Aalto University
School of Business

RISK AND SAFETY MANAGEMENT OF AUTONOMOUS SYSTEMS

Analysis of methods for use within the maritime industry

Master's Thesis
Ana María Manzur Tirado
Aalto University School of Business
Master of Information and Service Management
Fall 2020

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Ana Maria Manzur Tirado, Rowan Brown, Osiris Valdez Banda

Title of thesis Risk and Safety Management of Autonomous Systems: Analysis of methods for use within the maritime industry

Degree Master

Degree programme Information and Service Management

Thesis advisor(s) Markku Kuula

Year of approval 2020 **Number of pages** 84 **Language** English

Abstract

Maritime autonomous systems pose many challenges to their designers. A fully autonomous vessel must be able to handle everyday navigation and propulsion in addition to an extensive list of other tasks such as cargo handling, emergency maneuvering, ship-ship and ship-shore communications, situational awareness, and much more. If such systems are to be implemented for the sake of increased safety, their operational risk and safety must be managed and assured.

The goal of this thesis is to investigate how risk and safety of these systems can and should be managed. There are three categories of system modelling methods that can be used for this purpose. The oldest category is “sequential methods”, followed chronologically by the most popular category, called “epidemiological methods”, and then by the newest category, “systemic methods”.

This thesis contains an overview of the three categories. Followed by a literature review that investigates the approaches to risk and safety management of autonomous systems that are taken within four transportation industries (aviation, railway, automotive, and maritime). Next are three SWOT analyses, one for each category of methods.

For the role of autonomous maritime systems, the literature review and SWOT analyses indicate that STPA (a systemic method) is the optimal choice from the existing methods. This is because it is a comprehensive method that can handle complex socio-technical systems, such as those in question, while providing useful safety improvement recommendations.

However, no single method is better than every other in all situations, and STPA presents certain limitations and drawbacks. First, it is very resource intensive, demanding long time investments from expert personnel. Second, because few data on the proposed systems exist, it is very difficult to conclusively recommend a suitable method. Therefore, if practitioners decide to employ STPA, they should be open to considering other methods in case they can yield better results. Finally, STPA (and other systemic methods) cannot currently yield accident probabilities. This means that STPA, in its current form, is unable to entirely satisfy the IMO’s FSA, which is important for the future of autonomous ships. Conversely, the literature review and SWOT analyses indicate that methods that can satisfy the FSA are unsafe for this application. This is because they are too theoretically simplistic and not comprehensive enough to produce trustworthy results.

To solve this issue, one of the following should take place: (a) STPA (or another systemic method) is augmented to include probabilistic abilities; (b) STPA (or another systemic method) is combined with a sequential method to achieve the benefits of both categories (e.g. comprehensive and probabilistic results); or (c) a new systemic method is created that provides the depth of analysis of STPA as well as the required probabilistic capabilities.

However, barring the FSA issue, the enclosed analysis indicates that the optimal choice is a systemic method (specifically STPA) despite its heavy burden to resources. This may seem like a cavalier recommendation, but it is the most comprehensive method and it produces the most safety improvement recommendations, thereby making it the optimal choice. It is recommended that system analysis is performed from the design concept stage through to system operation, regardless of the method chosen. This is so that the analysis can be improved as more system data are produced.

Keywords safety and security, maritime industry, autonomous vehicles, STAMP,

PREFACE

This thesis started in the research project Smart City Ferries "AlyVesi" project and finalized in the Design for Value "D4 Value" program. ÄLYVESI was funded by the European Regional Development Fund (ERDF). Additional financiers are Finnish Transport Safety Agency and the cities of Helsinki and Espoo. The D4 Value program is partially funded by the Finnish Funding Agency for Innovation (TEKES). The goal of this thesis is to understand risk and safety management methods being used for autonomous transport systems, e.g. aviation, railway, and automobile; and to explore their application in the maritime industry. The final aim is to propose a method of risk and safety management for use with autonomous maritime systems.

Although this thesis is part of the Älyvesi project is not directly linked to other research work done in the project.

ÄlyVesi and D4 Value are the main beneficiary of this thesis, but other similar projects, companies, and research groups that are studying the viability of autonomous maritime vehicles may benefit as well. Other sectors/industries working with autonomous transportation systems may also benefit through furthered understanding of the different risk management methods currently in use and their relevance in the development of autonomous transport systems.

The authors want to thank the support received by ÄlyVesi and D4 Value, specially to all the experts who participated in the analysis carried out in the study.

Espoo 2019, Finland

Ana María Manzur Tirado
Rowan Brown
Osiris Valdez Banda

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATC	Air Traffic Control
ATSB	Australian Transport Safety Bureau
AUV	Autonomous Underwater Vehicles
BN	Bayesian Network
CAA	Civil Aviation Authority
CAST	Causal Analysis based on STAMP
CFC	Causal Factor Charting
CPA	Conventionally Piloted Aircraft
ETA	Event Tree Analysis
ETBA	Energy Trace and Barrier Analysis
FFA	Functional Failure Analysis
FMEA	Failure Mode and Effect Analysis
FRAM	Functional Resonance Analysis Method
FSA	Formal Safety Assessment
FTA	Fault Tree Analysis
GNSS	Global Navigation Satellite System
HAZAN	Hazard Analysis
HAZOP	Hazard and Operability analysis
HFACS	Human Factor Analysis & Classification System
HOT-PIE	Human, Organization, Technology, Process, information and Environment
ICAO	International Civil Aviation Organization
IMO	International Maritime Organization
MASS	Maritime Autonomous Surface Ships
MIT	Massachusetts Institute of Technology
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
SASWG	Safety of Autonomous Systems Working Group
SCM	Swiss cheese Model
SoS	System of Systems
STAMP	Systems Theoretic Analysis Model and Process model
STPA	Systems Theoretic Process Analysis
SWOT	Strengths, Weaknesses, Opportunities, and Threats
(s)UAS	(small) Unmanned Aerial System

1 Introduction

1.1 Background

Nowadays, we are seeing an increase in the use of automated systems and the trend is moving towards autonomous means of transportation. Currently, we can find many projects working on autonomous transportation systems in different domains such as air, water, roadways, and railways. There is an increased interest in developing these autonomous transportation methods due to the possibility of cost efficiency and improved safety.

An autonomous transportation system is a complex system that serves for transport of passengers and/or goods and in many cases will work in the same environment as manned vehicles. It is believed that by developing autonomous vehicles we can improve safety (by limiting the effect of human error) and reduce costs (Department of Transportation, 2016).

A contemporary example of autonomous systems in transportation is automated package delivery, with research and development being currently conducted by Project Wing (2018), Amazon Prime (2015), and others.

But however beneficial automation in transportation might be, there are inherent risks to safety that must be addressed before these systems are fully deployed. These include the safety of the vehicle itself, of the passengers and/or goods, and of the environment. It is therefore necessary to understand and determine a proper methodology to apply when conducting risk and safety management of an autonomous vessel.

1.2. Research Problems

In general, there are many methods used to approach risk and safety management. However, there is no method designed specifically for use with autonomous systems, and there is also no agreement on the best method to use when dealing with autonomous systems. Method choice therefore varies between different industries. Even within a specific industry, different institutions or individuals will use different methods depending on their knowledge, training, and confidence in the use of a specific method (Underwood & Waterson, 2013).

This extends to the maritime industry, in which there is no agreement on, or precedent for, risk and safety management methods of autonomous systems. This raises the following research questions:

1. What methods and frameworks are implemented for the management of risk and safety in the different industries involving autonomous transportation systems?
2. What are the key elements and issues for risk and safety management of autonomous transportation systems?
3. What is the optimum method of risk and safety management for autonomous systems within the maritime industry?

1.3. Aims of the Study

The aim of this project is to explore what methodologies and frameworks are implemented for the management of risk and safety in different industries for autonomous systems. An understanding will then be sought for the reasons behind the use and preference for some methods over others. To compare the methods, a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis will be developed for each category of methods.

After understanding the methods, the goal is to recommend a method or combination of methods for use in the risk and safety management of maritime autonomous ships. To achieve this, it is important to understand the key elements and issues pertaining to risk and safety management in the context of autonomous systems

A desired outcome of the thesis is that research and development groups can use it in pursuit of risk and safety management within the maritime industry. The aimed practical outcome is the recommendation of an approach to risk and safety management based upon sound comparison and analysis of the various possibilities, so that future teams can confidently employ a methodology in their own applications.

1.4. Scope, Limitations, and Outline

In the following section, the scope of the research as well as the limitations encountered will be presented followed by the outline of the thesis.

1.4.1. Scope

This thesis will consist of review and analysis of the various risk and safety management methods. These topics (review and analysis) will be conducted thoroughly and logically, and the result will be a recommendation of the method(s) optimally suited to automation in the maritime industry.

This thesis will not, however, feature in-depth discussion of how to employ the various methods. Rather, it will focus on their strengths, weakness, and applications. There will be no guide for implementation, and employment of the methods will not be exemplified. Further, quantitative comparisons of the methods will not be given.

1.4.2. Limitations

The literature reviewed in this thesis has been collected from a variety of contemporary sources and it contains a comprehensive list of the most popular risk and safety management methods used in various applications. The subsequent analyses have been conducted thoroughly using the SWOT system. This is a popular and reliable tool for qualitative analysis (Paraskevas, 2013).

But however thorough this work may be, the resultant recommendations and discussions will naturally be affected by opinions of the authors, and they should not be accepted without examination. Additionally, while the literature review concerns various autonomous systems within the broad transportation industry, the final recommendations are specifically focused on autonomous maritime systems. So, because the basis for these recommendations is the maritime domain, they will not necessarily apply in any other contexts.

1.4.3. Outline

This thesis will feature a literature review of both the standard risk and safety management methods and also of the methods used for autonomous transportation systems. This includes history, current uses, and future uses of the methods in industry.

Following that is a description of the analysis tool (SWOT) that will be used to compare the methods and identify the optimal choice. This is then followed by the analyses itself. Recommendations will be made, and conclusions will end the thesis.

2. Literature review

This section provides a critical review of relevant literature that will help explain the most important and most commonly used risk and safety management methods in the field of autonomous systems. The objective of the literature review is to substantiate the research questions. It will provide a background to understand the development of the different risk and safety management methods and it will indicate the range of applications and uses of the different risk and safety tools.

After explanation of these applications, we will focus on methods used in the transportation domains (railway, automobile, aviation, maritime) with special regard to autonomous systems. This part of the literature review will improve the understanding of what is currently used by the different industries, and it will provide insight into the reasons for using and/or preferring one method over another within the domain of autonomous systems.

2.1. Important Definitions

Throughout the following literature, there are several key words and phrases that will repeatedly surface. In this thesis, basic terminology is adapted from Aven (2015) and is defined as follows:

- **Systems** are combinations of components that can be physical, organizational, or human. Systems are simple if they are designed so that individual components affect other components in a progressive fashion to finally achieve desired results. As systems become more complex, interconnected components and variables create feedback loops that affect each other non-linearly.
- **Risk** is consideration of the probability and the impact of an unfortunate and negative occurrence with respect to something humans value. Impact could include injuries, deaths, loss of money, loss of time, or something else.
- **Safety** is the opposite of risk. Less risk means greater safety.
- **Models** of systems can be created using the methods referred to in this thesis. Some are probability based and some are qualitative. In this sense, “model” is broadly synonymous with “understanding”, and one of the purposes of using different methods is to gain understandings of different aspects of the system.

Automation in the maritime industry will include the implementation of autonomous systems. It is desired that risk and safety management of these systems is achieved. Risk and safety management can be considered a process of preventing and mitigating risks down to an acceptable level. The methods considered in this thesis are all tools that can be used in the management of risk and safety. Some of these methods identify hazardous situations, the sources of these situations, and probabilities of occurrence. Other methods give recommendations of ways in which hazards can be mitigated. And yet even more methods can give different combinations of the above. So, as the methods can yield different types of results, they themselves are not representative of the process of risk and safety management. Rather, they are part of the process, and the scope of risk and safety management is ultimately dependent on the situation.

This variation of results is because the methods have different theoretical underpinnings, and as such, they do not all yield results in the same "format". This is one of many reasons that different methods are picked for different situations; it depends on the user's goals. These differences will be considered in the analysis section of this thesis, with special regard to the case of autonomous maritime systems.

One final clarification is needed before the literature can be examined: The difference between accident analysis and risk and safety management. Much of the following literature is related to accident analysis rather than risk and safety management. This is because in some cases, methods that can be applied to risk and safety management can also be applied to accident analysis (this is more often true for older methods than for newer ones). Both processes involve hazard identification within the systems, and many cases of accident investigation are conducted with the intention of identifying hazards that can be remedied in similar systems. Therefore, a method's successes and characteristics in one application are likewise often exhibited in the other. However, these two processes are not the same, and not every method can be applied to both cases. For example, one can use "STAMP-CAST" for accident analysis, but should use "STAMP-STPA" in the process of risk and safety management. These methods will be further discussed later. The purpose of mentioning them here is to acknowledge that there are some differences between accident analysis and risk and safety management.

2.2. Risk and Safety Management Methods

Risk and safety management systems are used to support an organization to operate in the safest possible way while performing the tasks and/or processes that are required of it to function. The application of risk and safety management methods can vary in degree. It can be applied to an organization or company at a high level but can also be used for the design and/or operation of specific systems (e.g. machines).

According to Hollnagel (2008), safety is the absence of risk, and risk is usually associated with an event that could have a negative impact on the system being studied. The risk is possible, but it is not certain, and the extent of the negative outcome, before it happens, can only be estimated.

Research on this topic is quite vast, and there are over 100 models in existence that can be used for system analysis (Underwood & Waterson, 2013). Appendix 1 contains a list of some of the existing models. The first model, for accident investigation, was the Domino Model, developed by Heinrich in 1931. Since then, different accident investigation methods have been developed to suit requirements demanded by evolutions in accident "type" (Hollnagel & Speziali, 2008). According to Hollnagel & Speziali, due to continuous and fast development of the socio-technical system, and as a result of technological innovation, commercial opportunities, and user requirements, new methods for analyzing accidents are consistently needed. This has also been asserted by Robertson et al. (2015), who state that as socio-technical systems become more complex, better-suited methods to analyze systems will be required.

In comparison to investigation methods, risk assessment and safety management methods develop very slowly. This is because, historically, risk and safety management methods are only developed after new types of accidents have occurred and the corresponding investigation methods are developed. These methods either try to explain a certain type of accident in a specific industry or they try to be as comprehensive as possible by including the collective knowledge of accidents and industries (Hollnagel & Speziali, 2008). Most analysis models and methods have therefore developed in reaction to the trends and needs of the period.

Hollnagel & Speziali (2008) modified Perrow's 1984 coupling interactive diagram to offer a matrix that can better identify the type of socio-technical system under investigation and advise on the methods applicable to that type of system. Perrow defined the systems according to their coupling, tight or loose, and according to the manageability of the system being high (tractable) or low (intractable). Coupling refers to existing connections and dependability in a system between its components or subsystems in terms of functionality. A system has a tight coupling, for example, when one subsystem depends on another to function. Tightly coupled systems also exhibit little or no delay between one process and the next one, and once a process sequence is established, it cannot be easily changed. On the other extreme, a system with loose coupling is such that interactions between parts can be easily separated and there can be a delay when going from one process to the next.

Regarding manageability, a tractable system is one whose principles of functioning are known, and it is such that it can be easily described, and it will not change while doing so. Conversely, an intractable system can change while being described, and its functions can be partly or completely unknown. Hollnagel exemplifies an intractable system by citing activities in the emergency department of a hospital.

Figure 1 shows the systems characteristics according to their coupling and manageability and Figure 2 shows the accident investigation methods that are better suited to different types of socio-technical systems (Hollnagel & Speziali, 2008). Specifically, there are examples of some common socio-technical systems as well as some of the common accident investigation methods. These systems are positioned in the matrix depending on their manageability and their tractability, and one can, by inspection of the figure, decide which investigative method is best suited to the different socio-technical systems.

In 2013, Underwood & Waterson further developed the Hollnagel-Perrow matrix to classify the methods into three categories: Sequential, epidemiological, and systemic. This is a practical guide for safety professionals concerning the use of the different methods, showing under what circumstances each is preferable. Figure 3 shows the Underwood & Waterson matrix.

