

ORIGINAL ARTICLE

Open Access



Leading indicators and maritime safety: predicting future risk with a machine learning approach

Lutz Kretschmann

Correspondence: lutz.kretschmann@cml.fraunhofer.de

Fraunhofer Center for Maritime Logistics and Services CML, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany

Abstract

The shipping industry has been quite successful in reducing the number of major accidents in the past. In order to continue this development in the future, innovative leading risk indicators can make a significant contribution. If designed properly, they enable a forward-looking identification and assessment of existing risks for ship and crew, which in turn allows the implementation of mitigating measures before adverse events occur. Right now, the opportunity for developing such leading risk indicators is positively influenced by the ongoing digital transformation in the maritime industry. With an increasing amount of data from ship operation becoming available, these can be exploited in innovative risk management solutions. By combining the idea of leading risk indicators with data and algorithm-based risk management methods, this paper firstly establishes a development framework for designing maritime risk models based on safety-related data collected onboard. Secondly, the development framework is applied in a proof of concept where an innovative machine learning-based approach is used to calculate a leading maritime risk indicator. Overall, findings confirm that a data- and algorithm-based approach can be used to determine a leading risk indicator per ship, even though the achieved model performance is not yet regarded as satisfactory and further research is planned.

Keywords: Maritime safety, Accident prevention, Safety management, Risk prediction, Leading indicators, Machine learning

Introduction

Shipping accidents can lead to significant losses for both shipping companies and society at large (Jin et al. 2008). In recent years, weak economic growth, limited demand on many commodity markets and an oversupply of tonnage in almost all maritime transport segments have put shipping companies under considerable cost pressure. As a result, (safety-related) maintenance and training measures compete for tight budgets, and crew numbers are adjusted on many ships. In the “Allianz Safety and Shipping Review 2017”, experts warn of increasing risks in the shipping industry due to inadequate maintenance and repair (Dobie 2018). This raises concerns about the medium-term implications for ship and crew safety.



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

In order to address risks for ship and crew, it is necessary to develop and use effective methods, which allow an objective assessment of given hazards. The Formal Safety Assessment (FSA) of the International Maritime Organization (IMO) plays a central role in the identification and assessment of maritime risks (Montewka et al. 2014; IMO 2015). It defines risk as a combination of the probability that an adverse event will occur and the associated negative consequences. Beyond this technical definition used by the IMO, risk is a complex concept used in several ways. According to Cross (2012), two distinct meanings stand out in technical publications:

- risk as “a description of something that is uncertain and may not be an event or an outcome (it might be both, or it might be an exposure)”
- risk as “a measure to which a number or rank can be ascribed related to the extent to which potential outcomes are of concern to us.”

In the context of this research, risk is first of all understood as an inverse of safety and thus an instrument to identify opportunities to avoid accidents, which are harmful to property, life, or the marine environment. In order to do so, a quantitative approach common in the insurance industry will be adopted, which classifies risks of individual objects - which are ships in this case - based on historical data (e.g. past loss frequencies) by referring to characteristic features of these objects. This allows assessing risk levels (e.g. expected number of losses per time period) for objects with comparable features (Boodhun and Jayabalan 2018). Important indicators in this context are the loss, claims or accident frequency (in terms of the average number of claims or accidents per time) and the expected claims respectively accident expenditure (which also considers the monetary value of individual events) (Goelden et al. 2016). For a first scientific approach to the subject, it may also be sensible to consider only the most frequent type of accident/incident in constructing a realistic and practical risk exposure metric. Identifying and using such metrics can be of substantial benefit for different maritime stakeholders:

- Operators can use them for quantifying risks of individual vessels correctly to identify an optimal level of investment (both in human element, e.g. training of seafarers, and their asset, e.g. more comprehensive maintenance) which maximises profit.
- Maritime insurers can use them for quantifying risks of individual vessels correctly to price individual insurances appropriately.
- Maritime authorities (e.g. port state control) can use them for quantifying risks of individual vessels correctly to minimise accidents (e.g. through targeted inspections).

By contrast, currently “lagging indicators” dominate in the maritime sector in coping with risks once they are identified (Jalonen and Salmi 2009). In the course of maritime accident investigations, an ex-post identification of causes and contributing factors is carried out, which is of great importance in order to avoid comparable accidents in the future. However, this approach, with its focus on the past is not suitable for proactive risk management. This requires “leading indicators” which, in synergy with risk models,

create ex-ante transparency regarding the risk exposure of a ship. Intelligent risk monitoring and control systems thus help to ensure that disruptions in onboard operations are identified at an early stage, and actions can be taken once operational risks are no longer acceptable. Within the framework of an asset management system, proactive risk management with appropriate leading indicators can create significant benefit (ISO/PC 2012). An opportunity for taking up this approach lies in the adaptation of innovative algorithms and methods from risk research. Based on quantitative and qualitative data collected during onboard operation, which are becoming increasingly available, a predictive risk model can be used to determine leading maritime risk indicators.

Leading risk indicators adopt the idea of measuring conditions, attributes and states that affect the risk level of a system or activity and which are collectively referred to as risk-influencing factors. Accordingly, developing leading risk indicators requires an understanding of causal events and fault sequences leading to accidents in order to define suitable metrics associated with risk influencing factors (Toellner 2001). Once a change in such risk-influencing factors is detected, it can serve as an early warning of an increased risk of adverse events occurring and thus allows the implementation of adequate means to prevent accidents in advance (Oien and Sklet 1999). The idea of leading risk indicators has found its application in several methodologies in the field of risk management, with one example being “Tripod-DELTA”. This checklist-based approach was designed to identify conditions, attributes and states (risk-influencing factors) before “latent failures” can lead to an “active failure” (=adverse event). In practice, such “latent failures” or shortcomings can often be identified long before the actual accident (Hudson et al. 1994). Examples of leading indicators being used can be found in different sectors, including offshore oil and gas, energy and related process industries as well as nuclear safety (Tomlinson et al. 2011).

To this day, several scholars have already taken up the concept of leading indicators and applied it to the maritime industry. Examples include Balmat et al. (2009) who define a methodology to calculate a ship specific risk metric based on some risk factors making use of a fuzzy logic approach. Hänninen et al. (2014) use Bayesian networks (BN) as a modelling technique in combination with expert elicitation and historical data in order to obtain a ship-specific quantitative risk level assessment. Statistical methods form the basis of models used to predict future accident risk developed by Heij and Knapp (2018) and calculate safety leading indicators by Wang (2008). Moreover, extensive work done by ABS including statistical analysis on correlations between leading risk indicators and safety performance data has resulted in the publication of “Guidance Notes on Safety Culture and Leading Indicators of Safety” (American Bureau of Shipping 2014). As such, the idea to combine leading risk indicators with data and algorithm-based management methods is not new in the maritime domain. However, in contrast to existing work, this paper focusses on the application of machine learning (ML) methods to calculate a leading risk indicator based on safety-related data collected onboard. In doing so, concepts formulated for an ML-based approach that classifies construction projects according to their safety risk (Poh et al. 2018) as well as approaches for accident frequency modelling with ML algorithms in the context of non-life insurance pricing (Mendes et al. 2017; Zöchbauer 2016) are adopted.

