Wróbel, Krzysztof; Montewka, Jakub; Kujala, Pentti

Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels

Published in:
Reliability Engineering and System Safety

DOI:
10.1016/j.ress.2018.05.019

Published: 01/10/2018

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

Published under the following license:
CC BY-NC

Please cite the original version:
Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels

Krzysztof Wróbel\textsuperscript{a,b,⁎}, Jakub Montewka\textsuperscript{b,c,d}, Pentti Kujala\textsuperscript{b}

\textsuperscript{a} Faculty of Navigation, Department of Navigation, Gdynia Maritime University, Jana Pawła II av. 3, 81-345 Gdynia, Poland
\textsuperscript{b} Department of Mechanical Engineering, Marine Technology, Research Group on Maritime Risk and Safety, Aalto University, Tietotie 1C, 02150 Espoo, Finland
\textsuperscript{c} Faculty of Navigation, Department of Transport and Logistics, Gdynia Maritime University, Marszałka Sosnowskiego 81-87, 81-225 Gdynia, Poland
\textsuperscript{d} Finnish Geospatial Research Institute, Geodeetinrinne 2, 02430 Masala, Finland

\begin{abstract}
As the initiatives to develop and implement autonomous merchant vessels into the global shipping industry are gaining momentum, their safety remains in the spotlight. It is argued that every effort shall be taken to ensure that the safety of maritime transportation is not reduced in the process, but the question of how to achieve it remains open. Meanwhile, the systemic approach is more widely being used to analyse innovative systems’ safety. We therefore apply a System-Theoretic Process Analysis to develop a model suitable for safety analysis and design recommendations' elaboration for future autonomous vessels. Furthermore, we introduce a method of evaluating and communicating uncertainties pertaining to the method. The results indicate that the system-theoretic safety analysis’ outcome can be affected by manageable uncertainties despite the fact that the system in question is yet to be implemented.
\end{abstract}

\section{Introduction}
Recent R&D projects have investigated the feasibility of implementing a merchant vessel which would traverse the ocean without having any crew on board or even being controlled remotely. The results of such projects were encouraging, resulting in concepts and models attempting to quantify the safety of autonomous maritime navigation, see for example [1–3]. A body of literature focuses on safety quantification of prospective autonomous ships adopting risk assessment techniques that employ causal models [4]. To this end the Formal Safety Assessment has been utilised [5], resulting in the identification of major hazardous scenarios the unmanned ships can induce and their initial assessment in terms of safety as well as potential risk control options [4–7]. Although the apparent lack of data has been acknowledged, it was concluded that risk analysis is in favour of unmanned ships being generally safer than manned ones, provided that certain safety precautions are fulfilled [2]. Moreover, a general overview of potential failure propagation in case of an accident based on Bayesian framework is given in [8], arguing that particular aspects of a ships’ operations, such as navigation, stability or cargo conditioning are mutually related in a variety of ways and must not be analysed separately [9].

However, the proposed risk-based approaches feature several shortcomings. Firstly, they require empirical data, which are non-existent since the autonomous shipping is still in the development phase. Secondly, the modelling techniques adopted therein do not allow for the detailed analysis of potential interactions between system’s components but often assume simplified, causal, one-way relations instead. Thirdly, the safety is seen as a variable to be quantified rather than a feature to be controlled. This implies the limited applicability of the existing approaches to define the measures to effectively control the safety of the prospective autonomous ships. Therefore, the issue of ensuring such vessels’ safety remains open, as argued in [8,10,11]. As the aforementioned vessels’ expected implementation is a matter of years rather than decades [12], certain steps must be taken in order to ensure that the safety of marine transportation is not compromised in the process.

Thus another approach is needed to allow for the safety-driven design of the prospective autonomous merchant vessel’s (AMV) system, [13]. The method should be able to examine it holistically, mimicking all relevant interactions between its components and surrounding environment. Within the maritime domain quantitative methods prevail, pertaining to ship and waterways design, or accident response, see for example [14–18]. Rarely, qualitative methods to evaluate safety are used, [19,20]. The former requires numbers and quantifiable parameters, which often may be missing, or unknown. The latter allows the
incorporation of non-quantifiable (or difficult to quantify) factors such as organisational issues or human performance, [21].

Therefore, in this paper we delve into the safety of AMVs by applying the System-Theoretic Process Analysis (STPA), a tool rooted in System-Theoretic Accident Model and Processes (STAMP). STPA is a method of safety analysis that has been developed to elaborate on design recommendations for innovative technical systems, where safety is seen as a control problem, rather than an object of quantification [22].

The core of system-theoretic methods is to analyse interactions between a system’s components and ensure that these remain safe rather than focusing on the reliability of every single component. Such an approach is believed to better encompass potential hazards and help create feasible measures to mitigate them [22]. The method has been used for the safety assessment of systems of varying natures like vessels’ traffic management [23,24], automated driving vehicles [25] and offshore supply vessel dynamic positioning [26]. Despite various authors claiming that the system-theoretic approach delivers good predictions [21,27] of systems’ safety performance in the presence of limited knowledge regarding their actual design, little attention has been devoted to the evaluation of potential uncertainties which exist in the process of safety recommendations’ development and communication to a decision-maker, see for example [15].

The aim of this paper is two-fold. Firstly, it applies STPA to develop a model of autonomous ship’s safety during high-seas operations and to elaborate on safety recommendations for future developers of such a system. Secondly, we introduce a simple solution to assess and communicate the uncertainties pertaining to the safety control model developed. The latter supports the former, with the intention of providing the decision-maker, see for example [15].

By focusing on the interactions between the system’s components or their groups rather than a particular segments’ reliability, the developed model advocates various ways to control the safety: technical, organisational and operational patterns, providing the end-user with a set of hazard mitigation measures. Furthermore, the conducted uncertainty assessment gives the system developers a preliminary insight into the expected effectiveness of those mitigation measures, also informing about the areas of the system that call for more thorough investigation.

The proposed model can be used by various stakeholders, such as system developers to incorporate the holistic safety approach to ship design, ship operators to develop safer operational procedures and maritime administrations to facilitate the process of rule-making for the AMVs’ safety.