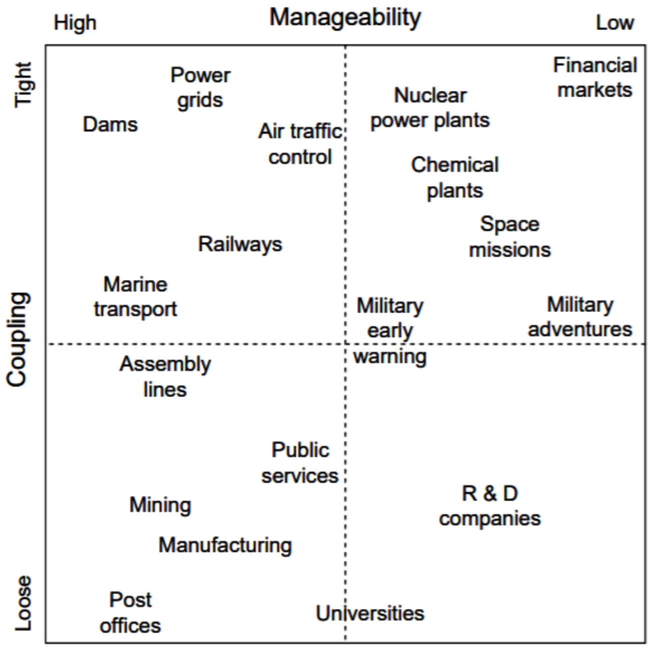


Figure 1: System Characteristics (Underwood & Waterson, 2013)

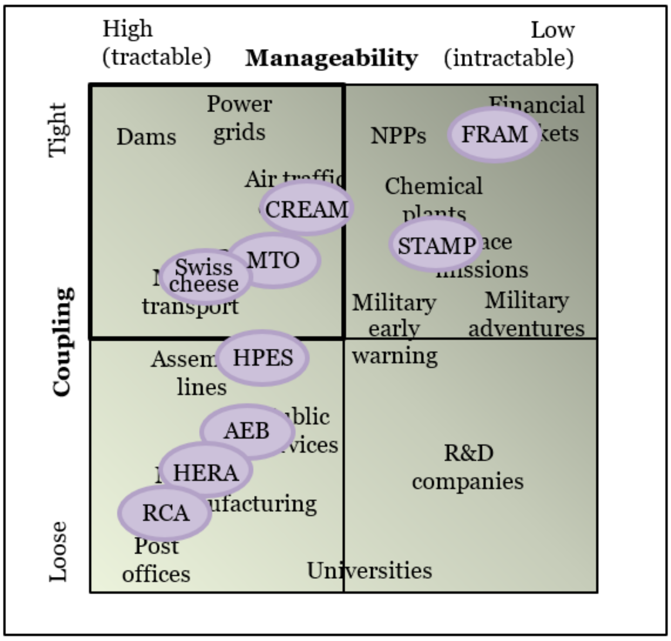


Figure 2: Characterization of accident analysis methods (Hollnagel & Speziali, 2008)

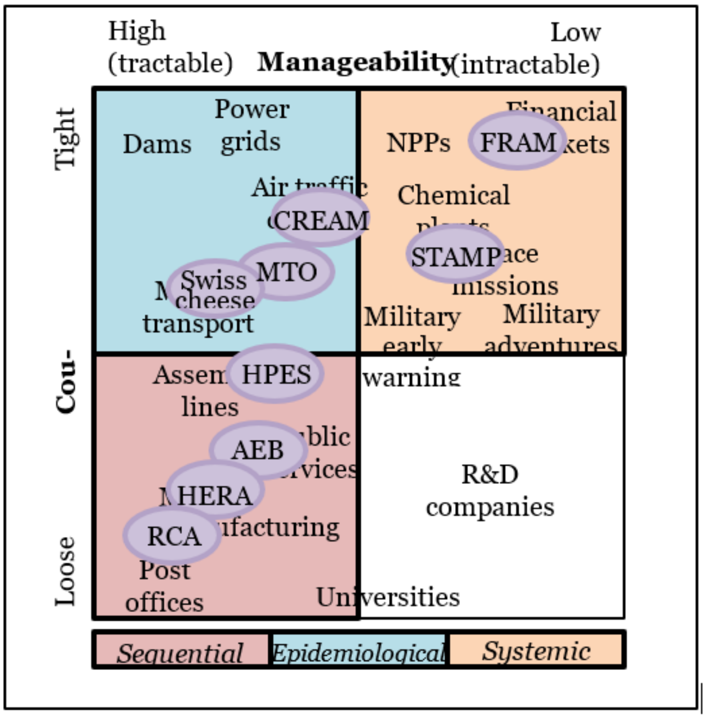


Figure 3: Categorization of accident analysis methods
(Underwood & Waterson, 2013)

2.2.1. Sequential methods

The first category is sequential methods. These describe a negative outcome as the result of a root cause event that produced a sequence of other discrete events which follow in a chronological manner. The result is a sequence of events that follow one after the other and that end up in an accident. There is therefore a linear correlation between the origin of the accident (the root cause) and the outcome (the accident). In sequential models, when the outcome becomes obvious, each step can be traced backwards to the origin (Hudson, 2014).

The first sequential model was the Domino Model developed by Heinrich in 1931. Other commonly used methods in this category include Fault Tree Analysis (FTA), the 5 Whys model, Failure Mode and Effect Analysis (FMEA), and Root Cause Analysis. These were the dominant models from the 1960’s to the 1990’s, and they are often recommended when analyzing simple systems where a physical component has failed or where human actions have caused an accident. However, as systems became increasingly complex, accidents became inexplicable using sequential methods. This is because system failures began to involve complex organizational components, and sequential techniques could not account for these factors. The need for new models therefore became apparent. This led to the advent of a second category of safety models called epidemiological methods.

2.2.2. Epidemiological methods

According to Underwood & Waterson (2013), these models view accidents as a combination of "latent" and "active" failures within the system. Latent conditions are the norms (e.g. management, working practices, and organizational culture) that influence the working styles and that have effects like not following the intended working procedures or having employees exhausted from working overtime.

These latent conditions set the scene for unsafe situations. Under such circumstances, errors can occur more easily and result in accidents or negative outcomes. Therefore, the latent conditions become obvious when combined with unsafe actions and breach of system defenses. The best-known epidemiological models are the Swiss Cheese Model (SCM) (Reason, 1990, 1997), the Human Factor Analysis & Classification System (HFACS), and the ATSB accident investigation model.

Epidemiological models, compared to sequential models, not only observe the connection between a root cause and a negative outcome, but also take into consideration how the organization and working practices (the latent conditions) affect or influence the appearance of an accident. Still, many of these methods are based on linear models and follow the cause-effects principle (Hollnagel, 2004). Therefore, as system complexity continued to increase, involving more social, software, and otherwise non-physical components, some authors began to argue that epidemiological methods were not capable of comprehensive system analysis (Rasmussen, 1997; Leveson, 2004). Hence, new models were desired that could consider all the socio-technical aspects of the system. To meet this demand, systems theory was used to create new methods, leading to the advent of the third category of risk and safety models: Systemic methods.

2.2.3. Systemic methods

Systemic methods are based on the application of systems theory where the objective is to understand the structure and behavior of any type of system. Systems Theory studies a system as a hole, not decomposing it into separate physical components and analyzing the behavior of each component and assuming that the operation of each component or subsystem is not distorted (Levenson 2004a). When dealing with complex systems, these assumptions are not valid since the interaction of the components affects how the system operates. System theory examines the system as a whole and studies how the parts interact and fit together. This approach is useful when studying complex systems that involve physical components, software and human interaction for example.

Systemic methods describe accidents as "unexpected behavior of a system resulting from uncontrolled relationships between its constituent parts" (Underwood & Waterson, 2013). For example, consider a large system of components consisting of humans, physical parts, and software. While each component operation may seem logical individually, they can still combine to create hazardous conditions.

The objective of systemic methods is to identify the causal factors that can lead to hazardous conditions (Leveson & Thomas, 2018) and provide guidance for preventing or mitigating them. Examples of systemic methods are the Systems Theoretic Analysis Model and Process model (STAMP) (Leveson, 2011), AcciMap (Rasmussen, 1997), and the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2004).

The systemic approach is advocated by many researchers who found that cause-effect approaches had theoretical limitations. For example, it was found that the process to discover the cause or combination of causes for an accident was stopped as soon as someone/something to blame was found (Allison, et al., 2017). The problem with sequential methods and finding someone/something to blame is that the opportunity to discover other possible unsafe actions and design problems, and to learn and improve the system, is lost (Leveson, 2011).

In Hollnagel's (2012) view, there is always some degree of uncertainty and ignorance about the way all complex socio-technical systems work. He developed the FRAM with this mind in order to have a method to emphasize that safety is something a system does and is not something it has. He believes that by focusing on what actually happens in the system we can reduce the unavoidable state of relative ignorance. In FRAM, the system is described by what it actually does and not by how it is imagined to work.

In addition to the three groups mentioned over the previous pages, there is one more noteworthy approach to risk and safety management: A combination of methods.

In the view of Alexander et al. (2009), to approach risk and safety in autonomous systems, which are complex socio-technical systems, using just one model/tool is often not enough. A combination is a more effective approach.

Likewise, Leong et al. (2017) propose the use of HOT-PIE, a technique to identify epistemic uncertainty, to augment other hazard analysis techniques such as STAMP-STPA (Systems Theoretic Process Analysis). By using HOT-PIE, they claim designers and safety engineers will be able to better track the uncertainties from design through the system life-cycle.

2.3. Risk and Safety Management Methods used by Industry

Research on accident investigation and hazard analysis methods is quite vast. There is abundant research on the use of many specific methods and how they can be applied to different types of socio-technical systems in different industries. There have also been some investigations where several methods are compared and evaluated at a high level, as well as some research on the use of specific methods for particular types of autonomous systems. However, there is currently no research that specifically reviews the risk and safety management methods utilized throughout the different industries that employ autonomous systems.

Previously, some researchers have approached the topic of identifying the best method of accident analysis by comparing the performance of a select few methods when analyzing the same accident (e.g. Salmon et al., 2011; Yousefi, Rodriguez Hernandez, & Lopez Peña, 2018). In general, the findings have been that each method offers advantages and disadvantages in different areas. For example, some are better for finding possible causal factors while others are better for identifying opportunities to improve the system. Some researchers have also proposed new tools for analyzing accidents based on combinations of existing methods, such as the Bowtie Diagram, which is based on a combination of FTA, Event Tree Analysis (ETA), and Causal Factor Charting (CFC) (Blaauwgeers et al., 2013). Other researchers have proposed new tools that can be added as extra steps or processes to existing tools (e.g. Leong et al., 2017).

2.3.1. General considerations

Traditional approaches to safety analysis assume that accidents are caused by component failures (Leveson, 1995; Roland & Moriarty, 1983). They therefore focus on reliability analysis techniques, particularly FTA or ETA. The goal of these traditional approaches is to determine scenarios of component failures that together will lead to an accident or loss event.

These approaches, by themselves, are becoming less effective because socio-technical systems are increasing in complexity and becoming more tightly coupled. This is in part related to an increased use of software in systems, which allows more complex and tightly coupled systems to be constructed. The potential for accidents arising from unsafe interactions among non-failed components (e.g. unplanned system and software behavior) is therefore increasing and traditional approaches to safety analysis are frequently inadequate.

Another reason that traditional analysis methods are no longer appropriate is because modern systems evolve too quickly for reactive analysis. In other words, since traditional, linear methods are only suitable for analysis of mature systems, they cannot be used with new technology. Ishimatsu et al. (2010) wrote, “But system designs have become so complex that waiting until a design is mature enough to perform a safety analysis on it is impractical. The only practical and cost-effective safe design approach in these systems is to design safety in from the beginning”.

There has been an increase in study within this topic in different industries over the previous few years. For example, the Safety of Autonomous Systems Working Group (SASWG) was created to study and provide guidance on how to manage safety methods/approaches when dealing specifically with the complex systems such as those alluded to by Ishimatsu et al., (2010) (Menon & Alexander, 2018). There has also been an increase in development, manufacturing, and field testing of these systems, such as the autonomous cars built by Ford (Ford Media Center, 2017).

Alexander et al., (2008) wrote that people fear the reliability of autonomous systems when it comes to safety issues. Those fears are founded by several accidents with autonomous aircraft and automobiles, such as the 2018 fatality in Arizona involving an Uber autonomous car (BBC, 2018). There has been a subsequent transition towards manufactures working together with authorities in order to certify the safety of their products in the case of autonomous systems.

To ensure an acceptable safety level for the operation of autonomous systems, they should "pass" or be certified against certain safety requirements (Alexander et al., 2009). These safety requirements need to be clearly stated for each autonomous system, and they should not be general requirements used for other "similar systems" that are manned or piloted (Alexander et al., 2008). This would ensure a low level of risk to safety, and it should eliminate the necessity of maturity in the system (as is currently required).

To this end, Alexander et al. (2009) propose the use of a combination of methods such as ETBA (Energy Trace and Barrier Analysis), FFA, and Hazard and Operability (HAZOP) analysis in order to identify and analyze the hazards from different perspectives. They then suggest using the Hall-May method to derive high level safety requirements.

This approach of combining methods has been proposed in other instances, some of which were described on the previous page (namely Bowtie Diagrams). Despotou et al. (2009) propose a combination of Dependability Deviation Analysis with Simulation-based Hazard Analysis (HAZAN) for hazard identification during system development. In other cases, authors have proposed the incorporation of one method within the process of another to expand its capabilities. For example, Leong et al. (2017) proposed employment of existing hazard analysis techniques like STPA or FMEA in addition to HOT-PIE to improve the overall capability of tracing unknown uncertainties.

In theory, combining methods would be effective because different methods present different strengths in analysis. For example, while one method might identify some types of hazards, it might miss other types of hazards that could be identified by another method. This tandem use is therefore effective in considering the largest possible number of hazards. However, one potential issue with this approach is that the use of multiple unrelated methods would most likely require more effort, time, skill, and data than using just one. This means it may not always be practical to use multiple methods.

In both cases, whether a single method or a combination of methods is to be used, an important aspect of the system must be identified. Namely, before one can choose an approach, one must specify if the system in question is a System of Systems (SoS) and delimit the scope accordingly (Rae & Alexander, 2011). Although the definitions for SoS vary depending on the author, Mitre (2014) provides the following definition, "a SoS is a collection of systems, each capable of independent operation, that interoperate together to achieve additional desired capabilities."

Systems of systems (SoSs) are complex and have many interacting components, so they are at risk from their environments. Additional risks pertaining to system security must also be considered. Causevic (2016) therefore suggests that in order to ensure safety and security, if a SoS is being analyzed, the systems should be analyzed together and not separately. While this is only a recent field of study, a handful of other authors have also broached the subject. Alexander (2007), Despotou et al. (2009), and Young & Leveson (2013) have all proposed methods to analyze the systems from the perspectives of both safety and security. The reason this topic is being mentioned is that, if autonomous maritime systems fit into this category, it will be an important indication that the scope of the chosen method will have to be comprehensive and not focused on individual components.

Now that general, practical considerations regarding risk and safety management have been introduced, we will focus on their use with autonomous systems in various industries related to transportation.

2.3.2. Railway

Historically, other means of transport have lagged rail in the implementation of automated systems. Early automation in the rail industry dates to the 1910's when magnetic track inspection systems were first developed to supplement human inspection (Boslaugh, 2013). Over the past century, railway industry in the United States has embraced automated inspection systems and they now cover over 48,000 km of track annually.

Early automatic train operations include the London Post Office Railway (1927) and the 42nd Street Shuttle in New York (early 1960's). Since then, the scope of automatic systems in train operations has markedly increased, and fully automated metros can be found in Paris, Barcelona, Guangzhou, Budapest, Nuremberg, and elsewhere (Boslaugh, 2013). Despite full automation, many of these systems do still employ a driver for emergency control.

There are a number of factors that make automation of urban metro systems easier than other means of transport: Tracks offer guidance to the trains and their exclusive use means collisions with other means of transport are unlikely; metros transport people between a finite number of stops; many subsystems have already been automated for years; some train speeds are too high for human reaction times (they must be controlled automatically); and the rail industry features a culture of development that supports new technology (Boslaugh, 2013).

However, there are some safety concerns about automated rail systems and train operation. Anecdotal accidents are the cause for a large amount of any public lack-of-confidence in the safety of automated trains. However, this does not mean these safety concern should be ignored, since there are inherent risks to automated operations.

To deal with these risks, publications from various railway regulatory bodies propose different approaches. The Federal Railroad Administration (2009) suggests the use of a linear method based on ETA. Likewise, the California High-Speed Rail Authority (2013) recommends linear methods like FTA and FMEA for hazard analysis. This is a noteworthy example because their purview is focused entirely on modern trains with relatively high levels of automated control.

Recommendations of linear methods are not unique to the US. HAZOP analysis, Process Hazard Analysis, ETA, FTA, and Failure Mode, Effects, and Criticality Analysis have all been specifically recommended by the European Railway Agency (2009). No single method was identified as the optimal choice.

While there are many current uses of linear methods, as exemplified above, some authors have argued that these are no longer acceptable (e.g. Klockner & Toft, 2015; An et al., 2011). They wrote that as automation in the rail industry causes increased system complexity and tightness of coupling, alternative methods of management and analysis must be employed.

Belmonte et al. (2011) applied the Functional Resonance Accident Method (FRAM) to one aspect of the rail industry: Traffic supervision. Their study used automatic train supervision systems linked to human operators within simulated accident situations. In other words, the experimental setup consisted of a complex socio-technical system within the rail industry. The purpose was to compare classic, linear approaches to safety analysis with systemic approaches like FRAM. They concluded that FRAM is a good addition to the classical approach because it promotes the analysis of “other situations” that were not originally considered by the linear methods.

In an attempt to study the interactions between different forces within a socio-technical system, Klockner & Toft (2015) used the SAFE-NET method to study railway safety occurrences. The SAFE-NET model was theoretically preceded by FRAM, and it was used by the authors to gain insight specifically into the relationships between contributing factors to major railway safety occurrences. This is another example of systemic approaches to risk and safety in the rail industry.

In another example, Dong (2012) applied STAMP to a rail accident that occurred on July 23, 2011, in which 40 people were killed and 120 people were injured. This example does not necessarily relate to automated train operation, but it does show a systemic method (STAMP) being used to identify relationships between the components of a socio-technical system that led to an accident. The result was a clear image consisting of a dangerous safety culture throughout the whole train organization. Managerial reorganization was strongly recommended. Another important recommendation from this thesis was that STPA should be included in the design of all safety critical railway systems.

The above academic examples of systemic methods being applied to the rail industry might be indicative of a shift towards these methods. Conversely, linear methods were also shown to still be very popular within industry and regulatory bodies. It is therefore difficult to predict which methods will dominate the future of risk and safety management of autonomous systems within the rail industry.

2.3.3. Aviation

Traditionally, aviation and aerospace have used sequential and epidemiological methods such as FMEA and FTA (Ishimatsu et al., 2010; Fleming et al., 2012; Silva Castilho et al., 2018). Many institutions, like the National Aeronautics and Space Administration (NASA), the International Civil Aviation Organization (ICAO), the Federal Aviation Administration, and Eurocontrol, have used or are using these methods.

One germane example is the Bowtie method. It is widely used in the aviation industry for identification and management of risk, with such users as airport operators, aircraft operators, ground service providers, militaries, and civil aviation authorities (Civil Aviation Authority United Kingdom [CAA UK], 2015). As previously explained, Bowtie combines three linear methods with some modifications: FTA, ETA, and CFC. It is barrier-based method following the idea behind Reason’s Swiss Cheese model. This method is also referenced in Annex 19 of ICAO’s Safety Management Manual (2013).

Some contemporary research also supports the use of linear methods. Ancel et al. (2017) propose the use of Bayesian belief networks to produce real time risk assessment of Unmanned Aircraft Systems (UAS). The starting point for constructing their model is to identify undesirable events and failures and their respective causal factors. For this purpose, they propose the use of ETA, FTA and/or FMEA.

The use of linear methods can be traced back through history. In the past, for piloted aircraft, the approach to risk and safety was based on lessons learned, and improvements to craft, pilot training, and maintenance regimes were only achieved after accidents had occurred. Because today's society is more risk-averse, this reactive approach to safety and risk is no longer acceptable (Clothier et al., 2006). Additionally, the rate at which Unmanned Aerial System (UAS) technology evolves might demand newer and faster approaches anyway because reactive control is simply very slow. Practically, reactive management of risks posed by UAS is also hindered by a lack of information about accidents and incidents, which is in part due to a lack of reporting.

