GNSS, which provides position and speed data, communication systems to send and receive information from the ROC, an AIS that obtains information about other vessels around the ship, and radar for sensing ships and other objects. The system is considered to have failed if the GNSS, communication systems, or both radar and AIS fail. GNSS is considered to be critical for obtaining the position and speed data that allow the ship to navigate. Communication is critical to maintaining the connection between the ship and the ROC. In this work, either AIS or radar is considered necessary to obtain information about other vessels around the ship. For an actual ship, this system should also include cameras and additional sensors to ensure sufficient situational awareness, as especially using only AIS can limit this. However, the fault tree and sensor package shown here is considered sufficient to show how such a system can work.

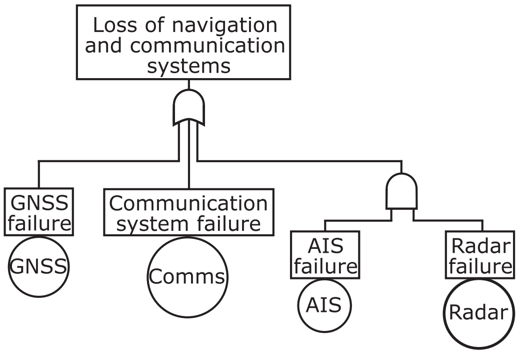
Loss of navigation and communication systems
GNSS failure
GNSS
Communication system failure
Comms
AIS failure
AIS
Radar failure
Radar

Figure 4. Fault tree showing a loss of navigation and communication systems onboard the autonomous ship.

Table 6. HUCAs used to identify scenarios that can lead to hazardous events.

HUCA	Description	Hazard(s)
HUCA-1	The human supervisor does not provide a notification to other ships.	H1
HUCA-2	The human supervisor is too late in notifying other ships.	H1
HUCA-3	The human supervisor does not initiate and organise emergency actions, including towing and rescue.	H1, H2, H3
HUCA-4	The human supervisor does not take remote control of the ship.	H1, H2, H3
HUCA-5	The human supervisor is too late in taking remote control of the ship.	H1, H2, H3
HUCA-6	The human supervisor takes remote control of the ship without the necessary understanding or time to safely control the ship.	H1, H2, H3
HUCA-7	The human supervisor hands over control to the autonomous ship when the autonomous system is unable to safely control the ship.	H1, H2, H3
HUCA-8	The human supervisor hands over control to the autonomous ship too early.	H1, H2, H3
HUCA-9	The human supervisor is too late in handing over control to the autonomous ship.	H1, H2, H3
HUCA-10	The human supervisor does not hand over control to the autonomous ship.	H1, H2, H3

4.2. Step 2: human STPA

The next step is focusing more specifically on the human supervisor, who is already included in the control structure (Figure 1) as a separate controller. In normal operation, the human supervisor is responsible for providing the plan(s) for the SRC to follow. When the autonomous ship is sailing, the human supervisor is in the ROC with a communication link to the ship. The human supervisor is responsible for following multiple ships at the same time and performing other tasks in the ROC. This means that the SRC must provide a notification in due time to allow the human supervisor to act. In this case study, the control system is not implemented with the ability to make new plans. The ship will therefore be operated at either LoA 3 or LoA 1.

The human supervisor can perform the following actions from the ROC:

- Notify other ships;
- Initiate and coordinate emergency actions, including contacting towing and rescue vessels;
- Take remote control of the ship;
- Hand over control to the autonomous system.

Based on these actions, the ten unsafe human control actions (HUCAs) shown in Table 6 were identified. HUCAs are used to differentiate between unsafe control actions related to the computer-based control system and unsafe control actions related to the human supervisor. From the HUCAs, a total of 24 scenarios in which these actions can occur are identified; they are shown in Tables 7–8.

This table includes both scenarios in which the control system fails to notify the human supervisor and scenarios in which the

Table 7. Scenarios in which a control action from the human supervisor can lead to a hazardous event.

Scenario	Description	HUCA
Sc-1	The human supervisor is not notified when the ship loses power and therefore does not notify other ships that the ship has lost power and is drifting without control.	HUCA-1
Sc-2	The human supervisor misses a notification due to exhaustion or tiredness when the ship loses power and therefore does not notify other ships that the ship has lost power and is drifting without control.	HUCA-1
Sc-3	The human supervisor is not notified when the ship loses propulsion and therefore does not notify otherships that the ship has lost propulsion and is drifting without control.	HUCA-1
Sc-4	The human supervisor misses a notification due to exhaustion or tiredness when the ship loses propulsion and therefore does not notify other ships that the ship has lost propulsion and is drifting without control.	HUCA-1
Sc-5	The human supervisor is notified too late when the ship loses power and therefore notifies other ships too late that the ship has lost power and is drifting without control.	HUCA-2
Sc-6	The human supervisor is too late to recognise a notification due to exhaustion or tiredness when the ship loses power and therefore notifies other ships too late that the ship has lost power and is drifting without control.	HUCA-2
Sc-7	The human supervisor is notified too late when the ship loses propulsion and therefore notifies other ships too late that the ship has lost power and is drifting without control.	HUCA-2
Sc-8	The human supervisor is too late to recognise a notification due to exhaustion or tiredness when theship loses propulsion and therefore notifies other ships too late that the ship has lost propulsion and is drifting without control.	HUCA-2
Sc-9	The human supervisor is not notified when the ship loses power and therefore does not initiate or organise towing and rescue.	HUCA-3

notifications are missed by the human supervisor. The rest of the paper focuses on the former scenarios since the aim is to design a control system that accounts for this possibility. Going into more detail on human factors, such as fatigue and boredom, is considered outside the scope of this work.

A challenge with integrating the human supervisor in the loop is providing enough time for intervention, i.e. to determine when it is necessary for the control system to notify the human supervisor. If the SRC is too late or does not provide a notification, the human supervisor will not be able to take the necessary action. However, if the SRC provides too many unnecessary notifications, the human supervisor may start neglecting these notifications. Over time, this can become a serious problem; the human supervisor may stop reacting to the notifications. The information given in the notifications can also affect the human supervisor's ability to react. Since the autonomous ship is not monitored continuously, the human supervisor will most likely not have a full overview of the situation when they receive a notification. The SRC should therefore provide the human supervisor with the information they need to react in addition to the notification. The results from the H-STPA are used to set up up the SRC.

4.3. Step 3: building the online risk model

The UCAs and scenarios shown in Tables 4 and 5, respectively, form the basis of the risk model. The risk model uses the first four UCAs. Changing the route is considered separately based on how much space the ship needs to maneuver depending on the wind and current. The MRC is entered when the risk cost becomes too high. When this happens, however, updating the risk model is paused until the ship exits the MRC. The two last UCAs are therefore not specifically added to the risk model. This also means that only the 15 scenarios based on UCA1–UCA4 are used to identify high-level RIFs (H-RIFs).

To reduce the complexity of the risk model, the scenarios are grouped together into the six H-RIFs shown in Table 9. The H-RIFs

Table 8. Scenarios in which a control action from the human supervisor can lead to a hazardous event.

Sc-10	The human supervisor misses a notification due to exhaustion or tiredness when the ship loses power and therefore does not initiate or organise towing and rescue.	HUCA-3
Sc-11	The human supervisor is not notified when the ship loses propulsion and therefore does not organise towing and rescue.	HUCA-3
Sc-12	The human supervisor misses a notification due to exhaustion or tiredness when the ship loses propulsion and therefore does not organise towing and rescue.	HUCA-3
Sc-13	The human supervisor is not notified that it is necessary to take remote control of the ship when the autonomous system is unable to control the ship.	HUCA-4
Sc-14	The human supervisor misses a notification due to exhaustion or tiredness that it is necessary to take remote control when the autonomous system is unable to control the ship.	HUCA-4
Sc-15	The human supervisor is notified too late that it is necessary to take remote control when the autonomous control system is unable to control the ship.	HUCA-5
Sc-16	The human supervisor is too late to recognise a notification due to exhaustion or tiredness that it is necessary to take remote control when the autonomous control system is unable to control the ship.	HUCA-5
Sc-17	The human supervisor takes remote control of the ship without the necessary situational awareness to safely control the ship.	HUCA-6
Sc-18	The human supervisor takes remote control of the ship while performing many other tasks in the ROC at the same time, which results in the nsafe control of the ship.	HUCA-6
Sc-19	The human supervisor receives incorrect information about the situation and therefore hands over control to the autonomous control system before it is safe to do so.	HUCA-7
Sc-20	The human supervisor has incorrect information about the system's autonomous capabilities and therefore hands over control to the control system before it is safe to do so.	HUCA-7
Sc-21	The human supervisor receives incorrect information about the situation and therefore hands over control to the autonomous control system when it is unsafe to do so.	HUCA-8
Sc-22	The human supervisor has incorrect information about the system's autonomous capabilities and therefore hands over control to the control system when it is unsafe to do so.	HUCA-8
Sc-23	The human supervisor must perform other tasks in the ROC before control is handed back to the autonomous system, which results in the unsafe control of the ship.	HUCA-9
SC-24	The human supervisor does not hand over control to the autonomous system while performing many other tasks in the ROC at the same time, which results in the unsafe control of the ship.	HUCA-10

Table 9. High-level RIFs used in the case study with the relevant UCAs.