The first objective of this paper is to establish a framework for designing leading maritime risk indicators, which can be used as a guide for systematic development of

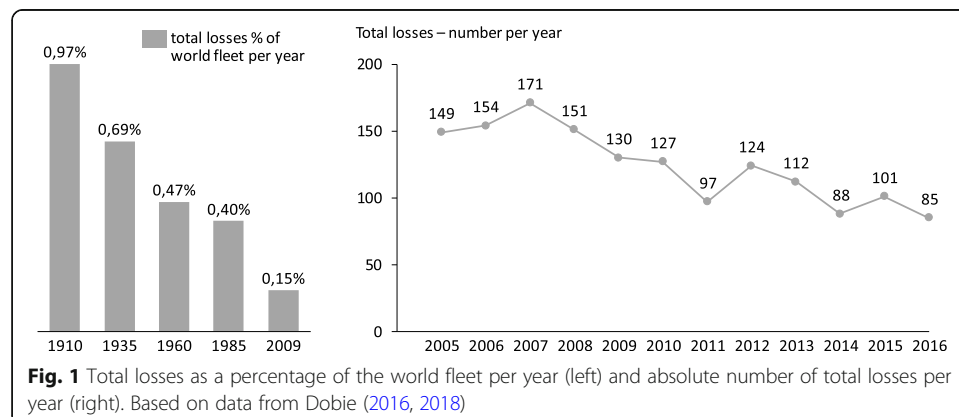
data-based risk models for ship operation. Secondly, the introduced framework is implemented in a proof of concept, which makes use of ML algorithms applied to safety-related data collected onboard in combination with data on maritime accidents. The goal is to quantify the risk associated with a particular operating condition on board using a risk model and further calculate a risk metric as a decision-making parameter. Compared to current practices in the industry, which are predominantly based on the results of ex-post maritime casualty investigations, the introduction of predictive data and algorithm-based leading risk indicators promises transparency ex-ante. Assessing the condition on board and analysing it with a risk model presents an opportunity to prevent maritime accidents more effectively than today by providing an objective basis for initiating suitable mitigation measures and keeping operational risks to an acceptable level.

The remainder of the paper is structured as follows. Section 2 gives an overview of the quantitative maritime risk picture by introducing and discussing key figures on significant risks for ships and crew. This is followed in Section 3 by the establishment of a conceptual framework for developing leading maritime risk indicators and operationalising them in a risk model. Subsequently, a proof of concept in Section 4 applies the framework to instantiate a data-based risk model and use this model to calculate a leading maritime risk indicator. Lastly, Section 5 summarises the results and provides a critical review of the outcome.

Quantitative maritime risk picture

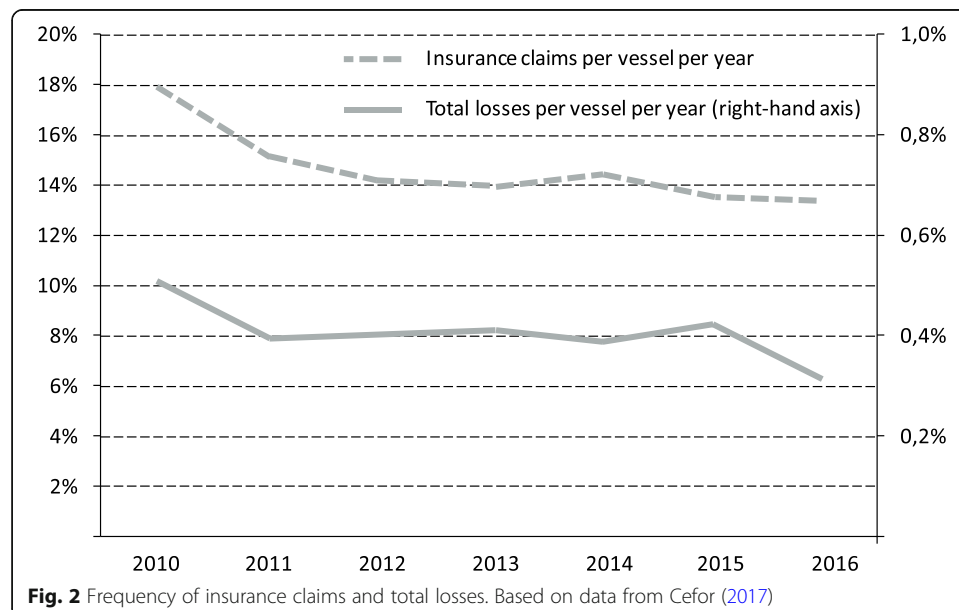
This section will give an overview of the overall maritime risk picture by presenting key figures and highlighting recent developments and characteristics of maritime accidents. In the context of this paper, an accident shall be defined as “an undesirable event that results in damage to humans, assets and / or the environment” (Kristiansen 2005). Fortunately, there has been an overall decline in the number of maritime accidents in recent years, first and foremost for particularly serious events (see Fig. 1). In consequence, the frequency of total losses has fallen by order of magnitude since the beginning of the last century, and the absolute number of total losses per year has dropped by almost half over the last decade.

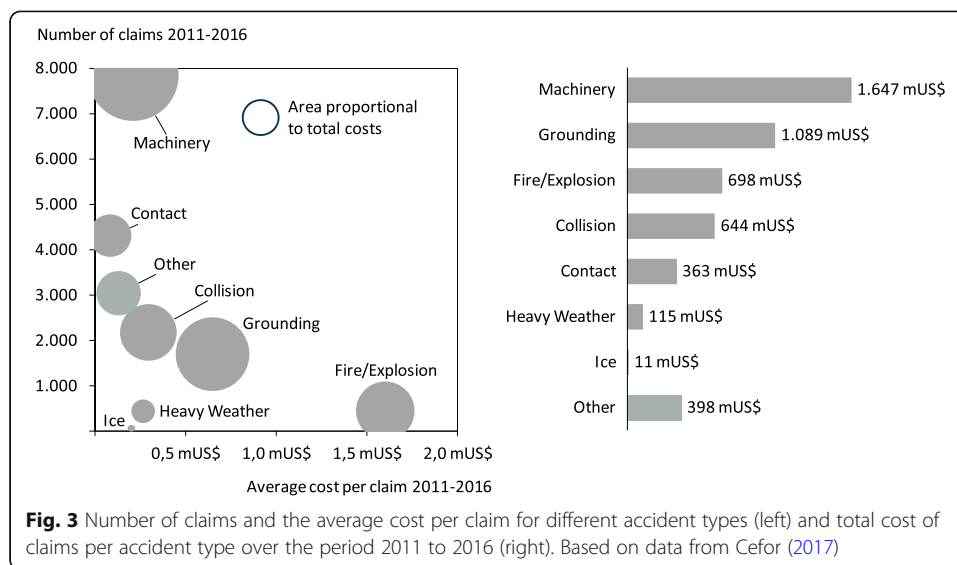
In order to assess the financial consequences of maritime accidents, the number of total losses reveals only part of the picture. Besides sinking and total economic losses,



other less catastrophic accidents need to be taken into account as well. Estimates of an accident rate per year are in the range of 10% for medium and severe events and 5% for accidents with damage greater than US\$ 100,000 (Schröder 2004). These figures correspond to publicly available statistics by the maritime insurance industry. According to the Nordic Marine Insurance Statistics (NoMIS), the number of insurance claims per ship was in the range of 13–18% over the period from 2010 to 2016. The share of total losses over the same period was between 0.6–0.4% (see Fig. 2). A recent decline in the number of claims is attributed to several different factors: Over the past few years, improved risk management by ship owners, the introduction of new regulatory measures, a declining average age of the fleet, technological advances in navigation and better incident management have all contributed to reducing the frequency of maritime accidents. (UK P&I Club 2018) As far as the annual average cost of accidents (claim value per ship) is concerned, these amount to approx. US\$ 40,000 in 2016 (Seltmann 2017).

Data from NoMIS also allow a distinction of accidents by category. In terms of frequency, machine damage is most common, with 39% of all incidents (Fig. 3). Damage attributable to navigational incidents (grounding, collision, contact) accounts for a further 41% of all incidents. Regarding the number of claims, Fire/Explosion, Heavy Weather and Ice are rather rare events. The picture is somewhat different for the average cost per incident. Here, fires or explosions on board take first place with an average cost of more than 1.5 mUS\$ per claim. Groundings are in second place with more than 0.5 mUS\$ on average followed by collisions and heavy weather damage. Machinery damage, on the other hand, is characterised by a comparatively low average cost per incident. Besides, Fig. 3 also illustrates the total damage costs per category over the period from 2011 to 2016. Here, comparatively inexpensive but frequent machinery damage events come in first place with about 33% of all costs followed by grounding with just under 22% and fire and explosions as well as collisions with about 15% of total damage claims respectively.