Risk and safety management of UAS is an important area of study. This is because there has been a marked growth in interest in UAS as more commercial research and development takes place and more applications for UAS are proposed. Possible uses include agriculture, surveillance, parcel delivery (Amazon Prime, 2015; Project Wing, 2018), humanitarian efforts (World Bank Group, 2017), ornithology (Wilson et al., 2017), and more.

However, an increased presence of UAS in the National Airspace System (NAS) will create a more complex environment and increase the risk of hazardous situations (e.g. higher risk for mid-air collision). There are therefore a number issues and considerations that must be addressed before widespread use of UAS is allowed. Examples include risk mitigation measures such as complex on-board technologies, procedural controls, and air traffic separation services, all of which would help achieve an acceptable level of airspace safety (Wilson et al., 2017).

There are other UAS-specific risks. Non-verbal communication with Air Traffic Control (ATC) services is one of them. Currently, communication is conducted between humans over the radio (verbal communication). Attention must be duly paid to ensuring that communication between UAS and ATC will be reliable.

Another issue is regulations for dealing with redundancies. For example, UAS will need to rely on external systems like the Global Navigation Satellite System (GNSS) in order to function as desired, but they will need to have a safety "backup" system that will allow continued functionality in case the GNSS connection is lost. Many other such redundant safety mechanisms will be required.

On the topic of regulations, there are many within the aviation industry that have been developed over decades of commercial air travel. While these regulations are detailed, well-understood, and widely applied throughout traditional sectors, they are not applicable to autonomous systems. The SASWG (2018) wrote that current “regulations and guidance are not well-suited to the type of software used in autonomous systems”. Despite this, there are cases of authorities applying the same regulations and certification methods to UAS as for Conventionally Pilot Aircraft (CPA). This is inadvisable because UAS and CPA are very different technologies used in different ways, and they cannot be treated the same. For example, consider the variation in size and functionality between different UAS. It will be so great that different classifications and regulations will be warranted for different types of UAS (Washington et al. , 2017). By extension, applying CPA regulations to UAS is clearly inappropriate.

The above regulatory issues are important topics of study for the aviation industry. Consequently, a large body of literature has been devoted to researching the presence that UAS will have in the NAS, recognizing the potential for mid-air collision with CPA and the risks posed to bystanders/third parties. Research of this kind has often tried to investigate ways in which the airworthiness of UAS can be satisfactorily certified, thereby allowing UAS and CPA to share airspace in a regulated manner. Examples of such literature includes Clothier et al., (2008), ICAO (2011), and Dalamagkidis et al. (2008). This is in line with Alexander et al. (2009), who wrote that autonomous systems should have to “pass” certain certification requirements to ensure they are safe to operate.

Creating proper regulations for UAS is important because they will define standard procedures for the design, production, maintenance, and operation of the aircraft. Currently, there are no such regulations specific to UAS, but they are expected to be at least as safe as CPA. To err on the side of safety, regulators tend to place high restrictions on their operation, e.g. limiting the areas where they can fly to low populated regions (Clothier et al., 2011). It has been contended that having these extra restrictions will impair their development for commercial use (Dalamagkidis et al., 2008).

Clothier et al. (2011) therefore proposed a possible framework for airworthiness certification of UAS. They also explained why the current CPA certification framework cannot be used for UAS (at least, not without modification). They wrote the following:

- CPA certification must ensure the security of passengers and crew whereas UAS certification must not.
- CPA certification must ensure the safety of third parties and bystanders. UAS are generally smaller than CPA (in some cases, much smaller). This means they pose a

lower threat to property and bystanders, and as such they should be categorized differently.

- When operating over uninhabited areas (such as sea/ocean), UAS pose no risk to humans but still pose risks to other aircraft. Regulations should reflect this.
- Imposing the same regulations on UAS as on CPA would create unjustified costs to the UAS industry due to over regulation.
- However, in some cases, using the CPA framework for UAS, when flying in particular regions, might present unacceptably high levels of risk.

They therefore propose the use of a risk matrix that evaluates different types of UAS and the areas over which they fly. The level of risk depends on to whom or what the risk is imposed, such as people on the ground, third party property, or other entities of value. It is then possible to assign different types of certification depending on the identified level of risk.

Another approach to certification was proposed by Gonçalves et al. (2017). They proposed the use of Petri Nets to prepare a safety assessment for UAS that can be accepted by certifying bodies to grant airworthiness. Their safety assessment process considers not only the airplane but also the control ground station, communication links, mission planning, air traffic control, and more, thereby helping with identification of the most feared events. Methods like FMEA, Failure Modes and/or Effects Summary, FTA, and Reliability Block Diagrams are suggested for evaluation of the failure conditions. Once the most feared events are found they proceed to apply Petri Nets. In order to use Petri Nets, data which comes from previous test flights for the specific UAS being evaluated is needed. It is worth noting that the above analysis methods are all linear, not systemic. Additionally, the amount of empirical data required may hinder its practical viability.

In addition to airworthiness certification, there is a lot of literature concerned with overall safety analysis. The main risks posed by UAS are to other aircraft (mid-air collisions) and to third parties such as bystanders and property. The latter problem is inherent for all craft flying over populated areas, both conventional and autonomous. On account of these risks, it has been predicted that regulations for autonomous systems in aviation will be stricter than for autonomous systems in other domains (SASWG, 2018).

Different approaches to generalized risk assessment of UAS have been proposed. Two such approaches, from Wilson et al. (2017), concern small Unmanned Aircraft Systems (sUAS). The first is a qualitative process, and the second is a probabilistic-model-based risk estimation methodology that uses Bayesian Belief Networks. The starting point for both methods is to identify risks posed by UAS. These are typically found using data from accident and incident reports, but this can be challenging for UAS due to lack of reporting. Therefore, to improve their estimates, the authors propose employment of FMEA and FTA.

Logan & Glaab (2017) also applied FMEA for small unmanned aircraft systems. With this approach they were able to find, with the help of subject matter experts, possible failure modes of the system in different operating conditions and define solutions to mitigate unwanted effects. In order to do an FMEA, the system was divided into subsystems that were then individually analyzed. Once the probable failure modes for each subsystem were identified, they were reviewed to find out how they could affect the overall system.

Melnyk et al. (2014) proposed a framework for predicting UAS safety levels consisting of a target level of safety and event tree framework. They wrote that many researchers support the use of event trees to predict casualties caused by UAS in the NAS. This is because event trees are well-established for helping to determine the probability and impact of specific failure events. What they propose is not a new method but the use of target level of safety and event trees together as a framework for analysis of sUAS which is more comprehensive than previous options. Theirs is a quantitative method and they used information from previous studies to populate the event tree. They also compared the results of using their framework with actual results from air carrier accidents and fatalities, and the model only differed by approximately 3%.

Hirling & Holzapfel (2017) introduced O.R.C.U.S., a tool designed to evaluate the risks from UAS flights in Germany. It generates predictions on possible fatalities according to the flight plan even in cases when not much information about the UAS model is available. The tool is based on the combination of two event trees; one for the overall probability of a catastrophic event, and the other for the percentage individual failure of a subsystem in the UAS.

One common trait shared between all the above risk and safety analysis methods is that they are all linear. This is not true throughout the whole aviation industry. For example, Rodrigues de Carvalho (2011) used FRAM to investigate a mid-air collision accident between two planes flying over Brazil. The purpose of the study was to analyze the resilience of the ATC system in Brazil. Although this study did not involve autonomous systems, it is a good example of the use of FRAM to analyze complex aviation systems. Specifically, by using FRAM, the author was able to demonstrate which controls were actually present in the Brazilian ATC system and how these controls changed overtime. These changes, which happened without being perceived, created a situation where an accident could have and did occur. Because the effects of the changes were non-linear, using traditional, linear methods may not have identified the issues.

Another example of systemic methods used within the aviation industry involves the plan to transform air traffic management from being ground-based to satellite-based. “NextGen” will be the new air traffic management system and it is paramount that at least the current level of safety is maintained. The expected benefits of NextGen will be reduced delays, improved environmental impact by reducing CO_2 emissions, and a safer airspace. But because it is a more complex system, it will need better analysis tools to ensure its safe operation.

Fleming et al (2013) propose the use of STPA for hazard analysis within NextGen, claiming that the subsystems will be complex and that failures will not only be due to component failure but also due to the interactions of software, which in a tightly coupled system can create unpredictable results. Accidents can occur in situations where all parts are non-failing, but unexpected interaction between them causes a failure. They wrote because STAMP is based on systems and control theory rather than reliability theory, it is therefore a better approach for such complex systems. In their comparison between system analysis using FTA and STPA, they found that STPA identified the same possible causal factors as FTA, but they also identified other factors not considered with FTA. This shows that STPA is a more “potent” tool for analysis of such systems, and according to the authors, it is easier to use. Results from the analysis can be used as requirements to improve the design of the ATC system. These results can also be used for training and to develop operational procedures.

Another example of systemic methods in the aviation industry is Allison et al. (2017), who applied the STPA method to an aircraft rapid decompression event. The purpose of their paper is to show the application of STPA in complex systems such as aviation. (The complexity of aviation systems means they can be collectively considered a “system of systems” because they are a group of separate systems interacting to achieve a common goal. This topic is explored further later in the thesis.) By using STPA, Allison et al. (2017) were able to find many unsafe control actions for which safety constraints were then found. The safety constraints varied according to individual needs, like redundancy in the system, improvement of operating procedures, placement of warning alarms, etc. By analyzing the resultant safety constraints, improvements to training and future development of the system could then be achieved.

Likewise, STPA was also used by a team of researchers testing a low-cost unmanned subscale blended-wing model demonstrator (Lu et al. , 2015). Their first approach risk and safety was a “fly-fix-fly” method, but after losing 3 demonstrators they realized it was not a viable approach because they were not learning enough from previous accidents and were thus unable to make improvements. Therefore, they adopted a systems theoretic approach that takes into consideration the social and technical aspects of their design, manufacturing, and operation processes. By using STPA during the design and manufacturing stages of the fourth demonstrator, they found that they were able to reduce the accident rate from between 65% and 100% to less than 5%.

STPA was also used by the Japan Aerospace Exploration Agency in a joint project with the Massachusetts Institute of Technology (MIT) to analyze system safety of a space vehicle in the early design phase (Ishimatsu et al., 2010). The vehicle in question is an unmanned visiting vehicle that takes commodities and components to the International Space Station and is launched by the H-IIB rocket. Previously, such analysis had been done by NASA using FTA. In this project, the researchers decided to use STPA because the system is software intensive and analysis techniques like FTA and FMEA tend to focus on component failure and are not efficient in evaluating software failure.

Silva Castilho et al. (2018) also used STPA to study take-offs for light aircraft in crosswind conditions. Considering only the aircraft alone, it can be said that these are complex systems with physical and software components. But during take-offs, this system will interact with the pilots, the external environment (wind), and ATC. These are also part of the system, and it is therefore quite complex. In light of this complexity, they found it advantageous to use STPA instead of other, linear methods. By using STPA, they identified new safety constraints that were not found with more traditional methods. The safer flight controls can be incorporated into use by manufactures, maintenance crews, instructors, and pilots.

In conclusion, over the past several pages, examples of systemic methods being used in the aviation industry have been given, both with UAS and with manned aircraft systems. Such methods can be useful for the complex socio-technical systems within the aviation industry. Conversely, linear methods have been used effectively for decades, so there are roles for all types of methods to play within the aviation industry.

2.3.4. Automotive

Automation in the automobile industry is a heavily researched topic within the private sector. However, proprietary research and trade secrecy mean that little knowledge of this work exists in the public domain. Still, there are some publications that indicate the general challenges of, and approaches to, risk and safety management in the automation of automobiles.

In its 2018 publication entitled Safety-Related Challenges for Autonomous Systems, the SASWG identifies several important challenges relating to safety risks posed by automation in the automotive industry.

First, low level autonomous systems require the option that control of the vehicle can be deferred to the driver. This includes both planned and unplanned (emergency) handovers, meaning the driver must be aware, at all times, of the possibility that they will be required to control the vehicle. This presents a human factors risk where driver readiness must always be maintained by the vehicle. To avoid this problem, systems capable of handling emergencies (or fully autonomous systems) must be developed.

Second, if autonomous vehicles could be connected to their environments, or in other words, if they could communicate with other vehicles and infrastructure, their operation could be optimized. But this communication poses a challenge since it must (a) involve many systems in a seamless fashion while also (b) providing acceptable levels of security to the vehicles.

Third, interconnection of autonomous vehicles will allow for cloud-based user analysis for the purpose of identifying and preventing hazards in real time. This poses a risk to personal information, since user data could be accessed through malicious attacks on such systems. Integrity and assurance requirements must therefore be implemented.

Fourth, learning by the autonomous system of individual habits and behaviors might allow the creation of more refined experiences for the user. This learning must not, however, compromised safety in any other situations, such as those involving other vehicles where “non-learned” software is involved. Ensuring such learning does not negatively affect safety is yet another challenge.

Fifth, while learning by autonomous vehicle systems could be invaluable for improving safety of the system in real world situations, these data cannot be gathered before the systems have been launched. Reactive improvements to safety that occur only after accidents and problems occur are unacceptable, so the SASWG suggests that simulations should be employed as a substitute for real world experience. The simulations themselves must therefore be suitably validated and true to real world situations to ensure the safety of autonomous vehicles.

Sixth and final, the overall number of risks and accidents can be expected to decrease with the use of autonomous vehicles. However, roadways and infrastructure used by these vehicles will also be shared with other demographics like pedestrians and property owners. Should any transfer of risk be experienced or expected (e.g. if overall risk decreases but risks to pedestrians increase), decisions must be made to accept or reject these transfers. This is another example of the challenges posed to autonomous vehicles involving risk and safety.

An important question can now be asked: How can manufactures and regulators handle the autonomous systems by which these risks will be posed?

The US Department of Transportation (2016) recommends that autonomous vehicles follow a “robust design and validation process based on a systems-engineering approach with the goal of designing highly autonomous vehicle systems free of unreasonable safety risks.” This is in line with recommendations from Alexander et al. (2009), who call for certification of autonomous systems based on hazard identification and analysis.

The Department of Transportation also recommends that the overall process should follow guidelines from the international regulations on the functional safety of road vehicles, ISO 26262.

ISO 26262 is an important regulation for the current generation of piloted vehicles. However, it does not specify any method for safety analysis. Sequential methods like FTA and FMEA have been used for hazard analysis in recent applications of ISO 26262, but with automated systems, these methods may be unable to identify issues caused by dysfunctional component interactions, software failure, and human error (Abdulkhaleq et al., 2017).

ISO 26262 is therefore an inadequate resource for safety standards in autonomous vehicles (Sabaliauskaite et al., 2018). To address this issue, it has been suggested that the safety scope of ISO 26262 should be extended to include support for the hazard analysis and risk assessment process of automated systems (Abdulkhaleq et al., 2017).

But this raises an important question: How should international standards and regulations like ISO 26262 approach risk and safety management in autonomous vehicles? Further, how should the private sector approach the safety of their own systems?

Some current publications focus on systemic methods for uses relating to autonomous vehicles. Specifically, STPA has been recommended in several instances by researchers at the University of Stuttgart and Continental AG. For example, Abdulkhaleq et al. (2017) applied STPA to a fully autonomous vehicle at Continental AG, and they concluded that it was an effective and efficient approach to safety support of autonomous vehicles. This was after discounting sequential methods for use with such a complicated system. They also recommend updating ISO 26262 to include provisions concerning the use of systemic methods for safety management of autonomous vehicles. Likewise, Abdulkhaleq & Wagner (2013) employed STPA in adaptive cruise control to achieve an acceptable level of risk, and they concluded that it is a powerful tool for safety analysis in autonomous automotive systems.

Sabaliauskaite et al. (2018) also considered safety and security of autonomous vehicles. They explained that with increasing levels of automation, ISO 26262 is not an appropriate reference point for contemporary safety systems. Rather, they proposed an approach to safety and security processes that considers higher levels of automation. To this end, they wrote that STPA is a viable tool for autonomous automotive systems.

There is therefore some reliable evidence that systemic approaches to risk and safety management can be, and are being, adopted by the automotive industry for use with autonomous systems. However, this trend is not necessarily valid across the industry, as much industry-led research is not published.

2.3.5. Maritime

Automation within the maritime industry has received increasing interest over the past few years. While the industry has historically adapted very slowly to technological advances, the possible benefits of automation have spurred research in both academia and industry. For example, it is thought that autonomous maritime projects have been undertaken by Finland, the European Union, the United States, Japan, the United Kingdon, and Norway (Wahlström et al., 2015).

Schröder-Hinrichs et al. (2019) predict that in the future, ships will contain a plethora of automated technology. This includes everything from assistive vehicle functions (e.g.

track keeping and speed control) to full automation. They also predict that ship control will be centralized and that fully autonomous ships will exist within special autonomous ecosystems. Hogg & Ghosh (2016) wrote that such ecosystems might consist of offshore ports, between which autonomous ships will sail under human supervision. While berthing will still be conducted by specialized crews, they predict that up to six ships could be remotely monitored by one shore-based officer, with all normal operations conducted without human input. Automated technologies will also be used for ship maintenance and port duties, where specialized robots and drones will eventually supplant human workers.

Proponents of automated maritime vessels foretell two benefits of automation. The first is an increased level of safety for the crew, the ships, the environment, and coastal infrastructure. This is because autonomous systems mitigate (but do not eliminate) the possibility of human error, such as fatigued officers making poor decisions. Additionally, eliminating ship-based crews will prevent on-board accidents, injuries, and sicknesses while allowing skilled mariners to work more typical, shore-based day jobs in shore control centers (Wahlström et al., 2015).

The second benefit is a drop in costs as ship owners will need to employ fewer or no crew (Wróbel et al., 2016). In fact, it has been predicted that autonomous ships will reduce operating costs by 40% (Hogg & Ghosh, 2016) for three main reasons: A reduction in crew size means there will be fewer salaries to pay; elimination of bridge and accommodations means weight will be saved (making more room for cargo); and lower fuel consumption will be achieved due to slower sailing speeds. These speeds will be afforded because longer voyage times are typically opposed by crews, but with no crew, ships can spend longer at sea, thereby using less fuel.