H-RIF	Description	UCA(s)
H-RIF-1	Machinery health status	UCA-1
H-RIF-2	Estimation of necessary power	UCA-2
H-RIF-3	Propulsion system health status	UCA-3
H-RIF-4	Navigational situation	UCA-4
H-RIF-5	Situational awareness of the control system	UCA-2, UCA-4
H-RIF-6	Control system reliability	UCA-2, UCA-4

Table 10. Input nodes derived from the H-RIFs used to build the risk model.

High-level RIF	Description	Input nodes
H-RIF-1	Machinery health status	ME state HSG state DG1 state DG2 state
H-RIF-2	Estimation of necessary power	PMS DP controller performance/ accuracy AP controller performance/ accuracy
H-RIF-3	Propulsion system health status	BT state AT state MP state ST state
H-RIF-4	Navigational situation	Traffic Obstacles Distance to closest grounding hazard Wind speed Wind direction Current Ship speed
H-RIF-5	Situational awareness of the control system	Wind speed Fog Rain Snow Cameras AIS Radar GNSS Communication system
H-RIF-6	Control system reliability	DP controller performance/ accuracy AP controller performance/ accuracy AIS Radar GNSS Communication system

are divided further into input nodes for the risk model, as shown in Table 10.

The risk model also has input nodes connected to the hazardous events and the consequences resulting from these events. The probability of collision/allision with another ship/obstacle depends on both the probability of violating the minimum separation distance and the ability of the other ship/obstacle to avoid the collision/allision. The consequences depend on different nodes and the hazardous event. If the ship collides/allides with another ship/obstacle, the damage to the ship depends on the size of the other ship/obstacle and the impact speed. If the ship grounds or has contact with the seabed, the consequences depend on the impact speed, the type of shore, and the seabed. Harm to humans depends on the number of people aboard the other ship/obstacle or the type of shore. Damage to other ships/obstacles depends on the impact speed and size of the other ship/obstacle. If the ship grounds or has contact with the seabed, the impact speed and type of shore affect the consequences. The conditional probability tables (CPTs) are built up based on data from Johansen and Utne (2022), DNVGL (2003), and Hassel et al. (2021) to obtain the likelihoods of the hazardous events shown in Figure 5.

The BBN is also used to monitor the machinery, propulsion, and navigation and communication systems based on the fault tree analysis with the nodes power status, propulsion status, and navigation status. This provides the SRC with the information it needs to assess whether the ship still has the necessary redundancy to continue sailing in a given situation. The power and propulsion systems have three states, ok, minimum, and failed, according to the fault tree analysis. Losing redundancy means that the node is set to minimum. The ship still has power and propulsion but will lose power and propulsion if another component fails. Each component, such as the ME and HSG, is modelled as either failed or working. The different navigation and communication systems are described using three states: poor, sufficient, and good. These systems are therefore only considered as failed or ok based on the fault tree analysis. In operation, the nodes describing power and propulsion are considered failed if the probability of losing power exceeds 0.3, and they are considered minimum if the probability of losing redundancy exceeds 0.3. The limit is set based on testing to find a balance between keeping the ship from stopping too often and also avoiding the situation in which the ship continues to sail when systems are not functioning.

The ENC module is used to find the presence and density of obstacles around the ship and the distance to the closest point the ship cannot safely navigate to. The module is set up such that anything shallower than 5 m is considered a shallow area that the ship must avoid in order to navigate safely. The obstacle density is based on the distance to the closest shallow point (i.e. areas with a water depth of less than 5 m) and on how much of the water around the ship is obstructed. The water depth of 5 m is the same as the maximum draft of the ship. Using this water depth is considered sufficient for assessing the proportion of obstructed water in this work. The ship must then avoid shallow areas with sufficient safety margins.

The percentage of obstructed water is calculated by considering a disk with a radius of 1400 m and finding the portion of the disk with land and shallow water. The radius is set through testing to ensure that the disk gives a good picture of the sea area surrounding the ship, considering that the ship is 80 m long. The ENC module is checked every 30 seconds to provide updated measurements to the risk model. Testing shows that this provides a good balance between the computation time and updated data. This information is provided to the risk model through the nodes 'Obstacles' and 'Distance to closest grounding hazard.'

4.4. *Step 5: building the SRC*

The SRC is set up using two sub-steps: the first involves setting up the actual controller and testing it to identify operational limitations. The second part involves implementing notifications to the human supervisor based on the results from the H-STPA.

The SRC calculates the expected cost using Equation (1) for each set of decisions (MSO-mode, SO-mode, and speed reference). The cost is the sum of the fuel cost $F(d)$, risk cost $R(d)$, operation cost $O(d)$, and potential future loss $L(d)$, depending on the decisions d.

The fuel cost is calculated using a look-up table with the specific fuel consumption (SFC) for different MSO-modes, speeds, wind conditions, and current conditions. This is multiplied by the planned sailing distance and the fuel price, as shown in Equation (2). This provides a good approximation for the fuel consumption, despite not accounting for all variations due to changing angles for wind and current in different places along the route. Calculating for each specific part of the route would also take much longer time due to the

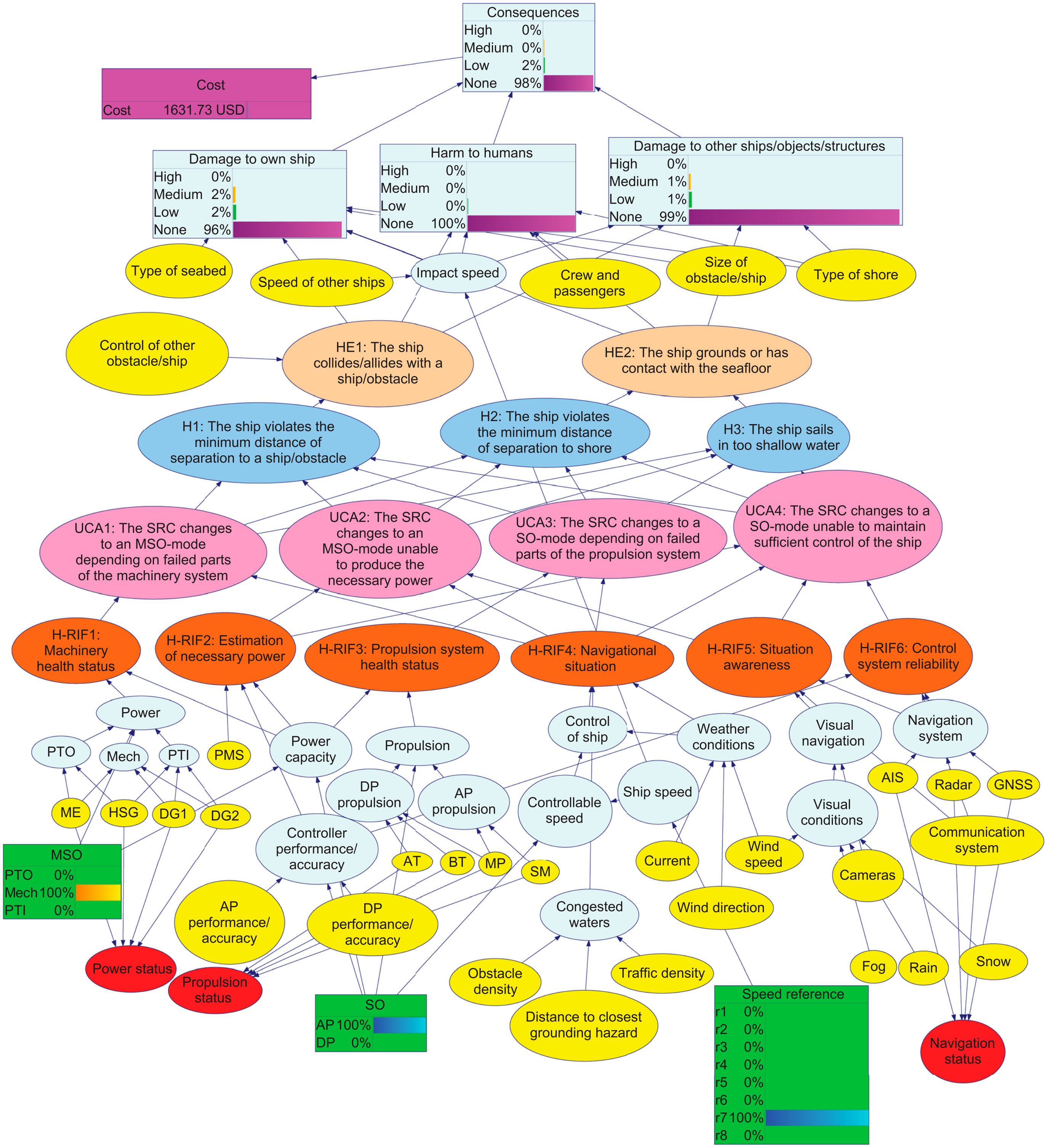
Consequences
High 0%
Medium 0%
Low 2%
None 98%
Cost
Cost 1631.73 USD
Damage to own ship
High 0%
Medium 2%
Low 2%
None 96%
Harm to humans
High 0%
Medium 0%
Low 0%
None 100%
Damage to other ships/objects/structures
High 0%
Medium 1%
Low 1%
None 99%
Type of seabed
Speed of other ships
Impact speed
Crew and passengers
Size of obstacle/ship
Type of shore
Control of other obstacle/ship
HE1: The ship collides/allides with a ship/obstacle
HE2: The ship grounds or has contact with the seafloor
H1: The ship violates the minimum distance of separation to a ship/obstacle
H2: The ship violates the minimum distance of separation to shore
H3: The ship sails in too shallow water
UCA1: The SRC changes to an MSO-mode depending on failed parts of the machinery system
UCA2: The SRC changes to an MSO-mode unable to produce the necessary power
UCA3: The SRC changes to a SO-mode depending on failed parts of the propulsion system
UCA4: The SRC changes to a SO-mode unable to maintain sufficient control of the ship
H-RIF1: Machinery health status
H-RIF2: Estimation of necessary power
H-RIF3: Propulsion system health status
H-RIF4: Navigational situation
H-RIF5: Situation awareness
H-RIF6: Control system reliability
Power
PTO
Mech
PTI
PMS
Power capacity
Propulsion
Control of ship
Weather conditions
Visual navigation
Navigation system
ME
HSG
DG1
DG2
DP propulsion
AP propulsion
Controllable speed
Ship speed
Visual conditions
AIS
Radar
GNSS
MSO
PTO 0%
Mech 100%
PTI 0%
Controller performance/ accuracy
AT
BT
MP
SM
Current
Wind speed
Communication system
Cameras
AP performance/ accuracy
DP performance/ accuracy
Congested waters
Wind direction
Power status
Propulsion status
Obstacle density
Traffic density
Fog
Rain
Snow
SO
AP 100%
DP 0%
Distance to closest grounding hazard
Speed reference
r1 0%
r2 0%
r3 0%
r4 0%
r5 0%
r6 0%
r7 100%
r8 0%
Navigation status

