Towards supervisory risk control of autonomous ships

Ingrid Bouwer Utne[1,2], Børge Rokseth[1,2], Asgeir J. Sørensen[1,2], Jan Erik Vinnem[2]
[1]Centre for Autonomous Marine Operations and Systems (NTNU AMOS), NTNU.
[2]Department of Marine Technology, NTNU Norwegian University of Science and Technology, 7491 Trondheim, Norway

Abstract

The objective of this paper is to outline a framework for online risk modelling for autonomous ships. There is a clear trend towards increased autonomy and intelligence in ships because it enables new functionality, as well as safer and more cost-efficient operations. Nevertheless, emerging risks are involved, related to lack of knowledge and operational experience with the autonomous systems, the dependency on complex software-based control systems, as well as a limited ability to verify the safe performance of such systems. The framework presented in the paper is the first step towards supervisory risk control, i.e., developing control systems for autonomous systems with risk management capabilities to improve the decision-making and intelligence of such systems. The framework consists of two main phases, (i) hazard identification and analysis through the systems theoretic process analysis (STPA), and (ii) generating risk models represented by Bayesian Belief Networks (BBN) based on the outcomes of the STPA. The application in the paper is aimed at autonomous ships, but the results of the paper have a general relevance for both manned and unmanned systems with different levels of autonomy, complexity, and major hazard potential.

Keywords

Autonomy; supervisory risk control; STPA; risk modelling; autonomous ships

Nomenclature

AUV – Autonomous Underwater Vehicles
BBN – Bayesian Belief Network
CPT – Conditional Probability Tables
DP – Dynamic Positioning
HMI – Human Machine Interface
LoA – Level of Autonomy
MASS – Maritime Autonomous Surface Ships
PMS – Power Management System
PV – Process Variables
RIF – Risk Influencing Factor
ROV – Remotely Operated Vehicle
SA – Situation awareness
SC – Safety Constraint

STPA – Systems Theoretic Process Analysis
UAV – Unmanned Arieal Vehicle
UCA – Unsafe Control Action

1 Introduction

The development towards maritime autonomous surface ships (MASS) is currently an important technolocial trend due to the potential for increased safety and efficiency, and optimized ship performance (DNVGL, 2018; DMA, 2017; LR, 2015). Autonomous ships are expected to become a cost-efficient alternative to conventional ships and improve safety and environmental impact at sea. It is expected that the introduction of autonomy will reduce the number of human injuries and fatalities (Department of Transport, 2019; Wrobel et al., 2017), which globally amounted to 8000 fatalities from 2008-2012 (IMO, 2016). Nevertheless, it is essential to ensure that autonomous ships have the desired level of reliability, availability, maintainability and safety to be acceptable for widespread use at sea (DNVGL, 2018). Hence, risk assessments are necessary to ensure safe operations (NMA, 2018).

An autonomous system includes improved perception, situation awareness, and planning/re-planning capabilities and may be characterized as deliberative control systems based on the feedback loops of sense, model, plan and act. Failures in critical ship functions, such as in the automatic sailing system or the dynamic positioning (DP) system, are not viable and may lead to loss of position and in the worst case; collision causing severe damage and human fatalities. Therefore, supervisory risk control is a dynamic functionality that needs to be designed and implemented into an autonomous ship's control system, providing the ship with the ability and system integrity to assess and control risks during the operation.

MASS may have functionality with different levels of autonomy (LoA), impacting the ship's operator dependency, communication structure, human-machine interface (HMI), intelligence, planning functionalities, and mission and operation capabilities. The LoA may, for example, be divided into: LoA 1: Automatic operation (remote control), LoA 2: Management by consent (teleoperation), LoA 3: Semi-autonomous or management by exception, and LoA 4: Highly autonomous during a misson or operation (Utne et al., 2017, Ludvigsen and Sørensen, 2016). Other catergorizations may distinguish between the LoA differently, depending on the specific application (Vagia et al., 2016). Motivated by NIST (2008), the four-level version used here is relatively general and aligned with other mobile robotic applications, such as NFAS (2017).

Conventional manned ships either have low LoA or are approaching with some functionality higher LoA. A ship may also have onboard systems with functionality in different LoA, and operators may be able manoeuver across different LoA, i.e., move the system from a high LoA into a manual mode and take over control (low LoA). Advanced ships in DP operation, for example, rely on the operator being onboard to take over control if the ship is in a situation that the control system cannot handle. In addition, LoA may change for the different operational modes, i.e., from departure, transit/sailing and docking.

Unmanned ships, on the other hand, may be implemented with a high degree of remote control and monitoring, and low LoA correspondingly, i.e.; remotely controlled by operators onshore,

or performing all operations autonomously (high LoA), but this requires a change in the current maritime regulation regime.

For systems with low LoA, situation awareness of both the exterior surroundings, as well as the integrity of the system itself are mainly related to relatively simple alarm systems associated with the ship control systems and the human operator's perception and understanding of the system and operation. Similarly, the ability for the system itself to plan and replan the mission may be limited. For systems with high LoA, situation awareness (SA) is to a large extent "transferred" from the operator to the autonomous system, including learning cababilities and decision making. To design and utilize systems with an acceptable risk level that cooperate, possibly replace, and outperform human capabilities, means that supervisory risk control is decisive.

Risk analysis consists of finding out what can go wrong, determine how likely is it, and what are the consequences (Rausand, 2011). Risk modeling is used to express risk qualitatively and/or quantitatively for a system or activity. Risk analysis employs risk modelling and is essential for risk management. Risk control can be defined as a "measure that is modifying risk" (ISO31000, 2009). Risk control of an autonomous ship should consider all relevant risk aspects to proactively avoid the need for activitating any contingency system. Generally, during operation of autonomous systems, risk control should be performed in two different but equally important "risk control modes" to support situation awareness and decision making (Utne et al., 2017):

i. By the human operator and the organization interacting, supervising and monitoring the autonomous system, and/or
ii. By the autonomous system, which means supervisory risk control.

In low LoA, the prevailing system risk control mode is (i), whereas in high LoA, the risk control mode is mode (ii), which we denote supervisory risk control. Hence, a system may switch between risk control performed by the human operator (supervisor) and supervisory risk control executed by the autonomous system, depending on the context, phase of operation, and LoA. For example, Vinnem et al. (2015) and Thieme & Utne (2017a) addressed mode (i). In this paper, the focus is on developing the basis for mode (ii), i.e., supervisory risk control by the autonomous system.

In general, the control system is divided into three main layers (Ludvigsen and Sørensen, 2016); (i) the control execution layer (the reactive control layer), (ii) the guidance and optimization layer, and (iii) the operation or supervisory layer (the deliberate control layer). In the mission layer, the mission objective is defined and planned (and possibly replanned). In the guidance and optimization level, the waypoints and reference commands to the controller are handled. In the control execution level, the plant control and actuator control occur. Risk must be considered in all three levels. The supervisory risk control "module", however, may be considered as a contribution to improved artificial intelligence, included in the operation/mission layer (iii) in the control architecture, supporting and enabling the autonomous system to model and plan its actions; i.e., making deliberate choices.

Most work related to safety of autonomous ships have so far focused on hazard identification and analysis, but not on risk modelling, even though Bayesian Belief Networks (BBN) have been developed for risk related to autonomous underwater vehicles (Hegde et al, 2018; Thieme & Utne, 2017; Brito & Griffith, 2016). Rødseth and Tjora (2014) discuss challenges with unmanned ships. Utne et al. (2017) clarify, categorize, and classify risk related to autonomous marine systems and autonomous ships, and establish a foundation for risk management of such systems. Wrobel et al. (2017) determine that the occurrence of navigational accidents may be reduced for autonomous ships, but the consequences from fire and structural failure may increase. Acanfora et al (2018) propose a method for route planning and execution by an autonomous ship, focusing on ship motion. Rokseth et al. (2017; 2018; 2019) demonstrate that the system theoretic process analysis (STPA) is feasible for risk analysis of systems with complex control functionality, such as DP systems. Montewka et al. (2018) propose research directions for safety and risk assessment and concludes that new risk analysis methods are needed. Thieme et al. (2018) review 64 existing ship collision and grounding risk models but find none directly suitable for risk assessment of MASS. Xiang-yu et al. (2018) present a novel ship domain model for autonomous ships, focused on collision risk. Wrobel et al. (2018a) use STPA to identify potential means for improving the safety of a remotely controlled merchant vessel. Wrobel et al. (2018b) apply STPA for analysing hazardous scenarios and determining design requirements to autonomous ships, and Rokseth et al. (2019) use STPA to derive a safety verification program for autonomous ships. They do not, however, apply STPA as a basis for developing online risk models as part of supervisory risk control, as we propose in this paper.

The objective of this paper is to outline a framework for developing online risk models as part of the deliberative layer of a control system for MASS. The framework is the first step towards supervisory risk control. The paper uses STPA for identifying hazardous events and corresponding scenarios, which provide direct input to the development of online risk models represented by BBN. The main focus of the paper is on the process of transforming the results from STPA into nodes and structure of a BBN. Constructing a BBN is usually performed using either subjective knowledge, the knowledge representation approach, or a machine learning approach (Darwiche, 2009). For risk analysis, typically the subjective approach is used. Hence, a systematic and structured approach bridging results from hazard identification into risk modelling is missing, and the framework proposed in this paper is an attempt to do so.

The main scientific contribution of the paper is related to how the outcome of STPA directly enhances the development of the BBN in two ways; (i) in the identification of nodes, and (ii) in the structuring of arcs connecting the nodes. A case study illustrates the proposed framework for an autonomous ship. The results of the paper create a basis for implementing built-in intelligent risk assessment during operation of complex software-based systems, such as MASS.

Fault tolerant control (Blanke et al., 2015) mainly aims at reducing the consequences of internal faults and includes methods for diagnosing on the control excecution level. Supervisory risk control, on the other hand, includes more than fault-tolerant control, related to the the capability of the autonomous systems to learn, adapt and improve.

The paper is structured as follows: Section 2 presents the methodological approach, Section 3

focuses on the case study, Section 4 includes the discussion, and Section 5 states the conclusions.

2 Methodology – the framework

2.1 Background and needs

A traditional risk model is typically represented by a bow tie, as shown in Figure 1. The left side represents the causes to the critical event, and the right side represents the consequences. A critical event may be caused by several different causes and lead to different consequences, which can be analyzed by fault trees, event trees, BBN, or a combination of these. The entire bow tie model represents an accident scenario.

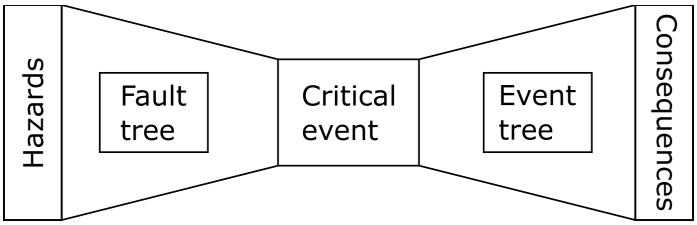

Figure 1. Bow tie model. Causes to the critical/hazardous event are represented to the left, and potential consequences to the right.