An important reason that automation in the maritime industry is emerging beyond the idea-formulation stage is that such vessels are foreseeable given current technological paradigms (Ding et al., 2012). However, there are problems that must be addressed before these systems can be implemented on a wide scale.

First, there is a lack of supporting infrastructure for autonomous vessels, including the ports and communication systems that will be necessary for these ships to function. Additionally, monitoring and emergency control of these ships will have to be conducted from shore control centers. This presents the possibility for human errors distinct from those possible on manned ships, and they must be prevented or mitigated before the centers become operational. Wahlström et al. (2015) identified the following human factors issues facing operation of these centers: Information overload; boredom/fatigue; changeover mishaps; a lack of vessel feel; constant reorientation of new tasks; delays in control and monitoring; and the ability to understand humans (E.g. to distinguish help-seekers from pirates). Maritime autonomous support infrastructure will therefore have to be built with special regard to novel technologies and human factors so as to mitigate these issues. The lack of such infrastructure is being addressed, for example, in the Baltic Sea, where industry (companies like Rolls Royce and ABB) is supporting an undertaking to prepare the area for autonomous shipping by 2025 (Haikkola, 2017).

But perhaps an even bigger problem presented by automation within the maritime industry is the question of risk and safety. While automation promises many possible benefits, it does not necessarily ensure greater levels of safety, and such complex systems can have revenge effects that actually increase accident severity. Risk and safety management of these systems is therefore paramount. Unfortunately, the novelty and uniqueness of these systems mean there is no empirical data on which to base their risk and safety management.

Hogg & Ghosh (2016) wrote that an increased prevalence of automation will create a rise in disastrous partnerships between people and automated systems. Similarly, Ding et al. (2012) predicted that increased system complexity and tightness of coupling might increase the risk of catastrophic accidents in certain abnormal situations. This prediction was supported by the results of a simulation-based study, where it was found that navigators frequently failed to notice subtle errors in their automated navigational system. This indicates that automation decreases situational awareness.

Another simulation-based study by Pazouki et al., (2018) yielded the same results: That automation diminishes situational awareness and the ability of operators to safely monitor vessel progress. Similarly, these concerns have also been recognized in the aviation industry, where there are anecdotal examples of complex human-machine interactions leading to situations where recovery was impossible. Designing systems (including shore control center systems) that do not allow an unsafe drop in situational awareness is therefore paramount.

The difficulties of implementing automation in maritime systems were also broached by Jalonen et al., (2017). In consideration of the feasibility of automating several common ship types, they explored high level concepts with regard to safety and security. They concluded that automation could increase safety, but it is not guaranteed. They give three recommendations to achieve this result: (1) The minimum level of confidence in safety should be equal to or greater than current levels; (2) automation should be introduced slowly and incrementally; (3) international co-operation (e.g. with the International Maritime Organization (IMO)) should be pursued.

Regarding this last point, Hogg & Ghosh (2016) also identified the lack of IMO regulations concerning autonomous systems as problematic, writing that its collision regulations, for example, will have to be substantially updated before autonomous systems are implemented. There are further legal questions as well, such as how liability will be placed and how to decide which types of cargo are too hazardous for unmanned transport.

Further, practical issues with automation have also been identified by the SASWG (2018). First, vessels navigate at long distances from land. For unmanned ships, this means vessel operators and the vessels themselves will be separated by extremely large distances. Delayed and/or intermittent communication would be problematic, so robust communication systems will be a necessary. Such systems will be costly because expensive, high-quality satellites will have to be operated in areas with few customers (open ocean), thereby increasing costs for ships operators (Hogg & Ghosh, 2016). Additional precautionary systems in case of failed communications may also be necessary, such as a secondary independent communication system and automatic emergency procedures (heaving-to, dropping anchor, or drifting (Hogg & Ghosh, 2016)).

Second, weather is another challenge that autonomous maritime systems will face. While some autonomous systems (like planes) can often avoid inclement weather (by flying around storms), this is not an option for relatively slow and large maritime vessels. The autonomous systems must therefore maintain their capabilities throughout a range of extreme weather conditions.

Third, maritime vessels, like other autonomous systems, will be exposed to hostile actions. Data security will be one important consideration due to the volume of data generated by autonomous systems. Additionally, maritime systems are also threatened by physical piracy. The vessels should therefore have some defense mechanisms against piracy situations and any other unwanted/unapproved boarding. Importantly, these systems must be able to differentiate between piracy and approved/authorized boarding, such as of maintenance teams.

Finally, access to remote or otherwise underdeveloped coastal areas that do not possess autonomous vessel infrastructure may be desired. Namely, this includes areas with poor communication/monitoring connections. Vessels accessing such areas must therefore be somewhat self-reliant, and adequate levels of reliability must be achieved and demonstrated across a range of situations.

The previous pages contain brief discussions of a handful of important problems and hurdles concerning autonomous maritime systems (Table 1) . But in addition to being a simple enumeration of issues, they also illustrate how little precedent autonomous systems have within the maritime industry (e.g. very few fully autonomous ships have been built). It therefore follows that there is little published literature focused on risk and safety management of autonomous maritime systems.

No.	List of system characteristics
1	Increased prevalence of automation will create a rise in disastrous partnership between people and automated system
2	Complex system
3	Tight coupling
4	Decrease of situational awareness due to automation
5	Reduction of operators ability to monitor vessel progress due to automation
6	Minimum level of confidence in safety should be equal to or greater than current levels
7	Robust communication systems are necessary
8	Maintain capabilities throughout a range of extreme weather conditions
9	Data security
10	Defense mechanism against piracy (unwanted/unapproved boarding)
11	Self-reliant

Table 1: List of characteristics of maritime autonomous vessels

However, one related industry in which automation is common is that of Autonomous Underwater Vehicles (AUVs). Because these systems are often used to explore hazardous environments, the risk of vehicle loss can be high. Brito et al. (2010) constructed a model for calculating the risk of losing an AUV by eliciting expert opinion on the probabilities of loss in different situations. Their risk-of-loss model is one of few in the maritime industry that is used in real operations (the authors apply their model to their own AUV operations). While their model is not one of the common sequential, epidemiological, or systemic approaches, it is one example of risk and safety management of autonomous maritime systems.

Other authors have also approached this topic. Wróbel et al. (2016) recognized two problems regarding risk and safety for maritime autonomous systems. First, there is a poor understanding of what operational circumstances can be expected for these vessels. And second, there is a poor understanding of how these vessels will be designed. Thus, considering how little is currently understood about these systems, to ensure an appropriate level of safety, the authors conclude that intensive risk analysis should be performed before anything is built. To this end, they present a hazard and consequence analysis performed using a Bayesian Network (BN). Their work illustrates the importance of risk and safety management models of maritime autonomous systems.

Likewise, Wróbel et al. (2018) consider a method for analysis of maritime autonomous systems, only they use a systemic method (STPA) instead of a BN. They wrote that the systems theory basis of STPA should allow it to handle complex systems, and that a lack of empirical data prevents the use of many other methods. They therefore used it to construct a model for safety analysis, and they found that it allowed them to make design recommendations, despite a lack of data. They concluded that STPA is a useful tool for hazard mitigation within the autonomous maritime domain, but they also highly recommended further analysis once more data on these systems is gathered, and more still throughout the systems' lifecycles.

Another example of systemic methods in the maritime industry was given by Aps et al. (2016), who applied STPA to maritime traffic management systems. By using STPA, they were able to identify hazards and unsafe speeds/maneuvers as identified by the IMO collision regulations. The goal of this model was to improve ship-level situational awareness and to help enforce safety constraints. While their focus was not on autonomous systems, the researcher concluded with the following: "STPA has proved to be an effective and efficient method to assess the safety management of a complex safety-critical sociotechnical system from the maritime domain."

And in a pertinent paper by Montewka et al. (2018), three approaches to ensuring maritime autonomous system safety were investigated. Namely, a framework for risk-based design was reviewed, as was a goal-based approach. A STAMP-STPA approach was the third method investigated. The authors wrote that the approaches are not interchangeable, and that they are actually all useful at different levels of risk management. As regards STAMP-STPA, they wrote that it (and systemic methods in general) present disadvantages such as non-intuitive presentation of results, a research-practice gap, and a low level of confidence among users. However, as supported by expert elicitation and a literature review, they also wrote that its flexibility in analyzing hazards in a range of scenarios and its ability to guide early design will be useful for the future of autonomous shipping. They additionally noted that it is proficient in formulating safety controls, and that "these safety controls represent the basis for initiating the safety management strategy of MASS [Maritime Autonomous Surface Ships] and the entire autonomous maritime system(s)".

In conclusion, there are many big issues concerning automation in the maritime industry that must be addressed. Additionally, the novelty, the high cost, and the complexity of large vessels mean there are few empirical data or past experiences on which engineers and safety professionals can base future designs and management approaches. Successful risk and safety management of these systems will therefore be a great challenge, and it presents many unknown variables to designers. However, there have been a few investigations into possible management and analysis methods, which seem to indicate a preference for systemic methods, at least in academia. It is difficult, however, to predict how industry will manage the risk and safety of these yet-to-be-designed systems, and what methods will present the optimal trade-off between cost and effectiveness.

3. METHODOLOGY

From section 1.2.1, this thesis seeks to satisfy three goals:

- Identify the methods and frameworks for risk and safety management of autonomous transportation systems;
- Identify the key elements and issues for risk and safety management of autonomous transportation systems; and
- Specify the optimum method of risk and safety management of autonomous systems within the maritime industry.

Having completed the literature review of risk and safety management of autonomous systems, both in general and within different transportation sectors, the methods used have thus been identified. Additionally, key elements and issues for management of these systems have been recognized in the context of different transportation sectors. It is now necessary to consider how these methods might be applied to future autonomous systems within the maritime industry, and to compare them with the intention of identifying the optimal method(s).

This comparison will be completed using a qualitative approach and following the inductive method. Material will be reviewed in order to find commonalities and patterns between methods and to analyze their relevance and usefulness for the particular case of autonomous systems in the maritime industry. Final recommendations of the optimal choice(s) will then be made.

A qualitative approach is often viewed as the antithesis of a quantitative approach. That is, it forgoes empirical inquiry and instead focuses on observations of the subjects' qualities (Donmoyer, 2008). Such an approach is useful because if offers the possibility to compare, in the case of this thesis, the range of different risk analysis methods used in industry. If this were undertaken using a purely quantitative approach, insights afforded by the analysis would be unlike those achieved over the following pages because they would be based on high level empirical study rather than on individual characteristics. Additionally, quantitative analysis might not even be possible because few empirical data exist for methods like FRAM and STAMP, thereby inhibiting analysis based on empirical study. Further, and perhaps more importantly, the qualitative approach offers the possibility to analyze reasons why certain methods were chosen. With this approach one can therefore obtain a working understanding of the various methods and their applicability to different circumstances.

An inductive method will be used, where one first observes, continues by trying to find patterns or similarities, and then offers a proposition (Blackstone, 2014). Thus, one does not start with a theory which they then seek to prove or disprove, but rather, they develop a theory based on patterns recognized within their observations.

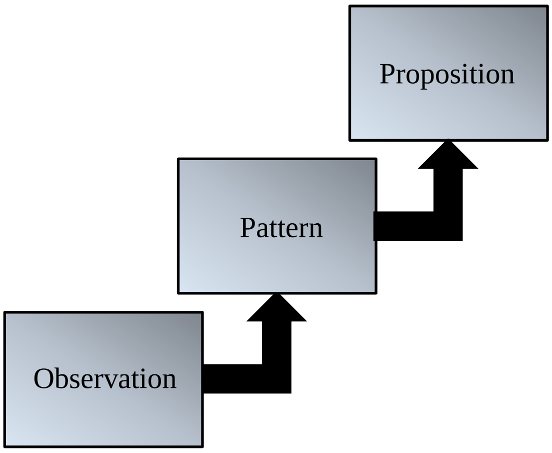


Figure 4: Inductive Method
Adapted from Trochim, William M.K.

The first step in making a proposal is to "observe", which has been completed in this case by conducting a literature review of the methods currently used for risk and safety management of autonomous systems. Next, an analysis is completed where "patterns" of use are identified within the previous observations. This will clarify the reasons for selecting one method over another as well as the advantages and disadvantages of one method compared to another. In this thesis, said analysis will be conducted using a SWOT methodology. Once the methods are understood, the final step is to "propose", based on these findings, an optimal risk and safety method (or methods) for autonomous maritime systems.

As mentioned above, the following analysis will be conducted using the SWOT methodology. Paraskevas (2013) defines SWOT as "an analytical tool for strategic management planning". Simply put, it is a tool for analyzing Strengths, Weaknesses, Opportunities, and Threats that allows for consideration of all the important internal and external aspects of a subject. Once these have been identified, informed decisions on the subject can be made.

In this case, this means risk and analysis management method(s) appropriate to autonomous maritime systems can be identified.

When considering what method to use, surveys and interviews were considered but decided against. On one hand, surveys are a fast and cost effective way to gather great amount of data but at the same time can be inflexible, because the questions that have been set cannot be modified in case they were not clear and also the validity of the answers can be questioned since the options for answers are limited (Blackstone, 2014). So surveys are not very useful when analyzing complex topics that need deep scrutinizing of the research topic. On the other hand, interviews are limited to the number of interviewees the researcher has access and who are knowledgeable in the research topic. For the case of this research, experts in risk and safety management as well as autonomous maritime vessels would have been required. Since the research on autonomous maritime vessels is still quite limited, there pool of experts is small and most probably the results would have been biased to their preference in risk and management methods. SWOT analysis was chosen as the best method to compare the different risk and safety methods. It eliminated the bias of the researcher and compares in equal way all the methods. Also the researcher doesn’t need to be an expert in either, risk and safety methods or autonomous maritime systems.

4. SWOT ANALYSES

As we have seen, the different risk analysis methods can be divided into three categories: Sequential, epidemiological, and systemic. Methods in each category share foundational concepts and have many similarities in the way they are applied. Methods in all three categories can be used in the process of risk and safety management.

A high-level SWOT analysis of each category has been conducted and the results are presented to better understand the use of each category, and when to use them and why. Within each of the three analyses, a set of general characteristics is considered so that the results can be fairly compared. That is, the same properties and trade-offs are considered for all three methods. For example, in the "strengths" section, sequential methods are said to be not resource intensive. Similarly, epidemiological methods are not resource intensive (at least, when compared to systemic methods). Conversely, systemic methods are very resource intensive, and it is therefore categorized as a "weakness".

The mentioned set of characteristics was compiled so that the three final arguments contain all the general and important points shared across each category. The list was compiled based upon the literature review, where instances of an author noting something about a method were recorded. After thoroughly reviewing the literature, all the most important and defining characteristics of the three categories had been recorded and compiled. Because there are over 100 methods (Appendix 1), not all possible characteristics can be considered, but the analyses are generally comprehensive so as to define the popular methods from each category.

In addition to the earlier example of "resource intensiveness", further considerations for strengths and weaknesses are as follows:

- What is the underlying theory? This affects the type of results produced.
- How fast is it to use?
- Are there clear guidelines and taxonomies available for use? This affects consistency of results and can have implications depending on the user's intentions. If consistency is required (e.g. for trend analysis), these are important considerations.
- How much expertise is required (both for the method and regarding the system itself)?

- How comprehensive and thorough are the results? How deep is the resultant understanding?
- How clear and easy to understand are the results, and are they presented in graphical format?
- Can it consider different types of system components? E.g. physical, human, software, and organizational.
- How popular is it?
- What kind of issues are identified? E.g. root cause, organizational problems, dangerous component interactions, or something else.
- Is apportioning blame possible/inevitable?
- Does it/can it yield probabilities of accident occurrence?

Regarding considerations of opportunities and threats, many are related to the applicability of the method to the system at hand. Additionally, consideration of the setting in which the method is being used must be made. They are as follows:

- What are the system characteristics? Some methods can be used with certain systems, and some cannot. Additionally, if a method can handle a system, it might be overly powerful and therefore more resources-heavy than necessary. The system characteristics are:
    - Complexity;
    - Component type (e.g. physical, software, human, and organizational);
    - Tightness of coupling;
    - Tractability (manageability);
    - How much, and what type, of data are known? Different methods require different types and amounts of data.
- What are the situational characteristics?
    - What is the desired level of thoroughness compared with resource use (this is essentially a trade-off between money/time and thoroughness of results)?
    - Is it desired that blame can be apportioned (or conversely, is apportioning of blame is undesired)?
    - Are probabilities of accident occurrence desired?
    - What are the industry regulations?

    - Are graphical and easy-to-disseminate results desired?
    - Is consistency necessary?
    - What is the level of user expertise and experience?

- Finally, is it popular in industry? Underwood & Waterson (2013) wrote that new methods take time to be proved and accepted by practitioners. This threatens the overall use of the method. It should be additionally noted that some users simply do not like to learn new techniques, further stymying acceptance of modern methods.

Having now explained the general considerations for the three categories of methods, the SWOT analyses can be found over the following pages. After these are the final recommendations and conclusions.

4.1. Sequential Methods

This category includes the Domino method and Fault Tree Analysis. It is the oldest category and these methods work well with simple systems consisting of mainly physical components. It is not effective for non-linear, complex systems that involve organizational and software components. According to sequential methods, accidents are resultant of a sequence of time-ordered, discrete events where a "root-cause" initiates the sequence in a deterministic and cause-effect fashion.

Strengths
• Simple, quick, and easy to use with guidance and taxonomies available.
• Useable by non-experts.
• First methods available; widely used and well-known.
• Can very effectively model simple systems with mainly component failures and human actions (Underwood & Waterson, 2013).
• Identifies root cause and allows apportioning of blame (Underwood & Waterson, 2014).
• Can yield clear and graphical results that can be easily disseminated to, and understood by, non-experts.
• Can yield probabilities of accidents (Alexander & Kelly, 2009).

- It is estimated that 96% of potential accidents can be captured by linear (sequential and epidemiological) methods (Hudson, 2014).