Figure 5. Online risk model BBN (adapted from Johansen and Utne (2022) and Johansen et al. (2023) and extended) showing an example of the risk cost.

increased complexity and need for more online simulations to estimate the fuel consumption. The following prices of LNG and diesel are taken from Ship & Bunker (2022): 1,326.50 USD/ton for LNG and 679.50 USD/ton for diesel. The price is therefore dependent on the MSO-mode, since this determines the type of fuel used:

$$F(d) = SFC(speed, wind, current, MSO) * distance * Price(MSO). \quad (2)$$

The risk cost is calculated from the risk model using Equation (3), which takes the probability of each consequence category from the STPA and multiplies it by the cost of the corresponding category. Severe consequences are given a cost of 4,550,640 USD, significant consequences have a cost of 455,064 USD, and minor consequences have a cost of 45,506.40 USD. These costs are estimated based on EfficienSea (2012), The Norwegian Agency for Public and Financial

Management (2018), and IMO (2018):

$$R(d) = Pr(severe)C_{severe} + Pr(significant)C_{significant} + Pr(minor)C_{minor} + Pr(none)C_{none}. \quad (3)$$

The operation cost is calculated by taking the cost per hour, multiplying it by the planned sailing distance, and dividing the resulting value by the speed reference, as shown in Equation (4). This cost includes personnel costs in the ROC and maintenance, insurance, lubrication, spare-parts, and logistics costs. These costs are estimated to be 341.30 USD/ht, based on costs from similar ships and data from Stopford (2009):

$$O(d) = Cost_{operating} * distance/speed. \quad (4)$$

The potential future loss is calculated similarly using Equation (5), with the expected loss of income per hour set to 910.10 USD/h. This cost represents the potential income if the ship was free and could start the next trip or mission earlier and calculated similarly as the operation cost. This can also be considered as a penalty cost to balance the risk, fuel, and operation costs.

$$L(d) = Cost_{future\ loss} * distance/speed. \quad (5)$$

The SRC considers a constant planning horizon equal to the remaining sailing distance at the start of the mission, d_0, used to calculate both fuel cost, operation cost, and the potential future loss. The costs considered in the SRC are then the costs of sailing another distance d_0. For the case study, this is equal to around 57 km or 30 nautical miles. By using a constant distance to calculate the cost, the weights of the risk, fuel, operation, and potential future loss are kept constant. Without a constant distance, the SRC would put more relative weight on the risk when the distance is small. This would cause the ship to go slower and use more energy, the closer it gets to the final way-point.

To check a route, the SRC goes through all the way-points to determine if the ANS can follow it with sufficient margins. Between each way-point, a set of intermediate points is used to check that the margin is sufficient along the whole route. In this work, the margin is set based on how accurately the ANS can control the ship in different wind conditions.

As identified in the H-STPA, finding the right balance between providing and not providing notifications to the human supervisor has a significant effect on the overall performance. To achieve this, the human supervisor should only be notified when the SRC expects that it will be unable to control the ship in the future. However, these notifications should be made before the SRC loses control so that the human supervisor has time to react. The human supervisor should also be notified when components, or sub-systems, fail without warning. Based on this, the human supervisor is notified when it become necessary to perform any of the control actions described in Subsection 4.2.

The SRC receives information from the risk model about the status of the machinery, propulsion, and navigation and communication systems, as described in Subsection 4.3. If any of these sub-systems fail, the human supervisor is notified that the ship is unable to continue. There is little the human supervisor can do in these situations, except for notifying nearby ships and the relevant authorities. The risk model is also used to assess redundancy in the machinery and propulsion system. If the autonomous ship loses redundancy in these systems, the human supervisor is notified, and the ship will enter the MRC. The ship will also enter the MRC if the risk cost becomes too high, e.g. due to changes in the environment or weather. In this case study, the cost limit for the SRC to enter the MRC is set at 5,119.47 USD, which is very low compared to the costs associated with the different consequences. However, testing shows that the risk cost very rarely exceeds this value, and this only occurs when the ship is unable to continue to sail safely. This is discussed further in Subsection 5.2.1.

Table 11. Verification objectives based on the STPA and H-STPA.

Verification objective	Description
VO-1	Verify that the SRC handles machinery failures by either changing the MSO-mode or entering the MRC-mode.
VO-2	Verify that the SRC selects a safe combination of the SO-mode and speed reference.
VO-3	Verify that the SRC enters the MRC with sufficient time and functionality for the ship to maintain its current position.
VO-4	Verify that the human supervisor is notified in the intended situations and avoid unnecessary notifications.
VO-5	Verify that the SRC provides notifications with the necessary information to allow the human supervisor to react to the situation.
VO-6	Verify that the SRC checks the route and, when necessary, either changes it or notifies the human supervisor that it is unable to change the route.

When the ship enters the MRC, it will try to maintain its current position until the human supervisor has checked the situation and decided how to proceed. In the MRC, the autonomous ship uses the DP-controller to maintain its position. The MSO-mode is chosen by checking the risk model to find out which mode has the lowest risk cost. If the ship is unable to change the active route, the human supervisor is sent a notification that explains why the route should be changed and why the SRC was unable to change the actively selected route.

4.5. Step 7: testing and verification of the control system

The SRC should be tested to check that it can control the ship in a safe and efficient manner before implementation and/or during updates/modifications. Setting up test scenarios starts with the different UCAs and HUCAs identified in the STPA and H-STPA. These are used as the basis for formulating high-level safety constraints and scenarios in which these constraints can be violated.

The STPA scenarios are mainly related to selecting an MSO-mode that is unable to produce the necessary power, a mismatch between the speed reference and SO-mode, using propulsion parts that have failed, or the speed being higher than it should be in confined or narrow areas. The scenarios describing insufficient power production are either caused by failures or due to the total load on the machinery system. Problems with setting the speed reference can involve setting it too low to use the rudder to steer the ship or setting it too high to use the tunnel thrusters. Scenarios identified in the H-STPA focus mostly on when the human supervisor is or should be notified. The H-STPA also identified scenarios in which the human supervisor has an insufficient understanding of the situation (mainly scenarios 17–24). Verification objectives are formulated based on the STPA and H-STPA scenarios; they are shown in Table 11.

The proposed control system is tested against the six verification objectives by simulating the ship and allowing random changes in the system and environment. The SRC must handle these changes, regardless of the timing and location of these changes. The simulator is based on the equations from Fossen (2011) with simplified dynamics and machinery models. The DP and autopilot controllers are PID controllers included as part of the simulator.

5. Results and discussion

5.1. Results

To demonstrate the proposed methodology, the SRC is tested using the verification objectives on a route close to Brønnøysund in a number of simulations with varying wind and current conditions, as well as random failures in the machinery and propulsion system. The wind speed is from 0–21 m/s from north, east, south, and west. The initial wind speed is increased by 0.5 m/s after each simulation, resulting in a total of 176 simulations to check. The wind is given an initial speed, with a 1×10^{-4} probability of changing at each time step during the simulation. The current is between 0 and 0.1 m/s. The current is given a random initial speed and direction that is then kept constant for the remaining time. Both the wind and current conditions are based on historical data from Norwegian Meteorological Institute (2021) and Barentswatch (2022) for the area considered, but they are assumed to be the same over the whole area.

The ship is simulated with random failures occurring in the machinery and propulsion system that the SRC must handle correctly. The ship has an original route passing through Brønnøysund (the yellow route in Figure 6) and an alternative route going around Brønnøysund (the white route in Figure 6). The alternative route provides more space for the ship to maneuver but is slightly longer.

The simulator is based on a simplified ship model without waves but with the wind and current affecting the ship. Failures are introduced using a random function in Python; there is a 1×10^{-5} probability of losing either power or propulsion, and losing redundancy, at each time step in the simulation. This is an artificially high probability to ensure that failures occur in order to test the controller. The wind is given an initial speed, which may both increase and decrease. The wind has a 1×10^{-4} probability of changing at each time step. The current is given a random initial speed and direction that is then kept constant for the remaining simulation time. The six verification objectives must be satisfied in each simulation for the SRC to pass the test.