With regards to casual events and contributing factors of maritime accidents, human error is a major aspect, directly linked to more than 50% of adverse events (EMSA 2018). Although human (erroneous) behaviour is not always decisive in causing the accident, it often plays an essential role in the developments leading to it. Against this background, other studies suggest an even greater influence with 75–96% of marine incidents and accidents caused by some form of human error (Hanzu-Pazara et al. 2008). Besides human error, another important contributing factor in many cases is the failure of technical equipment and systems (EMSA 2018).

Development framework

This section introduces a conceptual framework for developing leading maritime risk indicators and operationalising them in a risk model. After an introduction to relevant basic ideas and concepts, three successive steps of the development framework are described.

In practice, many problems are characterised by a high degree of complexity. One way to manage this complexity in problem-solving is to use models as a representation of reality limited to the most relevant aspects. In a more formal definition, a model is a representation of an object in a particular language and form that meets a particular need of the objects stakeholder (Pourret et al. 2008). In the context of this paper, the need of the objects stakeholder is to “ensure vessel safety” by managing relevant risks. This task is part of the safety management function of a shipping company with the objective of developing, planning and implementing measures aimed at accident prevention to minimise risks for seafarers, the environment and property (Hänninen et al. 2014). It follows that the purpose of a risk model, as it is used in this paper, is to represent all relevant aspects of maritime safety management in a suitable language and usable form. In doing so, the risk model simulates the dynamics and developments of risk prone, adverse events associated with the operation of the vessel in order to provide the ship-owner with information about the risk level based on an assessment of the operating conditions on board. For this purpose, a risk model can be both qualitative or take the form of a quantitative mathematical model, which enables the calculation of a

risk metric such as a leading risk indicator. A quantitative risk model should comprise all elements, which have a relevant influence on ship operation safety, and further reflect how individual elements interact with each other in defining the overall risk level. Elements in this context are all risk-influencing factors that affect the risk level of the ship in operation.

An objective and realistic assessment of an asset's condition is an essential prerequisite for implementing effective and efficient risk management (Beerboom 2016). Only where risks for the ship and crew are measured accurately will it become possible to carry out effective control and focused management of operational risks. Unfortunately, risk-influencing factors that have a significant influence on maritime accidents are often not directly measurable. In order to overcome this, it is necessary to identify suitable variables (indicators) that reflect the abstract concept of a risk-influencing factor (Oien 2001). Where risk-influencing factors are system states or attributes that affect the risk level of ship operation, indicators are observable and measurable variables that can be used to monitor risk-influencing factors (Oien and Sklet 1999). Indicators should be designed in a way to show where processes and conditions on board leave a defined "normal range". Hence, indicators provide signals of an undesirable, risk-increasing development and monitoring these indicators can support the process of controlling risks and ultimately preventing accidents (Wang 2008). Identifying appropriate indicators and combining them in a risk model requires the definition of selection criteria, which help to ensure that the monitored aspects (processes and conditions onboard) are relevant and unnecessary information is filtered out (Toellner 2001). As a key feature, indicators that monitor risk-influencing factors should be easy to quantify. Furthermore, it is crucial to define a systematic and identical process for measuring individual aspects on board.

In light of the above, the framework for developing leading maritime risk indicators and operationalising them in a risk model consists of three steps. A first step defines individual inspection points for the assessment on board (processes and conditions monitored onboard). The second step concerns the development of a classification scheme, which guides the evaluation per inspection point. Thirdly an aggregation method is specified, which defines how findings per inspection point are combined into one or more leading risk indicators (Beerboom 2016). With this last step, a risk model is operationalised by defining how results of onboard observations are incorporated in a suitable (mathematical) model, which calculates a risk level based on the assessment results.

Inspection points

The definition of appropriate inspection points for onboard condition assessment is of crucial importance. When selecting inspection points, both operating personnel and technical experts should be involved. Furthermore, domain-specific sources, such as the ABS Maritime Root Cause Analysis (American Bureau of Shipping 2005), or assessment guidelines, such as the "Guidelines on the marine assessment of F(P)SOs" published by OCIMF (2016), can be useful. A starting point for compiling and structuring a list of inspection points based on different industry sources is given in Table 1.

Table 1 Different elements, which can be considered when defining inspection points

Self Inspection Document for Ship Operator	Self Checklist for Shipboard Safety Management System	Guidelines on the marine assessment of F(P) SOs
Certification and manning	Shipboard tour	Regulatory compliance
Management and personnel	Company responsibilities	Crew and contractor management
Bridge	Designated person	Navigational equipment
Mooring	Master's responsibilities and authority	Safety and security management
Cargo operations	Resources and personnel	Electrical equipment
Engine department	Key shipboard operations	Pollution prevention and environmental management
Operational safety	Emergency preparedness	Structural condition
Health, safety and personnel protection	Deficiency control	Operations
Firefighting / Lifesaving	Maintenance	Mooring
Environmental protection	Document control	Communications
Security	Internal audit and system review	Navigation, propulsion and active heading control

Source: CDI (2016), ClassNK (2018), OCIMF (2016)

A promising opportunity to collect representative data about the condition on board is given in the context of mandatory ISM audits. Against the background of declining safety standards and an increase in the number of severe maritime accidents throughout the 1980s, the International Safety Management Code (ISM Code) was adopted by the IMO in 1993. (IMO 2018) As a consequence, shipping companies are obliged to establish a safety management system based on the ISM Code. This includes conducting ISM audits at regular intervals in order to identify and subsequently eliminate observations and non-compliances with the provisions of the ISM-Code. Thus, ISM audits provide a comprehensive picture of the condition of a ship. Although the audits primarily serve a compliance function, they can also be a simple yet economical option for collecting information about the condition on board for a data-based risk model. Beyond ISM audits there is a large number of additional investigations and audits carried out by the shipping company itself as well as other stakeholders (including maritime insurance companies, ship owners, financing banks, port and flag states, and classification societies), and several assessment guides and frameworks define which safety performance indicators can be measured how (Banda et al. 2014). In combination, all these reports provide a broad basis of quantitative and qualitative data covering important safety-related inspection points, which can be used in predictive risk modelling.

Another opportunity for obtaining information about the condition on board is by directly using digital sensor data. Today individual systems on the latest generation of vessels are increasingly equipped with extensive sensor technology. These various sensors generate a large amount of data during operations, which contain information about the condition of the ship and individual systems on board as well as the ship's environment. At a process control level (e.g. industrial control system (ICS) or supervisory control and data acquisition (SCADA) system), the measured values come together and the entire ship system is monitored. It is therefore logical to access field and sensor data at this level with the aid of appropriate data transmission protocols, e.g. as those being developed for the Industrial Internet Of Things (IIoT) (Bauernhansl and

ten Hompel 2014). Thanks to the digitalisation of maritime systems and by combining and evaluating the recorded sensor data, it is increasingly possible to monitor the technical condition of certain systems onboard (Condition Monitoring), to predict the further development (Condition Prediction) and to make decisions regarding the optimal timing of maintenance measures (Condition-based Maintenance) (Kretschmann et al. 2019). At the same time, this digital condition related information can and should be utilised when designing and developing leading maritime risk indicators.

Classification scheme

In order to determine the ship's condition by conduction an audit, the assessment at each inspection point needs to be carried out by qualified personnel on board, which can be either third party surveyors or own personnel of the shipping company. This assessment task can either make use of a discrete rating scale or, less commonly, a continuous scale. Easy-to-use and cost-effective rating scales with at least two predefined rating levels are usually specified in a uniform way for the whole assessment process. A special case is a dichotomous scale with only two predefined response options (usually approval or rejection). The advantage of a dichotomous design is good comprehensibility and comparatively short response time. On the other hand, acquiescence bias (tendency to agree) is a disadvantage and implementing a dichotomous design can become a challenge for complex situations (Moosbrugger and Kelava 2012). In case the assessment per inspection point is based on a multi-item rating scale, a reference point needs to be defined. Beerboom (2016), for example, proposes a scheme that uses the timeliness of corrective actions. The evaluation is based on four levels: "no noticeable defect", "long-term removal of defect required", "short-term removal of defect required", and "immediate removal of defect required". A multitude of other reference points is possible, including (operational) safety or operability. As far as the number of levels on a rating scale is concerned, seven is usually regarded as a sensible maximum (Moosbrugger and Kelava 2012). Industry-specific process models and assessment guidelines (see, e.g. CDI 2016; ClassNK 2018; OCIMF 2016) contain several pre-formulated questions, which can be used as starting point when developing a classification scheme, whether dichotomous or as a multi-item rating scale.