The paper is structured as follows: firstly, the materials and methods are described, including the System-Theoretic Process Analysis and prospective uncertainties assessment. Section 3 introduces the results of the study, which are then discussed in Section 4. Last but not least, the conclusions are drawn.

2. Materials and methods

2.1. Autonomous vessels’ concept

Recently completed research projects concluded that, from the technical and economical point of view, the implementation of autonomous cargo ships can be feasible [2,5], although some legal issues must first be resolved [28,29]. Throughout the projects’ deliveries [1,4,30,31] and in the increasing number of scientific papers based on the former, the general vision of an unmanned vessel is consistent [10,32–34], but the actual shape of the system remains in fact unknown. Therefore, we based our study on the literature review of the available sources pertaining to autonomous navigation and then on the experts’ opinions. The general view emerging from these is given in this Section.

Although unmanned by design, such vessels might be required to unberth and leave the harbour waters under a direct control of a ‘conning crew’, due to the relative complexity of such manoeuvres [35,36] and port authorities’ likely reluctance to accept unmanned vessels’ operations in restricted waters [37]. The vessel would therefore need to accommodate a crew of a size comparable to today’s ones for a limited period of time. Such crew would need to be capable of controlling the vessel and most of her equipment in a similar way to today’s.

The ‘conning’ crew would then disembark the vessel by a launch boat or helicopter as soon as she leaves the port and it is considered safe to leave her, and leave her under the supervision of a shore-based operator. Such a person would assume an overall command and navigate the vessel from the office-like facility located ashore. Both-way communication would most likely be provided by the satellite communication link, perhaps augmented by other means of short-range data transfer, if applicable [38]. The feasibility of such a solution has been to some extent proven in August 2017 when an Offshore Supply Vessel performed several operations in the North Sea while being remotely controlled from California [39]. The operator would make system-level decisions based on data received from the shipborne sensors. Both data and decisions would be relayed to the vessel by a communication link [40] and executed by actuators including thrusters and rudders. Nevertheless, certain fail-to-safe mechanisms would still be required in order to maintain system’s safety should the communication link fail for any reason.

The vessel would in such case be left on her own and would need to handle the situation autonomously. This might require going dead in the water or navigating to a safer area [6]. Since the autonomous navigation mode needs to be built-in to handle such emergencies, it can also be used to a greater extent – for the whole process of navigation itself. Herein, as soon as the vessel leaves high traffic density areas, operator’s attention could be directed to other unmanned ships facing more difficult conditions. The level of ship’s autonomy might be increased up to the point where she would require no more attention than a periodical check. Information provided by sensors would be analysed by highly sophisticated data fusion algorithms in order to automatically create decisions regarding virtually all aspects of navigation, cargo conditioning, machinery operations, stability and any other aspect of a vessel’s operation [33]. The human operator will retain a position of merely a passive supervisor, person in charge of strategic decision-making (i.e. general passage planning) or a trouble-shooter. It is the vessel’s control algorithms that will be responsible for making operational decisions and performing routine tasks.

Although the vision might look tempting, there are some social [41], legal [28,42] and technical issues that need resolving. The main source of potential problems is the performance of control algorithms and, as within a remote control mode, inability to manually operate any of the vessel’s equipment [11]. Nevertheless, this mode could be used until the condition is detected that would require more direct intervention of the human operator which would, by definition, need to be taken and executed remotely as there will be nobody on board. As soon as the situation is resolved, the system might switch back to the autonomous mode. With the vessel successfully reaching the port of destination, the ‘conning crew’ might be required again for berthing. After the cargo transfer or any other operations are completed, the cycle could repeat.

By that, the vessels are anticipated to follow an ‘adjustable autonomy’ scheme depending on the condition of the ship herself and a mission being executed. Expected levels of autonomy in the maritime industry as elaborated by Lloyd’s Register of Shipping [43] are listed in Table 1 below, although other frameworks have also been developed [44,45].

Fully autonomous mode of unmanned ship’s operation (AL-5) is expected to be the primary and the most extensively used one, particularly for ocean crossings. Autonomous operation can help exploit the
full potential of unmanned shipping by involving humans in the process to only a very limited extent and using automated systems instead. It is also the most challenging of all modes as less scientific and technical data is available to properly assess its actual feasibility and safety. The recent publication of numerous scientific papers on unmanned shipping does not change the fact that empirical data is required to validate statements contained within. Data pertaining to the safety of autonomous transportation is only available for other domains such as automotive or underwater [46,47]. It can, however, be considered incomplete as most of the technologies are still in their early stages of development.

In this paper, we focus on AL-5 and analyse the safety of vessel operating in this mode, using the methods presented in Sections 2.2.1–2.2.3. Security and cybersecurity issues in general remain outside of this study's scope, although it must be understood that they might pose a significant threat to system's integrity and negatively affect its safety in multiple ways [4,5].

### 2.2. Methods

Methods used to perform the safety assessment are described throughout this Section. Sections 2.2.1 and 2.2.2 present a brief description of System-Theoretic Process Analysis as well as a method of modelling the system, while Section 2.2.3 gives an overview of a mitigation potential elaboration. Section 2.2.4 in turn presents a method of assessing and communicating the uncertainties related to the safety analysis.

#### 2.2.1. System-Theoretic Process Analysis – STPA

STPA is a method of examining a given system's safety by analysing the interactions between its components [26] and the ways in which those can be unsafe [22]. The nature of such interactions shall ensure that the system as a whole remains within safety limits [48,49]. As a consequence of the above, any violation of the defined safety constraints leads to the emergence of a hazard (a system state or set of conditions that, together with a particular set of worst-case conditions, will lead to an accident). It is recommended to refrain from calculating the probabilities of a system transitioning to an unsafe state [15] due to lack of empirical data, particularly in initial phases of the system development [22].