Out of the 176 simulations, the SRC enters the MRC in 28 simulations, and the ship has a critical failure in three of these 28 simulations. The route is changed in 95 cases because of the current, wind, or a combination of both. The SRC manages to control the ship in a safe and efficient manner from start to finish by selecting the best MSO-mode, SO-mode, speed reference, and route according to the conditions. If any systems fail or the conditions exceed the operational limits of the autonomous ship, the SRC enters the MRC with sufficient time to stop and maintain its position. The human supervisor is then able to check the situation and decide how to proceed. Overall, the results show that the control system satisfies the verification objectives, but it is slightly conservative for the current setup. The following subsections show how the SRC works in some of the simulations.

5.1.1. Simulation 1: calm wind and current without any machinery problems

The first simulation has a wind speed between 0 and 2 m/s and a current speed of 0.07 m/s. The ship has no problems with the machinery and there is thus no need to change the route while the ship is underway. A timeline of the first simulation is shown in Figure 7. The ship starts with a speed of 7 m/s, as shown in Figure 8. This speed is reduced to 3 m/s after around 85 minutes because the route passes through a narrow strait. The route then passes through a more open area for a short while, and thus the speed is increased back to 7 m/s. As the autonomous ship enters the harbour area of Brønnøysund after around 95 minutes, the speed is reduced to around 3 m/s to account for speed limitations when sailing close to land. The ship keeps this speed through the harbour and increases the speed back to 7 m/s when it exits the harbour after around 115 minutes. The costs estimated by the SRC are shown in Figure 9. As long as the decisions d (MSO-mode, SO-mode, speed reference) and conditions stay the same, the fuel cost, operation costs, and potential future loss stay constant since they are calculated as the assumed cost of continuing to sail for a distance equal to d_0, as described in Subsection 4.4. When the ship enters the narrower parts of the route after 85 minutes, the risk cost starts to increase since the obstacle density increases and the distance to the closest grounding hazard decreases. After a short period of around five minutes, the risk cost is high enough for the SRC to lower the speed reference from 7 m/s down to 3 m/s. The ship will then use significantly more time to sail another distance d_0, which increases the operation costs and potential future losses. The fuel cost is reduced slightly since the ship uses less fuel and switches to PTI, which is cheaper with respect to fuel costs.

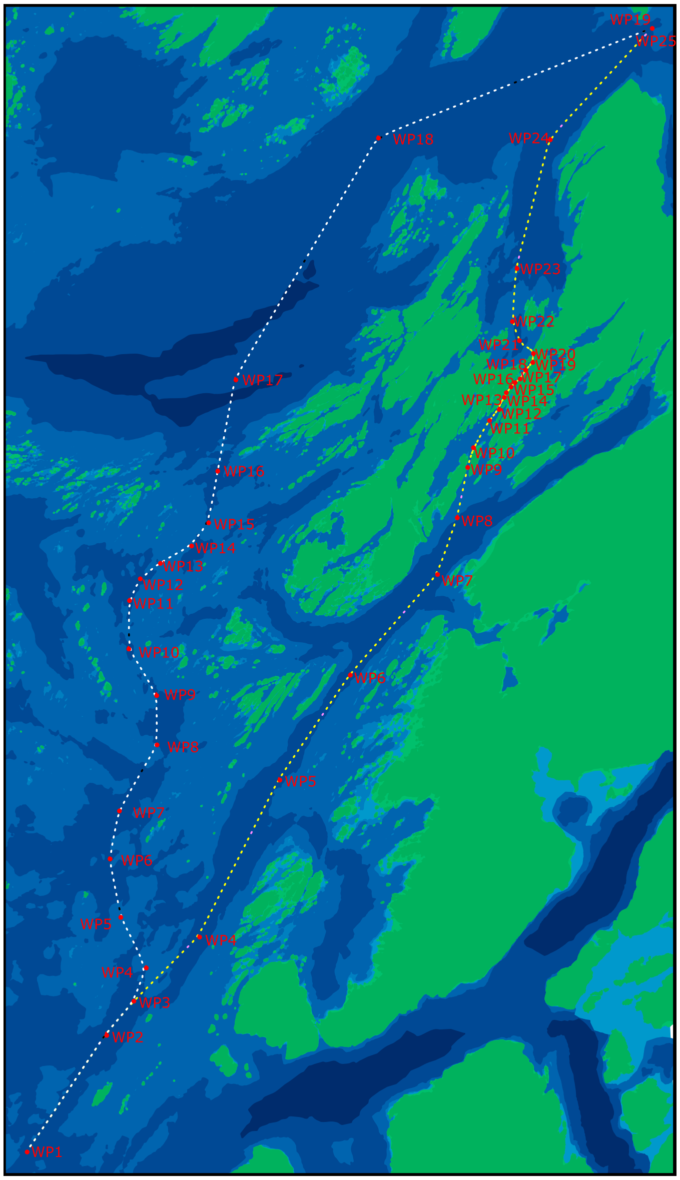

Figure 6. Map of the two routes sailed by the ship. The main route followed in simulation 1 is shown in yellow, and the alternative route followed in simulation 2 is shown in white.

The SRC changes the MSO-mode from Mech to PTI when it sails at a lower speed since the ship then needs less power. Operating in

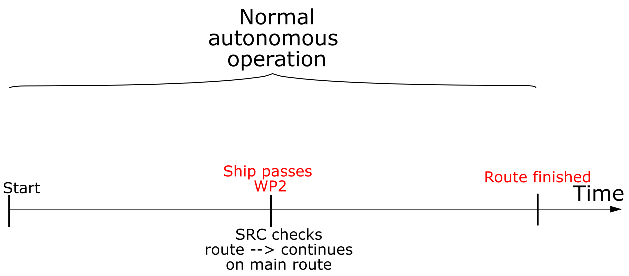

Figure 7. Timeline of the first simulation showing when the route is checked and the SRC decides how to proceed.

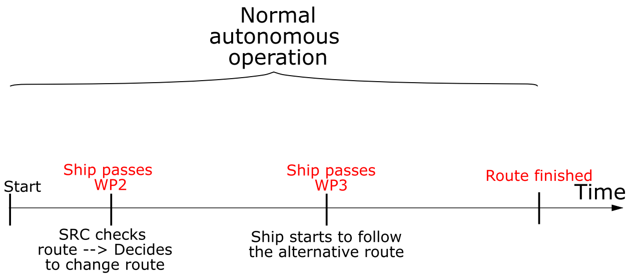

Figure 10. Timeline of the second simulation showing when the route is checked and when the ship starts to follow the alternative route.

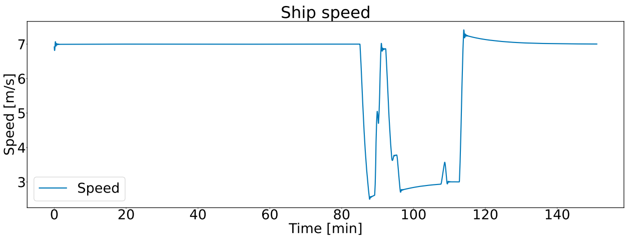

Figure 8. Ship speed in simulation 1.

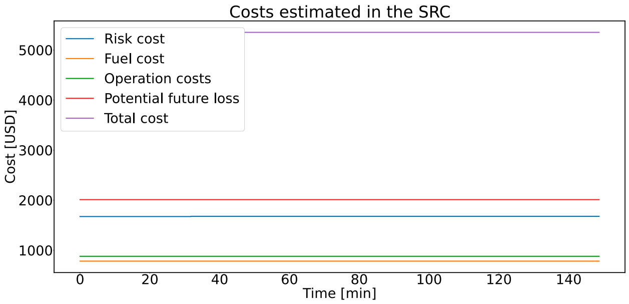

Figure 11. Cost in simulation 2 (d_0 = distance of the full route).

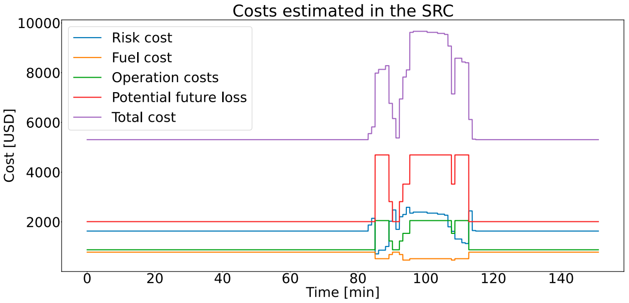

Figure 9. Cost in simulation 1 (d_0 = distance of the full route).

PTI also reduces the fuel cost slightly. The ship uses the autopilot for the whole simulation. The ship takes 150 minutes to sail the whole route and sails 57.7 km.

5.1.2. Simulation 2: strong breeze without machinery problems

In the second simulation, the ship is sailing in wind with a speed between 10 and 11 m/s. The route is first checked at WP2, where the SRC decides to follow the longer route (white route in Figure 6), where there are fewer obstacles to maneuver around and more space. The two routes split at WP3, where the ship then follows the white route. A timeline of the second simualtion is shown in Figure 10. On the alternative route, the distance to the closest grounding hazard and the obstacle density do not change enough to affect the risk cost. Combined with the constant planning horizon, this means that the costs stay constant throughout the whole simulation, as shown in Figure 11. Since the risk cost stays constant, the MSO-mode, SO-mode, and speed stay constant.

