Challenges, solution proposals and research directions in safety and risk assessment of autonomous shipping

Montewka J.[a,c,*], Wróbel K.[a], Heikkila E.[b], Valdez-Banda O.[c], Goerlandt F.[d,c], Haugen S.[e]

[a] Gdynia Maritime University, Poland
[b] VTT Technical Research Centre of Finland Ltd, Tampere, Finland
[c] Aalto University, Espoo, Finland
[d] Dalhousie University, Halifax, Canada
[e] NTNU Trondheim, Norway

Abstract: Maritime Autonomous Surface Ships (MASS) are becoming reality in the shipping industry. Besides numerous anticipated advantages this type of ships has there are new potential hazards as well that have to be addressed. This needs to be done at the design stage of transportation system that will encompass such vehicles, to make sure that the new system's safety can be guaranteed at the acceptable level. To this end numerous methods can be utilized pertaining to the field of safety and risk assessment. Those methods however are of different scope and may have different application areas.
Therefore in this paper we discuss selected methods suitable for safety assessment and quantification of transportation systems including goal-based safety case approach, system theoretic process analysis and risk assessment. Challenges and opportunities of those approaches are highlighted and the recommendations are given regarding the application areas of the methods.

Keywords: Safety and risk, maritime transport, autonomous ships.

1. INTRODUCTION

Maritime transportation safety is governed by global and local codes and practices, and a distillation of past experience [29]. These may become less and less relevant over time, especially when innovative solutions are involved. One of such innovation in the maritime world is autonomous surface shipping. Due to high diversity of potential design solutions, the prescriptive regulations are clearly unable to cope with them. Moreover, prescriptive regulations reflect the best engineering practice at the time they were written and may become obsolete where best practice is changing, e.g. with evolving technologies, shifting paradigms regarding ship operations.

Therefore it is of utmost importance to provide a proactive and systemic framework allowing safety assessment and/or quantification for innovative solutions. To this end the International Maritime Organization (IMO) offers two solutions: Formal Safety Assessment (FSA), and Goal-based Standards (GBS), see for example [18-20]. FSA is described as "*a rational and systematic process for assessing the risks associated with shipping activity and for evaluating the costs and benefits of IMO's options for reducing these risks*." It is often applied as a framework for risk-based design (RBD) for ships and offshore structures, [33,39].
GBS relates to "*high-level standards and procedures that are to be met through regulations, rules and standards for ships. GBS are comprised of at least one goal, functional requirement(s) associated with that goal, and verification of conformity that rules/regulations meet the functional requirements including goals*." In principle GBS is a high-level procedure, that may encompass the results of FSA at a stage of functional requirements, therefore those two are interlinked. Moreover, in the context of GBS and FSA safety is defined as the absence of unacceptable levels of risk to life, limb and health from unwilingful acts. To estimate risk the probability and consequences of anticipated accidents are sought. These are delivered through quantitative approaches, which are strongly preferred over qualitative in the context of FSA.

However, in order to ensure safety of complex and innovative socio-technical system a safety analysis should be conducted adopting appropriate methods, suitable for a given purpose, accounting for a available body of background knowledge. First, such analysis needs to point out the areas and processes that must be ensured in the system for its safe operations. Second, the potential solutions maintaining the safety in those areas need to be proposed. Third, the effect of limitations in the background knowledge should be evaluated and communicated to the end-user, [3,4].

The quantitative risk assessment as postulated by FSA, where the risk is defined as a combination of probability and consequences of an accident, can be informative for known systems, when the costs-benefit approach is adopted to justify the expenses related to the changes in system design. However it neither assists design process nor ensures the safety of the system in the situation where the amount of uncertainty related to the system design and operation is significant. To this end, the way how the risk is defined and what scientific perspective is taken need to be widened and qualitative approaches allowed.

Therefore, in this paper, we take closer look into three approaches to safety in the maritime domain, pointing out their strengths and weaknesses, in the context of novel technology such as autonomous ships.

The paper is structured as follows: section 2 discusses FSA and RBD, section 3 gives an overview of STAMP, section 4 introduces goal-based safety case, whereas Section 4 discusses potential future research directions and concludes the paper.

2. RISK-INFORMED DESIGN

2.1. Problem formulation

Risk-informed ship design, in the maritime domain usually referred to as risk-based design (RBD), is one of three major research lines in the domain of maritime transportation risk [28,39,52]. The other two are risk-informed management of the maritime transportation system [35,48,50,53], and risk-informed emergency preparedness and response planning [2,30,38]. Risk informed approaches are promising because they explicitly address risk, they contribute to a more uniform risk level and they allow more flexibility in choice of solutions to achieve acceptable risk. However, an important prerequisite for using risk informed approaches is that risk is communicated to decision-makers in a way that they can relate to and understand.

As found in [17], the view on risk adopted in the majority of studies within the maritime domain is based on the definition formulated by Kaplan and Garrick in [25]. It follows a perspective postulated firstly in Formal Safety Assessment (FSA), and continued in Goal-based Standards by the International Maritime Organization (IMO), [18-20]. Risk (R) is then defined as a combination of the probability (P) of an accident and its consequences (C), as follows:

$$R=f(P,C) \tag{1}$$

Under this definition, risk is normally expressed as the product of the probability (P) and consequences (C):

$$RI=PxC \tag{2}$$

This can also be regarded as an expression of the statistically expected consequences or loss. In FSA, RI is used for ranking hazards and identifying those that require detailed analysis in the later stage of FSA. This is useful and sufficient in many cases, but this definition can also hide information about situations that may be very different. Consider the following two examples:

- In one situation, risk is dominated by frequent events resulting in minor consequences - single or minor injuries and local equipment damage. Examples can be minor occupational accidents like hand injuries or minor falls.

