Ship Selection in Port State Control: Status and Perspectives

Abstract

Port state control (PSC) inspection is the safeguard of maritime safety and the marine environment. Due to the limited inspection resources and high inspection costs, the port states could only select substandard ships with higher risk for inspection. Therefore, efficient and accurate identification of substandard ships is of vital importance. This paper reviews ship selection methods currently implemented in practice and proposed in the existing literature and then discusses their pros and cons. Based on the review, a combined model that considers both ship deficiency and detention is developed and validated in this study. Reasonable and comprehensive comparisons between the proposed combined model and the ship selection method currently used at the Hong Kong port are conducted. The results of the comparison provide managerial insights and suggestions to the Memorandum of Understandings (MoUs). This study serves as the pioneer to review the ship selection methods implemented at the port states and proposed in current literature for PSC inspection. It is also the first study that combines the number of ship deficiency and detention in a unified model for ship risk calculation. This study is valuable for improving ship selection efficiency in MoUs and thus for safeguarding maritime transportation.

Keyword

Port state control, machine learning in maritime transportation, ship selection, ship deficiency, ship detention

1. Introduction

Maritime transport is the backbone of globalized trade and the manufacturing supply chain. According to the records of United Nations Conference on Trade and Development (UNCTAD), more than 80% of world merchandise trade by volume was carried by sea in 2018 (UNCTAD, 2019). Maritime transport is relatively safe and environmentally friendly compared with other transportation modes. However, as ships are vulnerable to extreme sea and weather conditions, ship accidents happen from time to time (Heij and Knapp, 2018; Jiang and Lu, 2020). Regarding environmental impacts of shipping industry, it is reported by the International Maritime Organization (IMO) that the total shipping carbon dioxide (CO_2) emissions increased from 962 million tonnes in 2012 to 1,056 million tonnes in 2018 (IMO, 2020). Meanwhile, the pollutants generated by the shipping industry, such as sulfur oxides, nitrogen oxides, and particulate matter, have serious adverse impacts on the marine environment and public health (Lun et al., 2014; Sampson et al., 2016; Shi et al., 2018). To effectively prevent ship accidents and reduce the pollutants from shipping industry, numerous international conventions and regulations are implemented, such as the International Convention for the Safety of Life at Sea (SOLAS) for maritime safety and the International Convention on the Prevention of Pollution from Ships (MARPOL) for environmental protection. Besides, to protect the rights and guarantee decent living and working conditions of seafarers, Maritime Labour Convention 2006 (MLC) is proposed by the International Labour Organization (ILO).

Ships whose hull, machinery, equipment or operational safety substantially below the standards required by the international maritime conventions and regulations or whose crew is not in conformance with the safe manning document are deemed as “substandard ships” (IMO, 2017). Ship flag state, which is a country where a commercial ship is registered or licensed with, is viewed as the first line of defense against substandard ships. However, traditional flag state control has its limits in terms of ensuring the implementation of maritime regulations by the ship owners, especially those choosing open registration (Li and Zheng, 2008; Wang et al., 2019). As a supplementary of flag state control, port states control (PSC) is proposed, which acts as the second line of defense against substandard ships. It authorizes port states to check whether foreign ships visiting their ports meet all the appropriate convention standards by the IMO and its impact on substandard ships is increasing. The objective of PSC inspection is to identify substandard ships and discourage ship owners from operating them. During a PSC inspection, a condition found not to comply with the requirements of the relevant convention(s) is called a deficiency. If critical deficiencies are found onboard, intervention action can be taken by the port state which is also called ship detention (IMO, 2017). To make PSC inspection more effective, regional Memorandum of Understanding on PSC (MoU on PSC) is established through co-operation of its members and harmonization of the members' activities. The development of regional cooperation for the PSC enables the exchange of information among the member states, prevents multiple inspections in the same region within a period of time, and eliminates the negative actions that reduce the commercial activities of neighboring ports (IMO, 2019). Currently, there are nine regional MoUs all over the world, namely Paris MoU, Tokyo MoU, Acuerdo Viña del Mar, Caribbean MoU, Mediterranean MoU, Indian Ocean MoU, Abuja MoU, Black Sea MoU, and Riyadh MoU.

2. Ship selection methods and basic inspection procedure in MoUs on PSC

One critical point faced by the MoUs on PSC is how to select ships for inspection, as port states recognize that inspecting all foreign ships would be both impractical due to the limited resources, and unnecessary since only part of the visiting ships are substandard (Intercargo, 2000). For one thing, conducting PSC inspection on a ship can be time-consuming and costly: a typical PSC inspection carried out by one or more PSC inspectors, who are also called PSC officers (PSCOs), usually takes two to three hours on board. If no deficiency is found, the cost per PSC is estimated to be 509 USD. If any deficiency is detected, the cost is estimated to reach 759 USD (Knapp, 2007). On the other hand, only part of the inspected ships has deficiency and much fewer ships are detained. According to the 2019 annual report of Tokyo MoU, which is the MoU in Asia-Pacific region and came into being in the early 1990's, only about 60% of all inspections from 2009 to 2019 detect deficiencies, and the highest annual detention rate is no more than 6% within this period (Tokyo MoU, 2020). It should also be mentioned that the deficiency rate and detention rate are calculated based on the inspection records of ships regarded to have higher risk and are selected to be inspected, and thus both ratios should be heavily reduced if taking all visiting ships into account. As a result, accurately targeting substandard ships not only guarantees the effectiveness and efficiency of PSC but also contributes to maritime safety and the marine environmental protection by ensuring the implementation of the related international and local conventions and regulations.

