Routledge
Taylor & Francis Group

Maritime Policy & Management

The flagship journal of international shipping and port research

ISSN: 0308-8839 (Print) 1464-5254 (Online) Journal homepage: https://www.tandfonline.com/loi/tmpm20

Shipping inspections, detentions, and incidents: an empirical analysis of risk dimensions

Christiaan Heij & Sabine Knapp

To cite this article: Christiaan Heij & Sabine Knapp (2019) Shipping inspections, detentions, and incidents: an empirical analysis of risk dimensions, Maritime Policy & Management, 46:7, 866-883, DOI: 10.1080/03088839.2019.1647362

To link to this article: https://doi.org/10.1080/03088839.2019.1647362

Published online: 29 Jul 2019.

MARITIME POLICY & MANAGEMENT
2019, VOL. 46, NO. 7, 866–883
https://doi.org/10.1080/03088839.2019.1647362

Shipping inspections, detentions, and incidents: an empirical analysis of risk dimensions

Christiaan Heij and Sabine Knapp

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, Netherlands

ABSTRACT

Inspections play a key role in keeping vessels safe. Inspection authorities employ different policies to decide which vessels to inspect, including type of vessel, age, and flag. Attention for vessel history is usually restricted only to past detentions. This paper demonstrates that the correlation between the probabilities of detention and (very serious and serious) incidents is very low and that proactive prevention of future incidents is improved by accounting for both risk dimensions, that is, by combining past incident and detention information for targeting high-risk vessels for inspection. Five combined methods are presented to classify vessels based on these two risk dimensions, each of which involves extensive sets of factors. These combined classification methods have predictive power for future incidents. Depending on the applied inspection rate, incorporation of incident risk improves inspection hit rates for vessels with future incidents by 30–50% compared to using only detention information. It is recommended to focus on vessels where both risks are relatively high. A practical example shows how the methods can be applied for inspection selection and for prioritizing inspection areas defined in terms of eight risk domains that include collisions, groundings, engine and hull failures, loss of life, fire, and pollution.

KEYWORDS

Maritime safety; inspection policy; vessel-specific risk; detention risk; incident risk; risk domains

1. Introduction

Although declining trends in detention and incident rates indicate better safety quality of vessels, there is still ample space for further improvement. It is particularly important that high-risk vessels are targeted for port state control (PSC) inspections to reduce so-called false-negative events. Such an event occurs if the inspection regime does not select a vessel for inspection because the employed targeting method indicates the vessel has low risk, whereas briefly afterwards it experiences an incident with serious consequences.

Current targeting protocols used by various Memoranda of Understanding (MoU) for PSC's mostly focus on detentions only. These protocols assume that detained vessels also have a high risk of future incidents. Notwithstanding the importance of detentions and solving deficiencies, the ultimate goal of inspections and detentions is not so much to prevent future detentions but to increase maritime safety by preventing future incidents. The research question studied in this paper is whether, by combining the two dimensions of detention risk and incident risk, PSCs can improve their chance to select those vessels for inspection that have the highest risk of future very serious and serious (VSS) incidents. This question is analyzed by comparing the targeting performance of a benchmark policy based on detention risk alone with four alternative methods,

CONTACT Christiaan Heij heij@ese.eur.nl Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Netherlands

one of which is based on incident risk alone, whereas the other three consist of various combinations (maximum, minimum, or average) of detention and incident risk. Another purpose of this paper is to help PSCs in ranking priorities across eight risk domains during their inspections: collisions, drift grounding, powered grounding, main engine failure, hull failure, loss of life, fire and explosion, and pollution.

The empirical analysis presented in this paper uses worldwide data from various sources including inspection data, ship-particular data, and incident data. Detention risk is found to be unrelated to past VSS incidents, to decrease with past inspections, and to increase with past detentions. The risk of VSS incidents is found to be unrelated to past inspections and past detentions and positively related to past VSS incidents. These results indicate that current inspection and detention policies do not relate well to incident risk. It is shown that the proposed alternative methods outperform the benchmark policy based on detentions alone. By incorporating incident risk in their targeting policies, PSCs can reduce their chance to miss vessels that have low detention risk but high incident risk. In practice, the fully data-driven approach presented here should be used in conjunction with expert knowledge and intelligence sources for guiding inspection decisions to reduce the risk of missing risky vessels.

We mention some related literature. Cariou, Mejia, and Wolff (2007), (2008) report positive safety impacts of PSC inspections and find that deficiencies are related mainly to type of vessel, age, and flag of registry. Emecen-Kara and Oksas (2016) give a comparative overview of inspections, detentions, and deficiencies for the various MoU for PSCs. The Paris MoU (2014, 2019) classifies risk profiles of vessels as high, standard, or low, based on age class, flag, performance of recognized organization and ISM company, and past inspections, deficiencies, and detentions. Yang, Wang, and Li (2013) discuss the importance of risk quantification to replace reactive policies by proactive ones. AMSA is currently the only maritime administration that uses an explicit statistical model for risk assessment (Mueller and Morton 2002; Mueller 2007). Knapp (2006) reports that all MoU's focus on detentions only, ignore past incident information, and only use their own inspection data for targeting vessels for inspection. Data-based policies involving rich sets of risk factors for targeting vessels for inspection have been proposed by Knapp (2006), Knapp and Franses (2007), and Li, Yin, and Fan (2014). Predictive approaches for PSC inspections can be found in Heij, Bijwaard, and Knapp (2011) to improve vessel survival gains and in Yang, Yang, and Yin (2018) to predict detention probabilities. Perepelkin et al. (2010) and Ji, Brinkhuis, and Knapp (2015) advocate using incident data besides detention data to evaluate the safety performance of registries. Luo and Shin (2019) conclude that the use of multiple data sources including incident data helps in developing preventive policies to improve safety.

The employed data are described in Section 2. Section 3 presents the methodology in terms of a data-driven inspection protocol. This protocol is evaluated empirically in Section 4, and Section 5 concludes.

2. Employed databases

The empirical analysis builds on three different databases that contain ship-particular characteristics, inspections, and incidents of the world fleet for three different periods. The first database has yearly data from 2010 to 2014 and provides vessel-specific risk formulas for detentions and incidents of various types (Knapp 2015). The second database applies for December 2017 and is used to compute risk scores for each vessel from the risk formulas of the first database. As it is important to evaluate the out-of-sample predictive performance of the various risk classification methods, the third database has data from January to March 2018 to check whether the risk scores for December 2017 of the second database have predictive power for detentions and incidents in the next quarter.

The data sources are IHS Markit for ship-particulars and inspections and LLIS, IMO, and IHS Markit for incidents. Inspection data are a combination of over seventy countries from eight

MoUs and incident data are combined from the four mentioned sources to reduce under-reporting. Since original data providers use different definitions for the seriousness of incidents, these data have been reclassified according to definitions of IMO (2000) for very serious (including total loss), serious, and less serious. Besides this reclassification, incident initial events were identified when possible to classify the type of incident. The analysis is restricted to very serious (including total loss) and serious (VSS) incidents and excludes less serious incidents and near misses because the latter are relatively less relevant for maritime safety and may be less well reported (Hassel, Asbjørnslett, and Hole 2011). Ship-particular data are available for about thirty variables (and more than 500 when counting dummies for categorical variables), including standard particulars such as ship type, age, size, and flag, as well as owner, classification society, engine designer and builder, proxy variables for maritime expertise such as years in existence, previous inspection and incident histories, and changes of ship-particulars. The included ship types are general cargo, dry bulk, container, tanker, passenger, and other types excluding fishing and tugs.

The incident database for 2010–2014 contains 376,508 observations with (at most) one observation per vessel per year (75,302 vessels on average per year, around 80% of commercial vessels worldwide), with 8,874 VSS incidents. This database also contains information on the types of incident shown in Table 1. The most common types for VSS incidents are drift grounding (25%), collision (24%), powered grounding (20%), and main engine failure (20%), and less common are fire and explosion (11%), pollution (7%), hull failure (7%), and loss of life (4%). Table 1 also shows risk domains associated with each incident type, which PSCs can use for selecting domains of special importance when inspecting a vessel. This database is used in Section 3.2 to estimate the probability of VSS incidents and associated incident types at the level of individual vessels.