- In another situation, risk is dominated by low frequency events with catastrophic consequences - multiple fatalities and total loss of a ship. An example can be a major collision between two ships.

RI may take the same value in both cases, because P is high and C is low in the first situation, while P is low and C is high in the second case. Based on the conventional view on risk, the two situations are thus equal. However, this misses a very important fact about the situation. Most likely, the first situation is based on much more information than the second one, and uncertainty about particular probability will be much smaller than in the second situation.
Even if we should have similar background knowledge, there is also another aspect that makes the two situations different. Catastrophic but infrequent losses are normally perceived as worse than frequent, more limited losses [36]. This information is also lost in the conventional expression of risk.

Thirdly, the measures to control the risks in these two situations may be different. In the first case, the focus might be given to reducing the probability of occurrence, whereas in the second case consequence reduction may be considered more important.

Finally, interpreting risk simply as a product of P and C can also lead to the misconception that risk is just a number, and becomes divorced from the scenarios of concern and available background knowledge. Applying this perspective, much of the relevant information needed for risk management is not properly reflected or even missing [3]. In many risk analyses, one sees that a lot of effort is put into producing as "accurate" risk numbers as possible. However, it is futile to calculate high-precision values in the risk analysis if other parameters essentially are "guesstimates" made by the analyst. In the extreme cases, the numbers obtained from databases and analysis are considered "the ultimate truth" about the probability of an accident in the analysed area, without proper reflection of the context and background knowledge.

This clearly also poses problems in relation to decisions about acceptable risk. When applying the ALARP principle, there is a need to define limits for the intolerable, ALARP and tolerable risk levels. If numbers are interpreted as the truth about risk, decisions will obviously also be made in accordance with this, not taking into account that limited background knowledge can introduce very large uncertainties in the numbers and thus also in the decisions.

The prevailing interpretation of risk in the maritime industry thus has some clear weaknesses. This is a general problem for decision-making, but in relation to autonomous shipping, the first item is particularly important. One characteristic of autonomous shipping is that this is new technology, which has not been available for a long time and has never been used in an application like this apart from very limited tests. Another important aspect is the extensive use of software to control the ships. There are several implications of this for risk analysis:

- There are no existing databases for autonomous ships. For important accidents types like collision and grounding, existing data are hardly relevant at all, and furthermore, it may also be questioned to what degree existing models are adequate.
- The inclusion of software increases the complexity of the systems and makes them harder to analyse. There is an increased possibility that we are unable to understand fully how the system works and that mistakes are made in the design of the software and hardware.
- New technology also implies that new types and accidents and in particular new causes of accidents are introduced. We may not be able identify these with our current methods for hazard identification (which often are based on checklists of different types).

For our purpose, all of these aspects are related to the background knowledge that we have and can use to perform risk analysis. Effectively, our background knowledge pertaining to autonomous ships is much less than for traditional shipping concepts. However, this can not be reflected in the way that risk is usually expressed. The uncertainty associated with new technology is effectively "hidden" for the decision-makers in this way.

In our view, risk informed ship design introduces some highly beneficial features into the decision process, but one has to be very aware of the underlying understanding of risk. We are proposing a different definition of risk in the following subsection, followed by some discussion of the benefits of this.

2.2. Solution proposal

To address some of the above-mentioned shortcomings of the existing approaches to risk in the maritime domain, a wider concept of risk should in our view be introduced to the field. This would allow: 1) a systematic and hierarchical description of the risk associated with a given system; 2) reasoning about risk control options (RCO) in light of available background knowledge (BK); 3) reflection of the effect of BK on the evaluated risk and proposed RCOs, [37].

As outlined in [17], a spectrum of scientific approaches to risk exist in the wider risk research discipline [7,42,47]. On one end of the spectrum sits the strong realist view, where risk perspectives consist exclusively of probabilistic risk measures. The evidences for these probabilities are based on data or models, such as typically applied in RBD [39]. At the other end of spectrum sits the strong constructivists view on risk, according to which risk is nothing but observer's perception about a given situation, and varies between observers [17]. The field of maritime risk analysis is dominated by the realist view, where risk is mainly understood and measured as the combination of P and C, based on historic accident data and engineering models, as found by [17].

In cases where the background knowledge is good and on a comparable level for all relevant risk contributors, the uncertainty related to the evidence is low, and in such cases the realist view may be adequate. It can be argued that this is the case for traditional design engineering in the maritime industry, where there are relatively large databases about accidents at sea, and numerous models and tools to facilitate ship design, and simulate maritime transportation systems. However, in numerous cases, there are also serious shortcomings related to the paucity and scatter of the available data, [15,34], underreporting of accidents and incidents [41,43], and issues related to reliability and validity of the models, see for example [16,45]. Obviously, all these shortcomings become more serious, when designing maritime transportation system encompassing autonomous vessels [54,55].