A standardised assessment procedure with a uniform checklist of inspection points and a clearly defined classification scheme reduces the subjective influence of the person conducting the assessment, but subjectivity cannot be eliminated. Personal experiences and preferences will always affect the collected information to some extent. The resulting bias must be taken into account when making decisions based on the collected information (Beerboom 2016).

The results of different audit regimes (regardless of whether ISM, class or PSC audits) are usually only available as written inspection protocols, which document individual observations and nonconformities. Of course, such reports are not suitable as an input for quantitative risk models and must first be quantified using a suitable methodology. Alternatively, the auditing process can be modified *ab initio* so that quantitative condition data is collected (e.g. via an assessment checklist and classification scheme) in a uniform and identical way for all audits. In order to ensure the highest possible efficiency of condition assessment within the audit process and to avoid data transmission

errors, it is furthermore recommended to record the data digitally on a mobile device (e.g. tablet) directly on-site (Beerboom 2016).

Aggregation method

Besides the selection of relevant inspection points and the definition of a classification scheme for measuring the condition per inspection point, the development framework further includes the design of an aggregation method. It defines how individual observations per inspection point are combined and summarised in an overall risk metric or leading maritime risk indicator. The critical aspect in selecting and designing the aggregation method is to ensure that it reflects the chain of events leading towards accidents as well as the (potentially non-linear) interaction between individual risk-influencing factors. One option to define an aggregation method is to adopt methodologies developed for multiple-criteria decision-making such as Utility Analysis or Analytic Hierarchy Process (Beerboom 2016). In the simplest case, the aggregation method is a weighted summation of individual condition measurements for all inspection points on board. The respective weighting factors can either be formulated directly by the decision-maker or determined indirectly by estimating a preference function (Geldermann and Lerche 2014). In contrast to the above, it is also possible to use different statistical or mathematical techniques to define a data-based aggregation method. Several options are discussed subsequently.

Data-based methods derive system risk information directly from historical condition data in combination with accident and incident figures. In order to accomplish this, correlations between onboard conditions and subsequent failure events are identified and analysed. These correlations are used for both defining a risk model structure and for parameterising the relationships between individual input parameters. As such, a prioritisation of different systems or operations on board in regards to their impact on the risk level (e.g. accident frequency) is deducted automatically from historical data. This ensures that historical data is the primary source for identifying the most common damage and the most vulnerable subsystems. Data-based methods can be further divided into statistical or stochastic techniques and ML methods (Sikorska et al. 2011).

Traditionally, statistical analysis is the backbone of quantitative risk assessment. The outcome of using statistical methods for modelling relevant risks can be rather simple risk assessment checklists, which are used today in aviation but also shipping. These are comparatively uncomplicated to use, but at the same time limited in terms of their explanatory power. A reason is that underlying linear accident models are only capable of limited extent in reproducing the highly complex structures of causal events and fault sequences leading towards maritime accidents (Hadjimichael 2009). Examples, where more complex statistical methods are used, can be found, inter alia, in Heij and Knapp (2018) and Wang (2008). Both studies develop a model that combines principal component analysis with regression. Heij and Knapp (2018) use Port State Control (PSC) observations to predict the risk of future accidents by calculating a ship-specific risk metric. Wang (2008) develops an approach to measure safety factors in shipping and use these to predict the future safety performance of shipping operations in terms of a leading risk indicator. Statistical approaches are also used outside the maritime domain in order to predict risks and the probability of accident events (Poh et al. 2018).

Concerning stochastic methods, BN is a frequently used method for modelling risks. BN allows the modelling of complex causal relationships by reproducing the conditional dependencies between considered variables in form of a directed acyclic graph (Jensen and Nielsen 2007). The BN structure and conditional probabilities can be based on human expertise or determined directly from fault event data. The use of algorithms for data-based learning of BN structure and conditional probabilities usually requires a large amount of representative data. Qualitative learning, which focusses on network structure, can be distinguished from quantitative learning, which determines the conditional probabilities between considered variables (Wagner 2000; Daly et al. 2011). Hybrid approaches allow combining data-based learning with (possibly fuzzy) expert estimates and are robust against partially incomplete data sets (Hänninen 2014). An advantage of BN is its capability to model complex systems with a large number of interacting factors. Overall, the use of BN for addressing maritime research questions has increased significantly in recent years (Banda et al. 2014). Hänninen et al. (2014) have used an approach that shows certain parallels with the risk model introduced in this paper. They use BN as a modelling technique in combination with expert elicitation and historical data in order to make a quantitative assessment of a ship-specific risk level. BN-based approaches are also used beyond the maritime industry to develop risks models (see, e.g. Pourret et al. 2008; Ayello et al. 2018).

Artificial intelligence and in particular ML have seen a significant increase in their application over the past few years. ML comprises various algorithms that learn dependencies through pattern recognition in data sets and use these relationships to make predictions (Nelli 2018). The basis for solving a task is a data set (e.g. labelled examples) in which ML methods automatically identify relevant statistical relationships and convert these into generalising rules used for completing a given task (Chollet 2018). ML methods thus determine correlations and patterns directly in the data without the need to specify a model, as is the case with statistical and stochastic methods. In order to accomplish this, ML requires a sufficiently large training data set. The most successful applications of ML algorithms today are in the field of “supervised learning” to automate decision processes. Based on a training data set comprising input (features) and the desired output (target), an ML algorithm independently learns existing correlations, which can subsequently be applied as decision rules on previously unknown input data (Müller and Guido 2017). In comparison to statistical methods, an advantage of ML is that it can represent both linear and non-linear relationships without being bound by restrictive premises and assumptions of some statistical tests. (Poh et al. 2018) ML methods are both suitable and promising for modelling and predicting risks. Poh et al. (2018), for example, have developed an ML-based risk model, which they use to predict the safety risk of construction projects, which shows several parallels to the approach developed and implemented for maritime risks in this paper. In order to predict the probability of accident events, Poh et al. (2018) use the results of safety audits as well as variables characterising the respective construction project (e.g. contractor type, project volume, project type) and compare the performance of different ML algorithms. Other interesting and relevant examples include the use of ML methods in the context of non-life insurance pricing (see, e.g. Mendes et al. 2017; Zöchbauer 2016).

Proof of concept

This section describes a proof of concept, which builds an ML-based approach to calculate a leading maritime risk indicator by implementing the conceptual guidelines set by the development framework introduced in the previous Section 3.

The objective of the proof of concept is to design and test a model, which simulates the cause-effect relationships of maritime accidents and thus enables the calculation of a data- and algorithm-based leading risk indicator. Against the background that (real-time) data from ship operation are expected to become increasingly available in the future (Kretschmann 2018; Kretschmann et al. 2019), data-based methods for operationalisation of the risk model are the focus of the study. These are based on the premises that the probability of certain error events occurring in the future is evident in specific conditions, attributes and states (risk-influencing factors), which can be measured on-board. Accordingly, the first step of building a data-based risk model is to identify correlations between data reflecting on onboard conditions and temporally subsequent accidents. The subsequent step is to build a maritime risk model that reflects the identified correlations.