Being rooted in STAMP, STPA shares all of its major features, both on advantages' and drawbacks’ sides. As for the former, the underlying assumption of probability-based thinking being not suitable for the comprehensive analysis of today's modern and complex systems is the major one [22]. Therefore, the interactions and mutual relationships between the system's components are studied instead of its reliability structure. For instance, system-theoretic analysis of the 'Sewol-Ho' ferry sinking helped identify numerous contributing factors that could be overlooked when analysing the ferry's reliability structure [50]. Evoking the classic Swiss cheese model [51], system-theoretic methods strive to keep the cheese slices in proper positions in relation of one to another rather than ensuring that their holes are sufficiently small. The latter approach, however, must not be neglected as the reliability remains one of the means of ensuring safety [52].

#### 2.2.2. Safety control structure elaboration and analysis

In a generic control process, as depicted in Fig. 1, system integrity depends on ensuring that interactions between components do not lead to safety constraint violations. The study follows the most frequently used classification of causal factors, where four potential ways of inadequate control can be distinguished (although some argue that this number can be increased to six, see for Ref. [49]):

(a) a control action required for safety is not provided or not followed;
(b) an unsafe control action is provided;
(c) a potentially safe control action is provided at the wrong time or in the wrong sequence;
(d) a control action required for safety is stopped too soon or applied too long [22,49,53].

As a preparation for the STPA, a model of the system's safety control structure, depicting mutual relationships between system's components, is created. This was achieved by reviewing the available literature pertaining to AMVs, see for instance [1,10,31,34,37,54-57] and a Delphi-based workshop with a wide selection of experts involved in the research project resulting in the design concepts of several unmanned vessels, including coastal ferry, costal container carrier and middle size, sea-going container carriers. The group of 15 experts was comprised of

![Fig. 1. A generic control loop, as given in [22].](image-url)
ship designers, naval architects, ship operators (navigators and captains), traffic controllers, representatives of the maritime authorities and researchers in various field of technical sciences (mechanical, telecommunications and electrical engineering, IT, maritime studies, law, naval architecture, system safety).

Similarly, a list of hazards (see Table 3) was created in order to systematise knowledge regarding the safety of an autonomous generic merchant vessel and identify potentially hazardous conditions that may be encountered by the system during its operations.

Thence, control loops within a safety control structure were investigated and a potential for inadequate control was sought. In the next step, each control action was examined with respect to above potential ways of inadequacy. Components involved and failure scenarios were identified and ways of mitigating the potential for inadequacy – recommended [21].

Such ‘generic’ ship is defined as not having any particularly demanding requirements pertaining to cargo stowage. This was based on the assumption that the autonomous shipping technology will first be tested on vessels carrying commodities not requiring a complicated conditioning. If the prototype tests prove the technology’s feasibility, more challenging cargoes could be accommodated by augmenting the design with new functionalities [30]. Bulk carriers, general cargo or container vessels can be good candidates for the prototype [58,59]. Although they can be technologically advanced even nowadays, their cargo conditioning equipment and technology as well as safety features are far simpler than those of tankers or passenger vessels, for instance. The notable exemption can be the river-crossing ferry as its mission’s relative simplicity and being close to river banks at all times also makes such a vessel a good candidate for the prototype of an autonomous ship [1].

As the entire concept of an autonomous vessel capable of crossing oceans is still at a relatively early design phase as this paper is being written, some vital information pertaining to the system’s actual shape can be lacking or incorrect. Therefore, the paper should be considered as merely an initial insight and elaboration of basic safety recommendations rather than a complete and final safety assessment. This is in line with the concept of safety-guided design, a process of an iterative cooperation between system developers and safety analysts [22]. Such recommendations are presented in the form of mitigation measures and evaluated by an assignment of a ‘mitigation potential’ value.

2.2.3. Mitigation potential analysis

Instead of calculating the probability of a hazardous event, a mitigation potential can be evaluated in a systemic approach, a parameter describing the effectiveness of a particular action (a mitigation measure in other words), aiming to restrict the accident’s likelihood or consequences. To this end, the following mitigation potential scale is used:

1. reduction of damage if an accident does occur;
2. reduction of the likelihood that the hazard results in an accident;
3. reduction of the likelihood that the hazard will occur;
4. complete elimination of the hazard from design [22].

The design process will involve safety-driven optimisation of a system aiming primarily at the reduction of an accident’s likelihood and then in confining its consequences. Thereby, it can be understood as searching for and implementing hazard control measures having higher mitigation potential assigned. Those with greatest mitigation potential are viewed as being more efficient and cost-effective when it comes to accident prevention and, in the worst-case scenario, damage reduction.

In our study, we reviewed the available literature in order to find all potentially feasible mitigation measures and recommend them as a protection against particular control action becoming inadequate. These measures were listed and their theoretical effectiveness was evaluated in the form of mitigation potential. As a result, we quantified the recommended measure’s capability of ensuring that the particular control action remains adequate. This was done instead of calculating the mitigation measure’s potential of preventing a hazard from leading to an accident and was caused by a low-detail level of the developed model.

Furthermore, we augmented our recommendations’ elaboration by the analysis of uncertainties.

2.2.4. Uncertainty assessment and communication

Kaplan and Garrick claim that the very purpose of risk analysis is to provide an input to the underlying decision making [60]. It is therefore the obligation of analysts to consider the consequences of their error, which can only be done if the uncertainties pertaining to the study’s results are identified and assessed. Wrong or weak assumptions, poor data or unreliable models may lead to unjustified conclusions within the safety assessment and wrong decisions [61-65]. With the presence of important uncertainties, decision-makers may justifiably opt for additional (and potentially superfluous) protective measures while accepting the increased costs [61].

The necessity of including the uncertainty analysis in safety assessments is therefore becoming widely acknowledged with different approaches applied [66-68]. Such a requirement was also raised with regard to system-theoretic methods of safety assessment [15]. As argued, STAMP and related tools help reduce the uncertainties by themselves as they offer a more insightful look into the system behaviour [21]. However, they must not be considered a perfect tool that eliminates the uncertainties completely. These will still exist on virtually all steps of STPA as depicted in Fig. 2. One of the main reasons for this situation is that the STPA is very often used to assess the safety of innovative endeavours, just as is the case of an unmanned merchant vessel. There are two potential solutions of this situation: (1) reduce the uncertainty by better modelling the system or (2) characterise uncertainty better [15]. The former can be difficult at the present stage of technological development as the only information pertaining to future autonomous ships’ system design can be extracted from the scientific and professional literature or elicited from experts involved in the works. This has been done in the course of the present study. There is no guarantee, however, that the AMV’s design will not dramatically change prior to implementation. On the contrary: the very purpose of the safety-driven design is to identify and suggest potentially beneficial revisions in order to improve future system’s safety performance. It is therefore the only choice to communicate the uncertainties to the decision-makers so as they could make informed decisions.