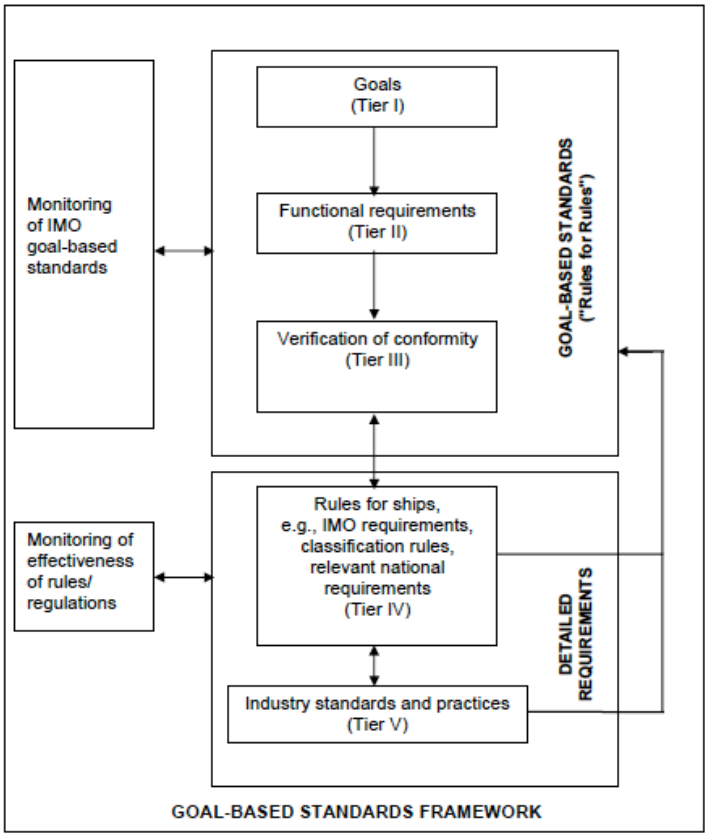
To allow a flexible framework for risk analysis for innovative systems with no or limited design or operational experience, it would be beneficial to differentiate between the concept of risk and ways to describe risk, as suggested by Society of Risk Analysis in [44]. Based on [17,24], the following terminology is adopted in this paper: 1) the risk concept concerns what risk means in itself, i.e. what risk "is"; 2) a risk perspective is a way to describe risk, a systematic manner to analyse and make statements about risk; 3) a risk metric is the assignment of a numerical value to an aspect of risk according to a certain standard or rule.

The risk perspective allows various scientific approaches to risk, depending on the available BK, utilizing the available sources of data and knowledge. This risk measurement should be performed in support of the most suitable risk metrics for the given decision problem, e.g. following guidance provided by [24]. Recently, there has been a shift in risk paradigm in the offshore oil and gas industry. The move has been made from PxC definition towards uncertainty based perspective, which stresses the relevance of uncertainty assessment in the process of risk analysis, informing thus the end-users about the quality of the obtained risk estimates, [22]. A similar development would be beneficial in the maritime industry, in particular in the context of innovative systems in general and autonomous shipping specifically.

In the context of GBS, the concept of risk is used at the stage of verification of conformity (Tier III in Figure 1). The risk level of a given ship design is confronted with the allowed risk levels as anticipated by the rules (Tier IV). The tolerable, intolerable and ALARP risk levels are defined by the the relevant stakeholders like IMO, authorities or classification societies. In the present FSA guidelines, the risk acceptance criteria are explicitly numerically defined based on risk measures combining probability

and consequence. However, in the presence of major uncertainty, which is not acknowledged nor their effect on the results evaluated, the calculated risk numbers can become unreliable. Enhancing the risk description with a qualitative uncertainty assessment, as suggested in [4], could be a way to alleviate the effects of this unreliability. Another way could be to adopt a different risk perspective, focusing only on probability of occurrence or consequences depending on the strength of available knowledge, and develop risk evaluation criteria to support decision making focusing on relative changes in derived risk metrics.

Figure 1. Goal-based standards framework by IMO, [19].



3. SYSTEM THEORETIC PROCESS ANALYSIS

System-Theoretic Accident Model and Processes (STAMP) is an approach to depict and review the function of safety from a systemic perspective. It analyses accidents by making a review of the entire socio-technical system [9,26]. Thus, it provides a more systemic way to model accidents and safety for producing a better and less subjective understanding about how accidents occur and how they can be prevented [14,46]. STAMP was initially developed as an accident-modelling framework, conceptualising socio-technical processes as systemic performances in a state of dynamic equilibrium. Therein, safety is considered to be a feature resulting from the performance of multi-layered feedback loops of information and control between different stakeholders, [1].

Gradually, STAMP evolved into few practical frameworks, such as *Causal Analysis based on System Theory* (CAST) and *System-Theoretic Process Analysis* (STPA). While the former is used in past accidents' investigations, the latter focuses on ensuring safety of existing systems or those in early phases of development. STPA is a hazard analysis technique that identifies accident scenarios that encompass the entire accident process by including design errors, component interactions, and other social, organizational, and management factors in the analysis [31]. STPA consists of four basic steps as given in Figure 2.