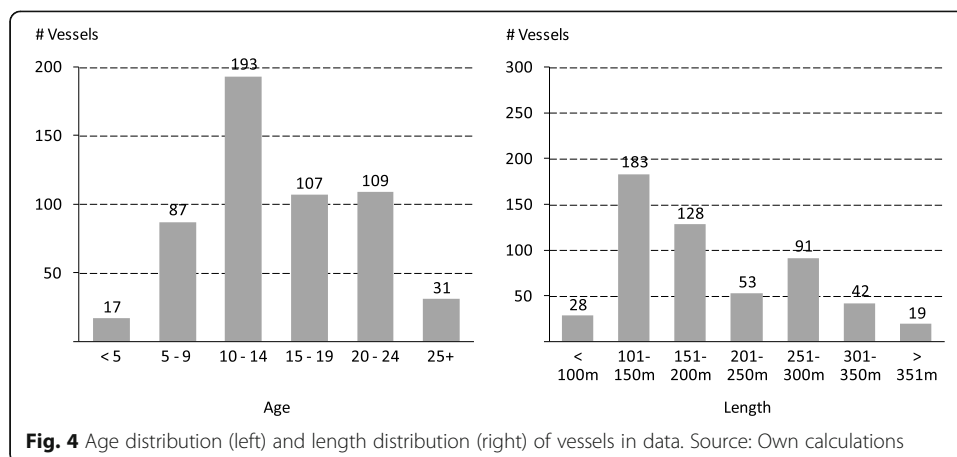
Principles of CRISP-DM (Shearer 2000), a process model for data mining projects, are adopted for the proof of concept. The CRISP-DM process model has six phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. Phase 1 comprises a specification of objectives and a problem definition, which in this case consists of subjecting the idea of a leading maritime risk indicator to a proof of concept. Phase 2 to phase 5 are discussed subsequently. Phase 6 (model deployment) is out of scope for a proof of concept.

Data understanding

According to the CRISP-DM process, the Data Understanding phase comprises collecting, describing and exploring data as well as verifying data quality (Shearer 2000).

Two data sets for several objects (ships) are used in the proof of concept. The first data set comprises maritime accident information, which represents the response variable or target for the analysis. Besides damage type, date of the event, the extent of damage (in US\$) and the involved object are given. The second data set represents the covariates, predictors or features used in the analysis. It contains observations made during past PSC on ships for which accident information is available. Furthermore, additional data is taken into consideration, which characterises the respective objects and which can be expected to have an impact on the accident risk according to previous studies (Knapp 2006, 2013). Data includes the size of the vessel in gross tonnage (GT) and tons deadweight (TDW), age of the vessel, country of the shipyard and length of the vessel.

A descriptive and explorative analysis was carried out for all available data. In order to protect the commercial interests of the companies that provided data for the proof of concept, only selected results can be presented here. The study is based on data of 544 container ships, which corresponds to about 10% of the world fleet in 2017 (Equasis 2018). Figure 4 illustrates the age and length of vessels in the data set. It shows a broad distribution, with smaller middle-aged ships making up the majority. For the mentioned 544 vessels, all accidents over a total period of 819 years of operation are

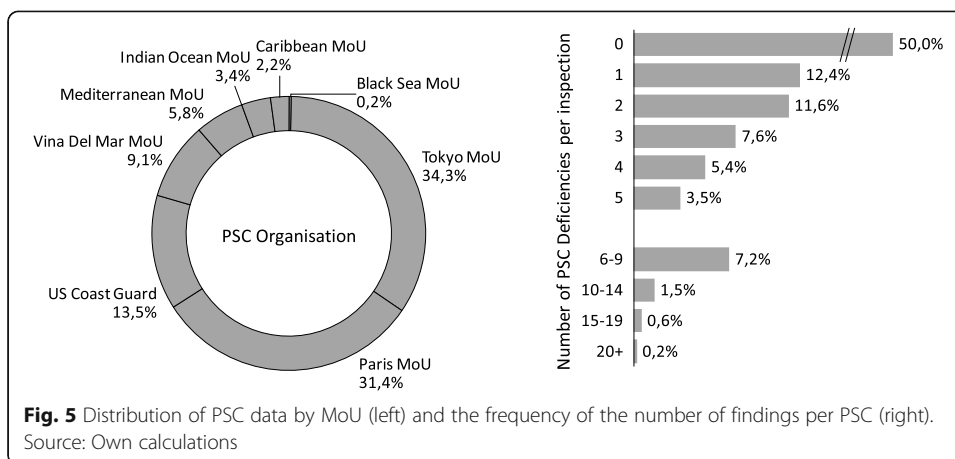


included in the data, which corresponds to 185 accidents. With about 22% fault events per ship per operating year, the frequency of accidents is slightly higher than the rate of 13–18% given by the Nordic Marine Insurance Statistics (Cefor 2017).

Observations made during PSC represent safety-related data used in the proof of concept. Several other authors have previously used PSC data in comparable studies (see, e.g. Heij and Knapp 2018; Hänninen et al. 2014; Wang 2008; Tsou 2018). PSC has been established as a “second line of defence” due to negligence in the control of international standards regarding ship safety, pollution prevention and the working and living conditions of seafarers by some Flag States (Directive 2009/16/EG). PSC inspections are carried out by the Port States according to a predetermined procedure with observations being documented uniformly. In total, PSC results of 5468 inspections were collected for the proof of concept covering the mentioned 819 years of operation plus the previous 2 years for each object (vessel) in the data set.

As part of a comprehensive check of data quality and integrity, the number of entries in the database was reduced from 5468 inspections to 4502. In particular, cases of “double reporting” were removed, as well as follow-up inspections whose results equal results of a prior initial inspection. At the same time, missing data was added, and false entries corrected as far as possible. Figure 5 shows the breakdown of PSC inspections data by Memorandum of Understanding (MoU) on PSC and further reveals how often different cases of “number of identified deficiencies per PSC” occur in the data. Overall, one or more deficiencies are reported in 50% of port State control inspections in the data set. This is in line with observations made by the BG Verkehr in German ports, indicating that every other ship inspected has deficiencies (BG Verkehr n.d.).

The documentation of deficiencies identified during a PSC is based on a catalogue of 555 Deficiency Codes, which in turn are assigned to 30 categories (Paris MoU 2017). Figure 6 shows how often deficiencies from different deficiency categories are included in the data set used in the proof of concept. Most frequent deficiencies come from the categories “Fire safety”, “Safety of Navigation” and “Life saving appliances”.



Data preparation

According to the CRISP-DM process, the Data Preparation phase comprises selection and cleansing of data and the preparation of the data set used for the model building including construction, integration and formatting of labelled examples (Shearer 2000).

First of all the preparation of raw data before the actual development of the model concerns defining time periods, during which PSC data (looking backwards) and accident events (looking forward) are combined in a labelled example. A long period of time during which PSC results are integrated into each labelled example is associated with the limited predictive value of information representing events far in the past. This is based on the assumption that PSC deficiencies documented long ago are less indicative regarding the current condition on board (and thus the probability of accidents) compared to recent PSC findings. Accordingly, it would make sense to include only rather recent PSC results in each labelled example. However, this also reduces the amount of data the algorithm can use for training since PSC inspections take place infrequently. Within the proof of concept, both a period reaching 1 year and 2 years into the past was considered. In the first case, results of 1.9 PSC are available per object on average to construct a labelled example. At the same time, no PSC data is available for about 10% of objects (since no PSC inspection had taken place in the previous 12 months). If data from the past 2 years is considered for each labelled example, results of 3.6 PSC are available per object on average, and less than 5% of objects remain without PSC data.

Accident frequency per object per year is chosen as the response variable or target in this analysis. Similar to PSC data, it is necessary to define a period during which accident events are considered when calculating the response variable and constructing each labelled example. The longer this period reaches into the future, the more accidents fall into it but, the smaller becomes the correlation between accidents relatively far in the future and onboard conditions observed during past PSC. Additionally, accident events are recorded only during a certain time span for each object in the database (average 400 days, minimum of 5 days, maximum 669 days). Accordingly, the number of labelled examples in the data set available for training of the ML-model gets smaller, if accidents over a longer period of time are taken into account. The reason is as follows: objects can only be used if they are represented in the data set with a time span



larger or equal to the chosen period of time, during which accidents are taken into account. At the same time, the accident frequency per labelled example increases, if accidents over a longer period of time are taken into account. On the other hand, a shorter period of time increases the number of examples available for training the ML-model, but the accident frequency is reduced. To demonstrate the effect: if the period of time is defined as 12 months, the accident frequency is 21%, and there are 608 labelled examples available for training. In case the period of time is set to 6 months, the accident frequency falls to 11%, but 1503 labelled examples can be used for training the ML-model.