Fig. 2. Uncertainties’ sources when using STPA.
Since the hazard mitigation measures’ effectiveness remains the most important of the systems’ features for designers, operators and maritime administrations, we focus on this stage of safety analysis and develop a method of assessing uncertainties pertaining to it.

The view that system-theoretic methods can be augmented by other frameworks in order to analyse the uncertainties existent therein was presented in [69]. One of the first attempts to include an uncertainty analysis in STAMP was then given in [15] where the strength of knowledge supporting the analysis was postulated as the most important factor to be included. In order to expand this approach, we modified the ‘degree of uncertainty’ scale as described in [68] and further polemised in [61]. The modification was carried out as a response to the on-going discussion in academia [70,71].

For each mitigation measure elaborated as per the framework given in Section 2.2.2, available information related to it has been assessed in five categories: understanding of phenomena, accuracy of the model, viability of assumptions made, availability of data, strength of consensus among experts. The framework is presented in Table 2. As seen, the assessment is qualitative in nature and was performed by the analysts themselves in the course of subjective assessment. The uncertainty in this case can be defined as an analysts’ degree of belief that the elaboration of a particular mitigation measure is supported by sufficient data, assumptions etc., and that the mitigation measure in question is feasible to implement.

After a mitigation measure has been identified as potentially valuable in ensuring adequacy of particular control action, additional information about it was collected and reviewed from available literature particularly in the field of unmanned shipping, autonomous operations and maritime transportation. Especially, the following has been assessed:

- Phenomena – what is the level of understanding of the mitigation measure’s functionalities? If it was successful in ensuring safety when applied in other domains, can the same be achieved in autonomous shipping?
- Model – is the model of a given mitigation measure’s interactions with the system available? Are the consequences of implementing the particular mitigation measure in autonomous shipping well-comprehended?
- Assumptions – do assumptions supporting the implementation of a mitigation measure have a strong basis?
- Data – is the empirical data addressing application of a mitigation measure published in a variety of sources? Are the results conclusive?
- Consensus – do authors of scientific and professional publications agree on the feasibility of a given mitigation measure? Is it mentioned as a potential solution in a considerable number of sources?

Based on the answers to above questions, an uncertainty level in each of five categories has been assigned to all the mitigation measures. This could be significant, moderate or minor. Thus, a subjective level of analyst’s confidence in the feasibility of particular mitigation measure is communicated.

Results of the above steps of autonomous merchant vessel’s safety assessment are presented in Section 3 and discussed in Section 4.

3. Results

This Section presents the results of analysis: the safety control structure of the autonomous ship, list of hazards as well as uncertainties pertaining to the mitigation measures’ elaboration.

3.1. Safety control structure of the autonomous ship

The autonomous vessel’s high-level safety control structure is presented in Fig. 3. Herein, the most discernible component is a ‘Virtual Captain’ (VC), a computer controlling all on-board equipment and processes. Data is fed by environmental sensors (those measuring parameters of the environment, i.e. radar with Automatic Radar Plotting Aid, Automatic Identification System, infra-red cameras, echo-sounder, log, gyrocompass, Global Navigation Satellite System receivers etc.) as well as internal ones (i.e. rudder angle indicator, main engine status indicators, tank gauges, fire sensors) [34]. VCs main objective is to ensure that the vessel follows the prepared passage plan, reaches the port of destination within the assumed time and without causing any hazard to herself, other assets, humans or environment.

Based on the information received and the ship’s control model, control algorithms formulate decisions, which are then executed by actuators in order to control shipborne processes. These actuators can include mechanisms of a diverse nature such as steering pumps, fuel system valves, fog horn and fire extinguishing system. Virtually all aspects of the ship’s operation must be controlled for a prolonged time without any involvement of human operators except periodical conditions check. Those aspects involve ‘Navigation’ (meaning vessel’s course and speed) and a large number of ‘Auxiliary processes’, the aim of which will be to ensure the vessels’ optimum performance and safety. These are not addressed individually as their list and characteristics would depend on the actual system’s design. Instead, they are only referred to in general.

Such an arrangement of the system would last until safety parameters exceed their limits. If that is the case, the VC would use the satellite communication link to call for the operator’s assistance and switch the entire system to a lower level of autonomy, e.g. AL-3, see Table 1. Such limits shall be adjusted by human operators, which is to be done via communication link.

The VC might be capable of coordinating certain actions with other
unmanned ships in the vicinity, but ways in which other third parties could influence vessel’s behaviour should be limited for security reasons. For instance, coastal states’ administrations ought to be capable of advising certain actions to be taken by the vessel, but the final decision should be the responsibility of the operators [31]. Similarly, the latter should have a convenient way of contacting the coastal states’ authorities in order to perform administrative work or coordinate certain actions. Search and Rescue operations can serve as a good example of these.

Furthermore, the autonomous vessel will operate within a certain organisational and legal framework as shaped by today shipping industry’s architecture. That will require following international regulations and rules for classification as well as cooperating with external organisations. The system’s position within such a framework is yet to be clarified [28,29].

The ‘system’ of an autonomous ship can therefore be defined as below:

‘all technical, organisational and human-based arrangements purposely designed or utilised in order to perform a safe navigation of a sea-going vessel operating autonomously’

The defined system will consist of each component that has been either designed or can be intentionally used as its part. That would include not only the ship itself, but also the shore-based control centre (SBCC), software, hardware and liveware involved, operational procedures and legislation. In other words, everything on which system’s designers can have certain degree of control.