Risk may be defined as the "effect of uncertainty on objectives" (ISO 31000 (2018)). An effect can be either positive or negative. Further, the standard states that "risk is often expressed in terms of the consequences of an event and the associated likelihood of occurrence". During operation of a ship, the focus is to prevent and reduce the likelihood of critical events and their causes and maximize the operational efficiency and output. Regarding "effect" in the above definition, safety would then refer to avoiding negative effects, whereas operational efficiency and cost optimization might include both positive and negative effects. Daily operation has usually a large focus on production efficiency and maintenance activities to prevent downtime that may follow from failures of critical technical equipment. If a critical event should occur, emergency response is activated and implemented to prevent and reduce the likelihood of serious consequences.

The framework in this paper focuses on avoiding hazardous events and their causes that may lead to negative effects. The main purpose is to enable the autonomous system to make decisions that mitigate or reduce the likelihood of critical events, i.e., the left side of the Figure 1. The term risk in a risk management context usually refers to 'major accident risk' because hazards like a fall [by an operator] on present or to a lower level deck or a minute excursion from an ideal route normally may be controlled by far less rigorous assessments and control actions. Thus, risk in this paper refers to major accident risks, including scenarios which are caused by sources on the ship itself (such as rudder failure or engine room fire), or from external causes, like a vessel on a collision course. This implies that control of major accident risks may be common with controls to optimize navigational efficiency.

In the industry, most current risk analysis methods are used during the design phase of systems, and not as tools for online risk control during operation, even though dynamic approaches to risk analysis have been developed in recent years. A dynamic risk assessment can be defined as a "method that updates estimated risk of a deteriorating process according to the performance of the control system, safety barriers, inspection and maintenance activities, the human factor, and procedures" (Khan et al., 2016). Generally, the increased availability of sensor data and improved computational capability provide enhanced opportunities for dynamic risk assessments (Zio, 2018). Examples of dynamic risk assessments using BBN that updates the risk when new information becomes available have been developed, for example, by (Li et al., 2019; Adedigba et al., 2018; Barua et al., 2016; Paltrinieri et al., 2014; Khakzad et al., 2013). Zheng and Zio (2018) present a dynamic risk assessment method, combining a hierarchical bayesian model with simulations and event trees, that allows for estimation of risk based on data collection during operation. This is more in line with the concept of "online risk management" which builds on data from different sources, such as historical data, sensors and measurements, and experience data (Vinnem et al, 2015). Neither of these approaches, even though they may be useful, are developed for supervisory risk control in general, nor for autonomous ships.

Autonomous systems depend to a large extent on software, which is highly complex for advanced systems. Physical separation and segregation of components, such as redundancy in ship machinery systems, may be overruled by software and control systems that operate across physical boundaries and separated systems. Several of the current risk analysis methods focus on decomposition of the system into components, which is challenging with complex systems, such as the DP system (Rokseth et al., 2017). According to Rasmussen (1997), risk management should be considered as a control function implemented to maintain system processes within the safe operation envelope. Leveson (2011) has proposed STPA, based on these ideas, in which safety is controlled by enforcing constraints on the system behavior, and accidents occur due to inadequate control or inadequate enforcement of safety constraints. STPA has been used in several applications (Leveson & Thomas, 2018), including for hazard identification of autonomous ships (Wrobel et al., 2018a; b), but not for generating risk models and BBN.

To establish supervisory risk control as part of the intelligence of a control system, the following aspects need to be adressed:

i. We need to know which hazardous events should be prevented and their causal factors in the system´s operation.
ii. We must be able to observe and verify the presence of the causal factors during operation (cf. left side of bow tie in Figure 1).
iii. We must know which combinations of causal factors that may lead to the hazardous or critical event. Hence, we need to structure the causal factors and create a foundation for gathering and assessing information and observations related to the causal factors in real-time. Such information may be of a qualitative, semi-quantitative, and quantitative nature, and must be collected during operation or from databases with historical and/or experience data.
iv. We must determine the effect of different combinations of causal factors on system level risk. If there is a high risk of safety constraints being violated, the system itself (or an operator when in low LoA) needs early warnings of a potential hazardous event.

These aspects, and in particular i. and iii, are focused on in the presented framework in this paper.

2.2 The phases and steps of the framework

To address the above-mentioned needs, we propose a two-phase process, combining:
(i) Hazard analysis through STPA.
(ii) Development of the risk model represented by a BBN.

In the first phase, STPA (based on Rokseth et al. (2017; 2018)) is used to identify and analyze the hazardous events, define unsafe control actions, scenarios and hazardous combinations of causal factors. The scenarios and causal factors form the basis for the structure and content of the risk model and BBN, developed in phase 2.

Figure 2 gives an overview over the phases and steps of the proposed framework, and how they are related to each other. Each phase and step are explained closer in the following Subsections.

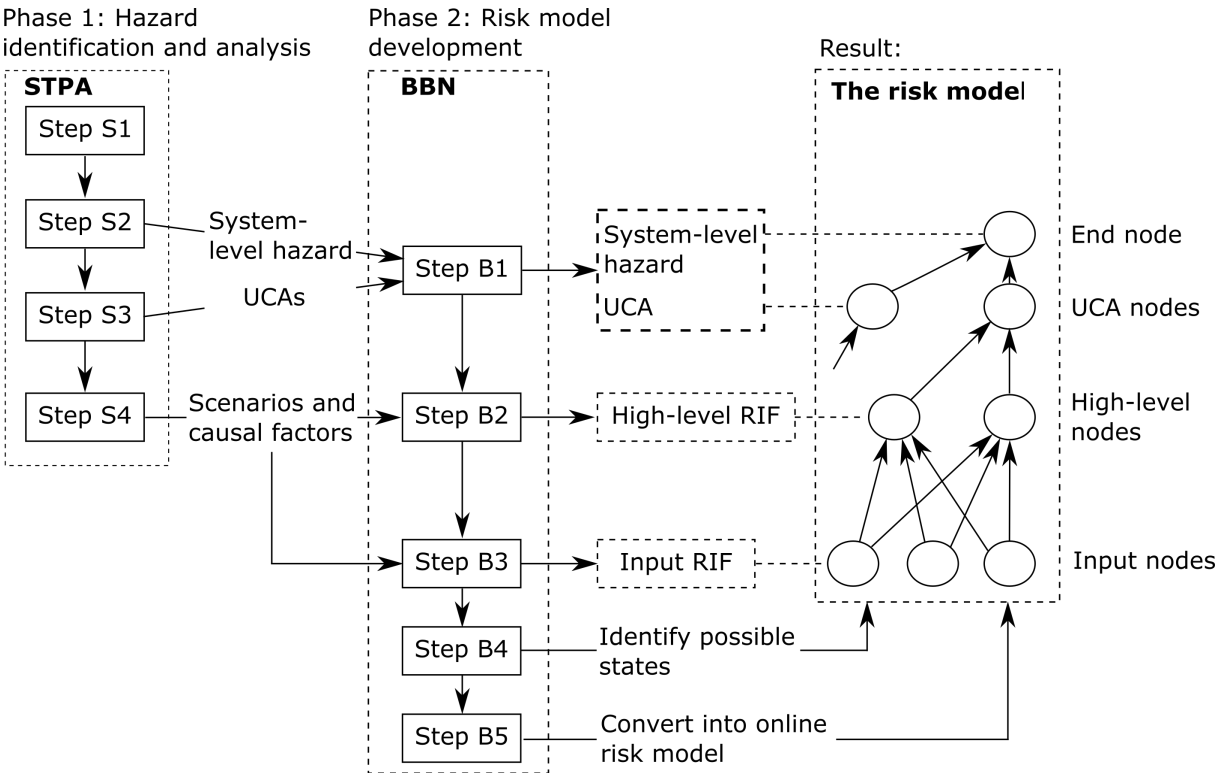

Figure 2: Overview of phases and steps in the proposed framework. UCA is unsafe control actions and RIF is risk influencing factor.

Phase 1 – Hazard identification and analysis - STPA

In this phase, the system accidents (major accident risks) and the system-level hazardous events are identified and analyzed. These are used to define corresponding system-level safety constraints that must be enforced to ensure that the system accidents do not occur. Next, we

identify how inadequate control may result in a transition into these hazardous states through violation of the system-level safety constraints (UCA-unsafe control actions) and scenarios in which the UCA may occur. Finally, causal factors (i.e., factors that are prerequisites or facilitators for the scenarios, or that makes the scenarios more likely to occur), are identified. Summarized, the process consists of the following steps:

- **Step S1: Define the system** with system boundaries and describe it as an autonomous control system. This includes identifying controller responsibilities and process variables.
- **Step S2: Identify hazardous events** at system level and safety constraints.
- **Step S3: Identify unsafe control actions** (UCA) that violate the safety constraints.
- **Step S4: Develop scenarios** in which the unsafe control actions may occur and identify their causes.

The first step is obviously important for achieving a feasible result of the analysis. In step S2, the system level hazardous events are identified, with safety constraints and process variables. Process variables are a set of variables that represent a controller's perception or belief regarding the relevant system states and are useful for identifying and specifying the hazardous context for UCA, scenarios and causal factors (for example, in which context is it hazardous not to provide a certain control action). In step S3, UCA that violate the safety constraints are determined, and these may occur in four different ways: a necessary control action is not provided/followed/executed, it is provided too early or too late, it is applied too long or too short, or an unsafe control action is provided (Leveson, 2011). In step S4, scenarios in which the UCA can occur and relevant causes to these are specified.

The outcome of the STPA is hazard identification, which provides the qualitative basis for developing the online risk model with BBN. The online risk model should enable monitoring, calculation of the likelihood of the hazards during operation, early warnings, and hence support regarding different action alternatives the control system of an autonomous ship can undertake during a voyage or operation.

Phase 2 – Developing the online risk model - BBN

BBNs are graphical models illustrating causal relationships consisting of nodes and arcs. The main objective of using BBN is to model risk influencing factors (RIFs) that influence a hazardous event or on an accident (Rausand, 2011). By monitoring the states of the RIFs, early warnings can be provided about possible deviations from the normal operating envelope of a system. Further, the likelihood of different hazardous scenarios can be calculated, providing decision support to the MASS. A RIF can be defined as "an aspect (event/condition) of a system or an activity that affects the risk level of this system or activity" (Øien, 2001).

The nodes in a BBN represent RIFs with states or conditions, and the arcs show the influence from a parent node on a child node. Nodes that do not have parents are input or root nodes, and the end node determines the outcome of the BBN. Conditional probability tables (CPT) determine the states of the child nodes based on the states of the parent nodes. The BBN can be updated using the Bayesian reasoning laws (Rausand, 2011). Dynamic BBNs most often consist

of a static model, but the information about the states of the nodes are updated at various time intervals.