5.1.3. Simulation 3: wind increases after the ship passes the alternative route

The third simulation shows the autonomous ship in winds with a speed of 6.5 m/s. The ship starts with a speed of 7 m/s, similar to the first simulation. The SRC first checks the route after passing WP2 in Figure 12. At that point, the SRC determines that it should follow the original route because the conditions are not too bad. As the ship continues, the wind starts to increase from the original speed of 6.5 m/s up to 8.5 m/s. At that point, the SRC reevaluates the route and determines that it would be best to be on the alternative route since it might encounter problems if it continues. However, the ship has passed the point where the two routes split, WP3, and turning around is not a possible decision for the current implementation of the SRC. Instead, the SRC enters the MRC-mode, starts slowing the ship down, and notifies the human supervisor about the situation. At this point, the SRC stops updating the cost since it stays in the MRC-mode until the human supervisor has decided how to proceed. In this case, the ship stops and the simulation is stopped without showing what the human supervisor decides to do, as shown in Figure 13. Since the SRC enters the MRC-mode because of the potential future situation, the costs shown in Figure 14 are constant.

5.1.4. Simulation 4: ship loses redundancy in power production

The fourth simulation shows how the SRC handles losing redundancy in the machinery system. The ship is sailing in calm weather with a wind speed of 2 m/s and a current speed of 0.07 m/s (Figure 16). The ship starts with a speed of 7 m/s, similar to the previous simulations. The ship passes WP3, where the SRC checks the route and decides to continue as planned. As the ship reaches WP9, the risk cost increases, as shown in Figure 17, at around 85 minutes. The SRC then starts reducing the speed reference to maintain sufficient control. At the same time, as the ship is slowing down, the SRC recognises that the main engine has problems. It then decides to switch the MSO-mode to PTI to avoid using the main engine. At the same time, the SRC decides to enter the MRC-mode and notifies the human supervisor since the fault tree showed that this main engine problem results in a loss of redundancy. While the human supervisor is notified, the ship stops and maintains its position. The simulation

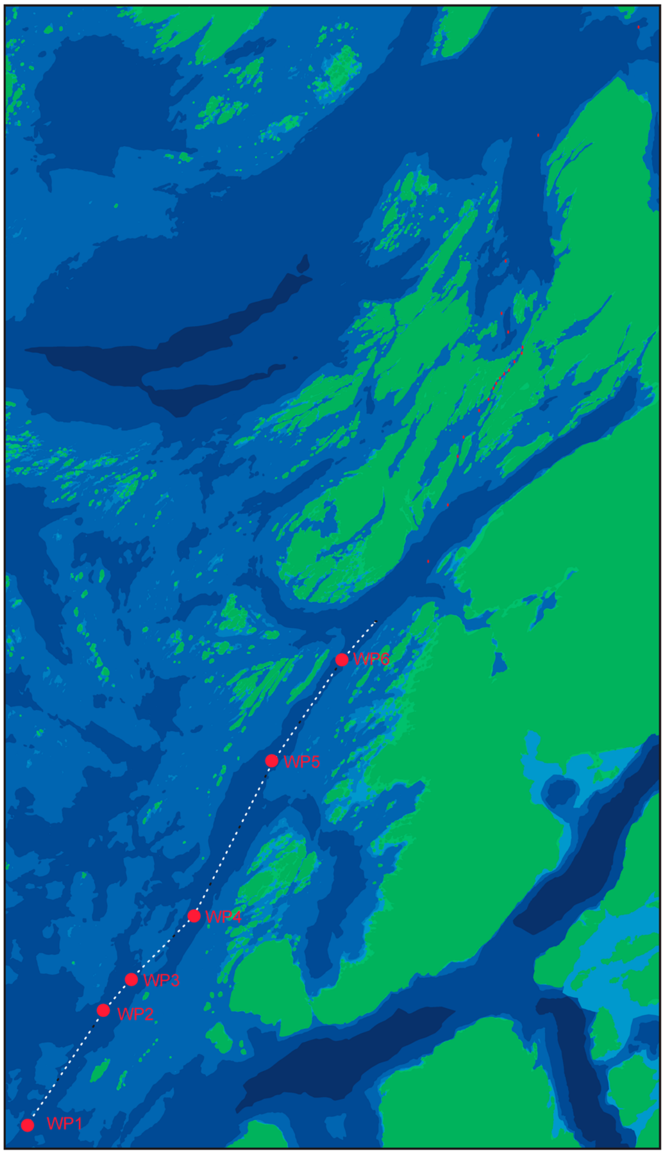

Figure 12. Map of simulation 3.

is stopped after the ship has stopped, without showing the decision made by the human supervisor. A timeline of the fourth simulation is shown in Figure 15.

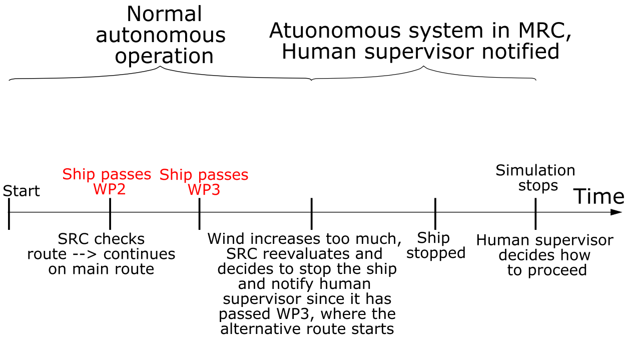
Normal autonomous operation
Atuonomous system in MRC, Human supervisor notified
Start
Ship passes WP2
Ship passes WP3
Simulation stops
Time
SRC checks route --> continues on main route
Wind increases too much, SRC reevaluates and decides to stop the ship and notify human supervisor since it has passed WP3, where the alternative route starts
Ship stopped
Human supervisor decides how to proceed

Figure 13. Timeline of the third simulation showing when the autonomous control system controls the ship and when the human supervisor is notified.

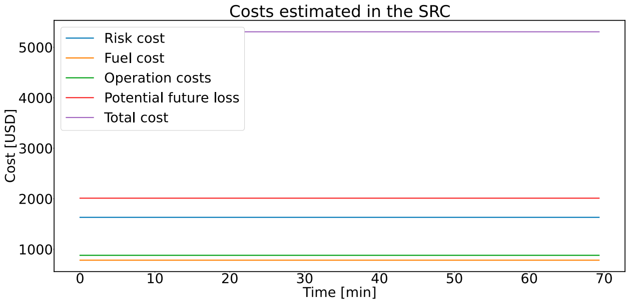

Figure 14. Cost in simulation 3 (d_0 = distance of the full route).

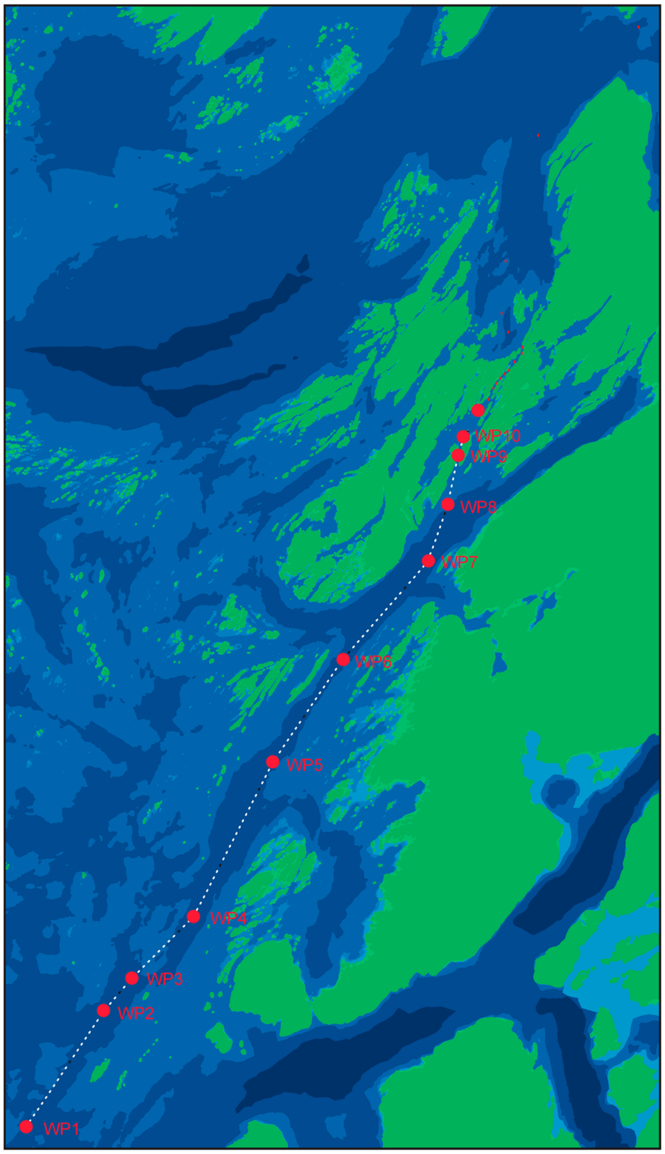

Figure 15. Map of simulation 4.