The natural environment and ships other than unmanned will to a large extent remain outside of the system, thus will be generally referred to as the ‘environment’. This will, unfortunately, also include illegal activities. These however remain out of this analysis’ scope.

After the safety control structure as well as hazards’ and constraints’ list had been created, we performed the actual STPA. Each of the control actions was investigated in order to identify potential consequences and causes of it being inadequate. Then, we sought mitigation measures that might reduce the potential for such inadequacies. These measures were assigned the mitigation potential.

Results of the above are presented in Section 4 as well as in Appendices.

3.2. STPA

Based on the system’s safety control structure presented in the former Section, a list of hazards and related safety constraints was compiled as given in Table 3.

As can be seen, an occurrence of certain hazards may propagate the emergence of others. Failure of propulsion, for instance (Hazard #2.2.) may lead to vessel’s grounding (#2.1), then loss of structural integrity (#2.6.). Therefore, mitigation measures capable of protecting against multiple hazards simultaneously can be characterised by greatest effectiveness.

The hazards’ list was then used as an aid in performing the actual STPA. In its course, a total of forty-eight control actions have been analysed with respect to their position within the system structure, potential scenarios leading to their inadequacy and consequences of such. Furthermore, potential ways of mitigating such inadequacies were elaborated and evaluated by the assignment of the mitigation potential. A total of 252 recommendations on mitigation measures’ implementation have been elaborated, each of them pertaining to one of three classes: covering liveware, software or hardware. By ‘liveware’ we understand all organisational, legal and operational factors in which a human plays a major and direct part.

The catalogue of control actions together with the results of STPA is presented in Appendices.

3.3. Uncertainties

Unfortunately, the process of elaborating recommendations on mitigation measures’ implementation is burdened with some uncertainties. These have been assessed in line with the method given in

Fig. 3. Autonomous vessel’s system safety control structure. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)
Table 3
List of high-level system hazards and safety constraints. Partly based on [4,11,72]. Repetitive hazards have been crossed out and omitted in further steps.

<table>
<thead>
<tr>
<th>#</th>
<th>Description of hazard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vessel’s physical interaction with manned structures results in death or injury</td>
</tr>
<tr>
<td>2</td>
<td>Vessel violates minimum CPA with another ship</td>
</tr>
<tr>
<td>3</td>
<td>Vessel enters a No Go Area</td>
</tr>
<tr>
<td>4</td>
<td>Vessel improperly interacts with other man-made structures</td>
</tr>
<tr>
<td>5</td>
<td>Vessel is incapable of containing dangerous chemicals or energy</td>
</tr>
<tr>
<td>6</td>
<td>Vessel is boarded by unauthorised personnel or such commodities are placed on board</td>
</tr>
<tr>
<td>7</td>
<td>System does not provide assistance to person in distress</td>
</tr>
<tr>
<td>8</td>
<td>Vessel’s inability to reach port of destination in expected time</td>
</tr>
<tr>
<td>9</td>
<td>Vessel’s interaction with other assets (including unmanned vessels) holds back other ships’ traffic</td>
</tr>
<tr>
<td>10</td>
<td>Vessel is incapable of properly containing dangerous chemicals or energy</td>
</tr>
<tr>
<td>11</td>
<td>Vessel does not meet stability criteria</td>
</tr>
<tr>
<td>12</td>
<td>Vessel’s watertight integrity is not maintained (due to shear forces, bending moments or puncture)</td>
</tr>
<tr>
<td>13</td>
<td>Vessel’s inability to deliver cargo in unchanged condition or in a condition that falls within industry standard</td>
</tr>
<tr>
<td>14</td>
<td>Vessel’s cargo is not loaded/stowed properly</td>
</tr>
<tr>
<td>15</td>
<td>Vessel is unable to maintain proper cargo stowage conditions</td>
</tr>
<tr>
<td>16</td>
<td>Vessel’s exposure to major damage or breakdown</td>
</tr>
<tr>
<td>17</td>
<td>Vessel violates minimum CPA with another ship.</td>
</tr>
<tr>
<td>18</td>
<td>Vessel does not meet fire safety precautions</td>
</tr>
<tr>
<td>19</td>
<td>Vessel’s watertight integrity is not maintained.</td>
</tr>
<tr>
<td>20</td>
<td>Vessel’s power supply is not provided or insufficient</td>
</tr>
<tr>
<td>21</td>
<td>Both-way communication with the vessel cannot be established</td>
</tr>
<tr>
<td>22</td>
<td>Vessel’s inability to prevent environmental pollution</td>
</tr>
<tr>
<td>23</td>
<td>Vessel’s ability to maintain integrity of tanks containing oils or oily mixtures</td>
</tr>
<tr>
<td>24</td>
<td>Vessel is unable to maintain proper fuel combustion parameters</td>
</tr>
<tr>
<td>25</td>
<td>Vessel is incapable of properly containing dangerous chemicals or energy</td>
</tr>
<tr>
<td>26</td>
<td>Vessel’s interaction with third-party assets causes reduction of their value or operational abilities</td>
</tr>
<tr>
<td>27</td>
<td>Vessel violates minimum CPA with another ship, runs into element of infrastructure or damages other man-made objects</td>
</tr>
<tr>
<td>28</td>
<td>Vessel contributes to delay of other ships’ traffic</td>
</tr>
<tr>
<td>29</td>
<td>System does not meet international, classificatory or national regulations</td>
</tr>
<tr>
<td>30</td>
<td>System’s communication subsystem unintentionally interferes with other assets</td>
</tr>
<tr>
<td>31</td>
<td>System’s interaction with other assets (including unmanned vessels) leads to the emergence of any of above</td>
</tr>
</tbody>
</table>

Section 2.2.4 and are presented within the Appendices where small symbols are placed as a reference to Table 2. Therein, a level of uncertainty in each of five categories is expressed for every mitigation measure. Grey shading within the symbols indicates that for each of five rows (corresponding to the uncertainty categories: Phenomena, Model, Assumptions, Data, Consensus in this order top to bottom), uncertainty has been assessed as either Significant, Moderate or Minor (in this order, left to right), see Table 4.