The advantage of BBN is the ability to present causal relationships, and to combine empirical data with expert knowledge, which often is necessary in risk analysis. The disadvantage is that the combination of n states grows exponentially for each additional node (Rausand, 2011). Hence, a BBN may soon encounter the trade-off situation with respect to computation resources available and model accuracy and complexity.

The risk model in phase 2 is developed as follows:

- **Step B1: Define end-node and UCA-nodes** based on the system-level hazard and the UCAs from the STPA.
- **Step B2: Identify high-level RIFs** related to the scenarios and the causes and connect these to the appropriate UCA nodes in the BBN.
- **Step B3: Identify the input RIFs** from the causes to the scenarios and connect these to the high-level RIFs.
- **Step B4: Identify states** for all nodes and build the **CPTs**.
- **Step B5: Convert the BBN** into an online risk model.

Step B1 consists of collecting the results from the preceding STPA and determining the end node of the BBN based on the system level hazard. Further, the UCAs from the STPA are defined as UCA nodes to be connected to the end node. In step B2, the high-level RIFs are identified from the scenarios and then connected to the UCA-nodes. In step B3, the causal factors associated with each scenario is examined to determine the parent nodes or input RIFs, which belong to each high-level RIF nodes. Figure 3 illustrates the emerging BBN structure based on the results from the STPA.

The result is a top down development of the BBN with relationships between the nodes represented by the arcs. The relationship between the system hazardous event represented by the end node, the UCAs, high-level RIFs and input RIFs are illustrated in Figure 3. Please also note that there may be direct causal relationships among high-level RIFs, and among Input RIFs (so that there may, for example, be a direct causal relationship from high level RIF 1 to high-level RIF 2).

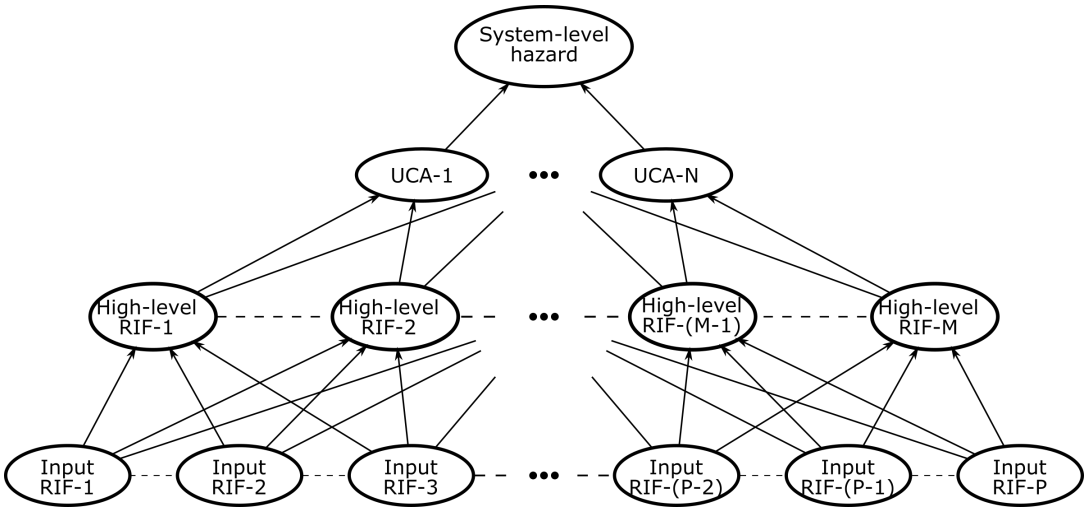

Figure 3. The overall structure of the BBN, which shows the link between STPA and BBN.

Both qualitative and quantitative information may have to be used to determine the probabilities for the CPT in step 4. We may use the safety constraints of STPA, requirements in regulations (if these exist), and sensor measurements, along with expert judgements, as a basis for defining states, and determining the probabilities for each state. Step B4 can be quite demanding, depending on the number of nodes and arcs in the BBN.

Step B5 consists of determining which nodes of the BBN can be measured in real-time by data from sensor systems, and which nodes are dependent on other types of data (historical, expert judgements, etc.). The online risk model can be used to calculate the likelihood of a given scenario based on real-time measurements, estimations and historical data. In addition, the time frame for the regular updates of the BBN must be determined. Some input data may be fed continuously, where as others may be available much more infrequently. In practice, the real-time data need to be stored in a database and fed back into the online risk model. It may also be relevant to consider if the CPTs should change with different operational conditions, and whether weights of the arcs should be implemented and altered dynamically to enable different influences of parents on child nodes.

The proposed process is tested in a case study, presented in the next Section.

3 Case study: Autonomous ship voyage

3.1. Introduction

In this case study, the proposed framework is applied to enhance the safe navigation and manoeuvering of an unmanned autonomous ship with a human supervisor in a shore control center. It is assumed that the autonomous ship navigates from one location to another, and that a preplanned route is provided for the ship. This route can be a set of waypoints with corresponding arrival times, which the ship should follow. During the voyage, the ship autonomously adapts its route based on information provided through the online risk model, with input based on prevailing weather conditions and other types of real-time data. Route

adaptation may, for example, include deviation from the planned route to avoid collision or grounding if there is an obstacle too close to the planned route, or altering the route or speed to avoid sailing into bad weather.

3.2 Phase 1 – STPA

Step S1: Define the system

In this step, we define the system and model it as a hierarchical control structure, as shown in Figure 4. This includes defining control responsibilities and process model variables. We start by defining three main control levels; a remote operator (a human supervisor not situated aboard the vessel, but at a shore control center), a guidance system, and the control execution.

The main responsibility of the remote human operator is to plan and supervise the voyage. In addition, the remote operator can alter (redefine) and, in an emergency, initiate a fallback strategy, such as manoeuvre around the closest waypoint or maintain the current position by means of a DP system. The guidance system is responsible for adapting the planned trajectoriy according to real-time data. This includes the responsibility of updating the planned trajectory to avoid obstacles, such as ships or structures that were not considered when planning the trajectory. This responsibility also requires the guidance system to predict the behavior of obstacles correctly and sufficiently early to find a trajectory that safely avoids them.

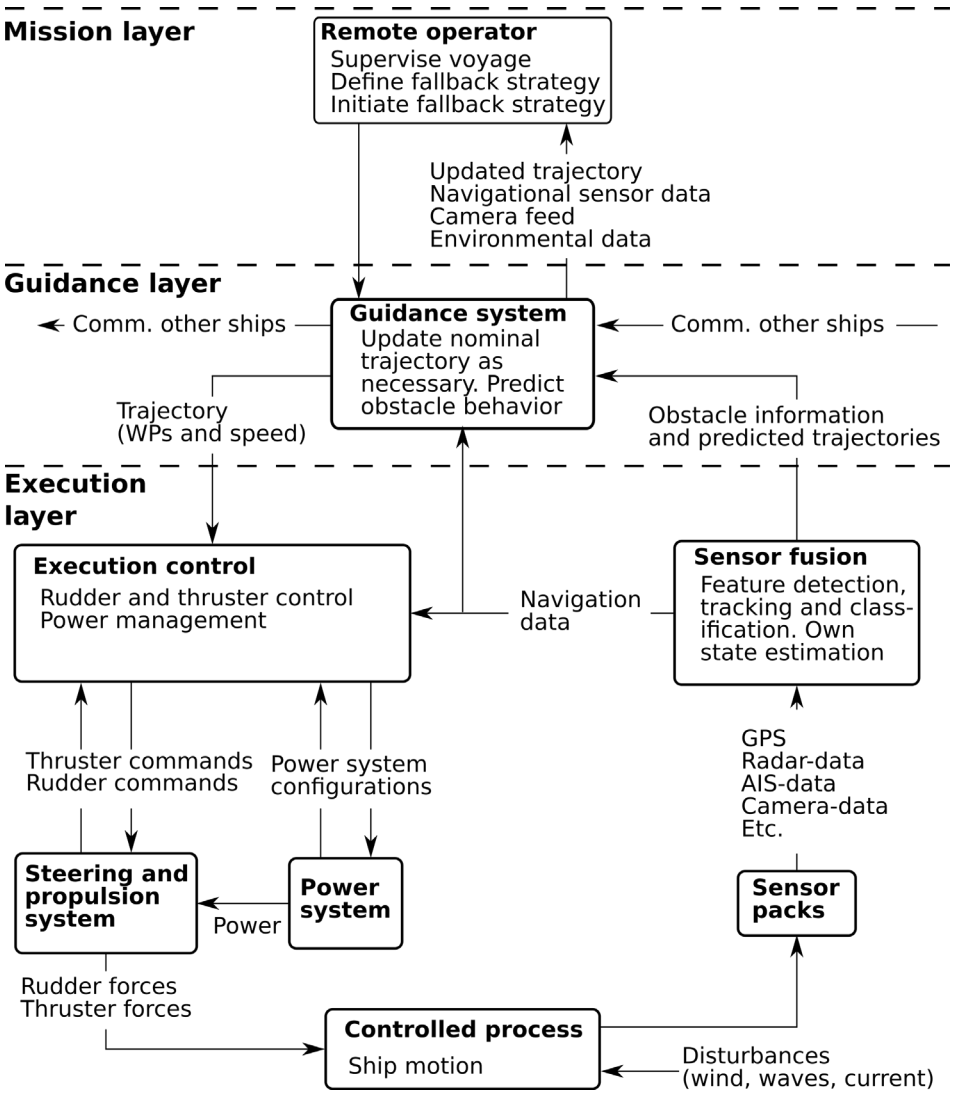

Figure 4: Hierarchical control structure for the unmanned autonomous ship. GPS refers to the Global Positioning System and AIS is the Automatic indentification System for ships. WPs is short for waypoints.

The execution control is the interface between the high-level control on one hand, and the actuator system and power system, i.e., low-level control, on the other hand. Execution control includes systems, such as an automatic sailing system, DP and the power management system (PMS). The main responsibilities are to control the actuator system and to implement configurations and reconfigurations in the power system, the actuator system, and various sensor systems, as requrested by the guidance system.

In addition to these controllers, the ship is equipped with a sensor package consiting of motion sensors, compasses, position reference systems such as global position system (GPS), wind sensors, automatioc identification system (AIS), radars, lasers, acoustic bottom tracking systems, computer vision systems, and a sensor fusion system (which all together may be referred to as the sensor system), as well as a steering and propulsion system, and a power system. The sensor system is responsible for:

1. Correctly measuring the own ship's navigational states. The sensor readings are also input to a state observer (Kalman filter or other type of observers) used to filter and estimate the states necessary to describe the motion, position and orientation, including speed of the ship.
2. Detecting obstacles.
3. Tracking obstacles (i.e., keeping track of each obstacle and estimating its navigational states such as position, heading and speed).
4. Classifying obstacles (i.e., categorizing obstacles into either ship and type of ship, or offshore structure, or a terrain feature including shallow water creating risk for grounding).

The steering and propulsion systems are responsible for actuating the vessel motion according to commands from the execution control, and the power system is responsible for providing the propulsion and steering units (as well as any other consumer of power aboard) with sufficient power.