The inspection database for 2010–2014 contains 158,187 inspections with 6,458 detentions and can have multiple observations for the same vessel in the same year. This database contains on average 21,117 inspected vessels per year, which corresponds to a vessel coverage rate compared to the incident database of 30%. This database is used to estimate the probability of detention at the level of individual vessels.

The database for January 2017 to March 2018 concerns 71,655 vessels. Inspection, detention, and VSS incident outcomes for 2017 and for the first quarter of 2018 (2018Q1) are shown in Table 2. The yearly VSS incident rate has dropped considerably compared to 2010–2014. The much lower inspection and detention rates compared to 2010–2014 are caused by the fact that the inspection database for 2010–2014 concerns MoUs with relatively high inspection rates and covers only 30% of the incident database. This causes no analysis problems, however, as the inspection database is used

Table 1. Incident types and associated risk domains.

Acronym	Meaning	Associated risk domains for inspection prioritization
VSS	Very serious (incl. total loss) and serious	All of below
COL	Collision and grounding	Passage planning, bridge management, crew qualification (includes navigation and communication failures)
DGR	Drift grounding	Main engine, emergency procedures (includes failing propulsion, steering gear, anchor, and mooring)
ENG	Main engine failure	Main engine, emergency procedures (includes camshaft, crankshaft, turbo charger, engine stoppage)
FIR	Fire and explosion	Fire related aspects, emergency procedures
HUL	Hull failure	Maintenace related issues including tanks and water integrity (includes rudder and tank problems)
LIF	Loss of life	Occupational safety, safety management, life boats
PGR	Powered grounding	Passage planning, bridge management, crew qualification (includes wrecked, stranded, grounded, navigation, communication)
POL	Pollution	Pollution prevention and emergency response

Table 2. Inspection, detention, and VSS incident rates in three periods.

Period	2010–2014	2017	2018Q1 (quarter)	2018Q1 (year)
VSS incidents				
# Vessels	75,302 (A)	71,655	71,655	71,655
# VSS incidents	1,775 (A)	1,178	174	696 (C)
% VSS incidents per vessel	2.36	1.64	0.24	0.97
Inspections				
# Vessels	21,117 (B)	71,655	71,655	71,655
# Inspections	31,637 (A)	81,085	15,868	63,472 (C)
% Inspections per vessel	149.82	113.16	22.15	88.58
# Detentions	1,292 (A)	1,977	480	1,920 (C)
% Detentions per vessel	6.12	2.76	0.67	2.68
% Detentions per inspection	4.08	2.44	3.02	3.02

*Code (A) means that 5-year total counts for 2010–2014 are divided by 5 to get average yearly ones.
*Code (B) gives the average of five yearly numbers of vessels in the inspection database.
*The column 2018Q1 (quarter) shows actual figures for this quarter, and 2018Q1 (year) inflates them to estimated yearly figures by multiplying the quarterly ones by 4, denoted by code (C).

only to relate inspection and detention probabilities to vessel-specific characteristics and not to predict inspection or detention rates. The database for January 2017 to March 2018 further contains estimated probabilities of detention and VSS incidents and associated incident types per vessel computed from information available in December 2017. The data for 2018Q1 are reserved for policy evaluation purposes.

3. Methodology

This section presents a data-driven inspection protocol and describes statistical methods to implement the four steps of this protocol.

3.1. Inspection protocol

The data-driven protocol consists of four consecutive steps.

Step 1. Estimate risk formulas that express probabilities of detention, VSS incidents, and eight incident types in terms of vessel-specific risk factors. Section 3.2 derives these formulas from logit models estimated from historical inspection and incident data for 2010–2014. As these formulas are relatively stable over time and the regulatory framework relevant for port state control does not change quickly, they can be updated at relatively low frequency, for example, once per 5 years. This allows the industry to adapt to changes in the legislative framework and to account for other changes such as the overall improvement of the industry.

Step 2. Use up-to-date information on current values of vessel-specific risk factors to determine current probability scores by means of the formulas of Step 1. Section 3.3 describes how this is done for December 2017. These scores can be computed relatively easily and can be updated frequently, for example, monthly, weekly, or even daily.

Step 3. Determine the set of vessels eligible for inspection, for example, vessels currently in port or also vessels expected to arrive shortly. Evaluate the relative detention and VSS incident risk scores of Step 2 for each of these vessels and rank their risk by one or several classification procedures, as described in Section 3.4 and illustrated in Section 4.3. Determine which vessels will be inspected, depending on their risk ranks, the available inspection capacity, and expert knowledge and other relevant intelligence. Sections 4.1 and 4.2 evaluate how well the proposed classification methods succeed in predicting future detentions and VSS incidents from January to March 2018.

Step 4. For each of the vessels selected for inspection in Step 3, use the probabilities for eight incident types determined in Step 1, combined with expert insights, to select priority areas for inspection. This is described in Section 3.5 and illustrated in Section 4.3.

3.2. Risk formulas

Step 1 of the protocol requires formulas that express probabilities of detention, VSS incidents, and incident types in terms of vessel-specific risk factors. The incident and inspection databases for 2010–2014 are used to estimate logit models for each of these probabilities according to the methodology described in Mueller and Morton (2002), Mueller (2007), Knapp (2006), Knapp and Franses (2007), and Heij and Knapp (2018). Knapp (2015) describes this method for the current database in more detail. Logit models express the probability (p) of an event (detention, VSS incident, or incident type) by means of the fraction $p = \exp(xb)/(1+ \exp(xb))$, where 'exp' denotes the exponential function, 'x' is the set of vessel-specific risk factors such as age, size, flag, and so on, and 'xb' is a weighted average of these factors with estimated weights per factor. Initial models contain some thirty factors and have more than 500 variables when counting dummies for categorical factors like flag, owner, engine designer and builder, and so on. These initial models are down-tested by removing insignificant factors until all remaining factors and dummy variables are significant (at 5% level). All models are estimated by quasi-maximum likelihood (Greene 2008) to allow for possible misspecification of the assumed underlying distribution function for logit models. As the fleet and maritime industry change only gradually and it is rather challenging and time-consuming to update this database, it is recommended to re-estimate the risk formulas, for example, once per 5 years.

Table A1 in the Appendix summarizes the 10 obtained models. The logit models for incidents are estimated for data with one observation per vessel per year. The logit model for detentions is estimated on a sub-sample of vessels and can have multiple observations per vessel per year if the vessel was inspected multiple times in that year. In general, models for more common events contain more risk factors because data on such events are more informative. The richest model is obtained for VSS incidents (8,874 events) and contains 15 scale variables such as age, size, and past incident history, and 172 dummy variables for factors like flag, engine designer and builder, and ship type. The smallest model is for loss of life (376 events) and contains 7 scale variables and 26 dummy variables. As incident types are restricted to VSS incidents, it is not surprising that past VSS incidents have positive effects on most incident type probabilities. Past inspections and detentions, however, are insignificant in most cases; the only exception, possibly by chance, is a positive effect of past inspections on powered grounding risk. This means that, given the other vessel specifics including past incident history, the past inspection and detention history of the vessel has no additional explanatory power for VSS incidents and associated incident types. Past inspections reduce detention risk and past detentions increase this risk, as expected. However, past incident history has no effect on detention risk. The effects of vessel age and size for VSS incident risk are opposite to those for detention risk, as detention risk is higher for older and smaller vessels whereas VSS incident risk is higher for newer and larger vessels. As the detention model reflects actual MoU decisions in practice, these outcomes indicate that inspection and detention policies in 2010–2014 did not incorporate incident information. This finding is directly related to the main message of this paper that past incident information is relevant for targeting vessels for inspection to reduce incident risk.