STAMP and STPA are based on a systemic and systematic approach which features certain methodological advantages. These consist in applying a non-linear, quantitative approach to analyse system's safety. Thus, accurate information on components' reliability is futile since it is interactions between components that constitute a central part of the analysis. Furthermore, these are analysed qualitatively with focus on potential causes, consequences and mitigation measures of their inadequacy. Neither probability of failure nor risk is calculated which allows inclusion of processes or components for which such calculations might be difficult or ineffective due to their nature or commonly lack of data. For the same reason, the method proved helpful in analysing safety of systems being in their early phases of development. This issue can cause quantitative analysis to be counterproductive as the exact structure of the system may be unknown, thus baning for its reliability analysis. Herein, an iterative cooperation between system's designers and safety analysts is postulated in a form of safety-driven design, see Figure 3 and [31].
On the other hand, STAMP's and related methods' disadvantages incorporate a rather non-intuitive visualization of results, research-practice gap, and a lower-than-satisfactory level of potential users' confidence with using non-linear methods of safety analysis instead of traditionally recognized ones, [13,49]. Additionally, in the application of this methodological approach, the decision about to what level of details the analysis needs to be concluded is another common limitation [21,51]. Thus, the implementation of these approaches can be mind challenging and time-consuming.

Nevertheless, the fact that system-theoretic framework can be applied to systems, development of which is yet to be completed, made it a potential method of safety analysis for MASS. Two studies have been performed recently [54,51], in which STPA was applied to preliminarily analyse safety concerns of MASS. Supported by experts' elicitation and literature review, STPA is helpful in elaborating a holistic model of autonomous ship's safety control structure and analysing it in order to identify potentially hazardous scenarios to which MASS could be exposed and elaborate feasible hazard mitigation measures. The results obtained through STPA did not differ from outputs of quantitative analyses, at least in the comparable aspects. Nevertheless, the following lessons have been learned:

- Holistic approach to MASS safety was not common to date with only specific functions being analysed, such as collision avoidance, [8], remote control, [40], and autonomy modes transition, [54]. Consistency of outputs may be a result of technology's innovativeness or the fact that they were based on similar literature sources or experts' experience.
- STPA is similar to other safety analysis frameworks in terms of being affected by analysts' background and previous working experience even despite the fact that experts were elicited to restrict the effects of framing.
- Refraining from calculation of risks pertaining to system's operation can be found difficult to accept or comprehend by some more conservative representatives of industry, who were taught of 'safety being the lack of risks, and the latter being a product of mathematical probability and potential consequences'.
- As noted above, decision-makers usually base their decisions on the evaluation of options they are presented with. Such presentation can be a result of quantitative analysis but not a qualitative one. It shall be underlined that these are distinct sets of methods that can provide different types of output, which does not necessarily mean that the results of one analysis are superior to another's.
- STPA offers a potential alternative to be used in risk-based design in existing approaches such as the FSA by IMO. It offers a hazard analysis approach capable of generate data for the analysis of a system under design.
- There is a major gap in maritime environment understanding between systems' potential operators and its designers. While the former are likely to originate from sea-going professionals, the latter often do not have any form of maritime operational experience, [10]. Although the very act of designing a vessel does not require the naval architect to be skilled as a seafarer, close links between those professional groups can prove beneficial. The gap can potentially be bridged iteratively through a safety-driven design process which gives an opportunity for including output of safety analysis in project-level decision-making.

- The application of the STPA provides specific safety controls with defined control logic principles which describe in details the safety demands for designing the MASS and the technologies linked to it. This provides the basis for establishing the initial safety and risk management strategy.
- The application of the STPA has clearly identified gaps regarding the interaction between the emergency response organizations (e.g. SAR services) and MASS in need of assistance. The current approach follows a coordination between SAR and the captain responsible for the vessel. The mentioned studies introduces the relevance of ship design, materials used and technology implemented to support the coordination of emergency situations.
- Equipment redundancy and testing seem to be crucial elements to ensure the safety performance of MASS. These two elements demand a more specialized analysis to ensure the efficiency in the design of MASS.
- Defining safety controls for MASS demonstrate the importance of having an efficient coordination among the stakeholders of the system. The safety responsibilities seem to be delegated among a higher number of system stakeholders compared to the current approach.
- Last but not least, the results of STPA shall be validated. This would not be possible until MASSs are implemented and their overall safety performance is tested. Similarly, results of a safety-driven design process as mentioned above are yet to be determined as the technology is still under development. Outputs of safety analyses performed to date shall be included in current system design works, but results of the latter are pending.

Figure 2. A simplified overview of STPA, [32].