In the rest of this case study, we focus mainly on the guidance system, and proceed by refining its control responsibilities as:

1. Updating the ship's nominal trajectory based on real-time data.
2. Configuring the steering, propulsion and power system to suitable modes of operation, ensuring sufficient reliability, low emissions, adequate manoeuverability, sustainable use of energy storage devices throughout the voyage, etc.
3. Configuring the sensor system to a suitable mode of operation.
4. Configuring the power system to a suitable mode of operation.

Due to limited space, the case study will furthermore focus on the first control responsibility above. Before proceeding to the next step, we define process model variables (PV) for the guidance system that are relevant with respect to the stated control responsibility. These are identified based on an assessement on what the guidance system needs to know to satisfy the control responsibility. The following variables can be identified:

- PV-1: Navigational intentions of potential obstacles. An assessment of this is necessary to determine whether any given trajectory is safe in an encounter with an obstacle.
- PV-2: Signalling from potential obstacles. In an encounter with an obstacle that is providing signals, it is necessary for the guidance system to know what message the obstacle is attempting to convey.
- PV-3: Classification of potential types of obstacles. To estimate future states (and intentions) of an obstacle, it is necessary for the guidance system to know which type of obstacle it is. Drift ice may, for example, be assumed to behave differently from a sailboat.
- PV-4: Obstacle´s current navigational states (i.e., position, heading, speed and turn rate of the own ship). These are necessary for the guidance system to know to assess whether a trajectory will pass too close to the obstacle.
- PV-5: Presence of obstacles in own ship´s nominal trajectory (or will be shortly). This is necessary for the guidance system to know to assess whether a trajectory is safe.

- PV-6: Own ship´s navigational states. This is necessary for the guidance system to know to assess an obstacle's relative position and whether execution control successfully follows the given trajectory.
- PV-7: Own ship´s nominal trajectory. This is necessary for the guidance system to know to assess whether it will interfere with an obstacle. Here, also COLREG (the IMO's rules for ship collision avoidance) may be considered (IMO, 1972).
- PV-8: Minimum safe distance to obstacles (ship safety domain). Necessary to know to assess what is an appropriate minimum distance to an obstacle.

The PVs are used for establishing the context for UCA, scenarios and causal factors, and have been identified in a brainstorming process by asking what the guidance system need to know to successfully update the ship's nominal trajectory based on realtime data.

Step S2: Identify hazardous events at system level

The system level hazards and high-level safety constraints need to be identified. The safety constraint is used to deteremine the unsafe control actions in the next step. The system accident and the relevant hazardous event in this case-study are:

- System accident: The autonomous ship collides with an obstacle.
- System level hazard: The autonomous ship does not maintain safe distance to obstacles.
- Safety constraint (SC): The autonomous ship must maintain a domain about itself (a ship safety domain), free of obstacles.

Step S3: Identify unsafe control actions

In this section, we identify how the guidance system's responsibility to update the planned trajectory based on real-time data can be executed in an inadequate manner so that the system level hazard can occur. This is achieved by considering the control responsibility together with the generic four types of UCAs, and different combinations of process model variables (PV1-8).

Table 1 presents the results, combining the generic types of UCA with the guidance system's control responsibility of updating the nominal trajectory based on real-time data. Applying the generic types of *too long* or *too short* are not applicable here because updating a trajectory is a discrete event.

Table 1. Control responsibilities and corresponding unsafe control actions.

Control responsibility	Provided	Not provided	Provided too early or too late	Provided too long or too short
Update the nominal trajectory based on real-time data	**UCA-1:** A trajectory update is provided that, if followed, will result in obstacle inside the ship safety domain	**UCA-3:** Trajectory update is not provided to avoid that an obstacle violates ship safety domain	**UCA-4:** A trajectory update intended to avoid that an obstacle violates the ship safety domain is provided too late	N/A
	UCA-2: A safe trajectory update is provided, but the ship does not follow the trajectory, and instead violates minimum separation distance of an obstacle			N/A

Step S4: Develop scenarios

In this step, we identify scenarios in which the UCAs can occur and the causes to the scenarios. Table 2 shows a selection of scenarios and causal factors identified for UCA-1: *A trajectory update is provided that, if followed, will result in obstacle inside the ship safety domain*.

The scenarios and causal factors presented in Table 2 represent the main results of the STPA. The scenarios are the conditions in which the UCA under consideration can occur, and the causal factors are prerequisites for the occurrence of the scenarios. Scenarios and causal factors have been found in this case study by inspecting the hierarchical control structure (Figure 4) developed in Step S1, following the process described in Leveson & Thomas (2018). This means that the controlled process in Figure 4 and the corresponding PVs are investigated to find ways the process can change into a hazardous state.

The first column in Table 2 specifies which PVs have been used to formulate each scenario. Anything that may impact the system state from normal to "not normal" can possibly be part of a scenario. Examples are failures or degradation of systems and components over time (e.g., technical condition of the ships), external disturbances from the operating environment (e.g., weather conditions and traffic density), direct inputs and outputs to the controlled process (e.g., the ship's manoeuverability), problems that are related to the controllers (see the execution layer in Figure 4, e.g., obstacle detection, sensor fusion problems, sensor reliability), operational modes of the ship (e.g., different LoAs), and transmission of information between the autonomous ship, other ships/obstacles, and the human operator/supervisor (e.g.,, interpretation of intentions and communication). Like other types of hazard identification methods and risk analyses, domain experts provide important input to the generation of the scenarios and the identification of the relevant causes. Leveson & Thomas (2018) do not recommend performing additional analysis, such as the failure mode and effect analysis (FMEA), to identify technical failures, but Rokseth et al. (2017) found this combination useful.

Table 2. Scenarios (Sc) and causal factors for UCA – 1.

UCA-1: A trajectory update is provided that, if followed, will result in obstacle inside the ship safety domain		
Relevant PVs	**Scenario**	**Causal factor**
PV 1, PV 2, PV 7	Sc – 1: A trajectory update is based on the anticipated navigational behaviors (intentions) of nearby obstacles. Sudden unanticipated behavior of one of the obstacles results in violation of ship safety domain.	Type of obstacle (e.g., fishing vessel, cargo vessel, pleasure yacht, small speed boats, terrain feature, bank/shallow, offshore structure)
		Traffic density (The more traffic, the more likely that another vessel must do some manoeuvering to avoid another obstacle hidden from own ship's sensor system)
		Type of traffic situation. (A disorganized traffic picture combined with high traffic density will result in more uncertainty than a unidirectional organized traffic flow with the same density)
		Whether obstacle signals its intentions
		Unclear, inprecise or incorrect signaling by obstacle. This may cause misunderstandings and result in incorrect anticipation of the obstacle's future behavior
		Technical condition of other vessels (a dead ship will not respond as required by COLREGs in most situations, a slow oil-tanker is not likely to change course quickly)
		Manoeuverability of other vessels (a slow oil-tanker is not likely to change course quickly – thus, high certainty can be obtained for the next minutes, only by estimating current states)
PV 6	Sc-2: Own ship's navigational state measurements/ estimates are imprecise or incorrect. Updating a trajectory based on incorrect or imprecise navigational states may result in violation of ship safety domain	Reliability of own ship's navigational state reference function (a function in the sensor fusion system)
		Weather conditions (In harsh weather, there will be more erratic movement of the own ship and the process of filtering and estimating the states will be more challenging and less precise)
PV 4	Sc-3: Estimates on navigational states of an obstacle are incorrect, resulting in incorrect or imprecise estimates on future behavior	Weather conditions and visual conditions
		Reliability of obstacle tracking function (a function in the sensor fusion system)
PV 3, PV 7, PV 8	Sc-4: An obstacle is incorrectly classified (e.g., iceberg is classified as a cargo ship), resulting in incorrect estimates on future behavior (e.g., if the iceberg is expected to act to avoid collison). A trajectory update based on incorrect beliefs regarding current and future navigational behavior of obstacles may result in violation of the ship safety domain.	Visual conditions
		Reliability of obstacle classification function (a function in the sensor fusion system)
		The type of obstacle (some types of obstacles may be harder to classify correctly than others)
PV 5	Sc-5: In a navigation scenario involving several obstacles, one obstacle is not detected, and the trajectory updated to handle the	Reliability of obstacle detection function (a function in the sensor fusion system)
		Traffic density (high density results in more chances to not detect an obstacle)

	scenario brings the undetected obstacles inside the ship safety domain	Speed of own ship (high speed reduces the time available to detect obstacles)
		Speed of obstacle (high speed reduces the time available to detect obstacles)
PV 5, PV 7, PV 8	Sc-6: A situation where all navigation options that the guidance system can formulate will result in violation of the ship safety domain, is encountered	Traffic density
		Narrowness of safe navigational area
		Own ship manoeuverability
		Weather conditions (bad weather may narrow down the feasible navigational options)
PV 7, PV 8	Sc-7: To avoid another ship, a sharp trajectory update is provided. The manoeuverability of the own ship is insufficient to follow the updated trajectory and consequently the ship safety domain is violated	Speed of own ship
		Weather conditions
		Min. turn radius in trajectory
		Own ship manoeuverability

3.3 Phase 2 – BBN

The system hazard, the UCA, the PV and the causal factors for the scenarios constitute the qualitative information and basis for developing a risk model represented by a BBN. The subsequent sections explain its evolvement in detail.

Step B1: Define end-node and UCA nodes

The system hazard and UCA from the STPA develops directly into the top level of the BBN:

System level hazard node: The autonomous ship does not maintain safe distance to obstacle.
UCA-1: A trajectory update is provided that, if followed, will result in obstacle inside the ship safety domain.

Step B2: Identify high-level RIFs

To identify the high-level RIFs related to UCA-1, each scenario with associated causes in Table 2 should be investigated. Table 3 shows the corresponding high-level RIFs for each scenario to be included in the BBN. The occurrence of one scenario can potentially cause the UCA. Hence, initially all high–level RIFs should be connected to the UCA node through arcs as a starting point for the BBN construction. The analyst, however, should then reconsider the initial arcs by assessing whether some scenarios are related or may impact each other, since this may influence the structure of the BBN.

In general, the high-level RIFs can be directly or indirectly related in three ways: first, when the high-level RIFs are connected to the UCA-node by arcs directed to the UCA-node, they are related only indirectly through the UCA-node (for example, that the occurrence of a combination of states in two high-level RIFs may cause a particularly high probability of occurrence of the UCA). In this case, their indirect relationship is defined in the CPT of the UCA-node. Second, a set of high-level RIFs may be related through direct causation. This is relevant when the occurrence of one scenario can affect the probability of occurrence of the UCA by affecting the probability of occurrence of another scenario. In the BBN this is realized

by connecting the high-level RIF for the first scenario to the high-level RIF for the second scenario by a directed arc. For example, from Table 3, in Scenario Sc-3, the future navigational states of an obstacle are incorrectly estimated. This can, according to the analysis, only cause UCA-1 through higher probability of occurrence of Scenario Sc-1 (i.e., through unanticipated behavior of the obstacle). Similarly, Scenario Sc-4, in which an obstacle is incorrectly classified, can only cause UCA-1 through Scenario Sc-1. Thus, rather than connecting the high-level RIFs associated to Sc-3 and Sc-4, to the UCA-node, they are connected to the high-level RIF for Sc-1, directly affecting the probability of occurrence of Sc-1. Finally, high-level RIFs can also be related indirectly through common input RIFs. For example, the input RIF traffic density affects the probability of occurrence of high-level RIFs 1, 5 and 6, as can be seen in Table 4.