Uniform ship selection procedure is adopted by the ports within an MoU on PSC. Take the Tokyo MoU as an example, the new inspection regime (NIR) is adopted to determine the inspection priority and inspection time interval of the visiting ships through the calculation of ship risk profile (SRP) (Tokyo MoU, 2014). The calculation uses criteria on an information sheet containing generic parameters (e.g. ship type, age, flag, recognized organization, and company) and inspection historical parameters (e.g. previous deficiencies and detentions), as is shown in Table 1. Each parameter is given a fixed weighting point and the SRP is determined by the total weighting points. Based on the total weighting points of the parameters, ships are divided into three risk profiles:

high risk ship (HRS), standard risk ship (SRS), and low risk ship (LRS). The selection scheme of the current inspection framework, which determines the scope, frequency, and priority of inspection is implemented by attaching the inspection time window to each profile. A time window is a period of time between the last and the current inspections out of which a ship must be inspected. The definition of SRP and the corresponding time window are presented in Table 2. There are two types of inspection priority in the current inspection framework: priority I includes ships whose time window is closed (i.e. time interval since last inspection exceeds the upper bound of the associated time window) and must be inspected; priority II includes ships within the time window and that may be inspected. Ships that do not enter the time window (i.e. time interval since last inspection is less than the lower bound of the associated time window) should not be inspected generally. Nevertheless, overriding factors, including but are not limited to ships being subject of report or notification by another authority and ships which have been permitted to leave a port state but are with deficiencies required to be rectified within a specific period, can trigger additional inspection between periodic inspections. In addition, ships without inspections before within the MoU are also given higher inspection priority (Tokyo MoU, 2014). Take an HRS as an example to illustrate the selection process. If its last inspection is no longer than 2 months, it should not be inspected unless overriding factors are reported; if its last inspection is between 2 to 4 months ago, it may be inspected as decided by the port state; if its last inspection is more than 4 months ago, it must be inspected.

After the ships for inspection are selected, the PSCO(s) will be assigned to the ships for inspection. As required by IMO, the basic inspection procedure is similar in the MoUs: an initial inspection is first conducted; if clear grounds of disconformity are detected, a more detailed inspection is carried out (IMO, 2017). Take the Port of Hong Kong as an example, usually one PSCO is assigned to a ship selected for inspection.

An initial inspection consisting of at least a visit on board a ship to check the certificates and documents will be carried out. If the overall conditions of the ship, including its equipment, machinery spaces and accommodation, and hygienic conditions on board, are deemed satisfactory by the PSCO, the inspection can be terminated. Otherwise, a more detailed inspection can be carried out. For the detained ships, a follow-up inspection should be conducted by attended PSCO on board before departure to check whether the deficiencies are rectified. The inspected ships without detention are also encouraged to apply for follow-up inspection (either onsite or remote) after rectifying their deficiencies.

Insert Table 1 here

Insert Table 2 here

Apart from the Tokyo MoU, other MoUs on PSC also adopt their own methods for ship selection. For example, Paris MoU, Abuja MoU, and Black Sea MoU also adopt NIR for SRP calculation and ship selection. However, the NIR used in these three MoUs is slightly different from that used in Tokyo MoU. For instance, the information sheet for SRP calculation in these three MoUs is different from the sheet used in Tokyo MoU (Abuja MoU, 2012; Black Sea MoU, 2016; Paris MoU, 2014), and the time windows attached to HRS, SRS, and LRS are 5–6 months, 10–12 months, and 24–36 months, respectively in Abuja MoU and Paris MoU (Abuja MoU, 2012; Paris MoU, 2014). Some other ports adopt more simple ship selection methods. For example, in the Mediterranean MoU, ships that have previous inspections during the last six months and that are found to be in compliance with the regulations are exempted from further inspection unless there are clear grounds to warrant further investigation (Mediterranean MoU, 2020).

3. Review of literature of ship selection methods for PSC

A recent literature review divides current literature of PSC into four main categories: factors influencing PSC inspection results, selection scheme of ships for inspection in PSC, effect of PSC inspection, and suggestions for MoU management (Yan and Wang, 2019). In this paper, we focus on literature that proposes ship selection scheme used in PSC inspection. Li (1999) is the pioneer who proposed an innovative risk score system to evaluate ship quality in PSC inspection. The author took several ship generic factors into account: ship age, flag, insurers, classification, and operators. The concept of risk score was also adopted by Degré (2007) to select high-risk ships for PSC inspection by combining three static physical ship factors, i.e. type, size, and age, with the criteria considered by Paris MoU, namely ship flag, recognized organization, and company. A vessel risk assessment system for PSC was developed by Xu et al. (2007a) based on support vector machine (SVM), and the system performance was further improved by combining with web mining technique for extracting new target features (Xu et al., 2007b). Another ship risk assessment system for PSC was proposed by Gao et al. (2007) which combined k-nearest neighbor with support vector machine (KNN-SVM) to remove noisy training samples and adopted bag of words (BW) to extract new features. Ship detention was the prediction target of all the above three papers: if a ship was predicted to be of high-risk and was detained, the prediction was regarded to be accurate. The highest accuracy of the three models is about 22% as a result of the highly imbalanced distribution of ship detention in the dataset: the number of inspection records with detention is much less than the records without detention. The imbalanced data has made the prediction a complex task. A self-evolutional ship targeting system for ship detention prediction was implemented by Zhou and Sun (2008) using generalized additive modeling (GAM) to relieve the negative Matthew Effect in PSC inspection at the Ningbo port. The negative Matthew

Effect is brought about by the currently used ship selection scheme as the current ship target system would unavoidably set those ships with bad history into a vicious circle by increasing their inspection frequency.