Weaknesses

- Do not identify issues pertaining to interactions of system components.
- Poor handling of managerial, organizational, human, and software components (Underwood & Waterson, 2013).
- Cannot comprehensively describe complex socio-technical systems, and the resultant understanding is not thorough. This is due to the linear thinking on which these methods are based (Yousefi et al., 2018).
- Can lead to incorrect/unjustified apportioning of blame, which additionally "represents a missed opportunity to learn important lessons about system safety" (Underwood & Waterson, 2014).
- Probabilities can be unrealistic and therefore dangerous.
- Assumes component failure modes are independent.
- Treats humans and software in the same way as mechanical hardware and assumes they fail in the same way (Fleming et al., 2013).

Opportunities

- These are optimal for systems with the following characterizes:
    - Simple (not complex);
    - Mainly physical components;
    - Loose coupling;
    - High manageability;
    - Empirical data are known.
- These are optimal for situations with the following characteristics:
    - Limited resources (time and money);
    - Identification of root cause/blame is desired;
    - Regulations require sequential methods;
    - Probabilities of accident occurrence are desired;
    - Dissemination of results to non-experts is desired;
    - Consistency of results is desired (e.g. clear guidance and taxonomies are beneficial);

    - The user has little expertise.
- These are popular methods in industry, meaning many users are comfortable with continuing to choose and use them.

Threats

- These are nonoptimal for systems with the following characteristics:
    - High levels of complexity in socio-technical systems;
    - Involving many human, organizational, and/or software components;
    - Tight coupling;
    - Low manageability;
    - Few empirical data known.
- These are nonoptimal in the following situations:
    - It is desired that no blame is assigned;
    - Thorough results are desired more than efficiency;
    - Regulations require other methods;
    - A comprehensive list of safety improvements to the system is desired.

SWOT: Sequential Methods

Strength	**Weaknesses**
• Simple, quick and easy to use • Useable by non-experts • Widely used and well known • Effectively model simple systems with component failure and human actions • Identifies root cause • Can yield probabilities of accidents	• Do not identify issues pertaining to interactions of system components • Poor handling of managerial, organizational, human and SW components • Cannot describe complex socio-technical systems • Can lead to incorrect appointing of blame • Probabilities can be unrealistic • Assumes component failure modes are independent • Treats humans and SW in the same way as mechanical HW assuming they fail in the same way
Opportunities	**Threats**
• Optimal for system with following characteristics: ○ Simple ○ Mainly physical components ○ Loose coupling ○ High manageability	• Nonoptimal for system with following characteristics: ○ High complexity in socio-technical systems ○ Involve many human, organizational, and/or SW components

○ Empirical data known • Optimal for situations where: ○ Limited resources ○ Need to identify rot cause ○ Regulation requires the use of sequential method ○ Probabilities of accident occurrence are desired ○ Consistency of results desired ○ User has little experience • Popular methods used in industry	○ Tight coupling ○ Low manageability ○ Few empirical data known • Nonoptimal for situations where: ○ Assignment of blame is not desired ○ Thorough results desired ○ Comprehensive list of safety improvements to the system is desired ○ Regulation requires other methods

Table 2: Sequential methods SWOT analysis summary

4.2. Epidemiological Methods

The most popular method in this category, and across all categories (Underwood & Waterson, 2014), is the Swiss Cheese method (Reason, 1990, 1997). Like sequential methods, these are based on cause-effect relationships. However, they do not only identify one root cause; they consider both active failures (such as failing equipment) and latent failures (such as a dangerous safety culture). Newer methods that are based on the Swiss Cheese Method (such as the ATSB investigation analysis model) are popular in industry and it has been shown that they can sometimes yield similar results to systemic methods (Underwood & Waterson, 2014).

Strengths
• Compared to systemic methods: ○ Simple, quick, and easy to use; ○ Yield fairly reliable and consistent results with guidance and taxonomies available; ○ Some more advanced methods can (in certain situations) approach similar results to systemic methods without the associated drawbacks (Underwood & Waterson, 2014; e.g. Sulaman et al., 2017); ○ Less resource intensive (money and time). • Identify both active failures (similar to sequential methods), and…

- Identify latent/organizational factors (thereby supporting a more comprehensive view than sequential methods).
- Widely used in industry (Underwood & Waterson, 2014).
- Can yield graphical outputs to aid in understanding/clarity of results and to aid in dissemination of results.
- Identify root cause and allows apportioning of blame.
- It is estimated that 96% of potential accidents can be captured by linear (sequential and epidemiological) methods (Hudson, 2014).

Weaknesses

- Not as comprehensive as systemic methods; cannot yield the same depth of results when considering complex socio-technical systems because they do not consider system dynamics and non-linear interactions between different components (Yousefi et al., 2018).
    - Underwood & Waterson (2014) wrote that some methods (namely the ATSB model) can blur the distinction between epidemiological and systemic methods by acting as a gateway for systems thinking. However, the closer one approaches to this effect, the more effort is required.
- Focusing on the root cause can lead to incorrect/unjustified apportioning of blame.
- Traditional epidemiological methods, like the Swiss Cheese Model, regard organizational mistakes as management errors.
- Accidents have more than one contributing factor.

Opportunities

- These are optimal for systems with the following characterizes:
    - Involve physical, human, and organizational factors;
    - Relatively few software components;
    - Tight coupling;
    - High manageability;
    - Empirical data are known.
- These are optimal for situations with the following characteristics:
    - Limited resources (time and money);
    - Identification of root cause/blame is desired;

    - Regulations require epidemiological methods;
    - Dissemination of results to non-experts is required;
    - Consistency of results is desired (e.g. clear guidance and taxonomies are beneficial).
- Popularity in industry means many practitioners are comfortable with these methods and that they will continue to be chosen and used by such practitioners.

Threats

- These methods are nonoptimal for systems with the following characteristics:
    - High levels of complexity in socio-technical systems;
    - Involving many software components;
    - Low manageability;
    - Few empirical data known.
- They are nonoptimal in the following situations:
    - It is desired that no blame is assigned;
    - Thorough results are desired more than efficiency;
    - A comprehensive list of safety improvements to the system is desired;
    - Regulations require other methods;
    - The user has no analysis experience.

SWOT: Epidemiological Methods

Strength	**Weaknesses**
• Compared to systemic methods: ○ Simple, quick and easy to use ○ Yield fairly reliable and consistent results ○ Some advanced methods can approach similar results to systemic methods ○ Less resource intensive • Can identify active failure • Can identify latent/organizational factors • Widely used • Can yield graphical outputs to clarify results • Identify root cause	• Cannot yield same depth of results in complex socio-technical systems compared to systemic methods • Focus on root cause identification can lead to incorrect blaming • Traditional epidemiological methods, like SCM regard organizational mistakes as management errors • Accidents have more than one contributing factor
Opportunities	**Threats**

• Optimal for system with following characteristics: ○ Involve physical, human and organizational factors ○ Relatively few SW components ○ Tight coupling ○ High manageability ○ Empirical data known • Optimal for situations where: ○ Limited resources ○ Need to identify rot cause ○ Regulation requires the use of epidemiological method ○ Dissemination of results to non-experts is required ○ Consistency of results desired ○ User has little experience • Popular methods used in industry	• Nonoptimal for system with following characteristics: ○ High complexity in socio-technical systems ○ Involve many SW components ○ Low manageability ○ Few empirical data known • Nonoptimal for situations where: ○ Assignment of blame is not desired ○ Thorough results desired ○ Comprehensive list of safety improvements to the system is desired ○ Regulation requires other methods ○ User has no analysis experience

Table 3: Epidemiological methods SWOT analysis summary

4.3. Systemic Methods

Two popular methods in this category are FRAM (Functional Resonance Analysis Method) and STAMP (Systems Theoretic Accident Model and Processes).

Both methods were designed based on systems theory. The strength of systemic methods is their ability to identify hazards stemming from interrelations of components, which is possible because they do not consider systems on a local, component basis. This allows identification of hazards that can occur even when all the components are functioning correctly on a local level but in such a way that they combine to create hazardous situations.

FRAM is a system modelling tool that can be used in the process of managing risk and safety. It identifies hazards by recognizing when variations in system functions affect each other and resonate to create noticeable hazards. This is something that sequential and epidemiological methods cannot detect because they are limited to cause-effect thinking.

While STAMP is also a systemic method, it has different theoretical underpinnings, means of application, and output than FRAM (Underwood & Waterson, 2012). It is the basis of STPA, which can be used in the process of risk and safety management, and CAST, which can be used for accident analysis (Thomas, 2014). It defines systems as “a hierarchy of control based on adaptive feedback mechanisms” (Leveson, 2004). The following SWOT analysis is focused more towards STPA than CAST because we are not interested in accident analysis.

Despite their theoretical differences, these two methods are based on systems theory, and this is what sets them apart from other methods like Swiss Cheese and Domino. Both FRAM and STAMP-STPA can be used within the process of risk and safety management, and so they are two main methods considered within this analysis.

Strengths

- Systems theory permits a greater depth of understanding of complex systems than is possible using sequential and epidemiological methods (Underwood & Waterson, 2013). This is because they consider non-linear relationships and the interactions of different components, recognizing that hazards and risks can arise even in situations where individual components are all acting correctly at a local level. They therefore encourage a comprehensive view of the system.
- They can handle all types of components (physical, human, organizational, software, etc.) by considering their interactions, not their specific behavior (e.g. see comparative study between STPA and FMEA by Sulaman et al. (2017)).
- These methods do not explicitly assign blame or identify singular root causes for accidents. Rather, they emphasize a wider view and a focus on safety improvements (Underwood & Waterson, 2014).
- They can be applied from an early stage of development onwards to eliminate or mitigate hazards (Abdulkhaleq et al., 2016; Ishimatsu et al., 2010).
- Handbooks exist for both STPA (Leveson & Thomas, 2018) and FRAM (Hollnagel et al. 2014) that were written by the creators of the two methods. They both provide guidance on how to employ the methods, although the FRAM handbook is specifically focused on healthcare whereas the STPA handbook is more general. Neither provide taxonomies.
- Both methods are relatively open-ended and flexible.
- They do not require empirical data.

- One instance of *accident analysis* (not hazard identification) by Yousefi et al. (2018) indicated that STAMP can identify far more safety recommendations and be more comprehensive than FRAM.

Weaknesses

- They are complicated to implement, meaning they can take longer to use and to learn, when compared with linear methods (Hollnagel & Speziali, 2008; Abdulkhaleq et al., 2013). Expertise in both the domain of study and safety analysis is required.
- They are more resource intensive (Underwood & Waterson, 2013; Abdulkhaleq & Wagner, 2013). They also require extensive system information.
- These methods do not focus on blame and they cannot (currently) yield probabilities of accident occurrence.
- Results are generally difficult to disseminate to non-experts, especially compared to linear methods. For example, Abdulkahaleq & Wagner (2013) wrote that STPA "has no detailed description, however, how to present the final argumentation about the hazards avoided and remaining risks".
- Analyst bias is more likely in systemic methods than in linear methods (Yousefi et al., 2018).
- These methods are open and developmental. While the handbooks are useful, these methods still require imagination and concerted effort.
    - Several authors have written that FRAM could benefit from a more structured approach (Herrera & Woltjer, 2010; Stringfellow, 2010). While the handbook might help with the above (Hollnagel et al., 2014), the author still maintains that the purpose of FRAM is to guide/control an analysis, not automate it. He explains that FRAM provides analysts with clues where to look, but not answers (Hollnagel, 2016).
    - Abdulkahaleq et al. (2013) wrote that "STPA needs a systematic method to notate the relation between the process model variables, control actions and hazards". However, they concluded that STPA generally has a systematic, step-by-step process of implementation. These steps can be found in books by Leveson (2011) and Leveson & Thomas (2018).

- Systemic methods can sometimes be less effective than some linear ones at finding pure component failures. For example, in a comparison study, Sulaman et al. (2017) found that FMEA identified more component failure hazards than STPA.
- These methods are qualitative and do not yield probabilities of accident occurrence. Alexander & Kelly (2009) contended that while quantitative modelling does have problems, abandoning it completely for qualitative methods is risky and unsafe.
- There are no taxonomies for either method.

Opportunities

- Systemic methods are optimal for systems with the following characteristics:
  - Complex;
  - Tightly coupled;
  - Low manageability;
  - Physical, software, human, and organizational components;
  - Information on the system is known.
- They are optimal for use in the following situations:
  - Ample resources are available;
  - Thorough results are desired;
  - System safety improvements are desired more than finding a root cause or assigning probabilities of accidents occurring;
  - Apportioning blame is not desired;
  - The user is very experienced in safety management and in the domain of analysis.
- Hudson (2014) estimates that 4% of possible accidents can only be captured using non-linear, non-deterministic thinking.

Threats

- These methods are nonoptimal for systems with the following characteristics:
  - Simple system;
  - Involving few human, software, and organizational components (e.g. mainly physical);
  - Loosely coupled;
  - High manageability;

    - Little system design information is known.
- They are nonoptimal in the following situations:
    - Efficiency is paramount (little time/money available);
    - Regulations require other methods (and they rarely recognize FRAM and STAMP-STPA);
    - The user is inexperienced;
    - If taxonomies are desired (e.g. for trend analysis);
    - If dissemination of results to non-experts is necessary (Underwood & Waterson, 2014);
    - Probabilities of accident occurrence are desired or necessary.
- These are less common than simpler methods, so their acceptance by safety professionals will be slow and it will be a long time before the methods can gain significant popularity in industry (Underwood & Waterson, 2012).

SWOT: Systemic Methods

Strength	**Weaknesses**
• System theory permits greater understanding of complex systems • Can handle all types of components • Focus on safety improvement instead of finding one root cause or assigning blaming • Can be used in early stage of development • Detailed handbooks written for both methods (STPA and FRAM) exist. • Methods are open-ended and flexible • Do not require empirical data	• Complicated to implement • Resource intensive • Cannot yield probabilities of accident occurrence • Results are difficult to disseminate to non-experts • Analyst bias • Require imagination and concerted effort • Can be less effective than some linear methods at finding pure components failures • Do not yield probabilities of accident occurrence • There are no taxonomies for either method
Opportunities	**Threats**
• Optimal for system with following characteristics: o Complex o Tight coupling o Low manageability o Physical, SW, human and organizational components o Information on the system is known • Optimal for situations where: o Ample resources available	• Nonoptimal for system with following characteristics: o Simple systems o Involve few human, SW and organizational components o Loosely coupled o High manageability o Little system design information known • Nonoptimal for situations where:

○ Thorough results desired ○ System safety improvement desired ○ Blaming not desired ○ Regulation requires the use of epidemiological method ○ User is very experienced in safety management and the domain of analysis	○ Few resources available (time and money) ○ Regulation requires other methods ○ User is inexperienced ○ Taxonomies are desired ○ Dissemination of results to non-experts is necessary ○ Probability of accident occurrence is desired

Table 4: Systemic methods SWOT analysis summary

5. RECOMMENDATIONS

In this chapter, the recommended risk and safety method will be presented. In order to choose the most suitable method to be used in the maritime industry with autonomous vehicles, was important to take into account some considerations. These considerations will be presented below. Then the recommended method will be nominated with explanation for the recommendation. Also an explanation on alternatives to satisfy the FSA is presented as well as possible ways to handle uncertainty.

5.1. Considerations

Before being able to recommend a risk and safety management method to be used when developing autonomous maritime vessels, it is important to take into account some considerations like the tractability and coupling of the system as specified by Underwood & Waterson (2013). Tractability and coupling of a system are not the only considerations, although are a clear starting point to understand the system being studied. It is also important to answer some understand other questions:

- Is the system to be analysed as a whole system, a system of systems or a subsystems?
- What is the system level of complexity and can the proposed method handle such system?
- What is the desired level of thoroughness in relation to the associated costs?
- What information is demanded by the method, and what information currently exist?
- What methods are currently in use?
- How should the results be used?

In the coming section the mentioned considerations will be explored in the case of autonomous maritime vessels.

5.1.1. Tractability and coupling

An important clarification must be made before a method can be proposed: What is the tractability and coupling of autonomous maritime systems? There is guidance in the literature for how to answer these questions. Underwood & Waterson (2013) describe a tightly coupled system as one whose components and subsystems are interconnected so that something affecting one can easily spread to the others. More formally, tightly coupled systems will include redundancies in their designs; will feature few delays in processes; will feature invariant process sequences; and will feature few substitutions of, and little slack in, supplies/equipment/personnel.

They also describe future system tractability. They explain that a system is intractable if its principles of functioning are unknown; it cannot be described simply and with few details; and it quickly changes over time. For example, a post office, with its regimented order, would be tractable. Conversely, a hospital emergency room, with its flexible and ever changing functionality, and would be intractable.

So, how can we describe maritime autonomous systems? While the vessels themselves do not yet exist, it is safe to say they will be intractable (low manageability) and tightly coupled. This is because these systems will be highly complex and self-reliant. They will also feature many redundancies and a wide range of advanced operational capabilities.

In other words, if a system will handle cargo loading, machinery maintenance, navigation, communication, security measures, emergency operations, and all the other necessary (sub)systems, it must be tightly coupled. And it will also be intractable for similar reasons. Its level of complexity, its number of interacting components, and its range of operational capabilities means it will be difficult to describe and that it will change quickly over time. These are the hallmarks of an intractable system.

Why is this clarification important? This question is clearly answered by Hollnagel & Speziali (2008). Axiomatically, they explained that if a system is tightly coupled and intractable, the methods used *must* be suitable for systems that are tightly coupled and intractable. You must therefore consider tractability and coupling in order to select the correct method. According to the Underwood & Waterson framework in figure 2, air traffic control, railways, and marine transport are all in the upper left corner, meaning these are tightly coupled systems with high manageability. For industries and processes in that space, epidemiological methods are considered adequate for explaining accidents and analyzing the system. Indeed, we can see from the literature review that the transportation industry has historically favored sequential and epidemiological methods for system analysis. But do autonomous transportation systems fit into this space?

Referring to the SWOT analyses, the opportunities/threats sections indicate the combination of tractability and coupling that is most suited to each category. It can be concluded that, on the basis of tractability and coupling alone, autonomous systems do not fit into this space. This indicates that systemic methods (and not epidemiological) are the best choice for analysis of maritime autonomous systems.

5.1.2. Other considerations

However, tractability and coupling are not the only important considerations. Following is a list of questions (adapted from Underwood & Waterson (2013) and from the SWOT analyses in section 4) concerning both the context of autonomous maritime systems and the provisional category of choice; systemic methods. Answering these questions will indicate the optimal choice.