The second part of data preparation concerns feature generation, by which data is transformed in a way to obtain input variables for the ML model. Depending on the specific ML methods used, the used attributes (data) must be subjected to a pre-processing. For example, artificial neural networks are very sensitive to particularly large or small input values and therefore require scaling and normalisation of input data (Frochte 2018). Another step of data preparation is the transformation of categorical attributes. In the proof of concept, this concerns the attribute “country of shipyard”. By far, the most common approach to handle categorical data is one-hot encoding, where the categorical attribute is replaced by several binary dummy variables, each representing one category. This, of course, increases the number of features, which in turn can have some negative effects.

When the number of features used to train an ML classifier is increased, model performance usually increases initially as well, but begins to drop beyond an optimum (also known as “Hughes Phenomenon”). This “curse of dimensionality” describes the phenomenon by which the feature space becomes more sparsely occupied if the number of dimensions (more features) increases but the size of the training data set remains

constant. The distance between two neighbours in the high dimensional feature space increases, and therefore it becomes more difficult for an ML-algorithm to calculate a meaningful model (Shetty 2015). For this reason, an optimal training data set contains several examples for each feature combination. Where this cannot be ensured, e.g. by collecting additional data, several options can help reduce the feature space, with Principal Component Analysis (PCA) being a commonly used approach. Especially for high-dimensional data with a small number of cases, PCA can positively influence the learning capabilities of ML algorithms (Frochte 2018).

In this proof of concept, the need for a dimensionality reduction concerns PSC data in particular. As previously described, PSC results are available subdivided into 30 individual deficiency categories. Since there are 7 observations per PSC inspection on average, only 15% of all possible entries in all 30 deficiency categories contain a value on average, while the remaining 85% are zeros. Accordingly, the PSC inspection results data set can be described as sparsely populated. Taking further into account that the overall number of labelled examples for training the ML model is rather small in the proof of concept, it appears appropriate to reduce the dimensionality of PSC data. This dimensionality reduction is accomplished by implementing a PCA, and only the first principal component is used as a feature when constructing labelled examples and calculating the risk model.

Modelling

According to the CRISP-DM process, the Modelling phase comprises the modelling technique selection, test design generation, as well as the design and implementation of one or more models (Shearer 2000).

Finding a suitable model that can accurately represent the level of safety on a ship is not a trivial problem. Reasons include a multitude of involved design, operation and maintenance factors, the difficulty of correctly modelling chains of events leading towards accidents and the need to build on subjective assumptions when quantifying risks, especially where insufficient data is available (Sii et al. 2004). This paper focuses on ML-methods for developing a data-based risk model. The advantage of this approach is that ML-algorithms can identify relevant correlations directly in the data without potentially constraining assumptions that characterise statistical predictive modelling techniques (Mendes et al. 2017). Several ML-algorithms would generally be applicable for the problem at hand, of which the ensemble method Random Forest (RF) was selected for the proof of concept. Ensemble methods combine multiple ML algorithms to form more powerful models. Two types of ensemble methods - RF and Gradient Boosted Decision Trees - have proven to be very effective for a variety of data sets and problems (Müller and Guido 2017). Since RF models are robust against overfitting and do not require extensive data transformation and scaling, the method is ideal during a proof of concept phase. Other models, such as artificial neural networks, may prove to be slightly more accurate but require significantly more data preparation and transformation (Müller and Guido 2017). Accordingly, it makes sense to start with a RF model and test other options to maximise performance, once a proof of concept has been established.

Based on the available data, the task “assign an appropriate risk level per ship”, which the RF algorithm is supposed to solve, can be formulated as a supervised learning classification or a supervised learning regression problem. The latter was chosen for the proof of concept. The RF model was implemented in python using scikit-learn library. Labelled examples used to train the RF model contained several features plus the target “accident frequency per object per year of operation”. Considered features were:

- PSC-indicator (first component of PCA)
- Number of PSC
- Number of deficiencies detected
- GT
- TDW
- Year of construction
- Country of shipyard (one-hot-encoded)
- Vessel length

Accidents over a period of 12 months into the future and PSC data for 24 months into the past were considered for constructing each labelled example. A cross-validation design ($n = 5$) was implemented. This approach divides the data set (labelled examples) into n equal subsets. Each subset is used as a test data set once to determine the performance of the model after the algorithm has been trained on the remaining data. Subsequently, the overall performance of the model is calculated by averaging the performance for each of the n splits (Cleve and Lämmel 2016).

Evaluation

According to the CRISP-DM process, the Evaluation phase comprises the evaluation and review of results achieved with the created models (Shearer 2000).

Mean squared error (MSE) is used as the primary performance indicator for model evaluation. A dummy regressor model that predicts the mean value of the target calculated in the training data set for all instances in the test data set serves as a performance benchmark. In case of a successful proof of concept, the MSE of the trained RF model turns out to be lower than the MSE of the dummy regressor. Additionally, a linear regression model serves as a second benchmark. Besides MSE, a coefficient of determination (r^2) and explained variance are calculated as additional performance indicators of the different RF models. In order to demonstrate the impact of the PSC indicator on performance, RF models were trained on two data sets. One contains the PSC indicator as a feature, and the other one does not. Results for the latter case are labelled “RF Model without PSC data”. Initially, two hyperparameters of the RF models (number of trees and the maximum number of levels per tree) were set manually. In order to improve RF model performance without overfitting on the data, a grid search was carried out to optimise these two hyperparameters. The results of the RF models with optimised hyperparameters are labelled “Optimised hyperparameters”. For all other model parameters, the default values were used. Performance indicators were calculated for ten RF models under each set up respectively, and the mean performance

value for these ten models is given as the overall performance. Table 2 summarises the achieved performance values of all models.

The dummy regressor model with an MSE of 0.3133 represents the benchmark for the RF models. A model based on linear regression does not perform better than the dummy regressor model on the tested data set. In comparison, both RF models (including and without PSC indicator) outperform the dummy regressor model, with an MSE of 96% and 97% respectively compared to the benchmark MSE. This is not considered to be a particularly large reduction in MSE. Thus, based on the available data, the RF models can only make a small contribution to predicting the likelihood of accidents correctly, which is also evident in the low r2 and explained variance values.

However, it needs to be borne in mind that the available data set is relatively small for using ML methods. A broader data set might make it easier for the ML algorithm to identify the most critical correlations between the available features and the target “accident frequency per object per year” and use them for prediction. Accordingly, the logical next step to substantiate the findings presented here should be to expand the data set and see if the model performance improves. Another related challenge is the imbalanced data set available in this proof of concept, where labelled examples with one or more accidents (21% of all examples) can be considered as comparatively rare events. By definition, rare events are incidents that occur at a significantly lower frequency than other more frequent events in the data set. Accurately predicting the rare class is known to be challenging with data mining techniques and classification algorithms (Maalouf and Trafalis 2011). Several problems associated herewith are highlighted in the machine learning literature (Weiss 2004), which are of concern in the proof of concept as well. This includes a small number of examples associated with the rare case in absolute and relative terms, uncertainty about differing distributions of the rare class in training and test sets or noisy data, which can make it difficult of to distinguish between exceptional (rare) cases and noise. Subsequent research will, therefore, also put a focus on appropriate methods to address problems related to rarity in this particular case.