In recent years, more advanced and accurate ship selection models are proposed and implemented. Based on the inspection data of bulk carriers from 2005 to 2008 in Paris MoU, a Bayesian Network (BN) based approach was implemented by Yang et al. (2018a) to predict ship detention. The key factors influencing detention in PSC inspections were also analyzed, namely the number of deficiencies, type of inspection, recognized organization, and ship age. Based on the BN model proposed by Yang et al. (2018a), a strategic game model incorporating the BN model outcomes was proposed to figure out the optimal inspection rate at the port states. Recommendations for port authorities were generated based on the results: when port authorities have sufficient resources, they should choose the optimal inspection rate; otherwise they can increase the punishment severity level to tackle the sub-standard efforts and illegal actions of ship owner (Yang et al., 2018b). BN model was also adopted by Wang et al. (2019) to predict ship deficiency number in PSC inspection. Besides, comparison between the proposed BN model with the currently used SRP ship selection scheme in Tokyo MoU was carried out and the superiority of the BN model was illustrated. To predict ship detention probability, a random forest based model was proposed by Yan et al. (2020). Comparison between the proposed detention prediction model and the SRP ship selection scheme was made and the results showed that the newly proposed model outperformed the SRP ship selection scheme regarding identifying detained ships. Apart from ship generic factors and inspection historical factors, ship involved casualties and incidents were also considered in some ship selection studies, as they could indicate higher ship risk and possible future incidents (Heij and Knapp, 2019; Knapp and Heij, 2020).

The ship risk prediction model proposed in the above mentioned studies are all calibrated on real PSC inspection records and data from other official databases (e.g. ship information database and ship incident database) and can thus enhance the accuracy of the prediction results. Besides, more comprehensive influencing factors are considered in the newly proposed models (e.g. ship generic factors, ship dynamic factors, and ship inspection historical factors) compared to the SRP ship selection scheme, which can also improve the quality of the prediction results and thus give better suggestions to port states on targeting higher risk ships for inspection. In addition, compared to the simple weighted-sum method used to generate the ship risk profiles, more advanced models are adopted to calculate ship risk level, and thus the mutual effect between the features and the overall model structure is much more reasonable. However, several drawbacks also exist in the proposed models in current literature. First, for models that purely generate ship risk scores, it can be hard to evaluate the results as we cannot observe the "risk" of ships. What we can observe is the inspection results: ship deficiency (both total deficiency number and deficiency category) and ship detention. Therefore, it would be more intuitive to develop prediction models for ship deficiency and detention, as they are also the main target of a PSC inspection. Second, for studies trying to improve inspection efficiency, very few of them consider both ship deficiency and detention. Even though ship deficiency and detention are highly related, they are not equivalent. As both are important and irreplaceable, if we can combine them in a unified model under a reasonable framework, ship selection efficiency can be further improved. Third, the methods used to compare the proposed models and the currently used ship selection model in MoUs in current literature usually ignore the general ship selection framework currently adopted within the MoUs (i.e. ships are divided into two inspection priorities, namely priority I and priority II). As a result, even if the proposed models are more accurate and efficient, its superiority over the currently used methods to calculate ship risk level remains to be verified. As the currently implemented ship selection framework is widely adopted and thus can be hard to change within the MoUs, figuring out the advantages of the proposed models over the current ship risk calculation method within the current inspection framework can further clarify the possibility of the application of the proposed models into practice.

To address these issues, a ship selection model considering both ship deficiency and ship detention is proposed and validated in this study using the PSC inspection records at the Hong Kong port from 2016 to 2018. Bothe ship deficiency and detention in PSC inspection predicted by advanced machine learning models are incorporated in a unified model reasonably (denoted by combined model for short), and comprehensive comparisons between the newly proposed combined model with the SRP ship selection scheme considering and ignoring the inspection priority used in the current ship selection framework of Tokyo MoU are conducted in a rational way. Based on the review of current ship selection models adopted by the port states and the selection methods proposed in the current literature, together with the numerical study results of the proposed combined model, managerial insights and suggestions are also proposed for the MoUs on efficient and effective ship selection in PSC.