Table 3. The relevant high-level RIFs for each scenario.

Scenario	**Relvant high level - RIFs**	**Remarks**
Sc-1: A trajectory update is based on the anticipated navigational behaviors (intentions) of nearby obstacles. Sudden unanticipated behavior of one of the obstacles results in violation of ship safety domain.	RIF-1: Prediction of obstacle intention	If the obstacle intention is correctly anticipated, the future behavior of the own ship trajectory can be updated such as to correctly account for future obstacle behavior.
Sc-2: Own ship navigational state measurements/ estimates are imprecise of incorrect. Updating a trajectory based on incorrect or imprecise navigational states may result in violation of ship safety domain	RIF-2: Measurement/ estimation of own ship's navigational states	Sc-2 will not take place if navigational states are correctely measured/ estimated.
Sc-3: Estimates on navigational states of an obstacle are incorrect, resulting in incorrect or imprecise estimates on future behavior	RIF-3: Estimation of obstacle's navigational states	This RIF affects the above scenario through RIF-1.
Sc-4: An obstacle is incorrectly classified (e.g., iceberg is classified as a cargo ship), resulting in incorrect estimates on future behavior (e.g. if the iceberg is expected to take action to avoid collison). A trajectory update based on incorrect beliefs regarding current and future navigational behavior of obstacles may result in violation of the ship safety domain.	RIF-4: Obstacle classification	This RIF affects the above scenario through RIF-1
Sc-5: In a navigation scenario involving several obstacles, one obstacle is not detected, and the trajectory updated to handle the scenario brings the undetected obstacles inside ship safety domain (PV5)	RIF-5: Obstacle detection time	
Sc-6: A situation where all navigation options that the guidance system is able to formulate will result in violation of the ship safety domain, is encountered	RIF-6: Feasible navigational options	
Sc-7: To avoid another ship, a sharp trajectory update is provided. The manoeuverability of the own ship is insufficient to follow the updated trajectory and as a consequence the ship safety domain is violated	RIF-7: Trajectory following performance	

Step B3: Identify the input RIFs

Finalizing the structure of the BBN is now straight forward, and the top down process of STPA directly appears in the structure of the BBN. Since the high-level RIFs in Table 3 represent each scenario, Table 2 can be used to find out which causal factors (input nodes) belong to each scenario (high-level RIF). Table 4 compiles the information from Table 2 and Table 3 and presents all high-level RIFs with corresponding input RIFs. Some input RIFs (causal factors) are included for more than one high-level RIF (scenario), which means that the input node needs to be connected to all the relevant high-level nodes.

Table 4. Scenarios with corresponding high-level RIFs and input RIFs, representing the emerging structure of the BBN.

Scenario	**High-level RIFs (nodes)**	**Input RIFs (nodes)**
Sc-1	RIF-1	Obstacle type
		Traffic density
		Type of traffic situation
		Obstacle's signalling
		Obstacle's technical condition
		Obstacles manoeuverability
Sc-2	RIF - 2	Reliability of own ship's navigational state
		Weather conditions
Sc-3	RIF - 3	Weather conditions
		Visual conditions
		Reliability of obstacle tracking
Sc-4	RIF - 4	Visual conditions
		Reliability of obstacle classification
		Obstacle type
Sc-5	RIF - 5	Reliability of obstacle detection
		Traffic density
		Speed of own ship
		Speed of obstacle
Sc-6	RIF - 6	Traffic density
		Narrowness of safe navigational area
		Own ship manoeuverability
		Weather conditions
Sc-7	RIF - 7	Speed of own ship
		Weather conditions
		Min. turn radius in trajectory
		Own ship manoeuverability

The resulting BBN is shown in Figure 5. For example, high-level RIF-1 represents scenario Sc-1 with the causal factors "obstacle type, traffic density, type of traffic situation, obstacle's signalling, obstacle's technical condition", and "obstacle's manoeuverability". These causal factors are represented as input nodes and linked with arcs to high-level RIF-1 in the BBN. Correspondingly, high-level RIF-2 is related to reliability of "own ship's navigational state reference function", and so on.

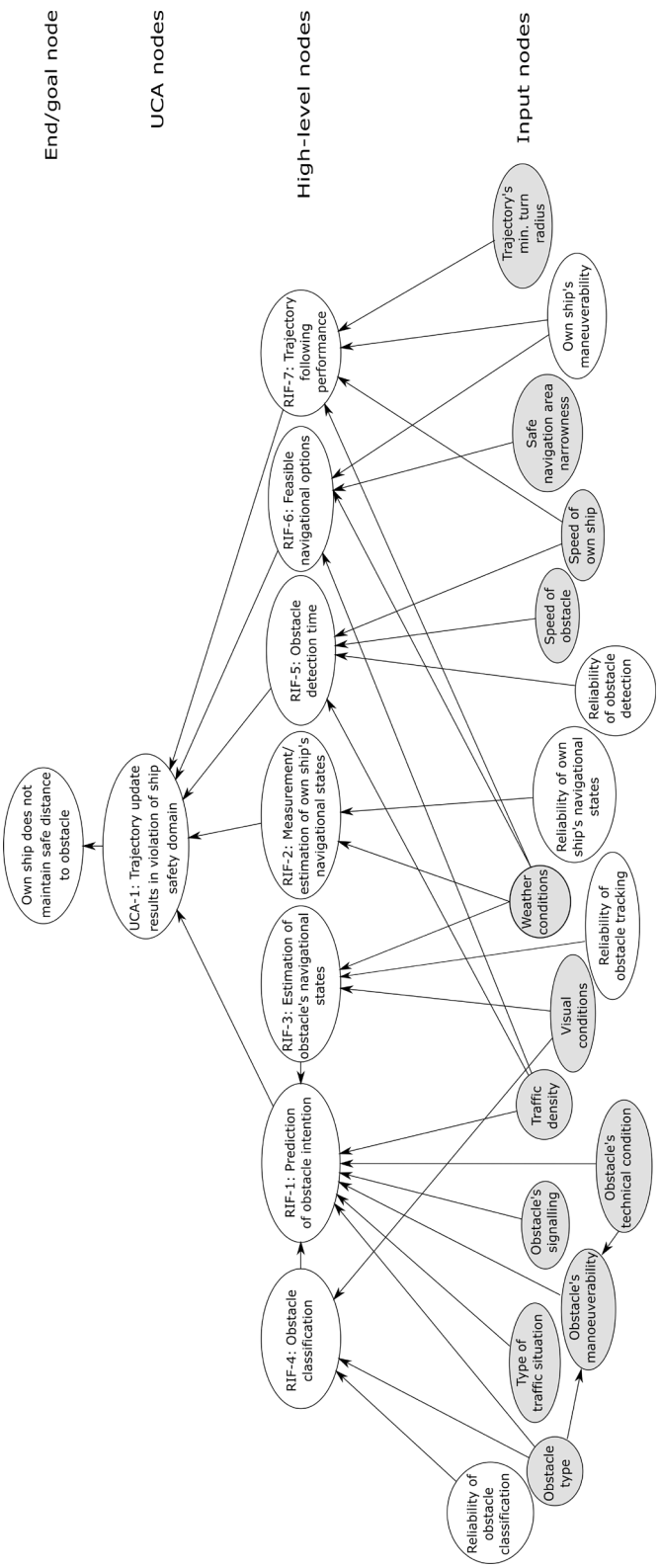

Figure 5. The risk model and BBN for UCA-1.

Step B4: Identify states and build the CPTs

To quantify the BBN, states and CPTs need to be defined. STPA provides only qualitative information, so the quantification of the BBN should follow the typical approach in BBN development. The results from the STPA, however, may provide initial input to the definition of states for the nodes. The states can be defined by closely considering each node, its corresponding scenario and the relevant PVs for that scenario, since these impact the condition of the system being studied. In the case study, the system accident is "The autonomous ship collides with an obstacle". Hence, collision risk studies (Chen et al, 2019; Bye & Aalberg, 2018; Sotiralis, 2016; Hassel et al, 2014; Kristiansen, 2013) have also been used as basis for further refinement of the states, with careful consideration of the difference between conventional ships and MASS. Details are provided in Table 5.

The states in Table 5 need to be further validated, when data are collected for the CPTs, to assess that the states produce meaningful results. Probabilities for the CPTs can be derived following the traditional methods of quantification of BBN, see, e.g., Fenton & Neil (2013). This is outside the scope of this paper.

Table 5. Suggested states for the nodes.

Type	**Nodes**	**States**	**Remarks and references**
System level hazard	Own ship does not maintain safe distance to obstacle.	Yes/no	A ship domain defines a safe space around the ship (Chen et al, 2019), and hence whether the ship performance is acceptable or not. Still, even if the MASS does not maintain a safe distance to other obstacles, a collision may not occur, since the focus here is on the causal side, and only one system level hazard and UCA-1 are included in the case study.
UCA-1	Trajectory update results in obstacle inside the ship safety domain.	Yes/no	This corresponds to "ship on collision course", which is a factor included in several collision risk models, see e.g., (Chen et al, 2019; Hassel et al, 2014). It is assumed that either the obstacle is on collision course or not.
High – level RIF 1	Prediction of obstacle intention	Incorrect/imprecise /correct	This node is somewhat related to what is often referred to as collision candidate detection, which is the focus of several methods (Chen et al, 2019), here influenced by e.g., traffic density, situation, signalling (communication) etc.
High – level RIF 2	Measurement/ estimation of own ship's navigational states	Incorrect/imprecise /correct	This would typically relate to the performance of the navigator/ bridge management team (BMT) in collision risk analysis. For MASS this function will be the responsibility of its control system (and technical system condition), as well as the performance and responsibilities of the shore control center. In addition, weather conditions will influence this node. Hence, an additional state is proposed here, compared to Sotiralis et al (2016) for higher fidelity.
High – level RIF 3	Estimation of obstacle's navigational states	Incorrect/imprecise /correct	Same as above remark.