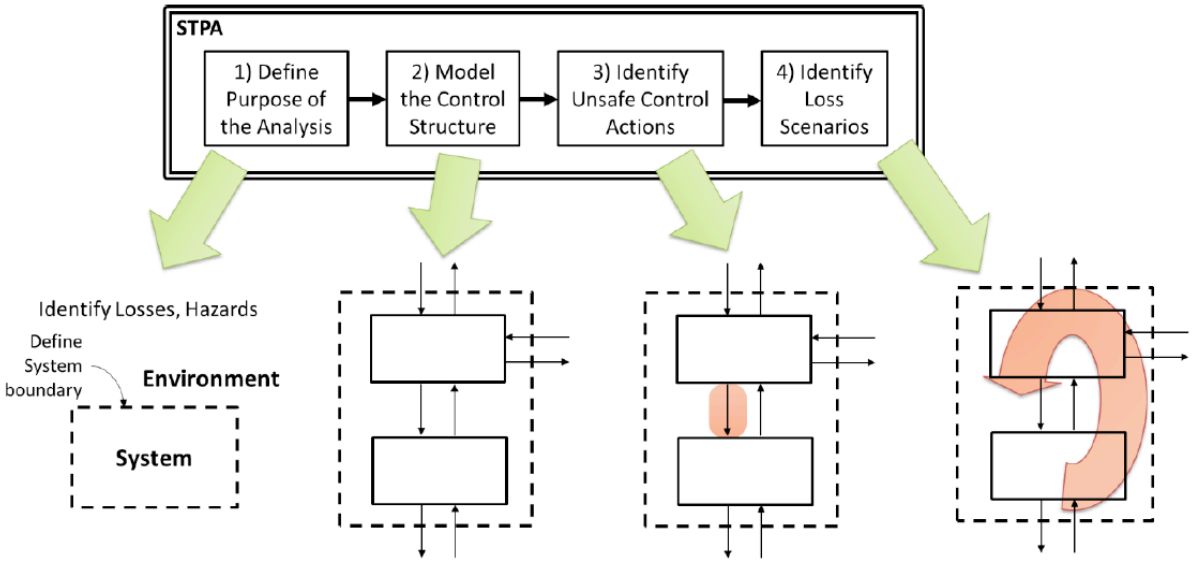
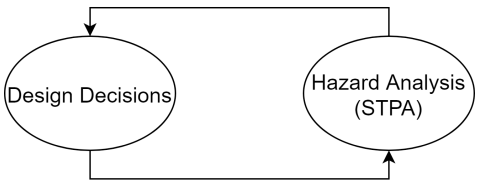


Figure 3. Safety-driven design concept, [31].



3. GOAL-BASED SAFETY CASE APPROACH

In addition to the actual risk analysis procedures, the documentation and communication of safety-related requirements, design choices and validation results are becoming increasingly important. The introduction of new technologies, such as situational awareness systems and autonomous decision-

making, are making ship systems increasingly complex. To receive required authority permissions to test and later operate autonomous vessels, a comprehensive demonstration regarding the safety of autonomous technologies is required [23].

Traditionally, designers have been able to follow a fairly well established set of design rules and standards. While standards and design guidelines also for autonomous technologies are slowly becoming available, they mostly focus on providing more general performance guidelines instead of providing prescriptive design rules [11]. Thus, technology developers can be seen to bear an increasing responsibility of ensuring and demonstrating the safety of new technologies. This will require a robust safety qualification procedure to be in place.

Safety qualification is the process of gathering and structuring the safety evidence required to credibly demonstrate that an appropriate level of safety is reached. Several guidelines exist for facilitating the safety qualification activities, and in the maritime sector these are mainly available from classification societies, e.g. [5,12,33]. In this section we present the general concept of a goal-based safety case approach to support safety qualification, and discuss its potential in activities related to ensuring the safety of new technologies that are developed for autonomous vessels. Additionally, we discuss experiences based on application of this methodology into an autonomous shipping prototype case study [23].

A simplified process flow for a safety qualification process with goal-based approach is presented in Figure 4. The qualification process should be of an iterative nature, so that major changes to the technology trigger a new round of qualification activities.

The goal-based safety case approach (not to be confused with the IMO-originated GBS discussed earlier) is a proposed extension to the regular safety qualification methodologies to help in structuring the results of qualification activities and especially in enabling communication between the different stakeholders involved in the safety design and qualification processes. In this approach, the safety requirements (represented as goals) and safety evidence (data created in the actual qualification activities) are presented together in a visual manner as a *structured safety case*. This provides a link showing which evidence items are provided to demonstrate the fulfillment of each of the safety goals. The structure of safety goals is a living documentation that is updated throughout the design and qualification processes. The resulting documentation can be represented as a tree-like model using various visualization languages, such as the Goal Structuring Notation (GSN) [27], a general example of which is presented in Figure 5.

Based on the authors' experiences, the major advantages in the goal-based approach are in the communicative power of the visual representation of safety goals and evidence, making the link between these easily comprehensible. This enables efficient communication regarding safety between the different stakeholders. By easing the communication, the approach potentially enables a faster approval of new technologies for autonomous shipping. The methodology, however, is new to the maritime sector and further case applications are needed to fully consider its benefits.

As a potential downside, the methodology is mainly designed with the communicational aspect in mind, and thus provides no direct tools for prioritizing the safety goals based on their safety impact. Neither does it directly provide tools for assessing the probabilities or uncertainties regarding the fulfillment of the goals. Thus, these aspects still need to be considered with relevant tools, such as ones described in this paper. Practically this means that while the communicational aspects are improved, the qualifying party still requires substantial understanding of the technical system and its safety-related design.

Several aspects in the goal-based approach also provide basis for future developments. One development path is a wider adoption of the goal-based documentation in different aspects of design, such as reliability and maintainability, to provide a more comprehensive assurance case of the autonomous vessel design. Another development path is seen in the potential of combining

probabilistic methods with the safety case approach to provide further justification for the safety-related design choices and to help prioritize risk controls in the system design phase.

Figure 4. A safety qualification procedure, resulting in safety argumentation documented as structured safety case.