4. Combined model for ship selection

4.1 Model structure

In this section, a combined ship selection model considering both the predicted ship deficiency number and the ship detention probability is proposed and validated. The ship deficiency prediction model is the Tree Augmented Naive Bayes (TAN) classifier developed based on Wang et al. (2019), and the ship detention probability prediction model is the balanced random forest (BRF) model developed based on Yan et al. (2020). More specifically, the TAN model is a type of Bayesian network which considers both the dependency between the class variable (i.e. prediction target) and the attribute variables (i.e. features) and the dependency among the attribute variables. In this study, the TAN classifier can predict the probability of ship deficiency number in each state (which is the class variable), and the total ship deficiency number can be further calculated as the sum of the product of probability of each state and the average deficiency number of that state in the training set. As ship detention probability is quite low (which is no more than 6% at the Hong Kong port in recent 10 years), we adopt the BRF model for ship detention probability prediction which is capable to deal with the imbalanced distribution of the prediction target in the dataset. The BRF model is based on the popular random forest model (Breimain, 2001) and it can generate balanced dataset to construct each tree in the forest. After a forest containing a certain number of trees is constructed, the probability of a ship to be detained can be calculated by dividing the number of trees predicting the ship to be detained by the total number of trees in the forest.

After obtaining the predicted ship deficiency number and detention probability, we calculate the risk score of a ship combining ship deficiency and detention which is denoted by $risk$. We denote the predicted deficiency number by def', $def' \in [0, +\infty)$, and the predicted detention probability by det', $det' \in [0,1]$. As the value ranges for def' and det' differ greatly, we further introduce α and β to unify their dimension. $risk$ in the combined model is calculated by using the following formula:

$$risk = \lambda(\alpha \times def') + (1-\lambda)(\beta \times det') \tag{1}$$

where $\alpha = 1 / \text{average deficiency number in the training set}$, $\beta = 1 / \text{average detention probability in the training set}$, and $\lambda \in [0,1]$. A larger λ indicates that more weighting points are given to the predicted ship deficiency number, and vice versa. In the combined model, $risk$ of a ship can be interpreted as the sum (joint effect) of the times of its predicted deficiency number over the average deficiency number (in the training set) and the times of its predicted detention probability over the average detention probability (in the training set). The ship risk scores generated by SRP ship selection scheme are calculated by using the method proposed by Wang et al. (2019).

4.2 Introduction of dataset

The dataset used in this study contains 1,974 initial PSC inspection records at the Hong Kong port from Jan 2016 to Dec 2018. The inspection records are downloaded from the web-based Asia Pacific Computerized Information System (APCIS) provided by the Tokyo MoU[1]. Ship related factors are searched from Word Register of Ships (WRS). 14 features highly related to ship deficiency and detention in current literature and based on domain knowledge are considered, namely ship age, ship gross tonnage (GT), ship length, ship depth, ship beam, ship type, ship flag performance, ship company performance, ship recognized organization (RO) performance, the number of times of changing flag, casualties in last five years, the number of previous detentions, last PSC initial inspection time, and the number of deficiencies in last PSC initial inspection in Tokyo MoU. Variable explanation and description (including the prediction targets and the features) are as follows.

(a) Ship deficiency number (prediction target in TAN model)

The number of deficiencies identified in the current PSC initial inspection. The value for this variable is between 0 and 51, with 4.39 as the mean value.

(b) Ship detention (prediction target in BRF model)

Whether a ship is detained (set the state to be “1”) or not detained (set the state to be “0”) in the current inspection. The detention rate over the whole dataset is 3.55%

(c) Age

The time interval (in years) between the keel laid date and the current PSC inspection date. The minimum value of age in the whole dataset is 0, the maximum is 47, and the mean value is 10.88.

(d) Gross tonnage (GT)

[1] http://www.tokyo-mou.org/inspections_detentions/psc_database.php

Ship GT is a nonlinear measure of a ship's internal volume, with 100 cubic feet as the unit. This variable ranges from 299 to 266,681, with mean value as 44,237.

(e) Length

Ship length is the overall maximum length of a ship (in meters). The value of the ships in the dataset is between 32.3 and 400.0, and the average value is 213.26.

(f) Depth

Ship depth is the vertical distance (in meters) measured from the top of the keel to the upper deck at side measured inside the plating. The minimum value of this variable in the dataset is 3.7 and the maximum value is 36.0, with the mean value as 17.65.

(g) Beam

Ship beam is the width of the hull (in meters). The minimum value of ship beam is 7.4, the maximum value is 60.0, and the mean value is 31.71 in the dataset.

(h) Type

Ships in the dataset are classified into the following types: bulk carrier, container ship, general cargo/multipurpose, passenger ship, tanker, and other.

(i) Flag performance

Ship flag performance is an indicator which evaluates the performance of ship flag state. It is calculated based on the flag Black-Grey-White list provided by Tokyo MoU (Tokyo MoU, 2018). The flag performance gets worse from "white", "grey" to "black". If the flag is not listed, the value for this variable is set to be "not listed".

(j) Company performance

Ship company performance is an indicator which evaluates the performance of ship International Safety Management (ISM) company. It is calculated based on company performance list provided by Tokyo MoU (Tokyo MoU, 2018). Company performance gets worse from "high", "medium", "low" to "very low". If the company is not listed, the value for this variable is set to be "not listed".

(k) Recognized organization (RO) performance

Ship RO performance is an indicator which evaluates the performance of ship recognized organization performance. It is calculated based on RO performance list provided by Tokyo MoU (Tokyo MoU, 2018). RO performance gets worse from "high", "medium", "low" to "very low". If the RO is not listed, the value for this variable is set to be "not listed".