3.3. Current vessel-specific risk scores

Step 2 of the protocol is to compute up-to-date probabilities for detention and VSS incidents and associated incident types using current information on vessel-specific risk factors. These probabilities are obtained by substituting the current data into the logit probability formulas of step 1, which requires data on the current age of the vessel, its current flag and owner, its recent history of inspections, detentions, and VSS incidents, and so on. These probabilities are computed from information up to December 2017 by specialized software, including filtering of vessels in service at that time and resulting in 71,655 vessels. This software processes the raw data feeds, cleans the data, and applies the formulas to estimate ship-specific probabilities.

The computed model-based probabilities can be interpreted as yearly probabilities only after calibration to correct for sample selection effects in the underlying databases. As the vessel database covers about 80% of the world commercial fleet and VSS incidents may be under-reported, this means that the incident database is incomplete with respect to both non-incident and incident observations. Incident data come from four different sources and are less under-reported than non-incident vessels, which implies that the model-based probabilities overestimate annual incident risk. The calibration needed to correct for these data defects is easily implemented for logit models. Let the database contain fractions 'f' of all non-incidents and 'g' of all VSS incidents worldwide. Then, correct levels of VSS incident probabilities are obtained from estimated logit probabilities $\exp(xb)/(1+\exp(xb))$ by adding the calibration factor $a = \ln(f/g)$ to xb, i.e., by using the formula $p = \exp(a+xb)/(1+\exp(a+xb))$, where 'ln' denotes the natural logarithm. This result is shown in Franses and Paap (2010, pp. 73–75) in case the database contains all events ($g = 1$) but under-reports non-events ($f < 1$) and is easily extended in case also the events are under-reported ($g < 1$). For our data, $f < g < 1$ so that a < 0, which results in downward correction of the estimated probabilities. Comparison of estimated VSS incident probabilities with empirically observed VSS incident rates shows that sample selection effects differ per ship type. The resulting calibration factors for VSS incidents and associated incident types are (rounded to two decimals) $a = -0.22$ for dry bulk vessels, $a = -0.08$ for passenger ships, $a = -0.04$ for general cargo vessels, $a = -0.03$ for tankers, and $a = -0.01$ for container ships and for other ship types.

Calibration along similar lines of estimated detention probabilities is hard because the inspection database covers only a limited part of the global fleet and contains multiple inspections of the same vessel in the same year. Estimated detention probabilities in December 2017 are 6.45% on average, whereas the empirical detention rate over 2017 is 2.76%. The logit detention probabilities are divided by 2 to match them roughly to empirical rates. This crude scaling is done only for intuitive interpretation of the estimated detention probabilities. This scaling does not affect the analysis in any sense because it is not based on numerical values of the probabilities but on percentile ranks that are scale independent.

3.4. Combining detention and incident risk

In step 3 of the protocol, vessels are ranked according to their risk scores computed in step 2. In practice, this ranking can be done for any subset of vessels of interest, but here this is done for the full set of vessels. To prevent difficulties in the interpretation of numerical levels of probabilities, they are transformed into relative scores by means of percentile rank scores. If, for example, a vessel has rank score 83 for VSS incident risk, this means that 83% of the vessels have lower risk and 17% have a higher risk for a VSS incident. Rank scores are computed for the full set of 71,655 vessels and rank from 0 for the least risky vessel to 100 for the most risky one.

The rank scores of each vessel are denoted by RDET for detention and by RVSS for VSS incidents. Two simple ranking methods are to rank vessel risk by RDET neglecting incident risk or by RVSS neglecting detention risk. Figure 1 shows three methods to combine these two risk dimensions, where the combined detention and incident risk is ranked according to the highest of the two ranks (method A), the lowest (method B), or their average (method C). By construction, percentile ranks of vessels are distributed uniformly across both axes. The figure also shows five quintile areas for each method, labelled from 1 for least risky to 5 for most risky. Each quintile contains 20% of all vessels if the two risk dimensions are independent. The fraction of the (100 × 100) square occupied by area 1A, for example, is 0.447 × 0.447 = 0.2, that of area 1B is 2 × 0.106–0.106 × 0.106 = 0.2, and that of area 1C is 0.632 × 0.632/2 = 0.2. RDET and RVSS computed in December 2017 for the set of 71,655 vessels are nearly completely uncorrelated (correlation 0.002) and the percentage of vessels in the five areas of Figure 1 ranges from 19.5% to 20.2% for method A, 19.6% to20.3% for method B, and 19.1% to21.0% for method C. The class of most risky vessels in the top quintile is of special interest and area 5A has 20.2% of the vessels, 5B has 20.1%,

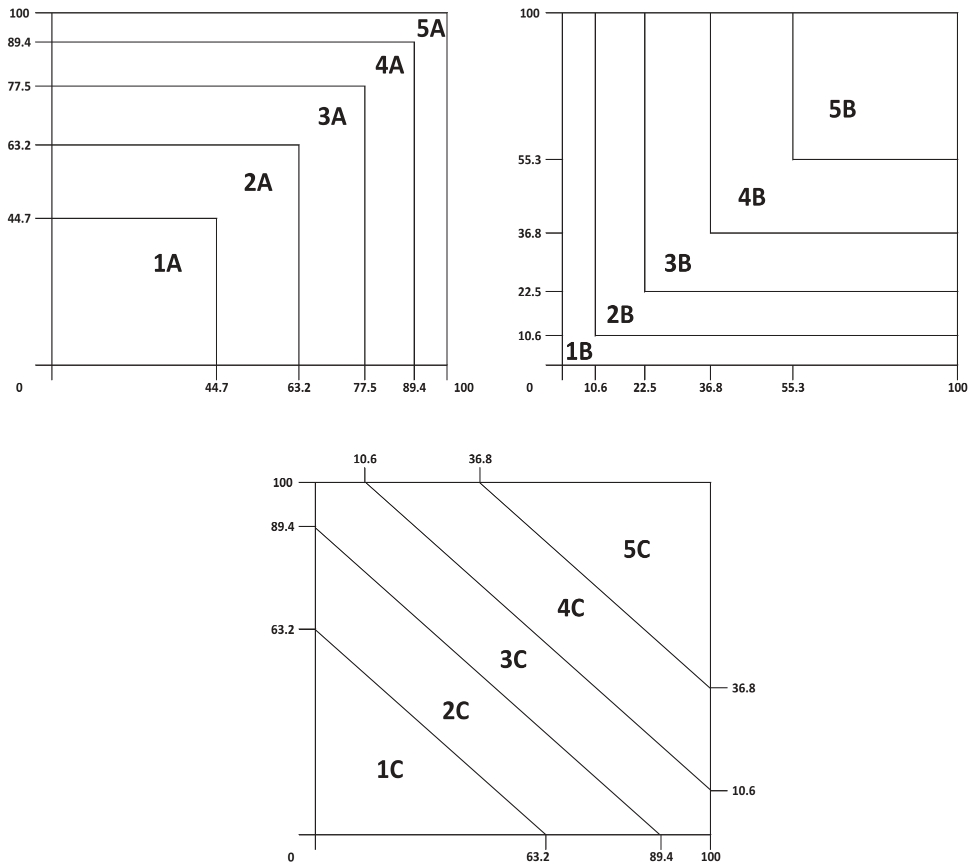


Figure 1. Three risk classification methods for combining the risk of detention and the risk of a VSS incident. The axes show the rank (scale from 0 for lowest to 100 for highest) of two risk indicators: VSS incident probability and detention probability.

and 5C has 19.6%. The empirical analysis in Section 4 will always use exact quintiles with boundaries that differ slightly from those in Figure 1.

As method C uses the average of rank scores of detention and VSS incidents, this corresponds to assigning equal weights to these two risk dimensions. Another option is to leave the selection of weights to regulators, see also Ji, Brinkhuis, and Knapp (2015). One can also follow an empirical approach. For example, a logit model for VSS incidents in 2017 assigns weights of about 2 for RVSS and −1 for RDET, a rather peculiar outcome that indicates once again that past inspection policies do not succeed well in targeting incident risk. This method is infeasible in real time, however, because the rank scores RVSS and RDET are computed with information of December 2017 and can hence not be used to predict VSS incidents in 2017. Estimation of weights based on past predictive performance becomes feasible when longer evaluation periods become available.