Table 2 Performance of the tested models

Modell		Score	r2	Expl. Var.	
		MSE			
DummyRegressor	I	0.3133			
LinearRegression	II	0.3115	0.0005		
RF Model including PSC data	n_estimators = 1000, max_depth = 4	III	0.3016	0.0247	0.0355
	Opti. hyperparam.	Comp. to I	96%		
		Comp. to III	100%	126%	
RF Model without PSC data	n_estimators = 1000, max_depth = 4	IV	0.3040	0.0202	0.0290
	Opti. hyperparam.	Comp. to I	97%		
		Comp. to III	101%	82%	82%
		Comp. to IV	99%	143%	

Source: Own calculations

At the same time, it must be taken into account as well that PSC observations represent principally suitable but by no means optimal data for developing a leading maritime risk indicator and operationalising it in a risk model. If input values were explicitly collected for a data-based risk model by implementing the development framework introduced in Section 3, it can be assumed that a larger contribution to correctly predicting the likelihood of accidents would be achievable. None the less, even with an optimal database a significant part of the probability of accidents can never be predicted using a data-based model due to the considerable influence of random events and external factors on the occurrence of many maritime accidents.

Based on the results of the proof of concept, it can further be concluded that PSC observations, represented by the calculated PSC indicator, contribute to a data-based prediction of the selected target “accident frequency per object per year”. This is evident in the results for the “RF model without PSC data” compared to the “RF model including PSC data”: The latter achieves both a higher r^2 and explained variance. In order to emphasise this further, the accident frequency of ships with an above-average PSC indicator can be compared to ships with a below-average value. It turns out that the accident frequency is higher by a factor of 1.3 for vessels with an above-average PSC indicator. If the overall risk indicator, calculated by the RF model for each object in the database, is used to subdivide two groups, the accident frequency per year of ships with an above-average risk indicator is higher by a factor of 3 compared to ships that have a below-average risk indicator value.

Conclusions

This paper represents a first step towards the development of an innovative data- and algorithm-based leading maritime risk indicator. Its contributions lie in two areas.

First, a conceptual development framework was introduced and described, which can guide the development of leading maritime risk indicators and their operationalisation in a data-based risk model. One challenge in this context is that relevant risk-influencing factors, which have a value in predicting the probability of maritime accidents and are thus crucial for establishing leading maritime risk indicators, are often not directly measurable. In order to overcome this, it is necessary to identify suitable variables (indicators) that point towards conditions, events, and sequences that precede accidents in ship operation and accordingly reflect the abstract concept of risk-influencing factors. As such, measuring these variables or indicators on board can be used as an alternative by providing signals for undesirable, risk-increasing developments and conditions. A second challenge lies in the high complexity of the system ship and its operation. A multitude of subsystems and elements work together to ensure efficient and safe ship operation, while at the same carrying the capacity to cause an accident or incident. Taking into account the above, the proposed development framework is divided into three consecutive steps: (1) the identification of appropriate inspection points for onboard condition assessment, (2) the development of a classification scheme which specifies the evaluation per inspection point, and (3) an aggregation method which operationalises a risk model by defining how findings per inspection point are combined into a leading risk indicator.

Maritime operations are extremely complex, involving several interacting human, mechanical, technological and environmental components and external influences.

Consequently, the risks associated with operating a vessel are equally complex and diverse. Overall human factors are thought to outweigh technical failures in their contribution to severe maritime accidents. Aspects associated with human errors like crew qualifications, working and living conditions, and compliance with working hours regulations and safety management are distinctly different from technical and management aspects related to maritime accidents. This should be considered in the development of leading risk indicators. When collecting new data, both the inspection points and the classification schemes should take into account differences between human factors and technical conditions. Moreover, regardless of whether existing data or newly collected data are used, the aggregation method should reflect the different importance of human factors and technical failures in the series of events leading to maritime accidents.

The second contribution of this paper was to show how the development framework can be applied to an actual case by calculating a data- and algorithm-based leading maritime risk indicator. In particular, the focus was on the operationalisation of a risk model by using ML algorithms. Based on a data set containing maritime accident information, which represents the response variable or target, plus a data set containing different covariates or features (in particular observations made during past PSC inspections as well as different characterising values for all considered objects), a proof of concept was attempted. The results obtained from the proof of concept indicate that a leading risk indicator can indeed be calculated with the chosen ML approach. However, the RF model only made a relatively small contribution to predicting the chosen risk metric “accident frequency per object” correctly. Accordingly, a next step will be to consider different strategies to improve ML model performance. First of all, it is a matter of increasing the number of labelled examples used to train the ML model, which is usually the first choice in order to improve performance. Further possibilities include testing of different data transformations and feature engineering strategies as well as using other ML algorithms. All the mentioned strategies will be considered in subsequent research.

Abbreviations

BN: Bayesian networks; FSA: Formal Safety Assessment; GT: Gross tonnage; ICS: Industrial control system; IIoT: Industrial Internet Of Things; IMO: International Maritime Organization; ML: Machine learning; MSE: Mean squared error; MoU: Memorandum of Understanding; NoMIS: Nordic Marine Insurance Statistics; PSC: Port State Control; PCA: Principal Component Analysis; RF: Random Forest; SCADA: Supervisory control and data acquisition; TDW: Tons deadweight

Acknowledgements

The article process charges of this work is supported by China Merchants Energy Shipping.

Author's contributions

Author was solely responsible for preparing the paper. The authors read and approved the final manuscript.

Authors' information

Lutz Kretschmann works as a senior scientist at Fraunhofer Center for Maritime Logistics and Services CML in Hamburg, Germany, where he is engaged in applied research and consulting projects in maritime logistics and shipping. A focus of his work is the digital transformation of the maritime industry, new data driven business models, and the economic viability of innovations such as autonomous ships. His own research interests lie in the field of machine learning and specifically applying deep learning to current problems in the maritime industry.

Funding

This research has been funded by the Hamburger Behörde für Wirtschaft, Verkehr und Innovation as part of the project “RIDERS”. Hamburger Behörde für Wirtschaft, Verkehr und Innovation had no role in design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The dataset about maritime accident events analyzed during the current study are not publicly available due to confidentiality agreements which reflect their proprietary and highly sensitive nature. Confidentiality includes providing details about the data owner.

The dataset about port state control results analyzed during the current study are published via EQUASIS and can be acquired from commercial maritime data provider.

Competing interests

The authors declare that they have no competing interests.