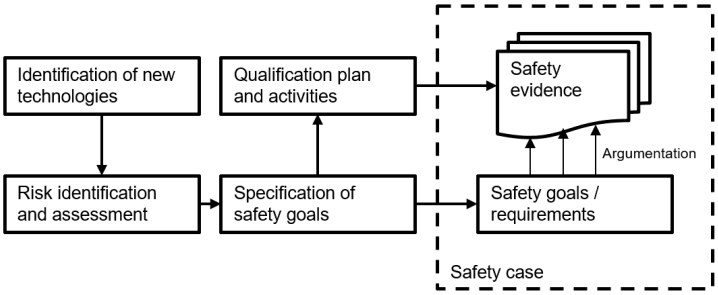
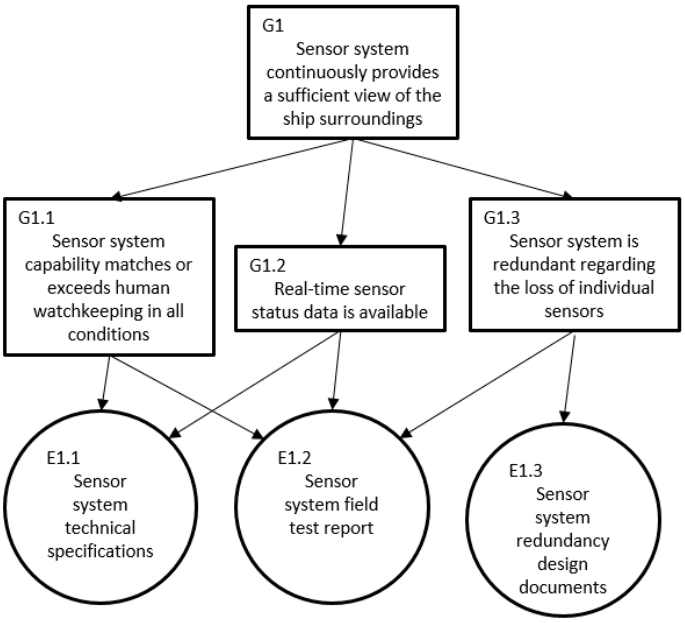


Figure 5. A simplified example of how the safety goals and evidence can be represented in the case of an autonomous ship sensor system. Goal-based safety case using the GSN visualization language is utilized to display the link between design requirements and safety evidence documentation.



4. CONCLUSIONS

In this paper, risk-informed design, system theoretic process analysis and goal based safety case approaches have been briefly discussed and described, in particular in relation to the discussion on ways of ensuring the safety of autonomous shipping. They have been discussed separately here, but if we look at the problem from the point of view of how to best manage risk, the approaches covered in the paper are all linked together. They cover different aspects of the process and they are also on different levels in the hierarchy of tools to manage risk.

4.1. What can be achieved with the present approaches?

Regarding risk analysis
Risk-informed design (and operation for that matter) is fundamentally different approach to design of a system compared to a traditional approach. In essence, traditional approaches are based on history only (experience embodied in regulations, rules and standards) while in a risk-informed approach we

also try to look into the future (in addition to using experience). For slowly developing technologies, a traditional approach may be adequate. We will then take small steps forward, gaining experience with the modified technology before taking it further and hopefully avoiding too many accidents and serious losses. However, the quicker technology develops, the more important it is that we can try to envisage future failures and not only rely on experience. For such an innovative technology, risk-informed approaches are therefore necessary.
Risk analysis can provide valuable information for designers of a novel system insofar it reflects properly the available background knowledge and relevant limitations. Since risk analysis is nothing but systematic organization of available background knowledge on a given system in order to seek suitable solutions to a potential problems assisting decision making process, therefore the effects of BK on the risk estimates and risk control options is mandatory. To this end an adequate scientific approach needs to be adopted and proper way of measuring and communicating risk is required. In the context of MASS the risk estimates that are obtained may be helpful in assigning risk control options for a given system, risk estimates assigned to various designs with respect to some baseline risk level, pointing to the areas in the analysed system that require more research, resulting in more reliable risk estimates.

Regarding STAMP
STAMP and STPA are capable of providing itemized information which can guide the initial design process of MASS. This system engineering approach supports the design and management of complex systems and maintaining these functional during its complete operational life [6]. This enables initiating the design of safety in the earliest conceptual design phase for engineering a safer system [31]. These methodologies provide a truly systematic and systemic approach which is capable of analysing accidents and hazards in different contextual scenarios and it is capable of formulating safety controls to prevent and or to react to those accidents and hazards. The implementation of the approach has previously been proficient for analysing hazards and proposing safety controls with a systematic and systemic approach that covers the operational context of MASS. These safety controls represent the basis for initiating the safety management strategy of MASS and the entire autonomous maritime system(s).

Regarding Goal-based approach
Goal-based approaches and risk-informed approaches go hand in hand. However, if we specifically talk about goal-based safety case approaches, this can also be regarded primarily as a documentation exercise. Clearly, risk-informed approaches require a different type of documentation than traditional approaches. A safety case may be a suitable way of doing this, documenting not just the end result, but also the goal-setting process, development of specific, lower-level requirements from this and the design process itself.

4.2. What can be proposed?

The discussed here methods can provide the input to the risk-informed design process and thus also for documenting the process and results in a safety case. Choosing the best method for a particular problem requires a good understanding of both available methods and the problem at hand. STPA is interesting because it approaches the problem differently than most other methods, regarding management of risk as a control problem (in fact the method focuses on controlling the safety, resulting in risk reduction). For autonomous shipping this may be a particularly relevant approach to take. On the other hand, we will need quantitative methods to support risk-informed design and at present STPA is not suitable for this purpose.