3.5. Inspection priorities

If inspection capacity is limited, then inspection efforts can be focused mostly on vessels in the highest risk quintiles. For example, if 20% of all vessels can be inspected, then vessels in the top quintiles 5A, 5B, or 5C are prime targets.

In step 4 of the protocol, for a given set of vessels selected for inspection, risk domains deserving special attention can be determined per vessel based on probabilities for the eight incident types of Table 1. For each vessel, inspectors can check percentile rank scores per risk domain and decide which domains get inspection priority. The risk domains listed in Table 1 include bridge management, passage planning, crew quality, maintenance issues related to engine and hull, occupational safety, emergency procedures, and pollution prevention. In principle, these risk domains can be extended and refined if more detailed incident type information becomes available with a sufficient number of events to construct reliable probability models.

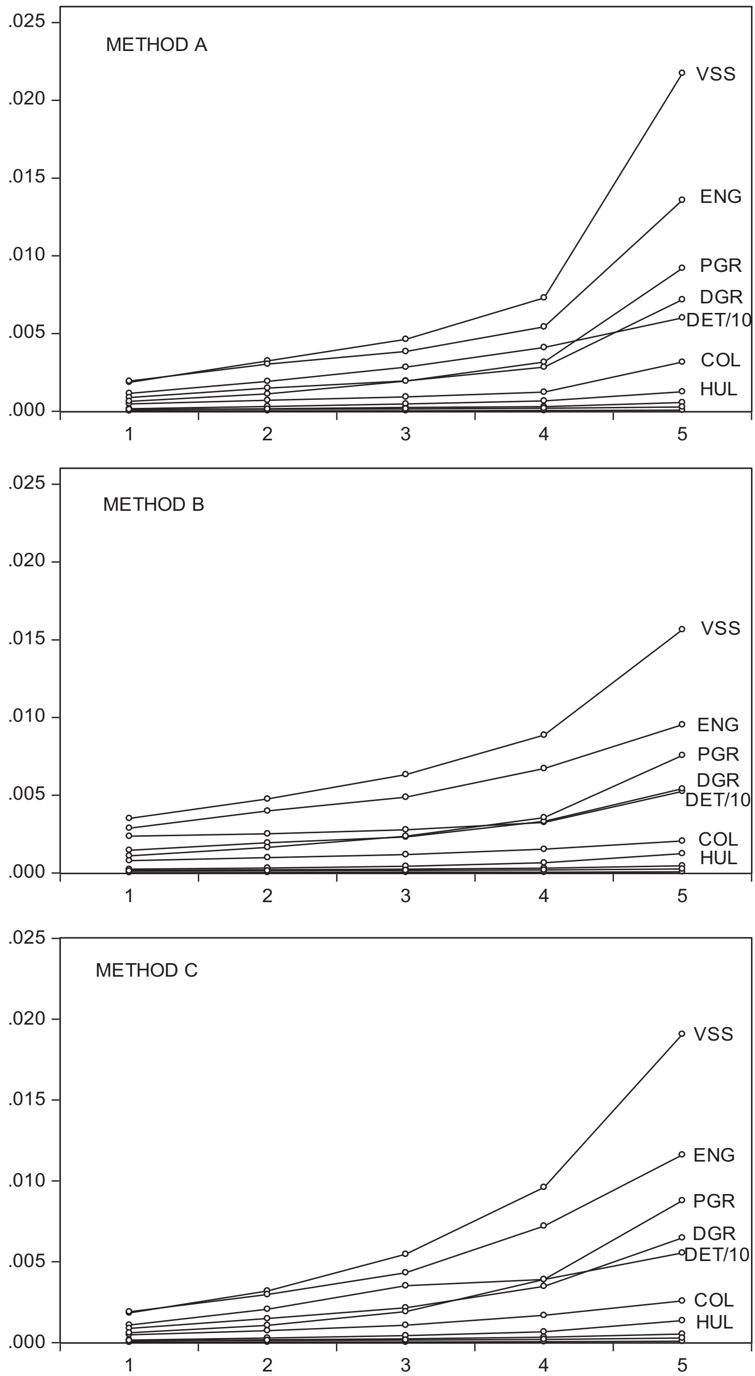
As the eight incident types of Table 1 are analyzed for VSS incidents, it is not surprising that the risk for each type increases per risk quintile. Figure 2 shows average yearly probabilities (on the vertical axis) for each incident type, computed in December 2017 for the set of 71,655 vessels, per risk quintile (on the horizontal axis, with 14.331 vessels per quintile) and for each risk combination method. It also shows average yearly detention probabilities per quintile, which are divided by 10 to get comparable vertical axis scales. All eight incident types have highest average probability in the top quintiles (5A, 5B, 5C), with roughly 2–3 times larger risk than average. Method A is more successful than method B in classifying vessels with the highest incident type risks in its top quintile, whereas method C takes a kind of middle position. The table in Figure 2 shows that engine failures and powered and drift groundings are expected to occur most frequently for future VSS incidents. Although the expected frequencies of loss of life, pollution, and fire and explosion are relatively small, these types of incident have of course serious consequences. The analysis in this paper will, therefore, transform probabilities into percentile rank scores that show the risk position of a vessel relative to the global fleet.

4. Results

This section investigates the predictive out-of-sample performance of the five risk classification methods of Section 3.4 in detecting future risk for VSS incidents. The risk scores of each vessel are determined from logit formulas with weights of risk factors estimated for 2010–2014 (see Section 3.2) and by substituting up-to-date values of these risk factors that apply for December 2017 (see Section 3.3). Performance of the five methods is evaluated for inspections, detentions, and VSS incidents in the forecast period January to March 2018 (2018Q1). It is shown that the four methods that include VSS incident risk (RVSS and methods A, B, and C) provide significant information for future VSS incidents, whereas the method based on detention risk alone (RDET) has no predictive value. This finding is confirmed by comparing the number of VSS incidents in high-risk classes with those in low-risk classes (Section 4.1) and by determining hit rates of vessels with future VSS incidents when each method is used to target vessels for inspections (Section 4.2). A small management example with 12 vessels illustrates how the proposed methods can be applied in practice to select vessels for inspection and to determine domains of attention during inspections (Section 4.3).

4.1. Predictive power of risk classification methods

To assist inspection decisions, the risk of each vessel is ranked relative to a reference group of vessels. In this application, the reference group consists of 71,655 vessels distributed as follows over vessel types: 26% general cargo, 22% tanker, 16% dry bulk, 9% passenger, 7% container, and 20% other excluding fishing and tugs. For each vessel, its probabilities of detention and VSS incidents are determined from data up to December 2017 (see Sections 3.2 and 3.3) and these probabilities are transformed into percentile rank scores RDET for detention and RVSS for VSS incidents. Incident risk is unrelated to detention risk, with correlation −0.03 for the two probabilities and 0.002 for the percentiles RDET and RVSS, showing that past inspection and detention decisions did not account for incident risk. Five vessel risk ranking methods are compared: RDET and RVSS that consider only a single risk dimension, and methods A, B, and C that combine these two dimensions (see Section 3.4 and Figure 1). For each method, vessels are classified into five risk groups, that is, five quintiles that each contains 20% of all vessels (14,331).



Quintile	VSS	DET	ENG	PGR	DGR	COL	HUL	FIR	POL	LIF
All (1-5)	0.78	3.22	0.56	0.32	0.29	0.13	0.06	0.03	0.02	0.00
5A	2.17	6.01	1.36	0.92	0.72	0.32	0.13	0.06	0.01	0.03
5B	1.57	5.25	0.95	0.76	0.54	0.21	0.12	0.05	0.01	0.03
5C	1.91	5.55	1.16	0.88	0.65	0.26	0.14	0.05	0.03	0.01
Highest	A	A	A	A	A	A	C	A	C	A
Lowest	B	B	B	B	B	B	B	B	B	C

Figure 2. Means of incident type probabilities per risk quintile for three risk classification methods. In each figure, the bottom three curves below 'HUL' are (from top to bottom) for FIR, POL, and LIF. The table shows the mean percentage risk (probability times 100) for vessels in the top quintile (top 20%) for each method, as well as the gross mean percentage over all vessels (all quintiles 1-5).