High – level RIF 4	Obstacle classification	Correct/wrong/no diagnosis	This is a function that would typically be performed by a ship's navigator/BMT onboard conventional ships, by use of visual observation, radar and. For the MASS itself, the technical condition of the sensor and data fusion systems will influence the outcome, along with any detection done by the shore control center. The route of the MASS will influence the potential type of obstacles that can be encountered. Kayaks, for example, are not likely to be in the middle of the ocean. The state definition here is similar to Sotiralis et al (2016).
High – level RIF 5	Obstacle detection time	Too late/late/early	Time available for detection and collision avoidance is crucial and is a factor included in other collision risk models (Hassel et al, 2014).
High – level RIF 6	Feasible navigational options	Yes/no	This is related to whether there are available navigational options for the MASS that will not result in vilolation of the safety domain which can be classified into yes or no.
High – level RIF 7	Trajectory following performance	Low/medium/high	For a conventional ship this would be related to the navigator/BMT's action. Human error may be divided into slips, lapses, mistake and violation (Reason, 1990), which cannot be directly transferred to MASS, even though human operators/supervisors in the shore control center can make errors. In principle, the performance could be discretized into an arbitrary number of states. However, due to the subjective nature of the variable, it is not seen as necessary to use more than three states.
Input RIF	Obstacle type	Cargo/offshore/ fishing/ passenger/ yacht/stationary	This RIF affects the "obstacle manoeuverability". Its states represent the type of obstacle, corresponding here to the categories in (Bye & Aalberg, 2018), including also stationary obstacles.
Input RIF	Traffic density	High/medium/low	This RIF is related to geometric collision probability in conventional collision risk studies. Traffic density could be defined based on the number of vessels in the area. See, e.g., (Chen et al, 2019; Kristiansen 2013).
Input RIF	Type of traffic situation	Overtaking, crossing, head on	The traffic situation could be specified based on the relevance of potential accident situations in the area, such as crossing, overtake or head on collision. See, e.g., Kristiansen (2013).
Input RIF	Obstacle´s signalling	Adequate/ inadequate	This would correspond to communication with ship or oil and gas platforms in conventional risk analysis (see e.g., Hassel, 2014; Kristiansen, 2013).
Input RIF	Obstacle´s technical condition	Low/medium/high	This RIF affects the "obstacle manoeuverability". The technical condition of the obstacle, given that it is a ship, and impact on safety has been explored in Baniela & Rios (2011). The definition of technical condition could be related to requirements to system availability, e.g., high $A \geq 0.97$.
Input RIF	Obstacle´s manoeuver-ability	Low/medium/high	This is a child node of obstacle type and obstacle's technical condition. The states would need to be quantified based on expected performance by the MASS.
Input RIF	Reliability of own ship's navigational states	Low/medium/high	This node is related to the functioning and performance of the sensor fusion system, which consists of software and sensor systems. The reliability of this system can be represented by a continuos variable R_d which can be discretized into an arbitrary number of states. Here, the variable is chosen to have three states from low to high.

Input RIF	Visual conditions	Low/medium/high	This node is related to precipitation and fog. Visibility can be defined in different ways, but "low" could correspond to dense fog/precipitation, "medium" to mist/fog, and "high" to clear conditions (Kristiansen, 2013). Sotiralis et al (2016) divide visibility into distance, i.e., >1 nm and <1 nm. Bye & Aalberg (2018) combine condition and distance and suggest five categories.
Input RIF	Weather conditions	Good, storm, rain, windy, fog	Weather conditions could be constituted by temperature, waves, current, wind etc. Instead of including all these nodes with different scales (e.g., Beaufort scale), one node for the weather is deemed enough. Possible states are good, storm, rain, windy, fog (Sotiralis et al, 2016).
Input RIF	Reliablity of obstacle tracking	Low/medium/high	This node is related to the functioning and performance of the sensor fusion system, which consists of software and sensor systems, meaning that it either performs adequately or inadequately. The reliability of this system can be represented by a continuos variable R_d which can be discretized into an arbitrary number of states. Here, the variable is chosen to have three states from low to high.
Input RIF	Reliablity of obstacle classification	Low/medium/high	Same remark as above.
Input RIF	Reliability of obstacle detection	Low/medium/high	Same remark as above.
Input RIF	Speed of obstacle	High/medium/low	This influences the obstacle detection time available. Whether speed can be characteristed as high/medium/low depends on type of obstacle. Generally, "low" could be >5 knots, due to this being a general speed limit in many ports.
Input RIF	Safe navigation area narrowness	Narrow/normal/ wide	This input RIF is related to the geometric collision probability in conventional collision risk studies. It could be calculated based on the waterway diameter and characteristics, and the beam of the MASS. See, e.g., Kristiansen (2013).
Input RIF	Own ship´s manoeuver-ability	Adequate/ inadequate	This node is related to the technical condition of the propulsion and steering systems and type of MASS. Either the performance is adequate for the type of ship or some systems have a degraded state meaning that manoeuverability is limited. In general, ship manoueverability influences the collision avoidance options available (Chen et al, 2019).
Input RIF	Speed of own ship	High/medium/low	This node influences the obstacle detection time available. Whether speed can be characterised as high/medium/low depends on type of MASS. Generally, "low" could be >5 knots, due to this being a general speed limit in many ports.
Input RIF	Min. turn radius in trajectory	Adequate/ inadequate	This node influences the ability of the MASS to follow the trajectory provided. Either the turn radius is adequate or inadequate. The definition of "adequate is related to the type of MASS and its performance characteristics.

Step B5: Convert the BBN into an online risk model

The last step is to convert the BBN into an online risk model. This means to establish real-time values, determine the need for empirical data and expert judgements to be included and used as a basis for the calculations in the model. Databases need to be created for collecting the relevant data and as the storage for updating the online risk model with new information. Bayes theorem is used to update the calculations. The dark gray input nodes in Figure 5 shows which nodes can be measured in operation. Further, it is necessary to determine how often the model needs to be updated. Traffic density for example, need a relative high update frequency.

There are different options for using the online risk model for supervisory risk control in an autonomous system. These are further discussed in Section 4.3.

4 Results and discussion

4.1 STPA

An important prerequisite for supervisory risk control is to know which hazardous events should be prevented and their causal factors. The latter is of particular importance for enabling early warnings of potential violations of safety constraints. STPA provides a comprehensive process to identifying hazards and revealing causal factors, which is beneficial for novel and complex systems, such as autonomous ships, for which there is limited experience available and lack of empirical data. The results of STPA are system level hazards, safety constraints, how unsafe control may cause violations of the system-level safety constraints in scenarios, and which causal factors may influence the scenarios.

The safety constraints identified in the STPA may influence the LoA for a given functionality in the ship´s operation. If the guidance system in the case study is merely relaying commands from the remote operator, the autonomy level may be set to 1 (Automatic operation), while if the remote operator only provides a destination for a leg of the voyage, and leaves route planning to the guidance system, the autonomy level is 4, or 3 if remote operator (supervisor) approval is required.

Please note that at this point, further refinement of the scenarios and causal factors could have been performed. An advantage is that we could more closely assess the underlying assumptions and factors impacting the states of the input RIF. For example, the causal factor of scenario Sc-2: *Reliability of own ship's navigational state (a function in the sensor fusion system)* could have been analyzed, by developing a separate hierarchical control structure for the sensor system and investigating which factors affect the reliability of the own ship's navigational state. How far such refinement should be performed, depends on the objective of the STPA analysis, the system definition and boundaries, as well as available data regarding the scenario and its causes. In risk analysis, typically only the most critical hazards and hazardous events are investigated further in detail (e.g., through fault tree, event tree and/or BBN). A current challenge with STPA is that no prioritization of the control actions can be performed. Wrobel et al. (2018c) and Gil et al. (2019) investigate this topic and provide some interesting results,

but without a conclusive recommendation. Hence, a refinement criterion cannot be provided yet, except to leave it up to the analyst to decide. Further exploration of this topic is not within the scope of the study in this paper.

Consider, for example, two extreme cases: in the simplest case (i), the input RIF *Reliability of estimation of own ship's navigational state* is given a static value, perhaps based on a verification process or the "quality" of the equipment used to satisfy the function of measuring or estimating the ship's navigational states. A more refined model (ii) would perhaps need to monitor signals from each relevant sensor, to assess, for example, noise to signal ratio and the frequency of signal wildpoints. In addition to an apriori assessment of the equipment quality, the refined model could provide a dynamic assessment of the state of the input RIF, including the effect of factors that are relevant only to a specific situation (for example, loss of satellite signals). The challenge is, however, that further refinement of scenarios and causal factors, always is a trade-off situation between the necessary level of detail, model outcome, and computational efforts.

One interesting result of the STPA for the case study is that the need for online risk models for autonomous ships is demonstrated. The case study shows, for example, that an error in the guidance system may lead to hazardous maneouvering of the ship. An online risk model, informing the guidance system of the MASS that the uncertainty related to its own position is high, could be useful to avoid such a hazardous maneouver.

4.2 A process for building and structuring the BBN

The BBN developed in the framework follows a top down approach, which is a natural outcome of using STPA since it is a top down process. Currently, there exists no well-defined way of structuring a BBN in risk analysis, even though there are different approaches, such as using accident investigation reports and accident models to identify nodes and defer relationships (Mazaheri et al. 2016). Instead of using such hindsight information, which is lacking for autonomous ships (and limited for autonomous systems in general), the proposed framework focuses on identifying and analyzing how hazardous events potentially may occur to enable early warnings to the control system of the autonomous ship. By using the results and the structure appearing in the results from the STPA analysis, the BBN can be systematically developed. The initial BBN can then be transformed into a dynamic BBN to be used as an online risk model in the mission layer of a control system of an autonomous ship.

One major advantage of a BBN is that, in addition to including the different causal factors that may lead to a scenario (and a system level hazard), their combinations and impact on each other are possible to monitor in operation. When the online risk model is developed, it is possible to identify which information is already gathered and/or used in real-time during operation of the system, and which new data needs to be collected. If new data must be gathered, system modifications should be considered, including installation of new sensor systems.

In general, sensitivity analysis and validation of the BBN model are important. Sensitivity analysis is performed to evaluate how the model responds to changes in the different parameters, i.e., to different combinations of states. Model validation means investigating if the

model complies with the initial purpose of its development (Mazaheri et al, 2016). In Section 2.1, four main requirements to the work were stated, of which the STPA and BBN combined are able to address. Since the main purpose of the paper is to demonstrate the link between STPA and BBN, sensitivity analysis and model validation have not been performed for the case study here.

To assess the overall risk for an autonomous system and/or operation, such as a MASS, means that there will be more than one UCA related to a system level hazard, and BBNs for each UCA should be developed. Then, the BBNs would have to be merged into one BBN to represent the system level hazard (see Figure 3). Several nodes may be relevant for more than one UCA or high-level RIF, which means that duplicated nodes must be considered carefully in the development of the CPTs and in the calculation process. This is not addressed further in this current paper but should be subject to futher work.

4.3 Online risk modelling for supervisory risk control

Integration with control optimization algorithms

Supervisory risk control for decision making and control under uncertainty could be based on optimization to determine policies that ensure that the risk level is acceptable. There are at least two ways in which the proposed framework for online risk modelling can be used to support such control.