5.2. Discussion

5.2.1. Risk-based control of autonomous ships

The control system proposed in this paper uses a BBN online risk model to assess the situation as the ship is sailing. The output from the risk model is a risk cost. This describes the expected cost related to potential hazardous events, given the current conditions and ship

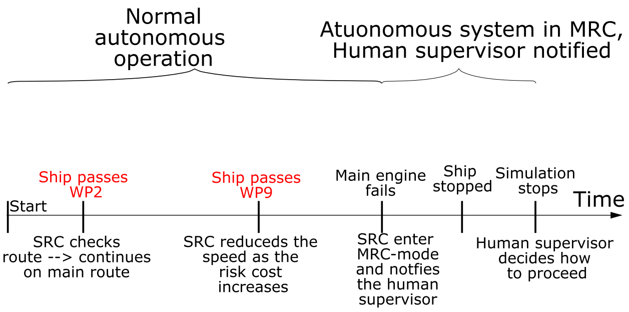
Normal autonomous operation
Atuonomous system in MRC, Human supervisor notified
Start
Ship passes WP2
Ship passes WP9
Main engine fails
Ship stopped
Simulation stops
Time
SRC checks route --> continues on main route
SRC reduceds the speed as the risk cost increases
SRC enter MRC-mode and notfies the human supervisor
Human supervisor decides how to proceed

Figure 16. Timeline of the fourth simulation showing when the autonomous system controls the ship and when the human supervisor is notified.

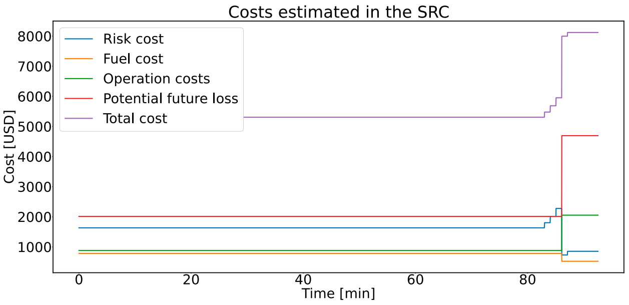

Figure 17. Cost in simulation 4 (d_0 = distance of the full route).

state. The risk model considers a constant time horizon equal to the initial estimated time needed to finish the route. The same time horizon is used to estimate fuel and operation costs. These estimates also assume constant conditions and a constant ship state. This approach provides a cost function that the SRC can use to assess the risk and reward of operating the ship, even if the reward is represented by the cost of operating the ship, i.e. fuel and operation costs. The proposed control system can therefore find a trade-off between reducing risk and minimising operation costs, since there will always be some risk related to autonomous ship operation. As shown in the case study, this enables the SRC to control the ship, similarly to how humans control conventional ships.

The current SRC uses the risk model to obtain a 'picture' of the current risk level and make decisions based on this picture. As shown in the case study, this results in a good performance, and the ship is controlled in an efficient and safe manner. However, computer-based controllers can also use simulations to predict the future state of the ship. This enables the controller to predict how decisions affect the ship before actually making them. This concept is already used in model predictive control (MPC): the controller can simulate the system and compute the optimum control inputs to drive the system towards the intended state. A similar approach could enable an SRC to plan multiple steps ahead, instead of just making decisions based on the current situation, which is done in the current paper.

The proposed control system enters the MRC if the risk cost becomes too high, if the power, propulsion, or navigation and communication systems fail or lose redundancy, or if the conditions worsen and cause the ship to be unable to follow the planned route with sufficient margins. As described in Subsection 4.4, the cost limit is set to the low value of 5,119.47 USD; this value is especially low compared to the costs estimated for the different consequences. However, the current cost limit ensured that the SRC entered the MRC when the ship was unable to continue safely while also limiting the number of times it could have continued sailing. The current limit is therefore considered suitable for the current controller, but it should be assessed further in future work. Assessing the MRC in more detail is also considered outside the scope of this paper. For the purpose of showing how the proposed control system works, it is deemed sufficient that the ship stops and maintains position. However, there might be cases where this is not the best way due to traffic and other conditions. Considering other ways to reduce the risk should therefore be considered in further work.

Deciding whether the power, propulsion, or navigation and communication systems have failed or do not have sufficient redundancy is done based on the fault tree analysis and the modelling of these systems in the online risk model. The nodes representing the power and propulsion systems calculate the probabilities that the systems have failed or do not have sufficient redundancy. The node representing the navigation and communication system only calculates the probability that the system has failed since these sub-systems are not modelled as binary systems. The threshold for when the systems are considered to have failed or to be without sufficient redundancy is set to 0.3 based on testing, similar to the cost limits. The controller works well with the current models and thresholds; it operates with sufficient safety margins. However, the fault tree analysis, models, and thresholds should be assessed in more detail in future work.

5.2.2. Human supervisors in the operation of autonomous ships

The human element is often overlooked or briefly mentioned as part of the technical development of the control of autonomous ships. However, since the operation of most ships under development and testing today still involves humans, this should still be accounted for when new control systems are designed. Situations in which responsibilities shift from the autonomous system to the human supervisor (shared control) are especially important to consider. This paper focuses on UCAs in which the human supervisor fails to react sufficiently that are caused by the poor design of the control system. Other important risk factors, such as the experience level of the human supervisor, human reliability, reaction time, and human-machine interactions, are not considered here to limit the scope of the paper, but they should be studied in future work.

In this paper, the ship can enter the MRC-mode when the SRC recognises that the ship performance may imply risks that are too high. This happens if the risk cost is too high, if any of the systems analysed with the fault tree analysis fails or loses redundancy, or if the ship is unable to follow the planned route. When it is in the MRC-mode, the ship stops and uses the DP-controller to maintain its position while the human supervisor is notified. In this way, the ship is in a safe and stable situation while the human supervisor has time to assess and make a good decision about how to continue. The work in this paper is therefore the first step towards developing a control system that actively supports the human supervisor. The control system should be further improved by assessing which pieces of information should be provided to the human supervisor in different situations. By offering better and more relevant risk-based information through efficient human-machine interfaces (HMIs), the safety of the systems and operations should improve. This is left as an important topic for future research.

Another challenge with existing control systems is that humans are sometimes notified so often that over time, it can become routine to cancel alarms without reacting further. Discussing this in detail is considered outside the scope of this paper, but the SRC is designed to avoid unnecessary notifications by allowing the autonomous control system to make more decisions without human input, such as changing routes, SO-modes, and MSO-modes; the system only notifies the human supervisor when it is unable to control the ship with the proper safety and efficiency margins. Setting these limits is still a potential challenge and a topic that should be addressed in more detail in future work, but the proposed SRC is a step in the right direction.

5.2.3. Testing and simulation setup

The proposed methodology is tested by simulating an autonomous cargo ship controlled by the SRC. The simulator is based on the models from Fossen (2011), but with some important simplifications. These simplifications make it easier to set up and run simulations, but they can also affect the accuracy. Not including wave forces is one such simplification. Wave forces are usually estimated using hydrodynamic programs in which 3D models of the ship are tested. However, the data necessary to make such models are not available for the considered ship. This affects both disturbances from waves and also waves made by the ship, which add damping.

Another simplification is related to the machinery and propulsion system. The machinery models provide the fuel consumption and power output but include no dynamics. The time necessary to change loads or start/stop parts of the machinery is therefore neglected. For the propulsion system, some simple dynamics are included by adding a slight time delay to the thrusters. The reduction in thrust from the tunnel thrusters at high speeds and the lack of force from the rudder at low speeds are, however, not included. As with wave forces, it is difficult to make an accurate model of these effects for the simulator. Therefore, the risk model is adjusted such that using the tunnel thrusters at high speeds and the rudder at low speeds increases the risk cost.

Including wind and current in the simulation also means some simplifications. Both wind and current will depend on the terrain around the ship when sailing close to shore and will change both speed and direction. However, the simulations done as part of this work assumes that wind and current are unaffected of the topography both over and under water. For the purpose of testing the proposed control system, this is deemed good enough. The testing include a limited number of simulations, 176 to be specific. This is done by selecting a combination of wind directions and wind speeds such that the ship is tested with wind from 4 different directions for each 0.5 m/s speed. The current is given a random direction and speed for each simulation. This mean that not all combinations of wind and current are tested. To ensure that all potential combination were tested, both wind and current would have to be varied in a systematic manner resulting in many more simulations. However, since the proposed control system is tested in a reasonable number of different combinations, it is deemed sufficient to show how it works and that it can handle a wide range of conditions.

Accuracy in the control system, especially for the motion controllers, affects the results. The motion controllers, i.e. the DP and autopilot controllers, are included in the simulator. The DP controller is a proportional integral derivative (PID) controller. The autopilot uses a PID controller for the heading and a PI for the speed. These have a base tuning that offers sufficient control of the ship to test the methodology and the SRC. However, since the SRC is a separate controller, both the DP and autopilot can be changed to more advanced and improved controllers later. Testing the SRC with more advanced motion controllers is left as an interesting topic for future research. Failures in the DP system and autopilot controllers, such as losing the position while in MRC mode, are also considered to be outside the main scope of this paper, and therefore they are left for future work.