The bottom quintile contains the 20% of vessels with the lowest risk and the top quintile contains the 20% with the highest risk.

Figure 3 shows empirical rates of inspections, detentions, and VSS incidents in 2018Q1 within each quintile (on the horizontal axis) and for each method. These event rates are shown as percentages on the vertical axis, so that a value of 1 means that 1% of the vessels in this quintile had this type of event at least once in 2018Q1. These percentages are most precise for inspections (12,791 vessels, 2,558 per quintile), followed by detentions (462 vessels, 92 per quintile), and they are least precise for VSS incidents (173 vessels, 35 per quintile). These vessel counts differ from the counts for 2018Q1 in Table 2 because vessels can have multiple events. For example, one vessel had two VSS incidents and 172 vessels had one such incident, so the total number of VSS incidents is 174 and the number of involved vessels is 173.

The graph for detentions shows that actual detention rates during the forecast period increase for higher quintiles. This agrees with the idea that the chance of detention should be higher for riskier vessels. Methods B and C have the highest detention rate in their top quintile, so these two methods are most successful in predicting high detention risk.

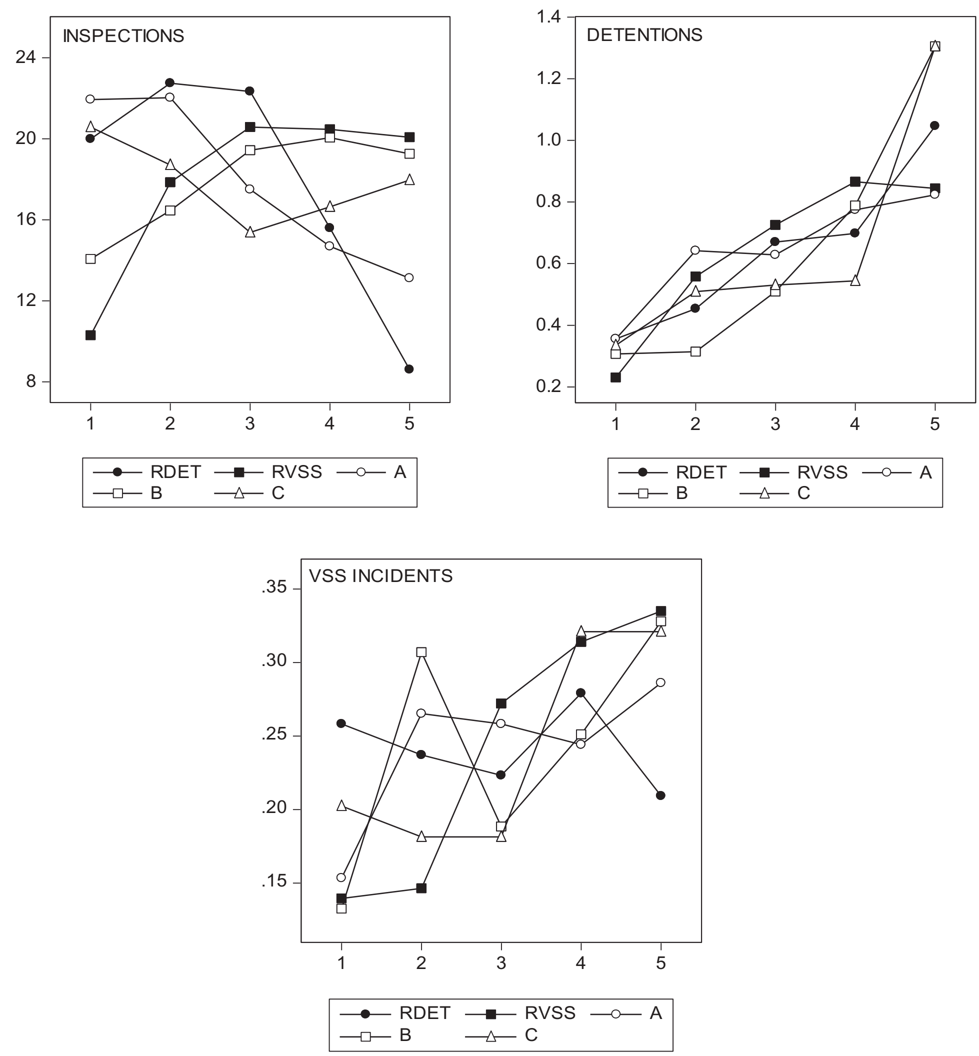


Figure 3. Rates of inspections, detentions, and VSS incidents in 2018Q1 for five risk classification methods.

The graph for inspections shows that inspection rates increase per risk class for methods RVSS and B, fluctuate for method C, and decline for methods RDET and A. The results for RVSS and B are rather weak, as inspection rates are flat across the highest three quintiles. The results for A and especially for RDET indicate that these methods do not correspond with actual inspections in 2018Q1. Stated otherwise, inspections in 2018Q1 were only weakly related to VSS incident risk and even negatively to detention risk. The actual inspection rate in the highest detention risk quintile for RDET is only 8.6%, less than half of the average inspection rate of 21.7% in the three lowest detention risk quintiles.

The graph for VSS incidents shows haphazard patterns due to the relatively small number of VSS incidents in 2018Q1. However, method RVSS shows a clear upward trend and all methods except RDET have clearly higher VSS incident rates in the top quintile compared to the bottom quintile. Incident rates for RDET are rather flat or even declining for higher risk, showing that detention risk alone is not well related to future VSS incident risk.

The visual comparisons in Figure 3 are supported by statistical tests on the research hypothesis that future inspection, detention, and VSS incident rates are higher for vessels falling in higher risk classes. Note that these risk classes have been determined 'ex ante', that is, from past data up to December 2017 and prior to the events in the evaluation period 2018Q1. The test outcomes are shown in Table 3. Outcomes under 'Event yes/no' compare median risk rank scores in two groups of vessels, those with the event of interest (inspection, detention, or VSS incident) and those without. A standard test for rank comparison between two groups is the rank-sum test of Wilcoxon (1945) for equal medians. The table shows the difference between the median rank score in the group with events from that without events, so that positive values are expected. The reported p-values are for the two-sided alternative that median ranks differ in the two groups.

The results for detentions are as expected, with a higher risk for detained vessels compared to non-detained ones. For all five methods, 'ex ante' detention risk signals future detention risk. This signal is strongest for method B, followed by C, RDET, RVSS, and finally A. The results for VSS incidents are also as expected, except for RDET. The strongest signal for future VSS incidents is provided by methods RVSS and C, followed by B, and that of A is small and only marginally

Table 3. Ex ante risk classifications and ex post outcomes for inspections, detentions, and incidents during 2018Q1.