(l) The number of times of changing flag

The total times of ship flag changing from keel laid date to the current PSC inspection date. The maximum value of this variable is 8 and the minimum value is 0. The average number of times of ship flag changing is 0.67 in the dataset.

(m) Casualties in last five years

A binary variable indicating whether the ship is involved in casualties (set to be "1") or not (set to be "0") in the last five years. The average value of this variable is 0.09 in the dataset.

(n) The number of previous detentions

The total times of detentions of a ship in all previous PSC inspections since the keel laid date. The maximum value of this variable is 18 and the minimum value is 0, with the mean value as 0.60 in the dataset.

(o) Last PSC initial inspection time in Tokyo MoU

The time interval between the last and current PSC initial inspections within Tokyo MoU (in months). For ships that are inspected for the first time (i.e. with no previous inspection records), the state of this variable is set to be "−1". For ships with previous inspections, the maximum value of this variable is 180.7 and the minimum value is 0, while the mean value is 10.30 in the dataset.

(p) The number of deficiencies in last PSC initial inspection in Tokyo MoU

The number of deficiencies identified in last PSC initial inspection within Tokyo

MoU. For ships that are inspected for the first time, the state of this variable is set to be "−1". For ships with previous inspections, the maximum value of this variable is 38 and the minimum value is 0, while the mean value is 2.51.

To calibrate and validate the combined model, we randomly divide the whole dataset into training set, which contains 1,580 samples (80%), and test set, which contains 394 samples (20%). The discretization and encoding methods used in the TAN model are the same as those used in Wang et al. (2019), and the encoding method used in the BRF model is the same as that used in Yan et al. (2020). The notation of the variables, states after discretization and encoding in the TAN model, and encoding method in the BRF model are shown in Table 3.

Insert Table 3 here

4.3 Comparison between the combined model and SRP

The visiting ships to the port of Hong Kong can be classified into four categories under the framework of SRP: ships that have not been inspected before (category I), ships that are out of the time window (category II), ships that are within the time window (category III), and ships that do not enter the time window (category IV). In practice, ships in category I have the highest inspection priority as they do not have previous inspection records in the database and thus have outstanding factors. Meanwhile, ships in category II are required to be inspected, and ships in category III may be inspected. Ships in category IV should not be inspected unless they have outstanding or overriding factors. If they should be inspected, they will be given the highest inspection priority. However, we cannot obtain the information of these factors from the public database provided by the Tokyo MoU, and thus we delete the ships in category IV in the test set. We also delete the inspection records in category I as they have the highest inspection priority. Only the ships that are not attached with the highest inspection priority are kept as we need to decide their inspection sequence (i.e. ships in category II and category III). After the processing, there are 317 records left in the test set.

We compare the proposed combined model with SRP in two schemes, where the ship inspection priority in the framework of SRP is considered in one scheme, while the framework of SRP is not considered in the other scheme to fully illustrate the superiority of the proposed combine model. More specifically, in scheme i, we compare the two models considering the framework of SRP, where the ships in category II are first inspected from higher to lower risk score in the combined model and SRP respectively, followed by inspecting the ships in category III from higher to lower risk score in the combined model and SRP respectively. In scheme ii, the current inspection framework is not considered and we just inspect the ships from higher risk score to lower risk score in both SRP and the combined model. It should be mentioned that the ship inspection sequence generated by the SRP is the same in both schemes, therefore we only use 'SRP' to indicate the ship inspection sequence in SRP under both schemes.

In the combined model, as the value for λ represents a trade-off between the predicted ship deficiency number and the probability of ship detention, when we change the value for λ from 1 to 0, we gradually give less weighting points to ship deficiency and more weighting points to ship detention. For intuitive representation, we first draw the curves of the accumulated detected ship deficiency number and ship detentions in the two schemes for $\lambda = 1, 0.5, 0$ as shown in Figure 1. We further calculate the improvement of the combined model by using Eq. (2) to Eq. (4) given $\lambda = 1, 0.9, \ldots 0.1, 0$:

$$\text{improvement} = \frac{\text{area(combined model)} - \text{area(SRP)}}{\text{area(SRP)}} \tag{2}$$

$$\text{area(combined model)} = \sum_{n=1}^{N} (t_n^{combine} + t_{n-1}^{combine}) / 2 \tag{3}$$

$$\text{area(SRP)} = \sum_{n=1}^{N} (t_n^{\text{SRP}} + t_{n-1}^{\text{SRP}}) / 2, \tag{4}$$

where $\text{area(combined model)}$ and area(SRP) are the areas under the curves of the combined model and SRP respectively, N is the total number of ships selected for inspection, and $t_n^{combine}$ (or t_n^{SRP}) is the total number of identified deficiencies or detention after inspecting n ships by using the combined model (or SRP). The comparison results are shown in Table 4.