The GNSS accuracy will affect the ship and its ability to navigate safely. The accuracy of GNSS has improved significantly over the last few years, but it is still assumed to be +/- 5 m. This can be improved using differential GNSS, but it can also be reduced by the environment around the ship. Sailing in narrow fjords with high mountains, where the satellite signal can be blocked and reflected by the mountains, can reduce the position accuracy. This uncertainty in the position data is something that future control systems for autonomous ships should account for. However, for the purpose of testing the methodology and the SRC, the accuracy is assumed to be sufficient for navigation in the case study. Investigating how to best account for this variation in the position accuracy is left for future research.

The proposed control system is tested using a set of verification objectives. These objectives are used to check that it can control the ship in a safe and efficient manner. However, the current verification objectives only consider high-level functionalities. This is deemed sufficient in this paper to verify the control system and show that the methodology works. However, further work should include more detailed verification objectives.

5.2.4. Uncertainty and sensitivity in the online risk model and SRC

The online risk model is used to assess the current situation and state of the ship to improve the decision-making capabilities of the control system. The SRC combines the risk cost, estimated using the online risk model, with fuel and operation costs to find the best way to sail the ship. Using a BBN is a good way to model different RIFs, especially when the exact relationship between all risk factors is not known. However, this also means that the model contains uncertainty. The structure of the BBN, states in the different nodes, and CPTs all contribute to uncertainty in the BBN.

The structure of the BBN is based on the STPA results, which describe the different RIFs. The STPA offers a good foundation for the BBN structure based on the different UCAs, scenarios, and hazardous events. Even though this reduces the model uncertainty, the STPA is a qualitative method for identifying hazards, and it provides less data for assigning states and building CPTs. These are therefore mostly based on the literature and expert judgement. The STPA provides some information that can be used to assign states for the different nodes and to determine what information is necessary in the risk model. In the case study, the nodes with the most uncertainty are the RIFs and UCAs; the main challenge is deciding how much each should affect the risk cost.

The effect of the different RIFs is assessed by conducting a sensitivity analysis on the BBN risk model to see how the risk cost is affected by the different nodes in the best and worst conditions. The results, shown in Figure 18, illustrate that the sensitivity varies significantly. The two RIFs that have the largest effect are the current and wind, followed by the power system, propulsion system, and PMS. Other important nodes impacting the risk cost are the obstacle density, traffic density, and navigation and communication systems. All these nodes can obtain good information from sources such as Norwegian Meteorological Institute (2021), Barentswatch (2022), Norwegian Mapping Authority (2021), and Marine Traffic (2021). The effect of failed machinery and propulsion systems is also thoroughly discussed in the STPA. However, the sensitivity analysis indicates that the system may be tuned more towards handling these nodes or factors, and it may potentially be neglecting other factors, such as visibility. Testing this idea is left as an interesting and important topic for future research.

The balance between the different terms in the cost function is also a source of uncertainty that affects decision-making. To reduce the uncertainty, the fuel cost and operation costs are based on simulation testing and historical data for similar ships, respectively. This helps reduce the uncertainty but may still affect the overall results. Basing the risk cost on the risk model, with the associated uncertainty, together with the potential future loss estimated based on the cost of hiring similar ships, will also add uncertainty to the total cost. Based on the performance over a wide range of conditions, however, the balance is assumed to be sufficient to test the proposed control

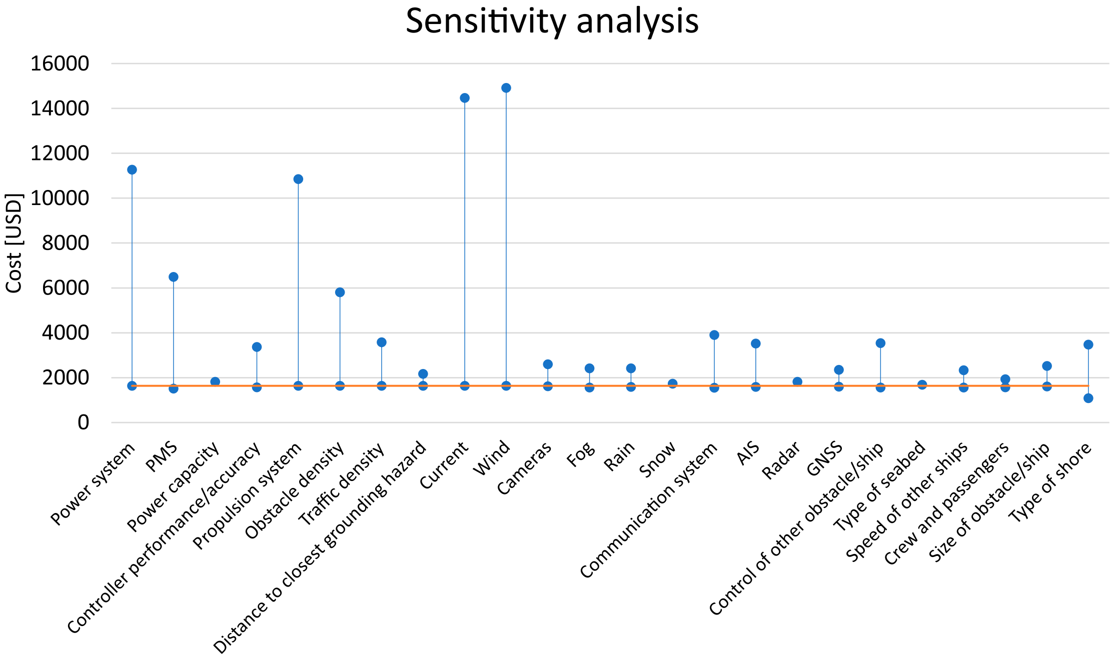

Figure 18. Sensitivity analysis showing the risk cost from the BBN with the nodes in the best and worst conditions.

system and show how the proposed control system functions. Reducing the uncertainty as improved data become available should be the subject of future work. This could be accomplished through model testing or running the proposed control system as a support system on an actual ship to see how its decisions compare to the decisions made by the crew.

6. Conclusion

The objective of this paper is to develop a methodology for building a risk-based control system for autonomous ships, designed with the ability to involve a human supervisor when potential operational challenges arise. The methodology uses an STPA as the basis for building an online risk model and for setting up the SRC, including the human supervisor. The BBN-based risk model is used to assess the current state of the autonomous ship and environment to obtain an estimation of the current risk. This is represented as a risk cost, describing the expected cost from potential consequences, given the current situation. The risk cost is combined with the cost of fuel, other operating costs, and potential future losses caused by the ship taking a longer amount of time to complete the current voyage. The SRC is then able to configure the ship according to the lowest total cost.

Since humans are still expected to be involved in the operation of autonomous ships, the proposed control system is designed with this factor in mind. The result is an autonomous control system capable of operating the ship in a safe and efficient manner, with the ability to assess its performance and determine whether it has the necessary control of the ship to continue safely on its voyage. If not, it will notify the human operator while transitioning to a minimum risk condition (MRC) to reduce the risk level and thereby reduce the probability of a hazardous event. By analysing the human responsibilities with an H-STPA, the SRC can be designed to make it safer and easier for the human supervisor to decide how the ship should continue. While the human supervisor uses time to react and decide how to proceed, the SRC is designed to keep the ship in an MRC to reduce the occurrence of hazardous events and serious accidents. In this way, both the autonomous control system and the human supervisor contribute to operating the ship safely and efficiently.

The proposed methodology and control system is tested in a case study involving an autonomous cargo ship sailing along the Norwegian coast. The human supervisor is in an ROC with a remote connection to the ship. The resulting control system is tested using a set of verification objectives based on the STPA and H-STPA. The shared control between the autonomous control system and the human supervisor enables the ship to pass the test for a wide range of conditions and situations, including calm winds, a strong breeze, machinery failures, and changing conditions that force the SRC to reevaluate decisions.

This study is the first step towards designing risk-based control systems that include the human supervisor in the loop. Future work includes improving the control system and the human-machine interface, as well as putting more of an emphasis on human reliability aspects and contingency situations. The current risk controller is designed to make decisions to gradually reduce the risk cost. However, if the risk cost is above a certain limit, the controller will go straight into the MRC-mode. Future work should determine if the controller can reduce the risk further before entering the MRC-mode, without compromising safety. This could enable the ship to continue sailing in more situations. A path planner capable of planning new routes while the ship is sailing would also improve the control system and make it capable of operating more autonomously.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work by Johansen and Utne is partly sponsored by the Research Council of Norway through the Centre of Excellence funding scheme (project number 223254), AMOS, and project ORCAS (project number 280655).

ORCID

Thomas Johansen http://orcid.org/0000-0002-9122-3861

References

Barentswatch. 2022. Wave forecast. https://www.barentswatch.no/bolgevarsel/?lang = en.

Blindheim S, Johansen TA. 2022. Electronic navigational charts for visualization, simulation, and autonomous ship control. IEEE Access. 10:3716–3737. doi: 10.1109/ACCESS.2021.3139767

Bremnes JE, Norgren P, Sørensen AJ, Thieme CA, Utne IB. 2019. Intelligent risk-based under-ice altitude control for autonomous underwater vehicles. OCEANS 2019 MTS/IEEE Seattle.

Bremnes JE, Thieme CA, Sørensen AJ, Utne IB, Norgren P. 2020. A Bayesian approach to supervisory risk control of AUVs applied to under-ice operations. Mar Technol Soc J. 54(4):16–39. doi: 10.4031/MTSJ.54.4.5

Brito M. 2016. Uncertainty management during hybrid autonomous underwater vehicle missions. Autonomous underwater vehicles 2016 (AUV 2016), p. 278–285.