Is the system to be analyzed a system of systems, a whole system, or a subsystem? Generally, automation in the maritime industry will encompass all three of these categories. In consideration of a fully autonomous vessel, it will contain many automated subsystems that will combine with other subsystems (e.g. navigation, propulsion, communications, etc.). Because these vessels will be operated in open ocean, they will interact with many other manned and autonomous systems (e.g. other ships, ports, and control infrastructure). In this respect, comprehensive analysis of autonomous vessels will mean analysis of a SoS, where the analyst must operate with a very wide scope.

What is the system's level of complexity and can the proposed method handle such systems? Maritime autonomous systems have the potential to be extremely complex, especially if considering their position within the broader maritime environment. In consideration of the maritime industry as a whole, the literature reflects this fact; that the size and complexity of these autonomous systems will be great and, in some aspects, revolutionary (e.g. Jalonen et al., 2017; Wróbel et al., 2018).

Even in consideration of only one independent vessel, the scope would include a range of physical, software, human, and organizational components. On this basis, some authors indicate that sequential and epidemiological methods would be insufficient (e.g. Stringfellow, 2010; Yousefi et al., 2018). Regarding accident models, Yousefi et al. 2018 wrote, "Sequential and epidemiological accident models do not fully capture the dynamics and nonlinear interactions between system components in complex socio-technical systems". This idea is mirrored in the SWOT analyses. Therefore, the complexity of autonomous maritime systems seems to demand analysis using systemic methods, which are the only ones capable of capturing a comprehensive and dynamic view of such systems.

What is the desired level of thoroughness in relation to the associated costs? As indicated in the SWOT analyses, systemic methods are the most costly in terms of time, effort, required expertise (in both the domain and in system analysis), and monetary expenditure. However, the complexity of maritime autonomous systems is unavoidable, and comprehensive, thorough, and safe analysis might necessitate the use of systemic methods, regardless of costs.

What information is demanded by the method, and what information currently exists? The SWOT analyses indicate that empirical data/evidence are necessary for many sequential and epidemiological methods, especially those that are probability based. Conversely, systemic methods require a complete understanding of the system, but no empirical data. While such a thorough understanding of the system might be difficult to grasp, it is a beneficial trade-off for systems that do not yet exist because there will be little associated empirical data (only expert elucidation).

Are there any regulatory constraints that prohibit or require certain methods over others? For ships on international voyages, it will be necessary that they comply with international regulations. This namely refers to the International Maritime Organization (Jalonen et al., 2017).

So, future autonomous ships will have to be built and operated in accordance with IMO (and other international) regulations. However, there currently are no such regulations for dealing with autonomous maritime systems. The IMO's Formal Safety Assessment (FSA) methodology (IMO, 2018) is meant to be employed in such circumstances, where it can guide the creation and proposal of regulations in terms of the costs/benefits associated with new maritime technologies. While it is not intended that the FSA is implemented in every proposal, its use is especially recommended in cases with far-reaching implications for cost to the industry and in terms of legislative/administrative burdens. In cases with unclear circumstances (such as the use of novel technologies), the FSA allows member states to more clearly understand the proposal and make the appropriate decision.

Currently, the FSA requires probabilistic modelling. This is perhaps the greatest practical hurdle when it comes to safety analysis and management of autonomous maritime systems. This is because systemic methods cannot currently be used for quantitative analysis and probability-based methods can seemingly not handle the wide scope presented by this new technology. This is a very noteworthy problem and it will be discussed later.

What methods are currently in use? From the literature review, we know that historical analyses of transportation systems employed sequential or epidemiological techniques. Autonomous transport systems in all industries are too young to conclusively identify preference for any one method. However, examples of systemic methods used for analysis of complex autonomous systems in other industries (e.g. automobile) indicate possible applicability in similar maritime systems.

How shall the results be used? If the results must be disseminated to non-experts, systemic methods are not as suited to this purpose as simpler methods. This is related to the FSA issue because the reason it requires probabilities is so that analysis results can be presented clearly and understood by non-experts. However, this issue of ease-of-dissemination is practical rather than safety based. If the ultimate goal of risk and safety management is to make safety improvements to the system, this should be the deciding factor of which method to use.

On that topic, another consideration is whether the goal is indeed to make safety improvements, or if it is identification of ostensibly clear root causes/issues that can be conveniently excised. Stringfellow (2010) wrote, "Many incident reporting and learning schemes used in organizations focus on identifying root causes. This is problematic because there is no such thing as a root cause. The selection of a root cause, like the selection of an initiating event in an accident, is a matter of convenience". So if the ultimate goal is to implement safety upgrades, desire of root cause identification should be ignored.

A final practical consideration is if a taxonomy is desired. If results classification is paramount, methods with published taxonomies and rigid guidelines would be better (in this respect) than more open-ended systemic methods. However, the number of hazards and safety improvements that one identifies would most likely suffer. And conversely, the open and developmental nature of systemic methods may be preferred by some analysts if traditional guidance is not well suited to their application.

5.2. Which Method Should be Used?

A wide majority of the important considerations mentioned above indicate that systemic methods are the optimal choice. It is additionally indicated that STPA would produce more comprehensive results than FRAM (or any other single systemic method). STPA is therefore the provisional recommendation. This does not, however, preclude future abandonment of STPA in favor of more effective alternatives. Rather, it has been shown that STPA is the optimal choice if one method (that currently exists) is to be used for establishing a risk and safety management strategy for autonomous maritime systems.

In addition to the open possibility that something else could supplant STPA as the optimal choice, two additional concessions must be made. First the issue of the IMO's FSA regulations will be discussed. Second, it is important to remember that "no specific method is the overall best in the sense that it can be used for all conditions" (Hollnagel & Speziali, 2008). In other words, while STPA is the recommended choice, and while this recommendation has been supported by various academic sources, it is not the necessarily optimal in all situations. Additionally, this recommendation does not mean that other methods could not be used for equal or even greater results, only that the analysis indicates that STPA is the optimal choice.

The reasons for this choice are numerous. Recommendations of systemic methods for autonomous system analysis in the maritime industry can be found throughout the literature (e.g. Thieme et al. 2018; Jalonend et al., 2017). For example, while referring specifically to autonomous maritime systems, Jalonen et al. wrote, "the System-Theoretic Process Analysis (STPA, see Leveson (2011)) model is believed to give the most promising result." Likewise, Thieme et al. postulated that the optimal risk model should be able to handle control and software systems, and that STPA or FRAM are two possible choices for this purpose.

In 2013, Underwood & Waterson published a review of 302 documents to find all explicit references of systemic models. They found that STAMP was referenced 52% of the time, while FRAM was mentioned 19.9% of the time. This relative popularity is not undeserved. In a 2018 comparative study of FRAM and STAMP accident analysis models, Yousefi et al. (2018) recommended STAMP for future accident modelling because it identified the most safety recommendations. They concluded that STAMP "is likely to be more instrumental and comprehensive in generation of recommendations" and urged its adoption in industry to demonstrate its capabilities. While STPA is not the same as STAMP, it is based on STAMP and so successes of STAMP do, to a degree, indicate the benefits of STPA.

There are many other proponents of STAMP and STPA. For example, STPA has been used successfully for hazard and safety analysis of autonomous vehicles at Continental AG (Abdulkhaleq et al., 2016) and within software-intensive aerospace systems at MIT and

JAXA (Ishimatsu et al., 2010). Likewise, STPA was recommended for safety assurance purposes in the aviation industry by Fleming et al. (2012).

STPA has also found some support within the maritime industry. For example, Aps et al. (2016) wrote, "STPA has proved to be an effective and efficient method to assess the safety management of a complex safety-critical socio-technical system from the maritime domain". Another example of its use, but concerning autonomous systems, is found within Wróbel et al. (2018), who successfully used STPA to compile a list of hazards and solutions related to the safety of autonomous vessels.

A final example can be found in Valdez Banda & Goerlandt (2018), who proposed a STAMP-based approach to designing maritime safety management systems. They applied their process to vessel traffic services in Finland in an attempt to systematically represent its functionality and controls. The outcome of this case study culminated with a defined safety management system along with the provision of a tool for monitoring its performance.

So, based on the above arguments, STPA is recommended over FRAM and all other sequential and epidemiological methods. This does not mean that STPA is objectively superior to FRAM, only that the literature and SWOT analyses indicate that it is the optimal choice for this application. The reasons systemic methods are recommended instead of sequential or epidemiological methods were explored in the previous subsection. To reiterate, it was shown that autonomous maritime systems will be tightly coupled and intractable. It was also concluded that to comprehensively and safely analyze the system, the scope of analysis must include the whole maritime environment. The system will therefore be a SoS and the analysis must have very wide boundaries. Further, the system will be highly complex and will include a variety of different components (human, physical, software, organizational, etc.). From the SWOT analyses, it can be seen that these system characteristics demand systemic tools of analysis. Therefore, STPA (or FRAM) is the logical choice.

However, there are associated drawbacks with analysis using STPA, hence the use of "provisional" to qualify earlier recommendations. First, systemic methods are the most costly in terms of time, effort, required expertise, and monetary expenditure (as shown in the SWOT analyses). The use of STPA will therefore require experienced, specialized professionals and greater financial investments than would be required for simpler methods.

Second, very little is known about the designs and operational concepts of autonomous maritime systems. This means that it is impossible to conclusively recommend one method over another because assurance of a method in the stated role cannot be given

(Thieme et al., 2018). This is why Wróbel et al. (2018) recommended that further analyses (using STPA) should be conducted throughout the design, construction, and operational phases; as more system data is revealed, more reliable results will be achieved, and it will become possible to ensure that the chosen method is appropriate.

Finally, and as previously explained, for the IMO to permit international use of autonomous systems, the FSA will have to be applied. And because the FSA requires probabilistic analysis, it cannot currently be satisfied by STPA (or FRAM). The important issue here is not that STPA cannot be used, but rather, that linear/probabilistic methods (some of which can satisfy the FSA) are not appropriate for this application.

If international regulations were already in place, and if industry practitioners wanted to perform system analysis, STPA could be used very effectively. But for analyzing the maritime autonomous ecosystem during application of the FSA, probabilistic methods must be used. And yet, neither do the necessary empirical data exist, nor can sequential/epidemiological methods appropriately handle a technological paradigm with such a wide scope.

5.3. Satisfying the FSA

To satisfy the IMO while still conducting a safe and comprehensive analysis, there are several possible courses of action. Namely, either the probability requirement in the FSA should be altered/ignored by the IMO or a systemic method should be augmented with probabilistic capabilities.

For the FSA to be altered or for IMO member states to deem its probabilistic stipulations unnecessary, it would (at least) have to be made very clear that probabilistic analysis is impossible and/or unsafe for autonomous maritime systems. In comparing, for example, STPA and the Domino Method, it is true that the former would be a safer choice than the latter. However, the distinction between methods is not always so obvious, and successfully arguing that quantitative modelling is categorically more unsafe than qualitative would be very challenging. Additionally, there are reasons to desire probabilistic results beyond simply satisfying the FSA. Montewka et al. (2018) wrote that quantitative analysis can provide valuable guidance to the designers of novel systems by reflecting available background knowledge and highlighting areas that require further research. Additionally, probabilities make prioritizing and disseminating results very clear because they show if something is more or less safe than something else. Quantitative system analysis is also supported by

Alexander & Kelly (2009), who argued that despite the drawbacks associated with probabilistic modelling, it would be dangerous to completely abandon it in favor of qualitative methods. So, foregoing the FSA requirements and avoiding probabilistic modelling might be very dangerous even if it were “permitted” by the IMO.

Alternatively, to satisfy the FSA by obtaining accident probabilities, perhaps a systemic method could be augmented with probabilistic capabilities. In this case, the model will most likely require empirical data, and efforts to obtain this data will be necessary (there are very few experts whose elucidation could be confidently relied upon, so empirical data are the only viable option for quantitative modelling (Wróbel et al., 2018)). One option is to use trial autonomous ships for beta testing to gather enough information for empirical analysis. These ships could be operated within national waters, bypassing IMO regulations altogether. Alternatively, the SASWG (2018) has proposed the use of simulations to gather empirical data for safety analysis within the automotive domain, thereby avoiding the need for test vehicles. Applying this idea to the maritime domain might help industry to obtain the necessary data more easily than by using test ships. In this case, the simulations themselves will require suitable validation to guarantee trustworthy results.

In addition to obtaining empirical data, finding a suitable probabilistic model for satisfying the FSA will be another challenge. This is because, as previously explained, neither STPA nor FRAM can currently handle quantitative analysis (Hollnagel et al., 2014; Leveson & Thomas, 2018). If such an ability could be added to a systemic method, it could then potentially satisfy the FSA guidelines. Alternatively, novel systemic methods with probabilistic capabilities might work, as could a combination of a systematic method with a more traditional, probability-based method. These three possibilities are hypothetical, and would require research and assurance before they are used. Still, there are reasons to support the use of combined methods, as discussed below.

Theoretically, combining methods is an effective way to ensure completeness of results. Because some methods have strengths in certain areas and weaknesses in others, choosing complementary methods allows the user to achieve more comprehensive results. A solution to the FSA problem might therefore be to combine the holistic modelling abilities of a systemic method with supplementary probabilistic modelling from a sequential or epidemiological method. This is not the first time that combining methods has been proposed. Underwood & Waterson (2013) explained that a combined sequential-systemic method would benefit from the sequential method’s ability to analyze technical failures and the systemic method’s strengths in analyzing wider issues. Similarly, for the purpose of hazard identification and management of general autonomous systems, Alexander et al. (2009) proposed a combination of ETBA, FFA, and HAZOP for comprehensive analysis from different viewpoints. Practitioners have also exemplified the benefits of using multiple methods. In the aviation industry, the Bowtie method has seen use for risk assessment by several national aviation regulators and air navigation service providers (CAA UK, 2015). This is noteworthy because Bowtie is essentially a combination of FTA, ETA, and CFC, therefore demonstrating that several different methods can be combined for greater effectiveness. Unfortunately, the specific combination of methods that could be used to satisfy the FSA and produce safe results is unknown. Whether it would consist of two separate methods applied consecutively or an integration of multiple methods is also unknown. However, as we have seen, it is a viable proposal that deserves future consideration.

5.4. Handling Uncertainty

One final recommendation for the use of STPA is that its users should remember to consider uncertainty in their results. If it is true that the purpose of risk and safety analysis is to help with underlying decision making, then it is logical that analysts should consider the consequences of their error. Uncertainties concerning their analyses must therefore be identified. Otherwise, unsafe decisions can inadvertently be made due to unsafe assumptions and untrustworthy data (Wróbel et al., 2018). By identifying uncertainties, better and safer decisions can be made, including implementation of protective measures. For methods to deal with uncertainty relating to STPA, see Wróbel et al. (2018) and Leong et al. (2017). Additionally, Montewka et al. (2018) wrote that for future quantitative and probabilistic analysis regarding the FSA, qualitative uncertainty in the data should be considered, and how risk is define and approached should be altered to suit the strengths of the data.

6. CONCLUSIONS

In conclusion, maritime autonomous systems are complex socio-technical systems that can pose severe risks to human safety. To handle such risks, three categories of system modelling methods exist that can be employed in the process of risk and safety management. They are sequential methods (the oldest category), epidemiological methods (the most popular category), and systemic methods (the newest category).

Their evolution over time has coincided with advances in technology and socio-technical interactions. Originally, it was thought that accidents were unfortunate and inevitable events (Perrow, 1984). Sequential models portrayed them as the result of individual component failures somewhere in the system. In recent years, the trend has moved towards viewing accidents as the result of dysfunctional interactions between system components. This marked the advent of systemic models. These models take better account of the complexity of modern socio-technical systems. In addition to physical components, these methods consider the effects that humans, organizations, and software can have, thereby incorporating them into the system.

A literature review was conducted to investigate both the general uses of each category of method as well as the different approaches to risk and safety management that are taken within different domains of the transportation industry. These domains are aviation, railway, automotive, and maritime. Additionally, the prevalence of automation within these domains was considered along with the contemporary approaches to risk and safety management of these autonomous transportation systems.

The literature review answered the first research question, “What methods and frameworks are implemented for the management of risk and safety in the different industries involving autonomous transportation systems?”. Although all categories of methods are used, there is a preference for the use of epidemiological methods due to them being simple to use, well know, no expertise required for their use and they are recommended by most regulatory organizations. It was also established, that in the case of autonomous systems, there is a tendency to use systemic methods which are newer and expertise is needed to apply them properly but are better able to handle such systems.

The second research question, “what are the key elements and issues for risk and safety management of autonomous transportation systems” was also answered by the literature review and summarized for maritime autonomous systems below:

- Complex system
- Tight coupling
- Decrease of situational awareness due to automation
- Reduction of operators ability to monitor vessel progress due to automation
- Minimum level of confidence in safety should be equal to or greater than current level
- Robust communication system necessary
- Maintain capabilities throughout a range of extreme weather conditions
- Data security
- Defense mechanism against piracy
- Self-reliant

Following this literature review are three SWOT analyses, one for each category of methods. The analyses contain the strengths, weaknesses, opportunities, and threats presented by each method. Following this is a consideration of the demands and requirements presented by automation within the maritime industry.

In consideration of these demands, the optimal method for risk and safety management of autonomous maritime systems was indicated by the literature review and the SWOT analyses to be STPA (Systems Theoretic Process Analysis). It was also shown that uncertainty analysis is an important addition to STPA. This answered the third research question, "what is the optimum method of risk and safety management for autonomous systems within the maritime industry?".

This recommendation is not made without certain acknowledgements and concessions. STPA (a systemic method) presents its own practical difficulties, including its resource-heavy nature and the necessity of analysts to be system experts. It also produces complex results that are qualitative and not probabilistic, which can be problematic if the results must be disseminated to laymen and non-experts.

Additionally, this recommendation is only provisional because little information on the system is known and there is little precedence on which to base the recommendation (only preliminary comparative studies and recommendations from other industries). Methods other than STPA should therefore not be excluded from consideration because they might be more effective choices for design, implementation and validation of future systems. It is also important that, regardless of the method chosen, continuous system analysis be performed, from the concept stage to the implementation stage, as more system data are revealed and as more literature on the use of the method is produced.