Insert Figure 1 here

Insert Table 4 here

Figure 1 shows that the performance of the combined model in scheme ii is the best, followed by the performance of the combined model in scheme i and then by the SRP ship selection scheme. Table 4 gives more detailed comparison between the combined model in scheme i and scheme ii with the SRP. Specifically, in scheme i, when $\lambda = 0.9$, i.e., we give 90% of the weighting points to the predicted ship deficiency number and 10% of the weighting points to the predicted ship detention probability, the combined model achieves the most significant improvement compared to the SRP by 14.85% considering the total number of ship deficiencies detected, whereas when $\lambda = 0$, i.e. we only consider ship detention, the combined model has the greatest improvement compared to the SRP regarding ship detention by 30.64%. The average improvement considering both ship deficiency and detention is maximized to be 22.38% when $\lambda = 0.8$. Similar pattern can be found in scheme ii. When setting $\lambda = 0.8$, most ship deficiency number can be identified and the improvement of the combined model compared to SRP is 25.87%. Meanwhile, most ship detentions can be found when $\lambda = 0$ and the improvement is 58.90%. When we choose $\lambda = 0.6$, the average improvement of the combined model considering both ship deficiency and detention can be maximized to be 41.61%, which is almost twice the average improvement of the combined model in scheme i compared to the SRP.

5. Suggestions for MoUs on ship selection in PSC

Based on the review of currently implemented ship selection methods in MoUs on PSC, the proposed ship selection models in current literature, and the comparison results between the combined model and the SRP, several suggestions are proposed for the MoUs' reference.

a) Apart from generating risk profiles of the foreign visiting ships based on the database of the inspection records conducted within the same MoU, the inspection records in other MoUs can also provide valuable information on ship risk level as ships can be inspected by any of their ports of call. Besides, ship inspection by other authorities or organizations, such as industry inspection and inspections conducted by ship flag states and recognized organizations, can also be considered. This can be achieved by establishing a global harmonized ship inspection database.

b) The widely adopted ship selection models in most MoUs (i.e. the SRP ship selection scheme) is heavily based on expert knowledge and is a simple weighted-sum model. Besides, although ships are divided into three risk profiles, there is no specific information about ship risk levels in the same profile even if those ships can have quite various conditions. To address these issues, historical inspection records can be leveraged to calculate more accurate ship risk scores and thus improve the selection efficiency. For example, inspection results and time of previous inspections, e.g. number of deficiencies in last inspection, previous detentions, and last inspection time can be incorporated. Apart from generating ship risk profile based on the database of the corresponding MoU, various databases consisting of more comprehensive ship factors can also be considered. For example, ship generic information provided by the Word Register of Ships[2] or the Lloyd's Register of Ships[3] can be considered, and databases of ship involved incidents and accidents[4], which are the intuitive indicator of ship risk level and need to be prevented from in the future can be used for ship risk profile generation. Instead of using the simple weighted-sum model, more advanced models for ship

[2] https://world-ships.com/
[3] https://www.lr.org/en/marine-shipping/
[4] https://gisis.imo.org/Public/Default.aspx

risk calculation, e.g. the statistical models and machine learning models should be developed and implemented.

c) Ship deficiency and detention, which are the main outcomes of a PSC inspection, should be given more attention when evaluating ship risk level for ship selection. As ship deficiency and detention conditions cannot be observed directly once the ships come to the port states, accurate prediction models for ship deficiency number and/or category as well as ship detention should be developed and their prediction results are worthy of being combined.

6. Conclusion

PSC inspection is the safeguard of maritime safety and the marine environment. As the pre-requirement of enforcement of PSC inspection, substandard ship selection plays a key role in achieving efficient and effective PSC inspection and MoU management. Through a thorough review of the currently implemented ship selection methods at the MoUs and ship selection models in existing literature, the pros and cons of these ship selection methods are analyzed. One main research gap in current literature is that ship deficiency and detention are not considered simultaneously to calculate ship risk score, even if both of them are the main outcome of PSC inspection and are irreplaceable. To bridge this gap, a combined model consisting of the TAN classifier for ship deficiency number prediction and the BRF model for ship detention prediction is proposed and validated by using the real inspection records at the Hong Kong port from 2016 to 2018. Both of the prediction models take a total of 14 parameters into account, including ship generic factors, dynamic factors, and inspection historical factors. To validate the performance of the combined model, reasonable and comprehensive comparisons between the combined model and the currently implemented SRP ship selection scheme are conducted considering and ignoring the ship inspection priority in current ship selection framework. Results show that the combined model performs better than the SRP in both cases regarding the total number of ship deficiencies and detentions detected. Especially, when 80% and 20% of the weighting points are given to the predicted ship deficiency number and ship detention probability respectively, the combined model reaches the highest improvement compared with the SRP within the current inspection framework. When we compare their performance out of the current inspection framework, if 60% and 40% of the weighting points are given to the predicted ship deficiency number and ship detention respectively, the highest improvement of the combined model can be achieved, which is 41.61% compared to SRP.

Based on the review of the currently proposed ship selection methods in practice and in literature as well as the numerical experiment results, three suggestions are proposed for the MoUs' reference. First, the PSC inspection records from other MoUs as well as from other authorities or organizations can also be considered for ship risk profile calculation in ship selection for PSC inspection. Second, historical inspection records and wider range of ship related factors (e.g. ship generic factors and ship involved casualty and incident factors) can be given more attention in ship risk score calculation. Besides, more advanced ship risk calculation models such as statistical and machine learning models should be adopted. Third, ship risk calculation should also account more for both ship deficiency and detention as they are the main outcomes of PSC inspection and are also the direct indicator of ship risk level.