These uncertainties have been summarised in Fig. 4, where a traffic light symbolism is utilised to describe uncertainties related to the mitigation measures belonging to one of three classes (liveware, software or hardware) and five major portions of the system: organisational environment, shore facility, communication, vessel and her direct environment (horizontal axis, also indicated by background colour in Fig. 3). Therein, red represents the number of instances in which significant uncertainty has been assigned to the process of elaborating given the mitigation measure. The latter pertains to the relevant portion of the system and the mitigation measures’ class. Similarly, green represents minor uncertainties whereas yellow denotes their moderate level.

In Figs. 5–7, the breakdown of uncertainties’ magnitudes for each of the three classes are depicted in more detail, taking into account the categories of uncertainty (phenomena, model, assumptions, data and consensus).

3.4. Case study

The application of the presented method is demonstrated through case studies. Two selected control actions, out of 48 that exist in the proposed model of autonomous ship safety, are analysed here.

3.4.1. Analysis of control action #31 Environment probing

#31: Environment probing consists of gathering environmental data by autonomous vessel’s sensors in order for the VC to create a situational awareness [73]. Those sensors can include GNSS receiver, radar, echosounder, log, infra-red camera, anemometer etc. The importance of this control action is based on the consequences of its inadequacy: should sensors fail to gather the data pertaining to weather and other ships’ traffic, the VC would become ‘blind’ and will not be capable of performing the navigation process safely and efficiently. This might cause improper decisions to be made and sent to the propulsion subsystem. It has been assessed that this could lead to the emergence of as many as thirteen different hazards as defined in Table 3.

Such inadequacy (failure to observe environmental conditions) can be caused by a variety of factors that have been identified. Those include sensors’ failures, installed sensors’ inability to measure a required feature, unsuitable sensors being installed or their sub-optimal performance. To counteract the above, the following mitigation measures have been elaborated:

1. implementation of redundancy or development of highly-reliable sensors,
2. use of sensors capable of measuring multiple features simultaneously (just as GNSS receiver can provide data pertaining to its position, speed and course over ground),
3. development and implementation of highly sensitive sensors with reduced sampling time.

All the above pertain to hardware solutions and interaction of the system with the environment. They are also intended to reduce the likelihood of the relevant control action’s inadequacy occurrence; therefore, they were assigned a mitigation potential value of ‘3’, as given in Section 2.2.3. As for protection against control degradation, the development of improved sensors can be named along with implementation of leading indicators detecting the worsening
performance of a particular sensor and prompting for its immediate replacement \[53,74\] (which would only be possible once the vessel calls at a port of convenience \[75\]).

Thence, uncertainties pertaining to the mitigation measures’ elaboration have been qualitatively and subjectively assessed based on literature review, sometimes pertaining to other domains than autonomous shipping, for which more experience has been gained in recent years and more information is available. Factors as listed within Table 5 have been taken into consideration while assessing the uncertainty.

3.4.2. Analysis of control action #21 Regulation

#21: Regulation (of auxiliary processes) incorporates control imposed upon phenomena not immediately related to the ship's movements. Those can include a variety of processes ranging from exhibition of navigational lights through the operation of AMV’s ballast system. Such diversity issues to be regulated by numerous types of equipment can lead to the emergence of many hazards, but also calls for multiple solutions and different measures of their mitigation. These vary from design-based and procedural to hardware.

Table 6 contains factors taken into consideration when evaluating

---

*Fig. 4.* A high-level breakdown of the uncertainties by class of mitigation measure and control action's position within the system.

*Fig. 5.* A detailed breakdown of uncertainties by categories for autonomous vessels' liveware solutions.
the level of uncertainties pertaining to particular hazard mitigation measures. The rationale behind the latter can be summarised as follows:

- Rigorous maintenance regime: since equipment is to operate in a prolonged maintenance-free mode, upkeep must be performed strictly as required to satisfy operational needs.
- Redundant equipment: certain system’s functions can be performed by secondary (spare) machinery should the primary one suffer from malfunction.
- Resilience-based design: the system shall retain ability to perform its basic (life-sustaining) operations in all circumstances for a period of time required for intervention, this can be achieved through resilience engineering.
- Procedures on consumables’ management: as various resources (fresh water, lubricants, oils etc.) can be required for machinery to operate, these must be available and procedures aiming in their supply shall be implemented.
- Capacity surpluses: since it may turn out that system meets previously unrecognised and demanding operational parameters, equipment shall be capable of flexibly adapting to such, for instance by operating with above-nominal capacity.
- Extensive testing: all machinery shall be thoroughly tested so as to demonstrate its fitness and interoperability with the rest of the system.
- Implementation of leading performance indicators: potentially dangerous conditions could be detected before they actually occur by implementation of leading indicators, measuring latent anomalies in given subsystem’s performance.

Herein, all mitigation measures except ‘resilience-based design’ are intended for the reduction of a potential hazard’s likelihood of occurrence.

Results of the performed analysis are presented mostly in Appendices for each control action as given in Fig. 3 and discussed throughout Section 4.2.

4. Discussion

Results obtained within the autonomous vessel's system preliminary safety assessment are discussed in Section 4.1. Thence, Section 4.2 elaborates on uncertainties pertaining to the former and the potential ways of addressing them.

4.1. Safety assessment results

Tables as given in Appendices shall be considered as one of the first steps in assisting designers of future autonomous ocean-going vessels in incorporating safety into the system’s design. Since the concepts of the system are still in the relatively early development phases as this paper is written, results of the study must inevitably be very general. Nevertheless, virtually all control loops and actions can and shall be addressed on a higher detail level as the system’s development progresses and more information is available. By reviewing the results of the study, few features can be highlighted.

4.1.1. Overview

For the reasons mentioned in the Introduction, instead of applying the methods of quantitative safety analysis, the system-theoretic, qualitative method has been used. Herein, safety was not actually evaluated since no specific statements describing the expected safety
performance of the system in question have been sought. Rather than that, the study consisted of seeking solutions by the implementation of which the safety can be ensured. This was done by reviewing a complex network of mutual interactions among the system's components. The advantage of such an approach over previously used ones lies in the possibility to perform the study in relation to subject, of which there is insufficient or no quantitative data, as is the case of AMV. Applying reliability-based methods to achieve it would mean the necessity to analyse a reliability structure of the system, which cannot be determined at this point. Nevertheless, the results of the study are in general consistent with those performed before - in their parts concerning potential hazards and solutions.