		INS (12,791 = 17.85%)		DET (462 = 0.64%)		VSS (173 = 0.24%)	
Risk method	Test	Diff	P-value	Diff	P-value	Diff	P-value
Event yes/no							
RDET	W	−10.87	0.000	14.71	0.000	−0.27	0.826
RVSS	W	**7.74**	0.000	12.19	0.000	**12.61**	0.000
A	W	−11.41	0.000	9.55	0.000	2.12	0.066
B	W	5.39	0.000	**21.67**	0.000	8.48	0.004
C	W	−3.85	0.000	18.66	0.000	12.43	0.005
Quintile 5/1							
RDET	T	−11.39	0.000	0.69	0.000	−0.05	0.392
RVSS	T	**9.78**	0.000	0.61	0.000	**0.20**	0.001
A	T	−8.81	0.000	0.47	0.000	0.13	0.017
B	T	5.18	0.000	**1.00**	0.000	**0.20**	0.001
C	T	−2.62	0.000	0.97	0.000	0.12	0.049
Quintile 4–5/1–2							
RDET	T	0.69	0.000	0.47	0.000	0.00	0.933
RVSS	T	**6.20**	0.000	0.46	0.000	**0.18**	0.000
A	T	−8.07	0.000	0.30	0.000	0.06	0.170
B	T	4.39	0.000	**0.74**	0.000	0.07	0.098
C	T	−2.34	0.000	0.50	0.000	0.13	0.002

*Test W is the Wilcoxon rank-sum test for equal medians and test T is the Satterthwaite test for equal percentages.
*For 'Event yes/no', 'Diff' shows by how much the median percentile rank score of vessels with event exceeds that of vessels without event, and for the two quintile comparisons 'Diff' shows by how much the percentage of events of vessels in the higher risk class exceeds that of vessels in the lower risk class.
*Bold numbers are for the classification method providing the sharpest contrast.

significant. Method RDET is not successful, as the median risk rank for vessels with VSS incident is even slightly lower than that of vessels without incident. Finally, the results for inspections are mixed because inspections in 2018Q1 were more likely for vessels with higher incident risk but less likely for vessels with higher detention risk. The median 'ex ante' VSS incident risk rank of inspected vessels (56.21) is higher than that of non-inspected ones (48.46), but the median detention risk rank of inspected vessels (41.61) is lower than that of non-inspected ones (52.48).

Another way to test the predictive power of risk classification methods is to compare event rates (of inspections, detentions, and VSS incidents) in high-risk classes with those in low-risk classes. The research hypothesis is that the event rate, measured as the percentage of vessels that has the event in 2018Q1, is higher for higher risk classes. Table 3 shows the difference of event percentages between the highest and lowest risk class, so that positive values are expected. A standard test for the comparison of percentages in two groups is the t-test of Satterthwaite (1946), and p-values reported in Table 3 are for the two-sided alternative that event percentages differ between the two groups. This test is implemented in two ways, one comparing the 20% most risky vessels (quintile 5) with the 20% least risky ones (quintile 1) and another comparing the 40% most risky vessels (quintiles 4 and 5) with the 40% least risky ones (quintiles 1 and 2).

The outcomes support earlier findings. For all five risk classification methods, detention rates are higher in the higher risk group with the largest difference for method B. As expected, the difference for quintile 5 compared to 1 is sharper than that for quintiles 4 and 5 compared to 1 and 2. For method B, for example, the detention rate, which is 0.64% for all vessels, is 1.30% in quintile 5 against 0.31% in quintile 1, and 1.05% in quintiles 4 and 5 against 0.31% in quintiles 1 and 2. The results for inspection rates are again rather mixed. The overall vessel inspection percentage in 2018Q1 is 17.85%, but only 8.60% of vessels in the highest detention risk quintile is inspected compared to 19.99% in the lowest one. Inspection rates are higher for vessels with higher VSS incident risk, 20.07% in quintile 1 against 10.29% in quintile 5 and 20.27% in quintiles 4 and 5 against 14.07% in quintiles 1 and 2. Finally, VSS incident rates are significantly higher for higher risk classes except for method RDET. At an overall incident rate of 0.24%, the rate in the highest (lowest) quintile is 0.33% (0.14%) for RVSS, 0.33% (0.13%) for method B, 0.29% (0.15%) for method A, 0.32% (0.20%) for method C, and 0.21% (0.26%) for RDET. This shows that all methods except RDET succeed rather well in signalling future risk for serious and very serious incidents, as this risk is about twice as large for vessels in the top quintile compared to the bottom quintile.

4.2. Hit rates for VSS incidents at various inspection rates

The main purpose of inspections is to prevent future incidents. The success of inspection strategies to target vessels with future VSS incidents can be measured in terms of the hit rate, defined as the targeted percentage of vessels with future VSS incidents relative to the overall percentage of vessels targeted for inspection. For example, if 50 out of 1,000 vessels have an incident and a strategy selects 200 vessels for inspection 30 of which have an incident, then the hit rate is equal to (30/50) divided by (200/1000), that is, 3. Random strategies unrelated to incidents are expected to have a hit rate of 1, and hit rates above 1 indicate successful targeting.

The five risk classification methods are applied to the set of 71,655 vessels with risk scores computed from information up to December 2017 (see Section 3.3). Now suppose that R% of these vessels can be inspected, then the best way to do this from a risk perspective point of view is to select the vessels with the highest R% of risk scores. The hit rates for VSS incidents in 2018Q1 are shown in Table 4 for each method and for quarterly inspection rates ranging from 5% to 100%. For a quarterly inspection rate of 20%, for example, all vessels are inspected that fall in the highest quintile, that is, with a risk percentile score above 80. The number of vessel hits in the table shows how many of the 173 vessels with a VSS incident in 2018Q1 are selected for inspection in this way. Method 'Random' is a benchmark rule that selects vessels arbitrarily for inspection,

Table 4. VSS incident hit rates during 2018Q1 for various risk-based inspection methods.

Inspection rate (%)		Expected inspections		Risk ranking method					
Quarterly	Yearly	Quarterly	Yearly	Random	RDET	RVSS	A	B	C
				Number of vessel hits					
5	18.5	0.05	0.2	8.65	10	10	11	12	13
10	34.4	0.1	0.4	17.3	17	23	19	25	26
20	59.0	0.2	0.8	34.6	30	48	41	47	46
40	87.0	0.4	1.6	69.2	70	93	76	83	92
60	97.4	0.6	2.4	103.8	102	132	113	110	118
80	99.8	0.8	3.2	138.4	136	153	151	154	144
100	100.0	1.0	4.0	173	173	173	173	173	173
				Hit rate					
5	18.5	0.05	0.2	1.00	1.16	1.16	1.27	1.39	1.50
10	34.4	0.1	0.4	1.00	0.98	1.33	1.10	1.45	1.50
20	59.0	0.2	0.8	1.00	0.87	1.39	1.18	1.36	1.33
40	87.0	0.4	1.6	1.00	1.01	1.34	1.10	1.20	1.33
60	97.4	0.6	2.4	1.00	0.98	1.27	1.09	1.06	1.14
80	99.8	0.8	3.2	1.00	0.98	1.11	1.09	1.11	1.04
100	100.0	1.0	4.0	1.00	1.00	1.00	1.00	1.00	1.00

*The yearly inspection rate is derived from the quarterly one with the assumption that selection for inspection is independent across quarters; for quarterly inspection rate r = R/100, the probability to be inspected at least once in a year is 1—(1-r)^4.
*'Expected inspections' is the expected number of inspections for each vessel.

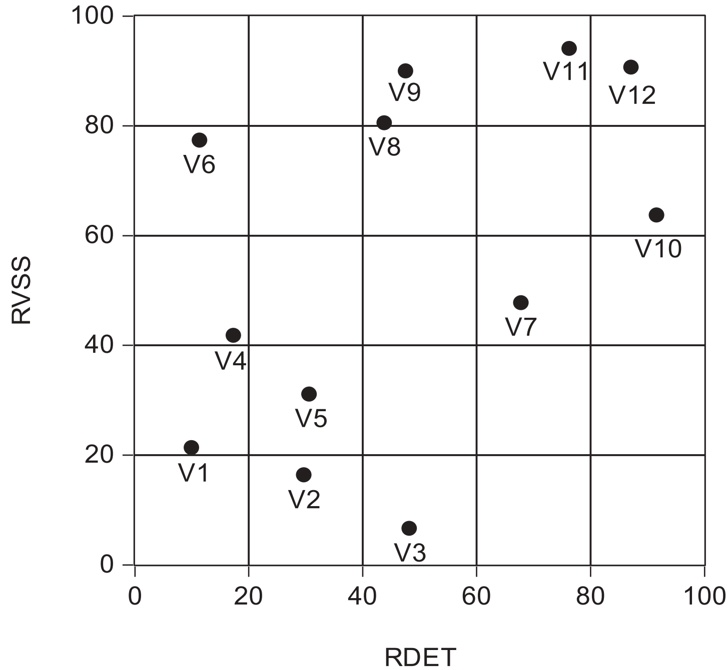
and the reported numbers of vessel hits for this method are expected outcomes. For a quarterly inspection rate of 20%, methods RVSS, B, and C have hit rates above 1.3, meaning they do more than 30% better than random. Methods B and C perform best for smaller inspection rates, with hit rates up to 50% better than random. Method RDET does not perform well and does not outperform the random benchmark, whereas methods RVSS, B, and C increase the hit rate by 30–50% for yearly inspection rates of about 20–60%. Method A is also better than random but performs less well than RVSS, B, and C.