There is one final issue: For the future international use of autonomous maritime systems, a Formal Safety Assessment must be conducted. As per guidelines from the International Maritime Organization (IMO, 2018), such an assessment must include probabilistic results. STPA in its current form is unable to satisfy this requirement, and it is believed that probabilistic methods (which can satisfy the FSA) are not comprehensive enough to produce trustworthy results. Unless the FSA guidelines are altered to permit purely qualitative analysis (which is not recommended), a new analysis method is necessary that can produce holistic safety improvements (like STPA) while also producing accident probabilities to satisfy the FSA. Because this new method will produce probabilities, it will require empirical data. Two possible ways to obtain this data are using beta test ships or validated simulations. And whether this new method should be an augmentation of a systemic method like STPA or FRAM, a combination of a systemic method with a probabilistic method, or an altogether novel approach, is unknown.

6.1 Recommendations for future research

The development of autonomous vessels is still at an early stage but there is research devoted to this topic and there are institutions and companies working towards the development of such systems. As future research, it would be useful to apply STAMP and FRAM during the design and development of autonomous systems and compare the outcomes with both models and the resource requirements. Going further, would be interesting to do a similar research applied specifically to maritime autonomous vessels

As development of autonomous systems increases, there will be more research done on the topic of risk and safety for such systems, therefore a comprehensive study of the methods used specifically for autonomous vehicles would improve the understanding of the strengths and weaknesses of the different methods classification and possibly support the conclusions of this thesis.

7. REFERENCES

Abdulkhaleq, A. & Wagner, S. (2013). Experiences with applying STPA to software-intensive systems in the automotive domain, 2013 STAMP Workshop, MIT, March 26-28. Retrieved from https://d-nb.info/106953319X/34

Abdulkhaleq, A., Lammering, D., Wagner, S., Röder, J., Balbierer, N., Ramsauer, L., Raste, T., & Boehmert, H. (2016). A Systematic Approach based on STPA for developing a dependable architecture for fully automated driving vehicles. 4th European STAMP Workshop, Zürich, Switzerland, September 13-15. Available from https://doi.org/10.1016/j.proeng.2017.03.094

Abdulkhaleq, A., Wagner, S., Lammering, D., Boehmert, H., & Blueher, P. (2017). Using STPA in compliance with ISO 26262 for developing a safe architecture for fully automated vehicles, 2017 STAMP Workshop, MIT, March 29. Retrieved from https://arxiv.org/abs/1703.03657

Alexander, R. (2007). Using simulation for systems of systems hazard analysis (Doctoral dissertation, University of York). Retrieved from https://www.cs.york.ac.uk/ftpdir/reports/2007/YCST/21/YCST-2007-21.pdf

Alexander, R., Kelly, T., & Herbert, N. (2008). Structuring safety cases for autonomous systems, 3rd IET International Conference on System Safety, Birmingham, UK, October 20-22. Retrieved from https://www-users.cs.york.ac.uk/~rda/iet_2008_paper_final.pdf

Alexander, R. & Kelly, T. (2009). Escaping the non-quantitative trap, 27th International System Safety Conference, Huntsville, Alabama. Retrieved from https://www-users.cs.york.ac.uk/~rda/non-quant%20paper%20final%20v2.pdf

Alexander, R., Kelly, T., & Herbert, N. (2009). Deriving safety requirements for autonomous systems, 4th SEAS DTC Technical Conference, Edinburgh, UK. Retrieved from https://www-users.cs.york.ac.uk/~rda/seas_paper_2009.pdf

Allison, C., Revell, K., Sears, R., & Stanton, N. (2017). Systems Theoretic Accident Model and Process (STAMP) safety modelling applied to an aircraft rapid decompression event. Safety Science, 98, pp.159-166. Available from https://doi.org/10.1016/j.ssci.2017.06.011

Amazon Prime. (2015). Revising the airspace model for the safe integration of small unmanned aircraft systems. Retrieved from https://utm.arc.nasa.gov/docs/Amazon_Revising%20the%20Airspace%20Model%20for%20the%20Safe%20Integration%20of%20sUAS[6].pdf

Ancel, E., Capristan, F., Foster, J., & Condotta, R. (2017). Real-time risk assessment framework for Unmanned Aircraft System (UAS) Traffic Management (UTM), 17th AIAA Aviation Technology, Integration, and Operations Conference, Denver, Colorado, June 5-9. Available from http://doi.org/10.2514/6.2017-3273

An, M., Chen, Y., & Baker, C. (2011). A fuzzy reasoning and fuzzy-analytical hierarchy process based approach to the process of railway risk information: A railway risk management system. Information Sciences, 181(18), pp. 3946-3966. Available from https://doi.org/10.1016/j.ins.2011.04.051

Aps, R., Fetissov, M., Goerlandt, F., Kujala, P., & Piel, A. (2016). Systems-Theoretic Process Analysis of maritime traffic safety management in the Gulf of Finland (Baltic Sea), 4th European STAMP Workshop, Zürich, Switzerland, September 13-15. Available from http://doi.org/10.1016/j.proeng.2017.03.090

Aps, R., Goerlandt, F., Fetissov, M., Kopti, M., & Kujala, P. (2016). STAMP-Mar based safety management of maritime navigation in the Gulf of Finland (Baltic Sea), 2016 European Navigation Conference, Helsinki, Finland, May 30-June 2. Available from https://doi.org/10.1109/EURONAV.2016.7530538

Aven, T. (2015). Risk Analysis, John Wiley & Sons, Ltd.

BBC. (2018, March 20). Uber halts self-driving car tests after death. Retrieved from http://www.bbc.com/news/business-43459156

Belmonte, F., Schön, W., Heurley, L., & Capel, R. (2011). Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway traffic supervision. Reliability Engineering and System Safety, 96(2), pp. 237-249. Available from https://doi.org/10.1016/j.ress.2010.09.006

Blaauwgeers, E., Dubois, L., & Ryckaert, L. (2013). Real-time risk estimation for better situational awareness, 12th International Federation of Automatic Control Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Las Vegas, Nevada, August 11-15. Available from https://www.doi.org/10.3182/20130811-5-US-2037.00036

Blackstone, A. (2014). Principles of sociological inquiry: Qualitative and quantitative methods. (n.p.): Saylor Foundation. Retrieved from https://saylordotorg.github.io/text_principles-of-sociological-inquiry-qualitative-and-quantitative-methods/index.html

Boslaugh, S. (2013). Railroad automation technology. In Garrett (Ed.), Sage Encyclopedia of Transportation: Social Science and Policy (pp. 1125-1127). Thousand Oaks, California: SAGE Publications, Inc.

Brito, M., Griffiths, G., & Challenor, P. (2010). Risk analysis for autonomous underwater vehicle operations in extreme environments. Risk Analysis: An International Journal, 30(12), pp. 1771-1788. Available from https://doi.org/10.1111/j.1539-6924.2010.01476.x

Civil Aviation Authority United Kingdom. (2015). Bowtie? Retrieved from https://www.caa.co.uk/Safety-Initiatives-and-Resources/Working-with-industry/Bowtie/

California High-Speed Rail Authority. (2013). Safety and security management plan (RFP No.: HSR 14-32). Retrieved from https://www.hsr.ca.gov/docs/programs/construction/HSR_13_06_B3_PtB_Sub6_Safety_Security_Management_Plan.pdf

Causevic, A. (2016). Risk assessment in autonomous system of systems - a review (Literature review, Mälardalen University, School of Innovation, Design and Engineering). Retrieved from https://pdfs.semanticscholar.org/6166/75236633a571dd6c72779e3ef120f1f16010.pdf

Clothier, R., Fulton, N., & Walker, R. (2008). Pilotless aircraft: The horseless carriage of the twenty-first century? Journal of Risk Research, 11(8), pp. 999-1023. Available from https://doi.org/10.1080/13669870802323353

Clothier, R., Palmer, J., Walker, R., & Fulton, N. (2011). Definition of an airworthiness certification framework for civil unmanned aircraft systems. Safety Science 49(6), pp. 871-885. Available from https://doi.org/10.1016/j.ssci.2011.02.004

Clothier, R., Walker, R., Fulton, N., & Campbell, D. (2006). A casualty risk analysis for Unmanned Aerial System (UAS) operations over inhabited areas, 12th Australian International Aeronautical Congress, Melbourne, Australia, March 19-22. Retrieved from https://pdfs.semanticscholar.org/5b3e/514dfd463d681526c4dda125bc195fd2197f.pdf

Dalamagkidis, K., Valavanis, K., & Piegl, L. (2008). Current status and future perspectives for unmanned aircraft system operations in the US. Journal of Intelligent and Robotic Systems, 52(2), pp. 313-329. Available from https://doi.org/10.1007/s10846-008-9213-x

Department of Transportation. (2016). Federal automated vehicle policy. Retrieved form https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016

Despotou, G., Alexander, R., & Kelly, T. (2009). Addressing challenges of hazard analysis in systems of systems, 3rd Annual IEEE Systems Conference, Vancouver, Canada, March 23-26. Available from http://doi.org/10.1109/SYSTEMS.2009.4815793

Ding, S., Duanfeng, H. & Zhang, B. (2012). Impact of automation to maritime technology, 2nd International Conference on Computer and Information Application, Taiyuan, China. Retrieved from https://download.atlantis-press.com/article/4040/pdf

Dong, A. (2012). Application of CAST and STPA to railroad safety in China (Master's thesis, MIT). Retrieved from http://sunnyday.mit.edu/safer-world/Airong-thesis.pdf

Donmoyer, R. (2008). Quantitative research. In Given (Ed.), Sage Encyclopedia of Qualitative Research Methods (pp. 714-718). Thousand Oaks, California: SAGE Publications, Inc. Available from https://dx.doi.org/10.4135/9781412963909.n361

European Railway Agency. (2009). Guide for the application of the commission regulation on the adoption of a common safety method on risk evaluation and assessment as referred to in article 6(3)(a) of the railway safety directive (ERA/GUI/01-2008/SAF). Retrieved from https://www.era.europa.eu/sites/default/files/activities/docs/guide_for_application_of_cms_en.pdf

Federal Railroad Administration. (2009). A practical risk assessment methodology for safety-critical train control systems (DOT/FRA/RDV-09-01). Retrieved from https://www.fra.dot.gov/elib/document/306

Ford Media Center. (2017, February 10). Ford invests in Argo AI, a new artificial intelligence company, in drive for autonomous vehicle leadership. Retrieved from https://media.ford.com/content/fordmedia/fna/us/en/news/2017/02/10/ford-invests-in-argo-ai-new-artificial-intelligence-company.html

Fleming, C., Spencer, M., Leveson, N., & Wilkinson, C. (2012). Safety assurance in NextGen (Nasa Technical Report NASA/CR-2012-217553). Retrieved from http://sunnyday.mit.edu/papers/NASA-CR-2012-217553.pdf

Fleming, C., Spencer, M., Thomas, J., Leveson, N., & Wilkinson, C. (2013). Safety assurance in NextGen and complex transportation systems. Safety Science, 55(3), pp. 173–187. Available from https://doi.org/10.1016/j.ssci.2012.12.005

Gonçalves, P., Sobral, J., & Ferreira, L. (2017). Unmanned aerial vehicle safety assessment modelling through petri Nets. Reliability Engineering & System Safety, 167, pp. 383-393. Available from https://doi.org/10.1016/j.ress.2017.06.021

Haikkola, P. (2017). Roadmap towards commercial autonomous shipping in 2025 [PowerPoint slides]. Retrieved from: https://www.oneseaecosystem.net/wp-content/uploads/sites/2/2017/08/onesea_roadmaps-august-2017_paivi-haikkola_rev.pdf

Herrera, I. & Woltjer, R. (2010). Comparing a multi-linear (STEP) and systemic (FRAM) method for accident analysis. Reliability Engineering & System Safety, 95(12), pp. 1269-1275. Available from https://doi.org/10.1016/j.ress.2010.06.003

Hirling, O. & Holzapfel, F. (2017). O.R.C.U.S. risk assessment tool for operations of light UAS above Germany. International Journal of Intelligent Unmanned Systems, 5(1), pp. 2-17. Available from http://doi.org/10.1108/IJIUS-08-2016-0006

Hogg, T. & Ghosh, S. (2016). Autonomous merchant vessels: Examination of factors that impact the effective implementation of unmanned ships. Australian Journal of Maritime & Ocean Affairs, 8(3), pp. 206-222. Available from http://dx.doi.org/10.1080/18366503.2016.1229244

Hollnagel, E. (2004). Barriers and accident prevention. Aldershot, UK: Ashgate Publishing Limited.

Hollnagel, E. (2008). Risk + barriers = safety? Safety Science, 46, pp. 221-229. Available from https://doi.org/10.1016/j.ssci.2007.06.028

Hollnagel, E. (2012). FRAM: The Functional Resonance Analysis Method: Modelling complex socio-technical systems. Farnham, UK: Ashgate.

Hollnagel, E. (2014). Safety-I and safety-II: The past and future of safety management. Farnham, UK: Ashgate.

Hollnagel, E. (2016). Strengths and weaknesses of the FRAM. Retrieved from http://functionalresonance.com/strengths-and-weaknesses/index.html

Hollnagel, E. & Speziali, J. (2008). Study on developments in accident investigation methods: A survey of the “state-of-the-art” (Report for Swedish Nuclear Power Inspectorate, SKI report 2008:50). Retrieved from https://hal-mines-paristech.archives-ouvertes.fr/hal-00569424

Hollnagel, E., Hounsgaard, J., & Colligan, L. (2014). FRAM - a handbook for the practical use of the method. Middlefart, Denmark: Centre for Quality.

Hudson, P. (2014). Accident causation models, management and the law. Journal of Risk Research, 17(6), pp. 749-764. Available from http://dx.doi.org/10.1080/13669877.2014.889202

International Civil Aviation Organization. (2011). Unmanned Aircraft Systems (UAS) (Cir 328), p.54. Montreal, Canada: ICAO. Retrieved from https://www.icao.int/Meetings/UAS/Documents/Circular%20328_en.pdf

International Civil Aviation Organization. (2013). Safety Management Manual (SMM), Third Edition (Doc 9859). Montreal, Canada: ICAO. Retrieved from https://www.icao.int/safety/safetymanagement/documents/doc.9859.3rd%20edition.alltext.en.pdf

Ishimatsu, T., Leveson, N., Thomas, J., & Katahira, M. (2010). Modelling and hazard analysis using STPA, Conference of the International Association for the Advancement of Space Safety, Huntsville, Alabama, May 19-21. Retrieved from https://dspace.mit.edu/openaccess-disseminate/1721.1/79639

Jalonen, R., Tuominen, R., & Wahlström, M. (2017). Safety of unmanned ships. Helsinki, Finland: Aalto School of Engineering. Available from http://urn.fi/URN:ISBN:978-952-60-7480-1

Klockner, K. & Toft, Y. (2015). Accident modelling of railway safety occurrences: The Safety And Failure Event Network (SAFE-Net) methods, 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, Las Vegas, Nevada, July 26-30. Available from https://doi.org/10.1016/j.promfg.2015.07.487

Leong, C., Kelly, T., & Alexander, R. (2017). Incorporating epistemic uncertainty into the safety assurance of socio-technical systems. Electronic Proceedings in Theoretical Computer Science, 259, pp. 56-71. Retrieved from https://www-users.cs.york.ac.uk/~rda/CREST%20paper%20(Chris%20Leong)%20Final.pdf

Leveson, N. (1995). Safeware: System safety and computers. Boston, Massachusetts: Addison-Wesley Professional.

Leveson, N. (2004a). "A systems-theoretic approach to safety in software-intensive systems," IEEE Transactions on Dependable and Secure Computing, vol. 1, no. 1, pp. 66-86, doi: 10.1109/TDSC.2004.1.

Leveson, N. (2004). A new accident model for engineering safer systems. Safety Science, 42(4), pp. 237-270. Available from https://doi.org/10.1016/S0925-7535(03)00047-X

Leveson, N. (2011). Engineering a safer world: Systems thinking applied to safety. Cambridge, Massachusetts: MIT Press.

Leveson, N. & Thomas, J. (2018). STPA handbook.

Logan, M. & Glaab, L. (2017). Failure mode effects analysis and flight testing for small unmanned aerial systems, 17th AIAA Aviation Technology, Integration, and Operations Conference, Denver, Colorado, June 5-9. Available from https://doi.org/10.2514/6.2017-3270

Lu, Y., Zhang, S., Tang, P., & Gong, L. (2015). STAMP-based safety control approach for flight testing of a low-cost unmanned subscale blended-wing-body demonstrator. Safety Science, 74, pp. 102-113. Available from https://doi.org/10.1016/j.ssci.2014.12.005

Macchi, L. (2011). A resilience engineering approach for the evaluation of performance variability: Development and application of the Functional Resonance Analysis Method for air traffic management safety assessment (Doctoral thesis, École Nationale Supérieure des Mines de Paris). Retrieved from https://pastel.archives-ouvertes.fr/pastel-00589633

Marhavilas, P.K., Koulouriotis, D., & Gemeni, V. (2011). Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009. Journal of Loss Prevention in the Process Industries, Volume 24, Issue 5, pp. 477-523. Available from https://doi.org/10.1016/j.jlp.2011.03.004

Melnyk, R., Schrage, D., Volovoi, V., & Jimenez, H. (2014). Sense and avoid requirements for unmanned aircraft systems using a target level of safety approach. Risk Analysis, 34(10), pp. 1894-1906. Available from http://doi.org/10.1111/risa.12200

Menon, C. & Alexander, R. (2018). Ethics and the safety of autonomous systems. UK: Safety-Critical Systems Club. Retrieved from http://eprints.whiterose.ac.uk/127572/

Mitre. (2014). System of Systems. Online. Available at: https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/systems-of-systems

Montewka, J., Wróbel, K., Heikkila, E., Valdez Banda, O, Goerlandt, F., & Haugen, S. (2018). Challenges, solution proposals and research directions in safety and risk assessment of autonomous shipping, Probabilistic Safety Assessment and Management PSAM 14, Los Angeles, California, September 16-21.

Paraskevas, A. (2013). Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. In K. Penuel, M. Statler, & R. Hagen (Ed.), Encyclopedia of Crisis Management (pp. 913-914). Thousand Oaks, California: SAGE Publications, Inc.

Pazouki, K., Forbes, N., Norman, R., & Woodward, M. (2018). Investigating on the impact of human-automation interaction in maritime operations. Ocean Engineering, 153, pp. 297-304. Available from https://doi.org/10.1016/j.oceaneng.2018.01.103

Perow, C. (1984). Normal Accidents: living with high risk technologies. New York, Basic books.