First, online risk models may take inputs from the actual sailing process to provide a real-time estimate of the current risk level. In this case, the controller needs a model that tells it how the controlled process, including the risk picture, can change state, and how a certain set of control inputs may affect the risk picture. Then this information must be used to design a control policy that minimizes cost or maximizes a utility function (cost/benefit). This enables a tradeoff between the objective of minimizing the output of the online risk model (the risk) and the objective of reducing additional costs, such as increased fuel consumption or delays. This trade-off is necessary since risk cannot be avoided at any cost, as any ship operation is associated with some risk. By designing a supervisory risk controller with the control objective of minimizing a cost function or maximizing a corresponding utility function, it is possible to identify control policies that minimizes the risk as far as possible without incurring unreasonable costs. Defining such functions, however, is a major challenge, but the risk model ensures a systematic foundation for developing and utlizing these.

A second option, instead of estimating the risk based on the actual sailing process, the online risk model may be used to predict the future risk, given a set of expected future control inputs and realization of stochastic disturbances. This can be achieved by simulating the sailing process using a mathematical model of the ship (digital twin) and the environment with the current states of the sailing process as initial states. To assess the effect of the selected set of control inputs on the risk picture, the dynamic risk model may run together with the simulation model, taking input from the simulated sailing process, and as such; predict the future risk in the simulated sailing process.

When simulations are utilized by the controller to identify a set of optimal future control inputs in the sense that they minimize a cost function, the implementation is usually referred to as model predictive control (MPC). MPC is currently a widely used and powerful control technology that has been highly successful in the optimization and automatic control of advanced industrial systems, see for example Qin & Badgwell (2003 and Johansen (2017). If the proposed online risk model is to be used together with simulations to include risk in the cost function, risk acceptance criteria, safety constraints and model uncertainty are decisive for selecting the parameters describing the cost function and the constraints. The selection of the control system parameters and objectives is therefore a very challenging computational task, and strongly influence the system functionality and performance. One example of this approach has been developed for collision-avoidance for autonomous ships, see (Johansen et al., 2016), but risk was not systematically analysed and included in terms of a risk model.

In structured and static environments and for relatively simple operations, autonomous systems may work efficiently. In most cases, however, the operating environment is unstructured and complex. MASS, for example, need to be able to cooperate with human operators and other autonomous and conventional ships. Testing and verification of autonomous systems with increased intelligence and the ability to learn will be even more challenging, because it is harder to predict and simulate everything that may occur and how they will behave in every situation. An existing "premature" example of online risk modeling for supervisory control is DP consequence analysis. This tool is used for advanced ships in DP operation (DNVGL, 2016; Sørensen, 2011). The online consequence analysis calulates the vessel's cabability to maintain position following any single failure in the thruster and power systems and is implemented in commercial DP systems as an alarm and advisory system. The current online DP consequence analysis assumes static conditions and does not include assessment of risk and the integration of risk models. The concept of online DP consequence analysis can be expanded to also consider navigation, acoustic, radar and visual sensors evaluating the consequences of the most severe failure impacting the ship's sitation awareness and navigation capabilities. An extension of such an online consequence system with risk models for improved online decision making (e.g., selecting the right LoA for the prevailing conditions) and verification of safe performance is a new and novel concept for autonomous ships. The proposed framework and following case study in this paper creates a foundation for such a concept.

Detailed development of the supervisory risk control system is a vast topic and is therefore outside the scope of this paper. It will, however, need to be based on the initial step of a systematic approach to identifying hazards (STPA) and the development of a risk model (BBN), as proposed in this paper. Industry development application and implementation for autonomous ships is the main subject of the industry research project ORCAS (IMT, 2018).

Risk acceptance

One of the most challenging issues with risk assessment is to determine what risks are acceptable. The standard NS 5814 (2008) defines risk acceptance criteria as criteria used as basis for decisions about acceptable risk. With MASS, the focus is to ensure high system availability, avoid mission abortion or disruption of the voyage, and prevent fatalities, environmental and material damage. Nevertheless, risk-based decisions need some criteria or

will be a trade-off between the potential loss of the MASS and the gain related to completing the mission successfully. For a control system to be able to utilize the results of an online risk model in decision making means that the "traditional" risk acceptance criteria need to be translated into meaningful safety constraints for the operation, which is highly challenging.

Johansen (2010) states that, based on (UK Health and Safety Executive, 1992; Fischhoff et al., 1983), the willingness to accept risk depends on the potential benefits, the extent it can be controlled (personally or institutionally), and the potential consequences that may follow. NORSOK Z-013N (2010) includes generic guidelines for choosing risk criteria, which are relevant to different industries, even though they are focused on preventing fatalities and major hazard risk. Utility (cost/benefit evaluation), equity (universal and unconditional right to a certain level of protection), and technology (use of state-of-the-art control measures) are the three "pure" criteria for judging risk acceptability (UK Health and Safety Executive, 2001). NORSOK Z-013 (2010) and Johansen (2010) provide four requirements to risk acceptance criteria that may be useful for MASS. The criteria need to:

- Support decisions and express the effect of risk reduction measures.
- Enable communication and understanding between users, operators and non-experts.
- Be clear and related to precise system and/or operational limits.
- Be independent of any concept solution through the way risk is expressed

IMO (2018) proposes acceptance criteria for fatality risk of 10^{-4} per year for crew and 10^{-5} for passengers and third parties for new ships. Criteria for cost-effectiveness of safety are also suggested. Such criteria do not directly provide meaningful information about the risk level in operation. The oil and gas industry use the term "loss of main safety functions" (Vinnem, 2014). An advantage is that such types of metrics involve less uncertainties since failures of such systems occur earlier in the event sequence. For MASS operation, examples of main safety functions are the hull integrity, collision avoidance system, etc. In practice, this means that risk acceptance is related to the reliability of the main technical functions of the MASS, as well as human reliability. In addition, the operational context, the design envelope of the systems, and environmental impact need to be considered.

Averages over time periods are often used to calculate risk metrics, and then assuming similar trends in the future (Vinnem, 2014). This might be applicable for risk trending but might not be possible to use for risk acceptance and risk measurement during an ongoing operation, in particular for MASS, as there is limited experience and data available for such systems. Safety integrity level (SIL) describes the amount of risk reduction that is provided by an electrical/-electronic/programmable electronic system (Marszal, 2001). IEC 61508 [2010, part 4, p. 19] defines safety integrity as: "The probability of a safety related system satisfactorily performing the required functions under all stated conditions within a specified period of time". Safety integrity is split into four discrete levels, from SIL 4 to SIL 1. The levels are distinguished by maximum tolerable failure frequency and the range of risk reduction required. Each SIL is quantitatively expressed by probability of failure on demand (PFD) and a risk reduction factor, derived from 1/PFD. To claim achievement of a specific SIL, also qualitative requirements must be adhered to (IEC 61508, 2010). SIL levels do not provide a meaningful representation of an acceptable risk level during operation.

According to DNVGL (2018), the tolerable risk level will most likely have to be defined by the International Maritime Organization (IMO) and flag states for specific operations. Further, a goal-based code is suggested stating that "autonomous and remote-controlled ships shall be as safe as conventional ships of the same type", but ultimately safety should be much improved. Translating the existing different types of acceptance criteria, as mentioned above, into specific operational constraints for a control system remain a challenge, but the qualitative safety constraints from STPA can be used as a basis.

5 Conclusions

This paper presents the first step towards supervisory risk control of MASS; namely providing a systematic process for identifying and analyzing hazards that directly can be used to develop the content and structure of a risk model to be used by the control system of an autonomous ship. Supervisory risk control means that the autonomous system is capable of risk management, enchancing its intelligence, through the integration of a risk model into the supervisory (mission) layer of the autonomous system's control hierarchy. Even though the main focus of the paper is a process for developing online risk models in terms of combining STPA and BBN, two general approaches to supervisory risk control are suggested; control based on real-time risk estimate feedback, and optimization through model predictive control.

The former means that online risk models may provide risk information that is used in a control policy that minimizes risk (cost) or maximizes a utility function (cost/benefit). The latter means that the online risk model may be used to predict the future risk by simulating the sailing process using a digital twin, i.e., a mathematical model of the ship the environment with the current states of the sailing process.

The main result of the paper is the proposed framework consisting of two main phases; i.e., (i) identifying and analyzing what and how things can go wrong with STPA, which is a feasible method for autonomous systems, and (ii) using the results of the STPA to develop nodes and a structure for a BBN that represents an online risk model to be used by the control system of an autonomous ship. Online risk models that can provide decision support to control systems of autonomous ships, subject to environmental and operational conditions and constraints, both proactively and reactively are needed. Proactively means early warnings on possible violations of the autonomous ship's operating envelope constituted by safety constraints. Reactively means that human operators and supervisors are given more time for efficient responses and crisis intervention through predictions of possible outcomes.

The framework presented in the paper is tested in a case study focusing on the guidance system of a MASS and demonstrating the feasibility for generating BBNs based on the qualitative results of STPA. Currently, there is no straight forward way of identifying nodes and structuring their relationship in a BBN for risk analysis and modeling. The paper presents how the unsafe control actions, the scenarios and their causal factors from STPA create a hierarchy that can be transformed into a BBN. This simplifies the development process of the BBN and ensures that the risk information contained in the model has been systematically derived. Even though the

framework focuses on MASS due to the limited operational experience available for such systems, the framework may also benefit other types of systems, such as other types of autonomous vehicles, conventional ships, etc. Future work should include developing a software tool for STPA which from the results also generates a BBN.

To utilize the online risk model based on the BBN in operation, risk acceptance criteria and operational safety constraints need to be determined. Qualititative safety constraints are developed in STPA but these need to be quantified to enable use by a control system in operation. Such constraints need to ensure that the MASS is at least as safe as conventional ships, but the aim should be on achieving safer solutions. Work remains to "translate" the typical overall risk acceptance criteria related to fatalities and costs per year into operational criteria that can be used for a control system of an autonomous ship to make decisions in operation.

Future work also includes to investigate the combination of several BBN into one online risk model for a control system, i.e., expanding the case study further. Finally, industrial implementation is promising for the future developments of control systems and the realization of autonomous ships, but there is still a gap from case study application to real practical integration that needs to be closed.

Acknowledgments

The work is partly sponsored by the Research Council of Norway through the Centre of Excellence funding scheme, project number 223254 AMOS, ORCAS with project number 280655 and UNLOCK with project number 274441.

References

Acanfora, M, Krata, P, Montewka, J, 2018. Toward a method for detecting large roll motions suitable for oceangoing ships. Applied Ocean Research, 79, 49-61.

Adedigba SA, Oloruntobi O, Khan F, Butt S, 2018. Data-driven dynamic risk analysis of offshore drilling operations Journal of Petroleum Science and Engineering 165, 444-452.

Baniela SI, Rios JV, 2011. Maritime safety standards and the seriousness of shipping accidents. The Journal of Navigation, 64, 495-520.

Barua S, Gao X, Pasman H, Mannan MS, 2016. BN based dynamic operational risk assessment. Journal of Loss Prevention in the Process Industries, 41, 399-410.

Blanke M, Kinnaert M, Lunze J, Staroswiecki M, 2015. Diagnosis and fault tolerant control. Springer, 3rd edition.