Based on the discussions in the previous sections and the above, it seems clear that there is a need to develop an improved framework for risk management in the maritime industry in general and in particular for autonomous shipping. Some of the key elements are:

- A more flexible perspective on risk, where in particular the aspect of background knowledge/uncertainty is incorporated in our presentation of risk to decision-makers to give them a better basis for making sound decisions.

- Goals-based and risk-informed approaches give flexibility in development of novel solutions, at the same time as retaining consistent and acceptable risk levels also for new technology.
- New risk analysis methods are better suited for analysing increasingly complex systems, with increased use of sensors, software, communication between ships and between ship and shore, very different demands on the humans involved etc. STPA may be one of such methods, but it is crucial to understand the system being analysed and its characteristics before committing to specific risk or safety analysis methods. Both method development and more guidance on choice of methods and combinations of methods is required.

Acknowledgements

First author appreciates the financial contributions from Polish Ministry of Science and Higher Education, grant number DS/442/2017, duration 2017-2019. The Merenkulun säätiö – the Maritime Foundation - from Helsinki is thanked for the travel grant.

References

1. Allison, C. K., Revell, K. M., Sears, R., & Stanton, N. A. (2017). Systems Theoretic Accident Model and Process (STAMP) safety modelling applied to an aircraft rapid decompression event. Safety Science, 98(October), 159–166.
2. Akbari, A., Pelot, R., Eiselt, H.A. 2017. A modular capacitated multi-objective model for locating maritime search and rescue vessels. Annals of Operations Research, doi: 10.1007/s10479-017-2593-1.
3. Aven, T. 2011. Selective critique of risk assessments with recommendations for improving methodology and practise. Reliability Engineering and System Safety 96:509-514.
4. Aven, T. 2013. Practical implications of the new risk perspectives. Reliability Engineering and System Safety 115:136-145.
5. Bureau Veritas. 2010. Guidance Note NI 525: Risk Based Qualification of New Technology – Methodological Guidelines. Neuilly sur Seine Cedex: Bureau Veritas.
6. Blanchard, B. S. 2004. System Engineering Management. John Wiley & Sons.
7. Bradbury, J.A. 1989. The policy implications of differing concepts of risk. Science, Technology, & Human Values 14(4):380-399.
8. Burmeister, H. C., Bruhn, W. C., & Walther, L. (2015). Interaction of Harsh Weather Operation and Collision Avoidance in Autonomous Navigation. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation, 9(1), 31–40.
9. Chatzimichailidou, M., and Dokas, I. 2015. The Risk Situation Awareness Provision Capability and Its Degradation in the Überlingen Accident over Time. Procedia Engineering 128:44–53.
10. Danish Maritime Authority. (2017). Analysis of Regulatory Barriers To the Use of Autonomous Ships. Copenhagen.
11. Danks D., London A. 2017. Regulating autonomous systems: Beyond standards. IEEE Intelligent Systems. Vol. 32:1. IEEE.
12. DNV. 2011. Recommended Practice DNV-RP-A203: Qualification of New Technology. DNV.
13. Findler, M. J., & Chalawadi, R. K. (2017). Teaching STAMP: High Level Communication Design Concerns for a Domestic Robot. In Procedia Engineering (Vol. 179, pp. 52–60). Zürich.
14. Fleming, C., Spencer M., Thomas, J., Leveson, N. and Wilkinson, C. 2013. Safety Assurance in NextGen and Complex Transportation Systems. Safety Science 55:173–87.
15. Goerlandt, F., Goite, H., Valdez Banda, O.A., Höglund, A., Ahonen-Rainio, P., Lensu, M. 2017. An analysis of wintertime navigational accidents in the Northern Baltic Sea. Safety Science 92:66-84.
16. Goerlandt, F., Kujala, P. 2014. On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. Safety Science 62:348-365.
17. Goerlandt, F., Montewka, J. 2015. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. Reliability Engineering and System Safety 138:115-134.
18. IMO. 2010. "Adoption of the International Goal-Based Ship Construction Standards for Bulk Carriers and Oil Tankers. Resolution MSC.287(87)." London.
19. IMO. 2011. "Generic Guidelines for Developing IMO Goal-Based Standards. MSC.1/Circ.1394." London
20. IMO. 2013. "Revised Guidelines for Formal Safety Assessment (FSA) for Use in the IMO Rule-Making Process. MSC-MEPC.2/Circ.12." London.
21. Hardy, K., & Guarnieri, F., 2011. Using a Systemic Model of Accident for Improving Innovative Technologies: Application and Limitations of the STAMP model to a Process for Treatment of Contaminated Substances. In The 15th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2011.
22. Haugen, S., Vinnem, J.E. 2015. Perspectives on risk and the unforeseen. Reliability Engineering & System Safety 137:1-5.