Project Wing. (2018). Retrieved from https://x.company/projects/wing/

Qureshi, Z. (2007). A review of accident modelling approaches for complex socio-technical systems. Defense Science and Technology Organization Edinburgh (Australia) Command Control Communications and Intelligence Div. Available from https://apps.dtic.mil/sti/pdfs/ADA482543.pdf

Rae, A. & Alexander, R. (2011). Is the "system of systems" a useful concept for hazard analysis? 29th International System Safety Conference, Las Vegas, Nevada, August 8-12. Retrieved from https://www-users.cs.york.ac.uk/~rda/sos_Rae-Alexander-Revised.pdf

Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem. Safety Science, 27(2-3), pp. 183-213. Available from https://doi.org/10.1016/S0925-7535(97)00052-0

Reason, J. (1990). The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society B, 327, pp. 475-484. Available from http://doi.org/10.1098/rstb.1990.0090

Reason, J. (1997). Managing the risks of organizational accidents. Farnham, UK: Ashgate.

Robertson, M., Hettinger, L., Waterson, P., Noy, Y., Dainoff, M., Leveson, N., Carayon, P., & Courtney, T. (2015). Sociotechnical approaches to workplace safety: Research needs and opportunities. Ergonomics, 58(4), pp. 650-658. Available from https://doi.org/10.1080/00140139.2015.1011241

Rodrigues de Carvalho, P. (2011). The use of Functional Resonance Analysis Method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. Reliability Engineering & System Safety, 96(11), pp. 1482-1498. Available from https://doi.org/10.1016/j.ress.2011.05.009

Roland, H. & Moriarty, B. (1983). System safety engineering and management. New York, New York: Wiley.

Sabaliauskaite, G., Cui, J., & Liew, L. (2018). Integrating autonomous vehicle safety and security analysis using STPA method and the Six-Step Model. International Journal on Advances in Security, 11(1 & 2), pp. 160-169. Retrieved from https://www.researchgate.net/publication/326504334_Integrating_Autonomous_Vehicle_Safety_and_Security_Analysis_Using_STPA_Method_and_the_Six-Step_Model

Safety of Autonomous Systems Working Group. (2018). Safety-related challenges for autonomous systems (SCSC-143). UK: Safety-Critical Systems Club. Retrieved from https://scsc.uk/scsc-143

Salmon, P., Cornelissen, M., & Trotter, M. (2011). Systems-based accident analysis methods: A comparison of Accimap, HFACS, and STAMP. Safety Science, 50(4), pp. 1158-1170. Available from https://doi.org/10.1016/j.ssci.2011.11.009

Schröder-Hinrichs, J., Song, D., Fonseca, T., Lagdami, K., Loer, K., & Shi, X. (2019). Transport 2040: Automation, technology, employment - The future of work. Retrieved from http://dx.doi.org/10.21677/itf.20190104

Silva Castilho, D., Soto Urbina, L., & de Andrade, D. (2018). STPA for continuous controls: A flight testing study of aircraft crosswind takeoffs. Safety Science, 108, pp. 129-139. Available from https://doi.org/10.1016/j.ssci.2018.04.013

Sklet, S. (2004). Comparison of some selected methods for accident investigation. Journal of Hazardous Materials. Volume 111, Issues 1–3, pp. 29-37. Available from https://doi.org/10.1016/j.jhazmat.2004.02.005.

Stringfellow, M. (2010). Accident analysis and hazard analysis for human and organizational factors (Doctoral dissertation, Massachusetts Institute of Technology). Retrieved from http://sunnyday.mit.edu/safer-world/MaggieStringfellowDissertation.pdf

Sulaman, S., Beer, A., Felderer, M., & Höst, M. (2017). Comparison of the FMEA and STPA safety analysis methods-a case study. Software Quality Journal. pp. 1-39. Available from https://doi.org/10.1007/s11219-017-9396-0

Thieme, C., Utne, I., & Heugen, S. (2018). Assessing ship risk model applicability to marine autonomous surface ships. Ocean Engineering, 165, pp. 140-154. Available from https://doi.org/10.1016/j.oceaneng.2018.07.040

Thomas, J. (2014). Systems Theoretic Process Analysis (STPA) tutorial [PowerPoint slides]. Retrieved from http://psas.scripts.mit.edu/home/wp-content/uploads/2014/03/Systems-Theoretic-Process-Analysis-STPA-v9-v2-san.pdf

Trochim, William M.K. research Methods Knowledge based. Available from https://conjointly.com/kb/deduction-and-induction/

Underwood, P. & Waterson, P. (2012). A critical review of the STAMP, FRAM and Accimap systemic accident analysis models. In N. Stanton (Ed.), Advances in Human Aspects of Road and Rail Transportation (pp. 385-394). Boca Raton, Florida: CRC Press.

Underwood, P. & Waterson, P. (2013). Accident analysis models and methods: Guidance for safety professionals. Loughborough, UK: Loughborough University. Retrieved from https://www.researchgate.net/publication/259339662_Accident_Analysis_Models_and_Methods_Guidance_for_Safety_Professionals

Underwood, P. & Waterson, P. (2014). Systems thinking, the Swiss Cheese Model and accident analysis: A comparative systemic analysis of the Grayrigg train derailment using the ATSB, AcciMap and STAMP models. Accident Analysis & Prevention, 68, pp. 75-94. Available from https://doi.org/10.1016/j.aap.2013.07.027

Valdez Banda, O. & Goerlandt, F. (2018). A STAMP-based approach for designing maritime safety management systems. Safety Science, 109, pp. 109-129. Available from https://doi.org/10.1016/j.ssci.2018.05.003

Wahlström, M., Hakulinen, J., Karvonen, H., & Lindborg, I. (2015). Human factors challenges in unmanned ship operations - insights from other domains, 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, Las Vegas, Nevada, July 26-30.

Washington, A., Clothier, R., & Silva, J. (2017). A review of unmanned aircraft system ground risk models. Progress in Aerospace Sciences, 95, pp. 24-44. Available from https://doi.org/10.1016/j.paerosci.2017.10.001

Wienen, H.C.A., Bukhsh, F.A., Vriezekolk, E., Wieringa, R.J. (2017). Accident Analysis Methods and Models — a Systematic Literature Review. 10.13140/RG.2.2.11592.62721.

Wilson, A., Barr, J., & Zagorski, M. (2017). The feasibility of counting songbirds using unmanned aerial vehicles. The Auk: Ornithological Advances, 134(2), pp. 350-362. Retrieved from https://doi.org/10.1642/AUK-16-216.1

World Bank Group. (2017). Unmanned Aircraft Systems technology. Retrieved from http://documents.worldbank.org/curated/en/895861507912703096/pdf/120422-REVISED-UAS-Web-final.pdf

Wróbel, K., Krata, P., Montewka, J., & Hinz, T. (2016). Towards the development of a risk model for unmanned vessels design and operations. TransNav: International Journal on Marine Navigation & Safety of Sea Transportation, 10(2), pp. 267-274. Retrieved from https://www.doi.org/10.12716/1001.10.02.09

Wróbel, K., Montewka, J., & Kujala, P. (2018). Towards the development of a system-theoretic model for safety assessment of autonomous merchant vessels. Reliability Engineering and System Safety, 178, pp. 209-224. Available from https://doi.org/10.1016/j.ress.2018.05.019

Young, W. & Leveson, N. (2013). Systems thinking for safety and security, 29th Annual Computer Security Applications Conference, New Orleans, Louisiana, December 9-13. Available from http://dx.doi.org/10.1145/2523649.2530277

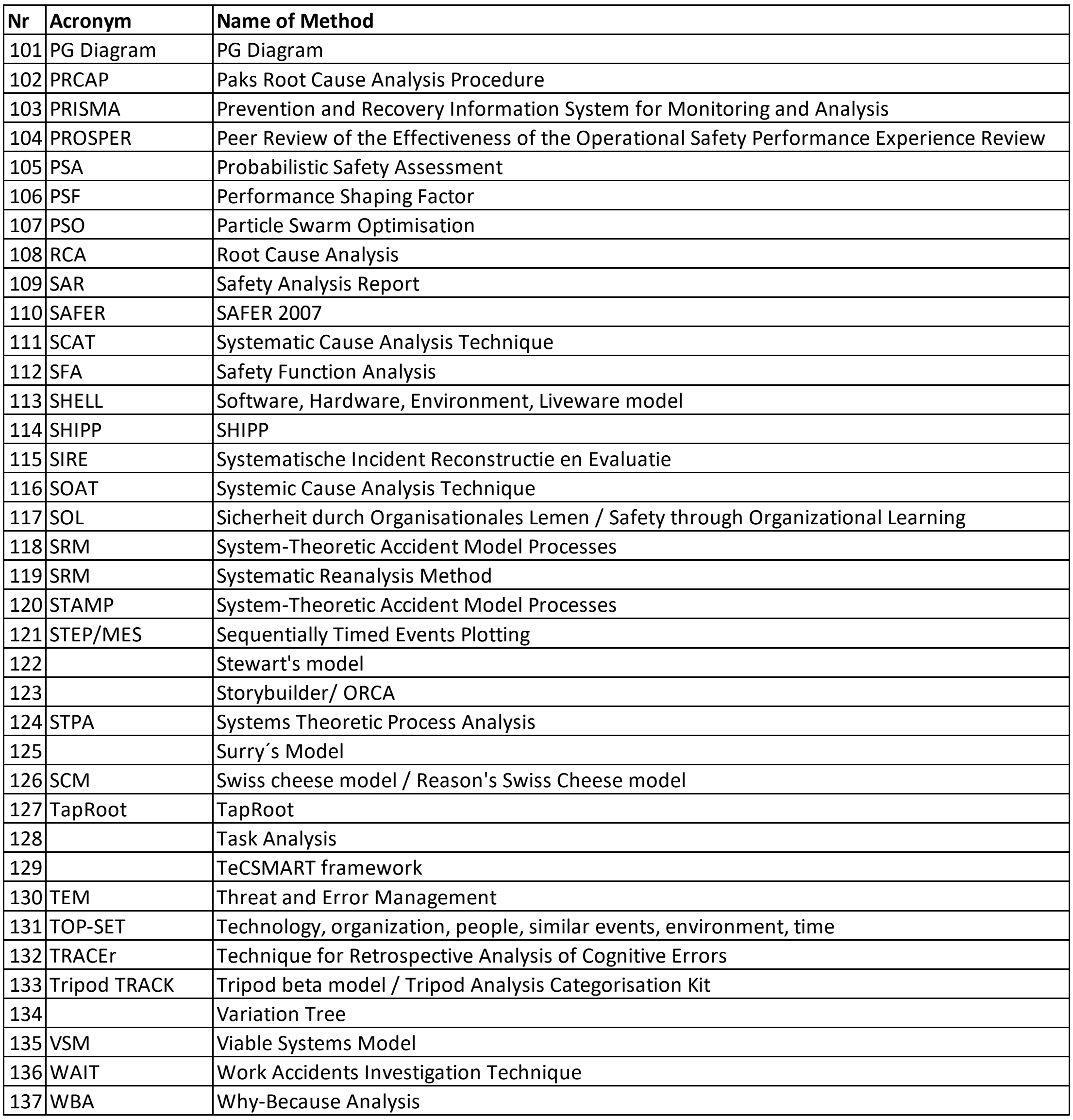
Yousefi, A., Rodriguez Hernandez, M., & Lopez Peña, V. (2018). Systemic accident analysis models: A comparison study between AcciMap, FRAM, and STAMP. Process Safety Progress, 2018. Available from http://doi.org/10.1002/prs.12002

8. APPENDICES

Appendix A: List of Risk and Safety Models

Nr	Acronym	Name of Method
1	24Model	24 Model (24.0)
2	3CA	Control, Change Cause Analysis, CCCA
3	3D Analysis	3 D Analysis
4	3 M / 5 M model	3 M / 5 M model
5	4M4E	4M4E
6	5 WHYS	5 WHYS
7		Accident Liability
8		Accident epidemiology
9	AcciMap	AcciMap
10	AcciTree	AcciTree
11	Achilles	Achilles
12	AEB	Accident Evolution and Barrier function
13	ArcGIS	
14	Apollo RCA	Apollo Root Cause Analysis / ARCA
15	APS	Accident Prototypical Scenario
16	APT	APT
17	ASSET	Assessment of Safety Significant Event Team
18	ATHEANA	A Technique for Human Event Analysis
19	ATSB	Australian Transport Safety Bureau (ATSB) accident investigation model
20	BA	Barrier Analysis
21		Birds model / Bird's accident causation model
22	BN	Bayesian Network
23		Bow-Tie model
24	CA	Change Analysis
25	CAS-HEAR	Computer-Aided System for Human Error Analysis and Reduction
26	CASMET	Casualty Analysis Methodology for Maritime Operations
27	CAST	Causal Analysis using STAMP
28	CBA	Cost Benefit Analysis
29	CCDM	Cause-Consequence Diagram Method
30	CCF	Common Cause Failure
31	CDM	Construction Design and Management Risk Assessment
32	C-HFACF	Complex Human Factor Analysis and Classification Framework
33	CFC	Causal Factor Charting
34	CIAF	Canadian Incident Analysis Framework
35	CIT	Critical Incident Technique
36	COA	Change Optimisation Algorithm
37	COCOM	Contextual Control Model
38	CREAM	Cognitive Reliability and Error Analysis Method
39	CRT	Current Reality Tree
40	CTM	Causal Tree Method
41	Domino	Domino Theory
42	DREAM	Driver's Reliability and Error Analysis Method
43	DWACN	Directed Weighted Accident Causation Network
44	ECFA	Events and Casual Factors Analysis
45	ECFA+	Ecents and Conditional Factors Analysis
46	ECFC	Events and Casual Factors Charting
47	ECFCA	Events and Causal Factors Charting and Analysis
48	EEA	Elementary Event Analysis
49	ESReDA	European Safety Reliability and Data Association
50	ETA	Event Tree Analysis

Nr	Acronym	Name of Method
51	ETBA	Energy Trace and Barrier Analysis
52		ETT/ EARM
53	FFA	Functional Failure Analysis
54	FMEA	Failure Mode and Effect Analysis
55	FRAM	Functional Resonance Analysis Method
56	FSA	Formal Safety Assessment
57	FTA	Fault Tree Analysis
58		Hale's model
59	HAZAN	Hazard Analysis
60	HAZOP	Hazard and Operability analysis
61	HEP	Human Error Probability
62	HERA	HERA
63	HF	Human Factors
64	HFACS	Human factors Analysis and Classification System / Human Reliability Analysis
65	HFIT	Human Factors Investigation Tool
66	HINT – J-HPES	HINT – J-HPES
67	HIP	Human Information Processing
68	HOE	Human and Organizational Errors
69	HOF	Human or Organizational Factors
70	HOT-PIE	Human, Organization, Technology, Process, information and Environment
71	HPEP	Human Performance Evaluation Process
72	HPES	Human Performance Enhancement System
73	HPIP	Human Performance Investigation Process
74	HRA	Human Reliability Analysis
75	HSE256	HSE256 -Health and Safety Executive
76	HSG245	Health and Safety Guidance
77	IAAM	Incident or Accident Analysis Method
78	INES	International Nuclear Event Scale
79		Influence Diagram
80	IPICA	IPICA
81	IRS	Incident Reporting System
82	ISIM	Integrated Safety Investigation Methodology
83	JAGMAN	Judge Advocate General method
84		Junior
85		Kitagawa's model / Kitagawa's accident causation model
86		Lawrence's model
87	LEADSTO	LEADSTO
88		Lee
89	LL	Lessons Learned
90	LOPA	Layer of Protection Analysis
91	MES	Multilinear Events Sequencing
92	MIA	Multi-Incident Analysis
93	MORT	Management Oversight and Risk Tree
94	MTO	Man, Technology and Organization analysis
95	OAC	OAC
96	OARU	Occupational Accident Research Unit
97	OOGM	OOGM
98	OIT	OIT
99		Orbit Intersecting theory
100	PEAT	Procedural Event Analysis Tool



Nr	Acronym	Name of Method
101	PG Diagram	PG Diagram
102	PRCAP	Paks Root Cause Analysis Procedure
103	PRISMA	Prevention and Recovery Information System for Monitoring and Analysis
104	PROSPER	Peer Review of the Effectiveness of the Operational Safety Performance Experience Review
105	PSA	Probabilistic Safety Assessment
106	PSF	Performance Shaping Factor
107	PSO	Particle Swarm Optimisation
108	RCA	Root Cause Analysis
109	SAR	Safety Analysis Report
110	SAFER	SAFER 2007
111	SCAT	Systematic Cause Analysis Technique
112	SFA	Safety Function Analysis
113	SHELL	Software, Hardware, Environment, Liveware model
114	SHIPP	SHIPP
115	SIRE	Systematische Incident Reconstructie en Evaluatie
116	SOAT	Systemic Cause Analysis Technique
117	SOL	Sicherheit durch Organisationales Lemen / Safety through Organizational Learning
118	SRM	System-Theoretic Accident Model Processes
119	SRM	Systematic Reanalysis Method
120	STAMP	System-Theoretic Accident Model Processes
121	STEP/MES	Sequentially Timed Events Plotting
122		Stewart's model
123		Storybuilder/ ORCA
124	STPA	Systems Theoretic Process Analysis
125		Surry´s Model
126	SCM	Swiss cheese model / Reason's Swiss Cheese model
127	TapRoot	TapRoot
128		Task Analysis
129		TeCSMART framework
130	TEM	Threat and Error Management
131	TOP-SET	Technology, organization, people, similar events, environment, time
132	TRACEr	Technique for Retrospective Analysis of Cognitive Errors
133	Tripod TRACK	Tripod beta model / Tripod Analysis Categorisation Kit
134		Variation Tree
135	VSM	Viable Systems Model
136	WAIT	Work Accidents Investigation Technique
137	WBA	Why-Because Analysis

The table above shows many of the existing risk and safety models, it is not a comprehensive list. This list was compile taking into consideration the models mentioned in the books articles and publications reviewed for this thesis like, Underwood & Waterson (2013), Marhavilas et al. (2011), Qureshi (2007), Hollnagel & Speziali (2008), Wienen et al. (2017), Sklet 2004, Despotou et al. (2009), Hudson (2014), International Civil Aviation Organization (2013) and Roland & Moriarty (1983).