These results confirm earlier findings that inspection policies based on detention risk alone are not successful in targeting risky vessels, whereas incorporation of VSS incident risk improves this targeting considerably. The above risk-based strategies depend solely on quantitative information, and the involved set of risk factors consists of all factors incorporated in the logit risk formulas described in Section 3.2 (see also Table A1 in the Appendix). In practice, inspection decisions involve an interplay of such quantitative information with qualitative expert knowledge and local considerations. The above analysis considers a simplified situation where inspections are based only on quantitative risk information and shows the potential power for improved inspection strategies apart from the evident importance of other information and considerations in the ultimate decision process.

The results in this and the previous subsection indicate that RDET does not perform well and that methods B and RVSS perform best on average, followed by C and A. It is, therefore, recommended to include incident risk in making inspection decisions and, if also detention risk is included, to focus on vessels where both risks are relatively high (method B).

4.3. Application example for inspection management

A small example with 12 real-world vessels is used to illustrate the proposed methods as a quantitative ingredient of the wider decision process which vessels to inspect. The risk formulas estimated from the database for 2010–2014 data and the vessel-specific information of December 2017 are used to calculate risk ranks relative to the world fleet. These risk ranks form the basis for selecting vessels for inspection and for determining inspection priorities. Figure 4 depicts the percentile rank scores RDET for detention and RVSS for incidents of the 12 vessels. The table shows for each vessel its two rank scores, whether it had an inspection, detention, or VSS incident in 2018Q1, the quintile scores of five risk classification methods, and the percentile scores



	Percentile		Event			Quintile					Percentile							
	RDET	RVSS	VSS	INS	DET	RDET	RVSS	A	B	C	COL	DGR	ENG	FIR	HUL	LIF	PGR	POL
Scale	(0-100)		(1 = yes, 0 = no)			(1 = lowest to 5 = highest)					(0 = lowest to 100 = highest)							
V1	10	21	0	1	0	1	2	1	1	1	36	40	53	47	10	61	14	48
V2	30	16	0	0	0	2	1	1	2	1	16	26	45	4	4	34	14	39
V3	48	7	0	1	0	3	1	2	1	1	2	35	28	0	41	31	1	51
V4	17	42	0	1	0	1	3	1	2	1	62	58	60	30	9	45	16	45
V5	31	31	0	1	0	2	2	1	3	1	39	38	58	14	11	32	50	39
V6	11	77	0	1	1	1	4	3	2	2	66	59	86	40	41	59	89	81
V7	68	48	1	0	0	4	3	3	4	4	11	17	4	58	69	89	42	40
V8	44	80	1	1	0	3	5	4	4	4	80	53	62	35	25	37	83	30
V9	48	90	1	1	0	3	5	5	4	5	90	86	90	82	64	49	85	39
V10	92	64	1	1	0	5	4	5	5	5	40	30	41	67	67	59	47	73
V11	76	94	1	1	0	4	5	5	5	5	78	58	68	76	71	67	73	93
V12	87	91	1	0	0	5	5	5	5	5	93	95	95	67	59	55	92	34

Figure 4. Risk indicators, inspections, and targeting strategies for a set of 12 vessels.

for eight incident types. It is now shown how this information can be used for steps 3 and 4 of the inspection protocol to identify high-risk vessels and inspection domains of interest.

Vessels 1 and 3 to 6 (two tankers, two dry bulk vessels, and a general cargo vessel) were in fact inspected in 2018Q1 and vessel 6 (general cargo) was detained, but none of these vessels had a VSS incident ('false positives'). The risk classification methods assign low risk quintiles to these five vessels, suggesting that these inspections had low priority. However, vessel 6 has a high risk for some incident types, especially for powered grounding, engine failure, and pollution, which may justify its inspection and detention. Vessels 2 (dry bulk), 7 (passenger ship), and 12 (general cargo) were not inspected, and the classification methods agree with this decision for vessel 2 but not for vessel 12 and, to a lesser degree, also not for vessel 7 which has a high risk for loss of life. The decision not to inspect vessel 2 is correct in the sense that it had no VSS incident ('correct negative'), but vessels 7 and 12 did have such an incident in 2018Q1 ('false negatives'). Vessel 12 has top priority for each of the five risk classification methods, and it would not have been missed for an inspection rate of 20% which is comparable to the actual inspection rate during 2018Q1, which is 22.15% on average and involves 17.85% of all vessels. Vessels 8 to 11 (two container ships, a tanker, and a general cargo vessel) were inspected but not detained and all of these vessels had a VSS incident in 2018Q1 ('correct positives'). For an inspection rate of 20%, four of the five risk classification methods would have signalled inspection for vessels 10 and 11, three for vessel 9, and one for vessel 8. For this inspection rate, methods RVSS, A, and C succeed in selecting four of the six vessels with a VSS incident, B selects three of them, and RDET two. Again, the incorporation of VSS incident risk helps in signalling vessels with high future incident risk.

Once vessels have been selected for inspection, the percentile scores of incident types can be used to identify inspection domains of special interest. Vessels 9–12 have top risk and would be selected for inspection by almost all classification methods. The incident type risk of vessel 9, a container ship, belongs to the top 20% of all vessels in the following five risk domains (see Figure 4): collision, engine failure, powered and drift grounding, and fire and explosion. Inspection of this vessel can put priority on these domains. Vessel 10 is a tanker and has high detention risk and quite high VSS incident risk, but incident type risks are quite flat across the eight domains so it is not easy to set priorities for this vessel. Vessel 11 is a general cargo vessel and has a top risk for pollution and quite evenly spread moderately high risk for all other incident types, which makes it hard to set inspection priorities apart from pollution. Vessel 12 is also a general cargo vessel with high risk in nearly all domains except pollution, with the highest risks for collision, powered and drift grounding, and engine failure.

5. Conclusion

Inspections play a key role in protecting the marine environment and in enhancing maritime safety. This paper proposes a four-step protocol for quantitative support of inspection targeting decisions that can be useful, for instance, for beneficial owner or safety management companies. The first two steps consist of using rich information sets to identify risk on the level of individual vessels for various risk dimensions, with detention risk and incident risk as the two primary dimensions and the risk of eight incident types as secondary dimensions. The last two steps consist of ranking vessels according to the first two dimensions to select vessels for inspection and of determining risk domains deserving special attention during inspections. The proposed methodology is illustrated by an empirical application using three databases on vessel particulars, inspections, detentions, and incident types. The first database for 2010–2014 is used to estimate the risk formulas for step 1 of the protocol, the second one for December 2017 gives up-to-date risk values for step 2, and the third one for the first quarter of 2018 evaluates the out-of-sample predictive performance of five risk classification methods for steps 3 and 4. In practice, the first database can be updated with relatively low frequency (e.g. once per 5 years), whereas this frequency should be higher for the second database (e.g. once per quarter) and for the third database (e.g. monthly, weekly, or even daily, depending on available data updates).

The outcomes show that current inspection strategies that are based mainly on detention risk can be considerably improved by also incorporating incident risk. Such extended methods produce significant signals for future risk of serious and very serious incidents, as the out-of-sample frequency of these incidents is twice as large for vessels with large (top 20%) risk compared to those with small (bottom 20%) risk. These extended methods also increase the hit rate of vessels with future VSS incidents, which for detention-based inspection methods are no better than random but which increase by 30–50% by incorporating incident risk for yearly inspection rates of 20–60%. The general recommendation is to include incident risk in making inspection decisions and, if also detention risk is included, to focus on vessels where both risks are relatively high (risk classification method B).