Received: 13 February 2020 Accepted: 20 August 2020

Published online: 19 October 2020

References

- American Bureau of Shipping (2005) Guidance notes on the investigation of marine incidents, Houston Updated 1 February 2014
- American Bureau of Shipping (2014) Guidance notes on safety culture and leading indicators of safety. Updated February 2014, Houston
- Ayello F, Sridhar N, Mosleh A, Jackson C (2018) Demonstration of a multi-analytic risk management tool for the California pipeline industry. California Energy Commission, Oakland
- Balmat J, Lafont F, Maifret R, Ressel N (2009) MARitime RiSk Assessment (MARISA), a fuzzy approach to define an individual ship risk factor. *Ocean Eng* 36(2009):1278–1286
- Banda OV, Hänninen M, Goerlandt F, Kujala P (2014) Bayesian networks as a decision making tool to plan and assess maritime safety management indicators. In: Probabilistic safety assessment and management (PSAM) 12, June 2014. PSAM 12 Proceedings, Honolulu
- Bauernhansl T, ten Hompel M, Vogel-Heuser B (eds) (2014) *Industrie 4.0 in Produktion, Automatisierung und Logistik*. Springer Vieweg, Wiesbaden
- Beerboom D (2016) *Objektive Zustandsbewertung von Mittelspannungsnetzen als Grundlage der Asset-Optimierung*. Dissertation Bergische Universität Wuppertal, Wuppertal
- BG Verkehr (n.d.) *Hafenstaatkontrolle Statistik und Beispiele*. [online] www.deutsche-flagge.de/de/psc/statistik [11.04.2019]
- Boodhun N, Jayabalan M (2018) Risk prediction in life insurance industry using supervised learning algorithms. *Complex Intell Syst* 4:145–154
- CDI (2016) *Self Inspection Document (SID) for Ship Operator*. Version 15.0.4
- Cefor (2017) *Annual report 2016*. The Nordic Association of Marine Insurers, Oslo
- Chollet C (2018) *Deep Learning mit Python und Keras*. mitp Verlags GmbH & Co. KG, Frechen
- ClassNK (2018) *SELF-CHK-SHIP-e Version 2018.3*
- Cleve J, Lämmel U (2016) *Data Mining*. De Gruyter, München
- Cross J (2012) Risk. In HaSPA (health and safety professionals Alliance), the Core body of knowledge for generalist OHS professionals. Safety Institute of Australia, Tullamarine
- Daly R, Shen Q, Aitken S (2011) Learning Bayesian networks: approaches and issues. *Knowledge Eng Rev* 26(2):99–157
- Dobie G (ed) (2016) *Safety and shipping review 2015*. Allianz Global Corporate & Specialty, Munich
- Dobie G (ed) (2018) *Safety and shipping review 2017*. Allianz Global Corporate & Specialty, Munich
- EMSA (2018) *Annual overview of marine casualties and incidents 2017*. European Maritime Safety Agency, Lisboa
- Equasis (2018) *The World Merchant Fleet in 2017 - Statistics from Equasis*
- Frochte J (2018) *Maschinelles Lernen. Grundlagen und Algorithmen in Python*. Carl Hanser Verlag GmbH & Co. KG
- Geldermann J, Lerche N (2014) *Leitfaden zur Anwendung von Methoden der Multikriteriellen Entscheidungsunterstützung*. Methode: Promethee. Georg-August-Universität Göttingen, Göttingen
- Goelden H-W, Hess KT, Morlock M, Schmidt KD, Schröter KJ (2016) *Schadenversicherungsmathematik*. Springer-Verlag, Berlin Heidelberg
- Hadjimichael M (2009) A fuzzy expert system for aviation risk assessment. *Expert Syst Appl* 36(2009):6512–6519
- Hänninen M (2014) Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Accid Anal Prev* 73(2014)
- Hänninen M, Banda OAV, Kujala P (2014) Bayesian network model of maritime safety management. *Expert Syst Appl* 41(17): 7837–7846
- Hanzu-Pazara R, Barsan E, Arsenie P, Chiotoroiu L, Raicu G (2008) Reducing of maritime accidents caused by human factors using simulators in training process. *J Marit Res* 5(1):3–18
- Heij C, Knapp S (2018) Predictive power of inspection outcomes for future shipping accidents – an empirical appraisal with special attention for human factor aspects, *Maritime Policy & Management*
- Hudson PTW, Reason JT, Wagenaar WA, Primrose M, Visser JP (1994) Tripod Delta: proactive approach to enhanced safety. *J Pet Technol* 46(1):58–62
- IMO (2015) *Formal Safety Assessment: Revised Guidelines for Formal Safety Assessment (FSA) for use in the IMO rule-making process (MSC-MEPC.2/Circ.12/Rev.1)*
- IMO (2018) *International safety management code*. International Maritime Organization, London
- ISO/PC 251/N183 (2012) *Asset management — Overview, principles and terminology*
- Jalonen R, Salmi K (2009) *Safety performance indicators for maritime safety management*. Helsinki University of Technology, Espoo
- Jensen F, Nielsen T (2007) *Bayesian Networks and decision graphs*. Springer, Berlin Heidelberg New York
- Jin D, Kite-Powell H, Talley WK (2008) *US ship accident research*. In: Talley WK (ed) *Maritime Safety, Security and Piracy*, Informa LLP, London
- Knapp S (2006) *The econometrics of maritime safety – recommendations to enhance safety at sea*, vol 96. ERIM Ph.D. Series Research in Management, Rotterdam

- Knapp S (2013) An integrated risk estimation methodology: ship specific incident type risk. In: Econometric Institute Report 2013–11. Econometric Institute, Erasmus University Rotterdam, Rotterdam
- Kretschmann L, Deichmann F, Jahn C (eds) (2019) White paper ship operation 4.0. Fraunhofer Verlag, Stuttgart
- Kretschmann L (2018) Digitalisierung des Service in der maritimen Zulieferindustrie: Datenbasierte Dienstleistungen im Kontext der maritimen Industrie 4.0. VDMA - Arbeitsgemeinschaft Marine Equipment and Systems
- Kristiansen S (2005) Maritime Transportation: Safety Management and Risk Analysis. Elsevier, Oxford
- Maalouf M, Trafalis TB (2011) Rare events and imbalanced datasets: an overview. *Int J Data Min Model Manag* 3(4):375
- Mendes J, de Vleriola S, Mahy S, Marechal X (2017) Machine learning applications to non-life pricing. *Reacfin White Paper*
- Montewka J, Goerlandt F, Kujala P (2014) On a systematic perspective on risk for formal safety assessment (FSA). *Reliab Eng Syst Saf* 127(2014):77–85
- Moosbrugger H, Kelava A (eds) (2012) Testtheorie und Fragebogenkonstruktion. Springer-Verlag, Berlin Heidelberg New York
- Müller CA, Guido S (2017) Introduction to machine learning with Python. O'Reilly Beijing Boston Farnham
- Nelli F (2018) Python data analytics. Apress Media, California
- OCIMF (2016) Guidelines on the marine assessment of F(P) SOs – Assessment Criteria and Questionnaire, 1st edn. Oil companies international marine forum, London
- Oien K (2001) Risk indicators as a tool for risk control. *Reliab Eng Syst Saf* 74(2001):129–145
- Oien K, Sklet S (1999) Methodology for development of risk indicators for offshore platforms, Stockholm. NKS/SOS-2 seminar on risk informed principles, 13.-14.4.1999, Bergendal
- Paris MoU (2017) List of Paris MoU deficiency codes. 1ST OF JULY 2017. [online] <https://www.parismou.org/list-paris-mou-deficiency-codes> [11.02.2019]
- Poh C, Ubeynarayana CU, Goh YM (2018) Safety leading indicators for construction sites: a machine learning approach. *Autom Constr* 93(2018):375–386
- Pourret O, Naim P, Marcot B (2008) Bayesian networks: a practical guide to applications. Wiley, West Sussex
- Schröder JU (2004) Datenerfassung bei Unfallursachen und begünstigte Faktoren für Unfälle in der Seeschifffahrt. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Dortmund Berlin Dresden
- Seltmann (2017) Global Marine Insurance Casualty Trends IUMI @ IMO
- Shearer C (2000) The CRISP-DM Model: The New Blueprint for Data Mining. *J Data Warehousing* 5(4):Fall 2000
- Shetty B (2015) Curse of dimensionality. [online] <https://towardsdatascience.com/curse-of-dimensionality-2092410f3d27> [28.03.2019]
- Sii HS, Wang J, Ruxton T, Yang JB, Liu J (2004) Use of fuzzy logic approaches to safety assessment in maritime engineering applications. *J Mar Eng Technol* 3(2):45–58
- Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. *Mech Syst Signal Process* 25(2011):1803–1836
- Toellner J (2001) Improving safety and health performance: identifying and measuring leading indicators. *Professional safety*. Bd. 46, Ausg. 9(Sep 2001): 42–47
- Tomlinson CM, Craig BN, Meehan MJ (2011) Enhancing safety performance with a leading indicators program. *Human Factors in Ship Design and Operation*, 16–17 November 2011, London, UK
- Tsou M-C (2018) Big data analysis of port state control ship detention database. *J Mar Eng Technol* 18(3):113–121
- UK P&I Club (2018) Review of the Year 2017
- Wagner M (2000) Bayes-Netze: Eine Einführung. Seminar "Bayes-Netze und Data Mining"
- Wang H (2008) Safety factors and leading indicators in shipping organizations: tanker and container operations. Dissertation Rensselaer Polytechnic Institute Troy, New York
- Weiss GM (2004) Mining with rarity: a unifying framework. *SIGKDD Explorations News* 6(1):7–19
- Zöchbauer P (2016) Data science in non-life insurance pricing. Master's Thesis ETH Zürich, Zürich

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