4.1.2. Human error
Firstly, a relatively high number of potential causes for control actions' inadequacy can be attributed to human error. Although it is understandable that human operators might have little control over a vessel operating in an autonomous mode, hazards can still result from human interactions with other system's components [22], see for example control actions #1-10c and 34 in Appendices. Those can be associated with the design process [56] (#14a,16,20,21), software development (#26,32,33), data interpretation (#9,40), limits' settings [46,108] (#34) or even illegal activities [6,109]. Humans' impact on the system's safety, although not evident from the safety assessment's results, will exist as humans will maintain an influence on its performance, one way or another [3]. Therefore, a relatively high number of mitigation measures are focused on liveware (see Fig. 4) and range from procedures on legislation implementation through operational trainings.

4.1.3. Technical considerations
Secondly, technical considerations will have a great importance to the safety of autonomous vessels. These pertain to both software and hardware, which must be reliable and efficient. Consider collision avoidance and assume that the applicable rules are not amended (some scholars raise concern that the implementation of autonomous vessels may require such amendments, see for example [29]). In order to prevent two ships from colliding, a set of conditions must be met. Two vessels shall not violate the minimum Closest Point of Approach (CPA), meaning that the distance between them shall at no circumstances be less than a certain value [110]. This limit can vary depending on circumstances, just to name a few: vessels' relative speed, area of navigation or weather conditions, [111–117]. The existence of a risk of collision must be determined by constantly calculating and monitoring the CPA and other proximity indicators that help to determine the situational awareness, e.g. the relative location of the target ship with respect to the own one, the rate of change for relative speed and course of the other ship. In order to achieve successful collision avoidance, the other vessel's presence must first be detected (#31) and its elements of movement must be calculated. This requires sensors to be reliable and data processing algorithms to be accurate (#37), accounting for the good seamanship practice. If more than one sensor is involved, data fusion issues apply. Assuming that the CPA and other indicators are calculated correctly and are below the minimum acceptable threshold (#35), certain action (#26) shall be taken by one of the vessels so as to reduce the risk of collision, as prescribed by COLREG,[118,119]. Therefore, the 'own' vessel shall analyse the data and determine if she is the 'give-way'. This will depend on many factors, just to name the relative bearing and speed or both ships' navigational status. Action required to avoid...
collision (e.g. heading or speed alteration) must be calculated together with its feasibility [120] (avoiding collision with one vessel might lead to colliding with another one or grounding). The decision must be made and executed by the actuation of either the rudder angle or the main engine’s revolutions. The effectiveness of the action taken shall be monitored [121,122]. On top of that, it might turn out that the object detected by radar was not in fact the vessel but a floating container for instance, and collision avoidance rules did not even apply for the situation.

This rather simple example highlights the importance of applying a holistic approach to an autonomous vessels’ system’s design. Here, all of its components must ‘cooperate’: humans set proper thresholds for proximity indicators (#34), sensors detect the object, algorithms process the data and create decisions, which are then executed by actuators. These interactions are sometimes extremely complex and must not be addressed on a linear basis [8,22].

### 4.1.4. Reliability and maintenance

Further on technical considerations, ensuring the sufficient reliability of equipment, including sensors as well as any other devices (#14a-17,20-22), can be a major issue. Nowadays, the crew on board a ship can perform maintenance and repairs, also as a contingency. This will not be possible for unmanned vessels, which must be adequately designed so as to survive any potentially hazardous mechanical breakdown or software malfunction [30]. These two issues include system’s inability to establish both-way communication between SBCC and the vessel. For such circumstances, a fail-to-safe mechanism shall be built-in to the system in order to prevent failure propagation [5,11] and allow for damage control with the assistance of other assets, salvage companies for instance. Efforts in damage control are reflected in

<table>
<thead>
<tr>
<th>Mitigation measure</th>
<th>Uncertainty category</th>
<th>Description</th>
<th>Result (uncertainty level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigorous maintenance regime</td>
<td>Phenomena</td>
<td>Maintenance procedures are well-known, however not for such complex systems with prolonged maintenance-free periods [5,44,86,87]</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Model of applying maintenance procedures can be established but will depend on data on equipment reliability, which might not be solid [22,30]</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Assumptions</td>
<td>It is assumed that decent procedures or algorithms can be elaborated [88], but this requires experience-based data</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Data</td>
<td>Data pertaining to maintenance needs of particular equipment are available, but assumes constant supervision of qualified personnel [89]</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Consensus</td>
<td>It is agreed that rigorous maintenance regime must be implemented [88]</td>
<td>Minor</td>
</tr>
<tr>
<td>Redundant equipment</td>
<td>Phenomena</td>
<td>Well-recognized as a way of ensuring particular subsystem’s reliability [44,75,76]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Consequences of applying it into autonomous shipping are rather well-comprehended: it will increase investment costs [76,77] but will also help providing necessary data, possibly by use of data fusion algorithms [78]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Assumptions</td>
<td>Redundancy is a well-explored subject, assumptions only refer to data fusion design [33], which is also well-known</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Data</td>
<td>Some data is only available for other domains (e.g. Dynamic Positioning) [79]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Consensus</td>
<td>Experts agree that this is necessary and profitable [2,10,75,80,81]</td>
<td>Minor</td>
</tr>
<tr>
<td>Resilience-based design</td>
<td>Phenomena</td>
<td>Resilience engineering is a relatively novel approach [89] and thus not widely applied [90]</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Models are lacking and those available usually pertain to domains other than shipping [89,91,92]</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Assumptions</td>
<td>Great number of assumptions must be made, relating to system’s actual structure [22]</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Data</td>
<td>Data on application of resilience engineering in shipping is lacking, some attempts have been made in other domains [92]</td>
<td>Significant</td>
</tr>
<tr>
<td></td>
<td>Consensus</td>
<td>Experts in autonomous shipping refer to concepts similar to resilience claiming that it must be built-in [30,86,93]</td>
<td>Minor</td>
</tr>
<tr>
<td>Procedures on consumables’ management</td>
<td>Phenomena</td>
<td>Such procedures have been successfully implemented in many shipping companies [94]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Model is relatively simple and well-developed [94,95]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Assumptions</td>
<td>These procedures would be based on existing ones so assumptions only pertaining to different model of operations need to be made [31,44,53]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Data</td>
<td>Data related to procedures on consumables’ management can be sensitive but is rather simple to calculate [95]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Consensus</td>
<td>Experts agree that this is necessary and profitable [96]</td>
<td>Minor</td>
</tr>
<tr>
<td>Capacity surpluses by design</td>
<td>Phenomena</td>
<td>Phenomena of including necessary surpluses by design is well recognized [97]</td>
<td>Minor</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Design surpluses are successfully applied to many modern systems</td>
<td>Minor</td>
</tr>
</tbody>
</table>
recently published requirements for passenger ships’ safe return to port, see for example [123]. These can be a starting point for the elaboration of future rules of classification for unmanned ships’ resilience engineering.