The methodology and empirical analysis presented in this paper can be extended in several ways if more data become available. In the application of the inspection protocol for the first quarter of 2018, the risk scores of each vessel stay fixed as determined in step 2 for December 2017. This can be improved by repeating step 2 more frequently, for example, monthly, weekly, or even daily to incorporate the most recent vessel information, in particular on recent inspections, detentions, and incidents. Further, the presented analysis is based on percentile ranks relating to the full set of vessels. It may be of interest to apply some or all steps of the protocol for smaller subsets of vessels, for example, per ship type or per inspection region, to tune inspection priorities more closely to the vessels of prime interest. No information on the incident types of Table 1 was available for the application in the first quarter of 2018, and it is of interest to add this information to evaluate how well the proposed methods succeed in signalling specific incident type risk. Another extension of interest is to refine the incident types and risk domains to deliver more detailed targeting suggestions for inspections, which would require more refined information on the precise causes of each incident.

Acknowledgments

We thank IMO, AMSA, LLI, and IHS Markit for providing their data to us for 2010-2014 and for January 2017 to March 2018.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Cariou, P., M. Q. Mejia, and F. C. Wolff. 2007. “An Econometric Analysis of Deficiencies Noted in Port State Control Inspections.” *Maritime Policy & Management* 34 (3): 243–258. doi:10.1080/03088830701343047.

Cariou, P., M. Q. Mejia, and F. C. Wolff. 2008. “On the Effectiveness of Port State Control Inspections.” *Transportation Research Part E* 44 (3): 491–503. doi:10.1016/j.tre.2006.11.005.

Emecen-Kara, E. G., and O. Oksas. 2016. “A Comparative Analysis of Regional Agreements on Port State Control.” *American Scientific Research Journal for Engineering, Technology, and Sciences* 18 (1): 259–270.

Franses, P. H., and R. Paap. 2010. *Quantitative Models in Marketing Research*. Cambridge: Cambridge University Press.

Greene, H. W. 2008. *Econometric Analysis, 6-th Edition*. Upper Saddle River, NJ: Pearson Prentice Hall.

Hassel, M., B. E. Asbjørnslett, and L. P. Hole. 2011. “Underreporting of Maritime Accidents to Vessel Accident Databases.” *Accident Analysis and Prevention* 43: 2053–2063. doi:10.1016/j.aap.2011.05.027.

Heij, C., G. Bijwaard, and S. Knapp. 2011. “Ship Inspection Strategies: Effects on Maritime Safety and Environmental Protection.” *Transportation Research Part D* 16: 42–48. doi:10.1016/j.trd.2010.07.006.

Heij, C., and S. Knapp. 2018. “Predictive Power of Inspection Outcomes for Future Shipping Accidents – An Empirical Appraisal with Special Attention for Human Factor Aspects.” *Maritime Policy & Management* 45 (5): 604–621. doi:10.1080/03088839.2018.1440441.

IMO. 2000. “Reports on Marine Casualties and Incidents - Revised Harmonized Reporting Procedures.” IMO Document MSC/Circ. 953, MEPC/Circ. 372. 14 December 2000. London: IMO.

Ji, X., J. Brinkhuis, and S. Knapp. 2015. “A Method to Measure Enforcement Effort in Shipping with Incomplete Information.” *Marine Policy* 60: 162–170. doi:10.1016/j.marpol.2015.06.015.

Knapp, S. 2006. *The Econometrics of Maritime Safety – Recommendations to Improve Safety at Sea*. Rotterdam: ERIM Ph.D. Series Research in Management.

Knapp, S. November 24, 2015. “Methodology and Implementation Aspects for Risk Formulas at Ship Level.” Technical Report for AMSA, Contract 15AMSA116.

Knapp, S., and P. H. Franses. 2007. “A Global View on Port State Control - Econometric Analysis of the Differences across Port State Control Regimes.” *Maritime Policy and Management* 34 (5): 453–482. doi:10.1080/03088830701585217.

Li, K. X., J. Yin, and L. Fan. 2014. “Ship Safety Index.” *Transportation Research Part A* 66: 75–87.

Luo, M., and S. H. Shin. 2019. “Half-century Research Developments in Maritime Accidents: Future Directions.” *Accident Analysis and Prevention* 123: 448–460. doi:10.1016/j.aap.2016.04.010.

Mueller, W. 2007. “Statistical Modelling of Ship Detention Based on Ship Inspection Data 1996 to 2005.” CSIRO Internal Report CMIS 2007/128, Commonwealth Scientific and Industrial Research Organisation, Canberra.

Mueller, W., and R. Morton. 2002. “Statistical Analysis of AMSA Ship Inspection Data, 1995-2001.” CSIRO Internal Report CMIS 2002/01, Commonwealth Scientific and Industrial Research Organisation, Canberra.

Paris Memorandum of Understanding on Port State Control. 2014. “Annex 7: Ship Risk Profile.” accessed 7 March 2019, https://www.parismou.org/system/files/Annex%207%20ship%20risk%20profile.pdf

Paris Memorandum of Understanding on Port State Control. 2019. “Ship Risk Calculator.” accessed 7 March 2019, https://www.parismou.org/inspections-risk/ship-risk-calculator

Perepelkin, M., S. Knapp, G. Perepelkin, and M. De Pooter. 2010. "An Improved Methodology to Measure Flag Performance for the Shipping Industry." *Marine Policy* 34 (3): 395–405. doi:10.1016/j.marpol.2009.09.002.

Satterthwaite, F. E. 1946. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (6): 110–114. doi:10.2307/3002019.

Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *Biometrics* 1 (6): 80–83. doi:10.2307/3001968.

Yang, Z., Z. Yang, and J. Yin. 2018. "Realising Advanced Risk-based Port State Control Inspection Using Data-driven Bayesian Networks." *Transportation Research Part A* 110: 38–56.

Yang, Z. L., J. Wang, and K. X. Li. 2013. "Maritime Safety Analysis in Retrospect." *Maritime Policy & Management* 40 (3): 261–277. doi:10.1080/03088839.2013.782952.

Appendix

Table A1. Logit model summaries.

Dependent variable		Number of					Past information				Ship particulars	
									VSS Incidents			
		Observations	Events	Variables	Dummies	McFadden R-squared	Inspections	Detentions	VS	S	Age	Size
DET	Detention	158,187	6,458	16	42	0.106	–	+	0	0	+	–
VSS	VSS incident	376,508	8,874	15	172	0.232	0	0	+	+	–	+
COL	Collision	376,508	2,109	12	97	0.165	0	0	0	+	–	+
DGR	Drift grounding	376,508	2,192	9	120	0.211	0	0	0	+	0	0
ENG	Main engine failure	376,508	1,773	8	74	0.198	0	0	0	+	0	0
FIR	Fire & explosion	376,508	965	11	16	0.159	0	0	+	+	0	+
HUL	Hull failure	376,508	600	10	17	0.217	0	0	+	+	0	+
LIF	Loss of life	376,508	376	7	26	0.164	0	0	0	0	+	+
PGR	Powered grounding	376,508	1,803	13	111	0.196	+	0	0	+	–	+
POL	Pollution	376,508	609	10	23	0.142	0	0	0	+	0	+

*'Dependent variable' shows the explained binary variable of the logit model, and 'Events' shows the number of observations with such an event (detention, VSS incident, or incident of given type).
*"'Variables' and 'Dummies' show the number of scale variables and the number of dummy variables in the final logit model after omitting insignificant variables (at 5% level).
*'Past information' shows the effects of past information variables in the logit models, positive (+), negative (–), or not significant at 5% level (0); considered past incidents are very serious (VS) and serious (S).
*Ship particulars' shows the effects of age and size (gross tonnage, GRT) of the vessel, measured in natural logarithms (ln) with variables ln(1+ AGE) and ln(GRT).