Brito M, Griffiths G. 2016. A Bayesian approach for predicting risk of autonomous underwater vehicle loss during their missions. Reliab Eng Syst Saf. 146:55–67. doi: 10.1016/j.ress.2015.10.004

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

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

Dittmann K, Hansen P, Papageorgiou D, Jensen S, Lützen M, Blanke M. 2021. Autonomous surface vessel with remote human on the loop: system design for STCW compliance. IFAC-PapersOnLine. 54(16):224–231. doi: 10.1016/j.ifacol.2021.10.097

DNVGL. 2003. DNV report no 2003-0277 Annex II FSA 2003. Technical report, DNVGL Group Technology & Research. http://research.dnv.com/skj/FSALPS/ANNEXII.pdf.

EfficienSea. 2012. Methods to quantify maritime accidents for risk-based decision making. Technical report, EfficienSea. http://efficiensea.org/files/mainoutputs/wp6/d_wp6_4_1.pdf.

Fan C, Wróbel K, Montewka J, Gil M, Wan C, Zhang D. 2020. A framework to identify factors influencing navigational risk for maritime autonomous surface ships. Ocean Eng. 202:107188. doi: 10.1016/j.oceaneng.2020.107188

Fossen TI. 2011. Handbook of marine craft hydrodynamics and motion control. John Wiley and Sons Ltd.

France ME. 2017. Engineering for humans: a new extension to STPA [PhD thesis]. Massachusetts Institute of Technology.

Gil M. 2021. A concept of critical safety area applicable for an obstacle-avoidance process for manned and autonomous ships. Reliab Eng Syst Saf. 214:107806. doi: 10.1016/j.ress.2021.107806

Hassel M, Utne I, Vinnem J. 2021. An allision risk model for passing vessels and offshore oil and gas installations on the Norwegian continental shelf. Proc Inst Mech Eng Part O: J Risk Reliab. 235(1):17–32. doi: 10.1177/0957650920907823

Hogenboom S, Parhizkar T, Vinnem J. 2021. Temporal decision-making factors in risk analyses of dynamic positioning operations. Reliab Eng Syst Saf. 207:107347. doi: 10.1016/j.ress.2020.107347

HSE. 2001. Reducing risks – hse's decision-making process protecting people. Technical report, The Health and Safety Executive (HSE). https://www.hse.gov.uk/managing/theory/r2p2.pdf.

Hu L, Naeem W, Rajabally E, Watson G, Mills T, Bhuiyan Z, Salter I. 2017. COLREGs-compliant path planning for autonomous surface vehicles: a multiobjective optimization approach. IFAC-PapersOnLine. 50(1):13662–13667. doi: 10.1016/j.ifacol.2017.08.2525

Huang Y, Chen L, Negenborn R, van Gelder P. 2020. A ship collision avoidance system for human-machine cooperation during collision avoidance. Ocean Eng. 217:107913. doi: 10.1016/j.oceaneng.2020.107913

IMO. 2018. Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process. Technical report, IMO. https:// m/en/OurWork/Safety/Documents/MSC-MEPC%202-Circ%2012-Rev%202.pdf.

ISO. 2020. ISO/TR 4804. ISO.

Johansen T, Blindheim S, Torben T, Utne IB, Johansen TA, Sørensen AJ. 2023. Development and testing of a risk-based control system for autonomous ships. Reliab Eng Syst Saf. 234:109195. doi: 10.1016/j.ress.2023.109195

Johansen T, Utne I. 2020. Risk analysis of autonomous ships. e-proceedings of the 30th European safety and reliability conference and 15th probabilistic safety assessment and management conference (ESREL2020 PSAM15), p. 131–138.

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

Leveson N. 2011. Engineering a safer world: systems thinking applied to safety. MIT Press. (Engineering systems). ISBN 9780262016629.

Li M, Mou J, Chen L, He Y, Huang Y. 2021. A rule-aware time-varying conflict risk measure for mass considering maritime practice. Reliab Eng Syst Saf. 215:107816. doi: 10.1016/j.ress.2021.107816

Liu C, Chu X, Wu W, Li S, He Z, Zheng M, Zhou H, Li Z. 2022. Human–machine cooperation research for navigation of maritime autonomous surface ships: a review and consideration. Ocean Eng. 246:110555. doi: 10.1016/j.oceaneng.2022.110555

Loh T, Brito M, Bose N, Xu J, Nikolova N, Tenekedjiev K. 2020. A hybrid fuzzy system dynamics approach for risk analysis of AUV operations. J Adv Comput Intell Intell Inform. 24(1):26–39. doi: 10.20965/jaciii.2020.p0026

Loh T, Brito M, Bose N, Xu J, Tenekedjiev K. 2020. Fuzzy system dynamics risk analysis (FuSDRA) of autonomous underwater vehicle operations in the Antarctic. Risk Anal. 40(4):818–841. doi: 10.1111/risa.v40.4

Lyu H, Yin Y. 2019. COLREGS-constrained real-time path planning for autonomous ships using modified artificial potential fields. J Navig. 72(3):588–608. doi: 10.1017/S0373463318000796

Marine Traffic. 2021. Marine traffic. https://www.marinetraffic.com/.

Norwegian Mapping Authority. 2021. Norgeskart. https://norgeskart.no.

Norwegian Meteorological Institute. 2021. Met. https://www.met.no/en/weather-and-climate.

Parhizkar T, Hogenboom S, Vinnem JE, Utne IB. 2020. Data driven approach to risk management and decision support for dynamic positioning systems. Reliab Eng Syst Saf. 201:106964. doi: 10.1016/j.ress.2020.106964

Pedersen TA, Neverlien Å., Glomsrud JA, Ibrahim I, Mo SM, Rindarøy M, Torben T, Rokseth B. 2022. Evolution of safety in marine systems: from system-theoretic process analysis to automated test scenario generation. International conference on maritime autonomous surface ships.

Porathe T. 2021. Human-automation interaction for autonomous ships: decision support for remote operators. TransNav. 15(3):511–515. doi: 10.12716/1001

Ramos MA, Thieme CA, Utne IB, Mosleh A. 2020a. A generic approach to analysing failures in human-system interaction in autonomy. Saf Sci. 129:104808. doi: 10.1016/j.ssci.2020.104808

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

Rausand M, Haugen S. 2020. Risk assessment: theory, methods, and applications. 2nd ed. Wiley

Rødseth Ø, Lien Wennersberg L, Nordahl H. 2021. Towards approval of autonomous ship systems by their operational envelope. J Mar Sci Technol. 27:67–76. doi: 10.1007/s00773-021-00815-z

Rokseth B, Utne IB, Vinnem JE. 2018. Deriving verification objectives and scenarios for maritime systems using the systems-theoretic process analysis. Reliab Eng Syst Saf. 169:18–31. doi: 10.1016/j.ress.2017.07.015

Ship & Bunker. 2022. Rotterdam bunker prices. https://shipandbunker.com/prices/emea/nwe/nl-rtm-rotterdam#LSMGO.

Stopford M. 2009. Maritime economics. 3rd ed. Routledge

The Norwegian Agency for Public and Financial Management. 2018. Guide in socio-economic analysis. https://dfo.no/fagomrader/utredning/samfunnsokonomiske-analyser/verdien-av-et-statistisk-liv-vsl.

Thieme CA, Rokseth B, Utne IB. 2021. Risk-informed control systems for improved operational performance and decision-making. Proc Inst Mech Eng Part O: J Risk Reliab. 237:1748006X211043657.

Torben T, Glomsrud JA, Pedersen TA, Utne IB, Sørensen AJ. 2022. Automatic simulation-based testing of autonomous ships using Gaussian processes and temporal logic. Proc Inst Mech Eng Part O: J Risk Reliab. 237.

Utne IB, Rokseth B, Sørensen AJ, Vinnem JE. 2020. Towards supervisory risk control of autonomous ships. Reliab Eng Syst Saf. 196:106757. doi: 10.1016/j.ress.2019.106757

Utne IB, Sørensen AJ, Schjølberg I. 2017 June. Risk management of autonomous marine systems and operations. International conference on offshore mechanics and arctic engineering, Volume 3B: structures, safety and reliability.

Valdez Banda O, Kannos S, Goerlandt F, van Gelder P, 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. 191:106584. doi: 10.1016/j.ress.2019.106584ISSN 0951–8320.

Wang H, Guo F, Yao H, He S, Xu X. 2019. Collision avoidance planning method of USV based on improved ant colony optimization algorithm. IEEE Access. 7:52964–52975. doi: 10.1109/Access.6287639

Woo J, Kim N. 2020. Collision avoidance for an unmanned surface vehicle using deep reinforcement learning. Ocean Eng. 199:107001. doi: 10.1016/j.oceaneng.2020.107001

Wróbel K, Montewka J, Kujala P. 2017. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliab Eng Syst Saf. 165:155–169. doi: 10.1016/j.ress.2017.03.029

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. doi: 10.1016/j.ress.2018.05.019

Wu B, Yip TL, Yan X, Guedes Soares C. 2022. Review of techniques and challenges of human and organizational factors analysis in maritime transportation. Reliab Eng Syst Saf. 219:108249. doi: 10.1016/j.ress.2021.108249

Yang R, Utne IB. 2022. Towards an online risk model for autonomous marine systems (AMS). Ocean Eng. 251:111100. doi: 10.1016/j.oceaneng.2022.111100