4.1.5. Hazards

Recent research reveals that some accident causal categories can have a greater impact on potentially reducing the safety of autonomous vessels than others, [11,124]. For instance, software or hardware malfunctions can be more vital to safety than errors occurring within resource management. This can be attributed to the fact that wrongdoings made at lower levels of organisation hierarchy can be more difficult to timely identify and correct (especially in autonomous operations). In this context, errors occurring within legal or organisational framework can have an impact on the occurrence of technical malfunctions, but not necessarily result in an immediate danger to the system. On the other hand, in daily operations, software or hardware malfunctions can propagate on fewer other components, but the results of such propagation can be both immediate and devastating.

Moreover, one can notice that for numerous control actions, their inadequacy can lead to the emergence of a large number of hazards (#1-7c,9-10c,14a-17,34-41). This can be attributed to the relatively low detail level of the analysis and the complexity of the system in question where, under conditions of autonomous operation, failures can propagate rapidly.

Analysing the control actions’ catalogue did not help identify any new hazard to be added to the list as given within Table 3. This may be due to the fact that the analysis has been performed in a very low level of detail and that the initial list has been refined in cooperation with experts in the field who have included all the system-level hazards they have ever encountered. Possibly, the list of hazards could be extended once more information about the system layout is available and its specific processes can be analysed more thoroughly.

4.1.6. Systemic approach summary

The innovativeness of the AMVs’ system and confusion regarding its actual, future layout force the safety analysis to be based primarily on a literature review. Unlike conventional shipping, there are very few experts who can be elicited in order to gain their impressions regarding autonomous shipping. And even if this can be achieved, such persons are in majority involved in ongoing commercial projects on system development and thus can be biased or unable to share their expertise. Moreover, most of the safety analyses performed to date were based on a probabilistic paradigm. The study herein is based on a different approach, a systemic one, although built on the foundations of previous ones as most of the content is inspired by the results of a ‘probabilistic’ literature review. Therefore, results are to a large extent consistent with those achieved previously.

However, an application of systemic approach helped view the system holistically, systematise the mutual relationships between its components of a different nature and elaborate some solutions on preventing or otherwise handling the potential inadequacies of these interactions. To date, research focused on mitigation measures was rather scarce or limited [30,37,125,126] with authors generally focusing on hazard identification [5–7] rather than on seeking diversified solutions to thereby-defined problems. In this context, the control actions’ descriptions as given in Appendices can be viewed as a compilation of recommendations on the implementation of safety-critical solutions for specified problems.

Nevertheless, some uncertainties exist herein and must be discussed.

4.2. Uncertainties

Within the framework introduced in Section 2.2.4., no resultant uncertainty is calculated based on the magnitudes of uncertainty within a particular category. Instead, the degree of belief is communicated in such a way that future system developers and decision-makers can easily recognise aspects requiring more attention in order to reduce the uncertainties. For instance, a more detailed inspection of the control actions’ catalogue (Appendices) and uncertainty analysis results (Figs. 5–7) leads to the conclusion that software covering operations within the vessel or communication require further study. The
The opportunity of analysing a system that is in an early phase of development was also used to refine the method of assessing some of the uncertainties present in system-theoretic approach. The magnitudes of uncertainty were assigned as supported by background knowledge available at the present stage of AMVs' technology development. As it progresses, more information would become available to analysts, and uncertainties could be re-evaluated. The purpose of the presented uncertainty analysis method is to communicate which parts of the future system require further, more detailed study with respect to the reduction of the uncertainties and improvement of safety. Communicating the magnitude of uncertainties pertaining to particular aspects of system's operation can attract future system developers' attention to the need of collecting additional data or the improvement of some models.

The results of our study indicate that the developers of future AMVs might wish to concentrate their effort on software development and validation as this part of the system appears to be hampered by the most significant uncertainties pertaining to its safety performance. One of the yet-to-be resolved issues with safety analysis and resulting uncertainties assessment is their potential incompleteness. Furthermore, the usefulness of the hereby elaborated safety recommendations and uncertainties-related data should be evaluated as soon as the system in question is in fact designed and more empirical data becomes available. Within these aspects lies the potential for further study.

Acknowledgements

The authors appreciate the financial contributions from Finnish Funding Agency for Technology Innovation (TEKES), since this research was co-funded by the Advanced Autonomous Waterborne Application Initiative (AAWA) project (project number 5166/31/2014; duration 01.01.2015–30.06.2017). The views expressed remain solely those of the authors. Additionally, the second author appreciates the financial...
We are grateful to Mr. Risto Tuominen and Mr. Risto Tiusanen as well as other experts for their valuable advice on the model of safety control structure and a high-level hazards’ list. Mr. Owen Jones assisted us in the necessary English proofreading.

We also wish to express our gratitude to three anonymous reviewers whose valuable comments allowed us to improve the initial version of the manuscript.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ress.2018.05.019.

References


[126] Rødseth ØJ, Burmeister H-C. Developments toward the unmanned ship 2012.