23. Heikkilä E., et al. 2017. Safety Qualification Process for an Autonomous Ship Prototype – a Goal-based Safety Case Approach. Marine Navigation: Proceedings of the 12th International Conference on Marine Navigation and Safety of Sea Transportation (TransNav 2017).
24. Johansen, I.L., Rausand, M. Foundations and choice of risk metrics. Safety Science 62:386-399.
25. Kaplan, S., Garrick, J.B. 1981. On the quantitative definition of risk. Risk Analysis 1(1):11-27.
26. Karanikas, N. 2017. Human Factors Science and Safety Engineering. Can the STAMP Model Serve in Establishing a Common Language? 32nd EAAP Conference, 132-149.
27. Kelly, T. & Weaver, R. 2004. The Goal Structuring Notation – A Safety Argument Notation. York: University if York, De-partment of Computer Science and Department of Man-agement Studies.
28. Klanac, A., Varsta, P. 2011. Design of marine structures with improved safety for environment. Reliability Engineering & System Safety 96(1):75-90.
29. Kristiansen, S. 2005. Maritime transportation: Safety management and risk analysis. Elsevier Butterworth-Heinemann, 508p.
30. Lehikoinen, A., Luoma, E., Mäntyniemi, S., Kuikka, S. 2013. Optimizing the recovery efficiency of Finnish oil combating vessels in the Gulf of Finland using Bayesian Networks. Environmental Science & Technology 47:1792-1799.
31. Leveson, N. G. (2011). Engineering a Safer World - Systems Thinking Applied to Safety. Cambridge, MA: MIT Press.
32. Leveson, N. G., & Thomas, J. P. (2018). STPA Handbook. http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf
33. Lloyd's Register. 2014. Guidance Notes for Technology Qualification. London: Lloyd's Register.
34. Mazaheri, A., Montewka, J. Nisula, J., Kujala, P. 2015. Usability of accident and incident reports for evidence-based risk modeling - A case study on ship grounding reports. Safety Science 76:202-214.
35. Merrick, J.R.W., van Dorp, J.R., Mazzuchi, T., Harrald, J.R., Spahn, J.E., Grabowski, M. 2002. The Prince William Sound Risk Assessment. Interfaces 32(6):25-40.
36. Meyer, T., Reniers, G. 2016. Engineering Risk Management. De Gruyter Graduate, 284p.
37. Montewka, J., Goerlandt, F., Kujala, P. 2014. On a systematic perspective on risk for formal safety assessment (FSA). Reliability Engineering and System Safety 127:77-85.
38. Neves, A.A.S., Pinardi, N., Martins, F., Janeiro, J., Samaras, A., Zodiatis, G., De Dominicis, M. 2015. Journal of Environmental Management 159:158-168.
39. Papanikolaou A. 2009. Risk-based ship design: Methods, tools and applications. Springer-Verlag Berlin Heidelberg, 376p.
40. Porathe, T., Prison, J., & Man, Y. (2014). Situation awareness in remote control centres for unmanned ships. In Human Factors in Ship Design & Operation. London.
41. Psarros, G., Skjong, R., Eide, M.S. 2010. Under-reporting of maritime accidents. Accident Analysis & Prevention 42(2):619-625.
42. Schrader-Frechette K.S. 1991. Risk and rationality: philosophical foundations for populist reforms. Berkeley: University of California Press.
43. Sormunen, O.-V., Hänninen, M., Kujala, P. 2016. Marine traffic, accidents, and underreporting in the Baltic Sea. Scientific Journals of the Maritime University of Szczecin 46(118):163-177.
44. SRA. 2015. Society for Risk Analysis Glossary, Available online: http://www.sra.org/resources [Accessed 14.06.2018]
45. Ståhlberg, K., Goerlandt, F., Ehlers, S., Kujala, P. 2013. Impact scenario models for probabilistic risk-based design for ship-ship collision. Marine Structures 33:238-264.
46. Stringfellow, M.V., Leveson N. and Owens B. 2010. Safety-Driven Design for Software-Intensive Aerospace and Automotive Systems. Proceedings of the IEEE 98(4):515–25.
47. Thompson, P.B., Dean, W. 1996. Competing conceptions of risk. Risk Health Safety and Environment 7:361-384.
48. Uluscu, Ö.S., Özbas, B., Altiok, T., Or, I. 2009. Risk analysis of the vessel traffic in the Strait of Istanbul. Risk Analysis 29(1):1454-1472.
49. Underwood, P., Waterson, P., & Braithwaite, G. (2016). ' Accident investigation in the wild ' – A small-scale , field-based evaluation of the STAMP method for accident analysis. Safety Science, 82, 129–143.
50. Valdez Banda, O.A., Goerlandt, F., Kuzmin, V., Kujala, P., Montewka, J. 2016. Risk management model of winter navigation operations. Marine Pollution Bulletin 108:242-262.
51. Valdez Banda, O.A., Goerlandt, F., 2018. A STAMP-based approach for designing maritime safety management systems. Saf. Sci. 109:109-129.
52. Vanem E., Skjong R. 2006. Designing for safety in passenger ships utilizing advanced evacuation analyses - A risk-based approach. Safety Science 44(2):111-135.
53. Weng, J., Meng, Q., Qu, X. 2012. Vessel collision frequency estimation in the Singapore Strait. Journal of Navigation 65:207-221.
54. Wróbel K., Montewka J., Kujala P. 2018. "Towards the Development of a System-Theoretic Model for Safety Assessment of Autonomous Merchant Vessels." Reliability Engineering & System Safety, June. Elsevier. doi:10.1016/J.RESS.2018.05.019.
55. Wróbel K., Montewka J., Kujala P. 2017. "Towards the Assessment of Potential Impact of Unmanned Vessels on Maritime Transportation Safety." Reliability Engineering & System Safety 165 (September): 155–69. doi:10.1016/j.ress.2017.03.029