Brito M, Griffiths G, 2016. A Bayesian approach for predicting risk of autonomous underwater vehicle loss during their missions. Reliability Engineering and System Safety 146, 55-67.

Bye RJ, Aalberg AL, 2018. Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. Reliability Engineering and System Safety, 176, 174-186.

Chen P, Huan Y, Mou J, van Gelder PHAJM, 2019. Probabilistic risk analysis for ship-ship collision: State-of-the-art. Safety Science, 117; 108-122.
Danish Maritime Authority (DMA), 2017. Analysis of regulatory barriers to the use of autonomous ships. Final Report, Rambøll, Core, Denmark.
Darwiche, A, 2009. Modeling and Reasoning with Bayesian Networks. Cambridge Unversity Press, New York.
Department for Transport, 2019. Maritime 2050. Navigating for the Future. Report, OGL, London, UK.
DNVGL, 2018. Remote-controlled and autonomous ships, DNVGL group technology & research, position paper 2018 in the maritime industry.
DNVGL 2016. Rules for classification of ships – Dynamic positioning systems with enhanced reliablity.
Fenton, NE, Neil, M, 2013. Decision analysis with Bayesian Networks. CRC Press, Taylor & Francis Group, US.
Fischhoff B, Slovic P, Derby SL, Keeney RL, 1983. Acceptable Risk. Cambrige, USA: Cambridge University Press.
Gil, M, Wrobel, K, Montewka, J, 2019. Toward a method evaluating control actions in STPA – based model of ship-ship collision avoidance process. Journal of Offshore Mechanics and Arctic Engineering, 141 (5), 051105.
Hassel M, Utne IB, Vinnem JE. 2014. Analysis Of The Main Challenges With The Current Risk Model For Collisions Between Ships and Offshore Installations On The Norwegian Continental Shelf. In proc. of the Probabilistic Safety Assessment and Management PSAM 12, June 2014, Honolulu, Hawaii.
Hegde J, Utne IB, Schjølberg I, Thorkildsen B, 2018. A Bayesian approach to risk modeling of autonomous subsea intervention operations, Reliability Engineering and System Safety, 175, 142-159
IEC 61508: Functional Safety of Electrical/Electronic/ Programmable Electronic Safety Related Systems. Geneva, Switzerland: International Electrotechnical Commission; 2010.
IMO: International Maritime Organization. 1972. Convention on the international regulations for preventing collisions at sea (COLREGs).
IMO: International Maritime Organization, 2016. Seafarers rights 2016. Deaths and injuries at sea. http://seafarersrights.org/seafarers-subjects/deaths-and-injuries-at-sea/ (Accessed: 2017-05-12)
IMO: International Maritime Organization, 2018. Revised guidelines for formal safety assessment, MSC-MEPC.2/Circ.12/Rev.2, London, UK.
IMT 2018: Online risk management and risk control for autonomous ships. KPN ORCAS: Research project with industry participation, Department of Marine Technology, NTNU: https://www.ntnu.edu/imt/orcas (Accessed: 2019-02-08).
ISO 31000, 2018. Risk Management – Guidelines. International Standardization Organization.
Johansen IL, 2010. Foundations of Risk Assessment. ROSS Report, NTNU, Trondheim, Norway.
Johansen TA, Perez T, Cristofaro A, 2016, Ship collision avoidance and COLREGS compliance using simulation-based control behavior selection with predictive hazard assessment, IEEE Trans. Intelligent Transportation Systems, 17, 3407 – 3422.

Johansen, TA, 2017. Toward dependable embedded model predictive control. IEEE Systems Journal, 11, 1208-1219.

Khakzad N, Khan F, Amyotte P. 2013. Dynamic safety analysis of process systems by mapping bow-tie into Bayeisan network. Process Safety and Environmental Protection 91, 46-53.

Khan F, Hashemi SJ, Paltrinieri N, Amyotte, P, Cozzani V, Reniers G, 2016. Dynamic risk management: a contemporary approach to process safety management. Current Opinion in Chemical Engineering, 14, 9–17.

Kristiansen S. 2013. Maritime transportation: Safety management and risk analysis. Routledge.

Leveson N, 2011. Engineering a safer world. Systems thinking applied to safety. The MIT Press, Cambridge, USA.

Leveson N, Thomas JP. 2018. STPA Handbook. March 2018. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf (Accessed: 01-08-2019).

Li X, Chena G, Khan F, Xua C, 2019. Dynamic risk assessment of subsea pipelines leak using precursor data Ocean Engineering 178, 156-169.

LR, 2015. Global Marine Technology Trends 2030, Lloyd's Register, QinetiQ and University of Southampton.UK.

Ludvigsen M, Sørensen AJ, 2016. Towards integrated autonomous underwater operations for ocean mapping and monitoring, IFAC Journal of Annual Reviews in Control, 42, September 1-13, Elsevier Ltd.

Marszal EM, 2001. Tolerable Risk Guidelines. ISA Transactions, 40, 391-399.

Mazaheri A, Montewka J, Kujala P, 2016. Towards an evidence-based probabilistic risk model for ship-grounding accidents, Safety Science, 86, 195-210.

Montewka J, Wrobel K, Heikkila E, Valdez-Banda O, Goerlandt F, Haugen S, 2018. Challenges, solution proposals and research directions in safety and risk assessment of autonomous shipping. In Proceedings for the Conference on Probabilistic Safety Assessment and Management PSAM 14, September 2018, Los Angeles, CA.

National Institute of Standards and Technology (NIST). 2008. Autonomy levels for unmanned systems (ALFUS) Framework. Volume I: Terminlogy, version 2.0, NIST Special Publication 1011-I-2.0.

NFAS. 2017. Definitions for Autonomous Merchant Ships. Draft. http://nfas.autonomous-ship.org/resources/autonom-defs.pdf (Accessed: 2018-06-25).

NORSOK. Z013: Risk and Emergency Preparedness Assessment. Lysaker, Norway: Standards Norway; 2010.

Norwegian Maritime Authority (NMA), 2018. Requirements to documentation for constructing autonomous, unmanned and/or remotely operated vessels. Guidelines, Draft, Norway. (In Norwegian: Krav til dokumentasjon i forbindelse med bygging av autonome, ubemannede og/eller fjernstyrte fartøy).

NS5814:2008: Requirements for Risk Assessment. Norway: Standard Norge; 2008.

Paltrieri N, Scarponi GE, Khan F, Hauge S, 2014. Addressing dynamic risk in the petroleum industry by means of innovative analysis solutions. Chemical Engineering Transactions, 36, 451-456.

Qin SJ, Badgwell TA, 2003. A survey of industrial model predictive control technology, Contr Eng Pract 11, 733-764.

Rasmussen J, 1997. Risk management in a dynamic society: a modelling problem. Safety Science, 27(2-3):183-213.

Rausand M. 2011, Risk assessment. Theory, methods, and applications. Wiley, Hobroken, USA.

Reason, J. (1990). Human Error. Cambridge University Press, Cambridge, UK.

Rokseth B, Utne IB, Vinnem JE, 2017. A systems approach to risk analysis of maritime operations. Proceedings of the Institution of Mechanical Engineers. Part O, Journal of Risk and Reliability 231 (1), 53-68.

Rokseth B, Utne IB, Vinnem JE, 2018. Deriving verification objectives and scenarios for maritime systems using the systems-theoretic process analysis. Reliability Engineering & System Safety, 169, 18-31.

Rokseth B, Haugen OI, Utne IB, 2019 Safety Verification for Autonomous Ships. MATEC Web of Conferences. 273: 02002.

Rødseth ØJ, Tjora Å, 2015. A risk-based approach to the design of unmanned ship control systems. Maritime-Port Technology and Development – Ehlers et al. (Eds). Taylor & Francis Group, London.

Sotiralis P, Ventikos NP, Hamann R, Golyshev P, Teixeira AP. 2016. Inforporation of human factos into ship collision risk models focusing on human centered design aspects. Reliability Engineering and System Safety, 156, 210-227.

Sørensen, AJ, 2011. A Survey of Dynamic Positioning Control Systems. IFAC Journal of Annual Reviews in Control, Volume 35, Issue 1, April, Pages 123-136, Elsevier Ltd, ISSN: 1367-5788.

Thieme CA, Utne IB, 2017. A risk model for autonomous marine systems and operation focucing on human-autonomoy collaboration. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 231 (4), 446-464.

Thieme CA, Utne IB, Haugen S, 2018. Assessing ship risk model applicability to Marine Autonomous Surface Ships. Ocean Engineering, 165, 65-75.

UK Health and Safety Executive. The Tolerability of Risk from Nuclear Power Stations. In: Ballard G, Broadbent D, Clarke R, Dunster CB HJ, Littlewood B, Slater D, et al., editors. Norwich, United Kingdom: Health and Safety Executive UK; 1992.

UK Health and Safety Executive. Reducing Risks, Protecting People - Hse's Decision Making Process. Norwich, UK: Health and Safet Executive UK; 2001.

Utne IB, Sørensen AJ, Schjølberg I, 2017. Risk management of autonomous marine operations and systems. Conference on Ocean, Offshore & Arctic Engineering, OMAE 2017-61645, Trondheim, Norway.

Vagia M, Transeth AA, Fjerdingen SA, 2016. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Applied Ergonomics, 53, Part A, pp. 190–202.

Vinnem JE, 2014, Offshore Risk Assessment Vol. 1. London, Heidelberg, New York, Dordrecht: Springer London.

Vinnem JE, Utne IB, Schjølberg I, 2015. On the need for online decision support in FPSO-shuttle tanker collision risk reduction. Ocean Engineering 101, 109-117.

Wrobel K, Montewka J, Kujala P, 2017. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliability Engineering and System Safety, 165, 155-169.

Wrobel K, Montewka J, Kujala P, 2018a. System-theoretic approach to safety of remotely-controlled merchant vessel, Ocean Engineering 152, 334–345.

Wrobel K, Montewka J, Kujala P, 2018b. Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels, Reliability Engineering and System Safety, 178, 209-224.

Wróbel, K, Gil K, Montewka J. 2018. Towards A Method Evaluating Control Actions In STPA-Based Model of Ship-Ship Collision Avoidance Process. In Proceedings of the ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering, OMAE2018 June 17-22, 2018, Madrid, Spain OMAE2018-77790.

Zeng Z, Zio E, 2018. Dynamic Risk Assessment Based on Statistical Failure Data and Condition-Monitoring Degradation Data, IEEE Transactions on Reliability, 67 (2), 609.

Zhou X-Y, Liu Z-J, Wang F-W, Ni S-K, 2018. Collision risk identification of autonomous ships based on the synergy ship domain. The 30th Chinese Control and Decision Conference (2018 CCDC), IEEE.

Zio, E, 2018. The future of risk assessment. Reliability Engineering and System Safety, 177, 176-190.

Øien K, 2001. Risk indicators as a tool for risk control. Reliability Engineering and System Safety, 74 (2), 129-145.