This study is the very first study that gives a comprehensive review on ship selection methods applied in the MoUs on PSC and developed in current literature. It is also the first study that proposes a unified model considering both ship deficiency and detention for ship risk calculation. Several managerial insights and suggestions are proposed based on the review and the combined model, which will be valuable in improving ship selection efficiency in MoUs and thus preventing substandard shipping and safeguarding maritime transportation.

References

Abuja MoU, 2012. Memorandum of understanding on port state control for West and Central African Region, 1999. Accessed 16 Aug 2020. http://www.abujamou.org/post/90.pdf.

Black Sea MoU, 2016. Information sheet of the BS MoU new inspection regime. Accessed 16 Aug 2020. http://www.bsmou.org/downloads/info-sheets/InfoSheetBSMoUNewInspectionRegime.pdf.

Breiman, L., 2001. Random forests. Machine Learning 45(1), 5–32.

Chi, Z., Jun, S., 2010. Automatically optimized and self-evolutional ship targeting system for port state control. In Proceedings of 2010 IEEE International Conference on Systems, Man and Cybernetics, 791–795.

Degré, T., 2007. The use of risk concept to characterize and select high risk vessels for ship inspections. WMU Journal of Maritime Affairs 6(1), 37–49.

Gao, Z., Lu, G., Liu, M., Cui, M., 2008. A novel risk assessment system for port state control inspection. In Proceedings of 2008 IEEE International Conference on Intelligence and Security Informatics, 242–244.

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 45(5), 604–621.

Heij, C., Knapp, S., 2019. Shipping inspections, detentions, and incidents: an empirical analysis of risk dimensions. Maritime Policy & Management 46(7), 866–883.

IMO, 2017. Resolution A.1119(30): Procedure for port state control, 2017. Accessed 17 May 2019, http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Assembly/Documents/A.1119%2830%29.pdf.

IMO, 2019. Port state control. Accessed 15 Dec 2019. http://www.imo.org/en/OurWork/MSAS/Pages/PortStateControl.aspx.

IMO, 2020. Reduction of GHG emissions from ships-fourth IMO GHG study 2020 – final report. Accessed 11 Sep 2020, http://www.imo.org/en/MediaCentre/HotTopics/Pages/Reducing-greenhouse-gas-emissions-from-ships.aspx.

Intercargo, 2000. Port state control: a guide for ships involved in the dry bulk trades. Accessed 20 Nov 2018. http://www.mardep.gov.hk/en/others/pdf/psguide.pdf.

Jiang, M., Lu, J., 2020. Maritime accident risk estimation for sea lanes based on a dynamic Bayesian network. Maritime Policy & Management 47(5), 649–664.

Knapp, S., 2007. The econometrics of maritime safety: recommendations to enhance safety at sea. Ph.D thesis at Erasmus University Rotterdam.

Knapp, S., Heij, C., 2020. Improved strategies for the maritime industry to target vessels for inspection and to select inspection priority areas. Safety 6(2), 1–21.

Li, K. X., 1999. The safety and quality of open registers and a new approach for classifying risky ships. Transportation Research Part E 35(2), 135–143.

Li, K. X., Zheng, H., 2008. Enforcement of law by the port state control (PSC). Maritime Policy & Management 35(1), 61–71.

Lun, Y. V., Lai, K. H., Wong, C. W., Cheng, T. C. E., 2014. Green shipping practices and firm performance. Maritime Policy & Management 41(2), 134–148.

Mediterranean MoU, 2020. Selection of ships for inspection. Accessed 16 Aug 2020. http://197.230.62.214/Basic_Principlse.aspx#3.

Paris MoU, 2014. Annex 7 of Paris Memorandum of Understanding on Port State Control. Accessed 20 July 2019. https://www.parismou.org/system/files/Annex%207%20ship%20risk%20profile.pdf.

Sampson, H., Bloor, M., Baker, S., Dahlgren, K., 2016. Greener shipping? A consideration of the issues associated with the introduction of emission control areas. Maritime Policy & Management 43(3), 295–308.

Shi, W., Xiao, Y., Chen, Z., McLaughlin, H., Li, K. X., 2018. Evolution of green shipping research: themes and methods. Maritime Policy & Management 45(7), 863–876.

Tokyo MoU, 2014. Information sheet of the new inspection regime (NIR). Accessed 15 Mar 2019. http://www.tokyo-mou.org/doc/NIR-information%20sheet-r.pdf.

Tokyo MoU, 2018. Memorandum of understanding on port state control in the Asia-Pacific Region. Accessed 19 October 2019, http://www.tokyo-mou.org/.

Tokyo MoU, 2020. Annual report on port state control in the Asia-Pacific Region 2019. Accessed 1 Aug 2020, http://www.tokyo-mou.org/doc/ANN19-f.pdf.

UNCTAD, 2019. Review of maritime transport 2019. Accessed 30 Nov 2019. https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2563.

Wang, S., Yan, R., Qu, X., 2019. Development of a non-parametric classifier: Effective identification, algorithm, and applications in port state control for maritime transportation. Transportation Research Part B 128, 129–157.

Xu, R., Lu, Q., Li, W., Li, K. X., Zheng, H., 2007a. A risk assessment system for improving port state control inspection. In Proceedings of 2007 International Conference on Machine Learning and Cybernetics, 818–823.

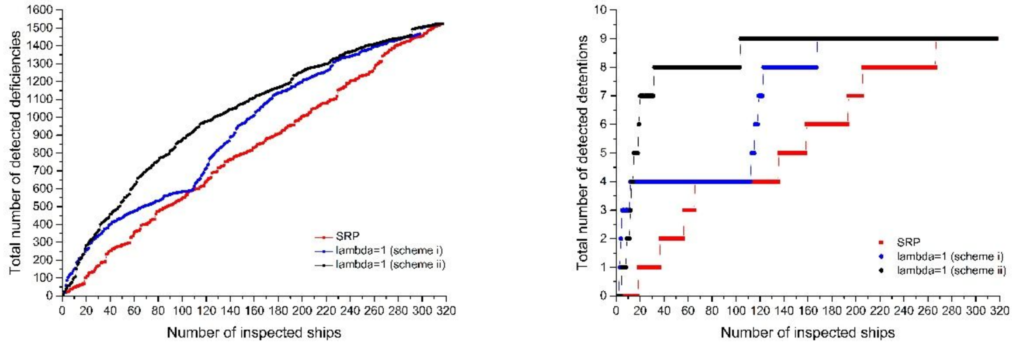
Xu, R., Lu, Q., Li, K. X., Li, W., 2007b. Web mining for improving risk assessment in port state control inspection. In Proceedings of 2007 International Conference on Natural Language Processing and Knowledge Engineering, 427–434.

Yan, R., Wang, S., Peng, C., 2020. Prediction of ship detention in port state control considering data imbalance. Journal of Computational Science, under review.

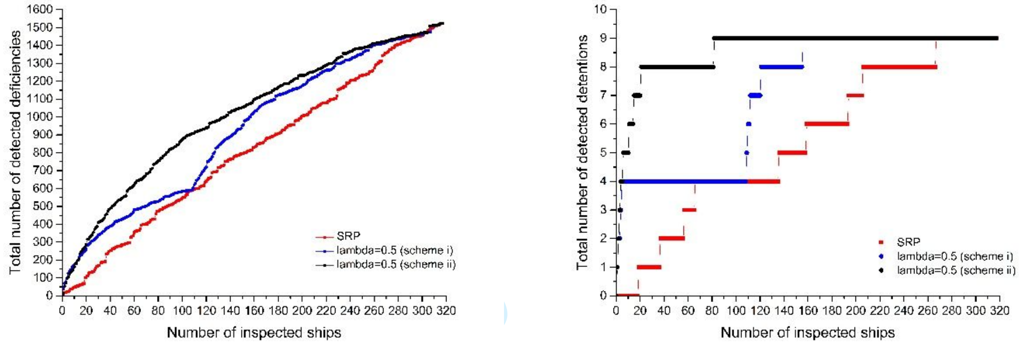
Yan, R., Wang, S., 2019. Ship inspection by port state control—review of current research. Smart Transportation Systems 2019, 233–241.

Yang, Z., Yang, Z., Yin, J., 2018a. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transportation Research Part A 110, 38–56.

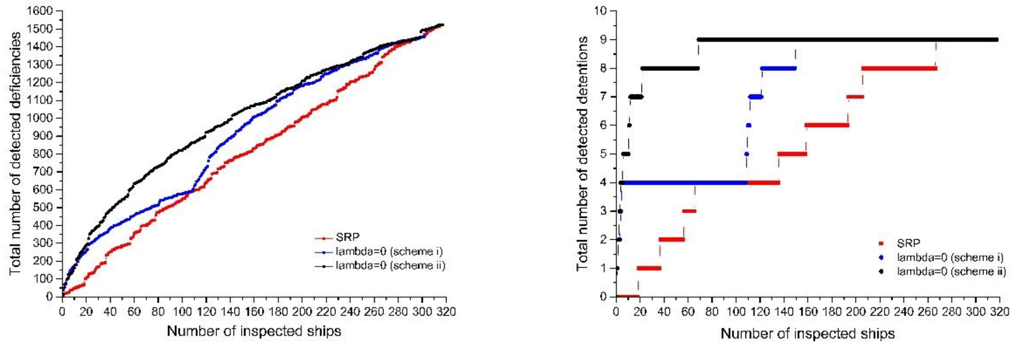
Yang, Z., Yang, Z., Yin, J., Qu, Z., 2018b. A risk-based game model for rational inspections in port state control. Transportation Research Part E 118, 477–495.



(a) Total number of detected deficiencies and detentions when $\lambda = 1$



(b) Total number of detected deficiencies and detentions when $\lambda = 0.5$



(c) Total number of detected deficiencies and detentions when $\lambda = 0$

Figure 1. Comparison between the combined model and SRP regarding the total number of detected deficiencies and detentions

Table 1. Ship risk profiles in the ship selection scheme of Tokyo MoU (2014)

Factors	**Values**	**Weighting points**	**High-risk ships**: Total weighting points is at least 4.	**Standard-risk ships**: Total weighting points is at most 3 and the ship does not meet all the criteria for low-risk ships.	**Low-risk ships**: Total weighting points is at most 3 and the ship meets all the criteria below.
Ship type	Chemical tanker, gas carrier, oil tanker, bulk carrier, passenger ship, container ship	2			All types
Ship age	All types with age > 12 years	1			All ages
Flag performance in Black-Grey-White list	Black	1			White, and should be audited by the IMO
Recognized organization performance	Low/very low	1			High, and should be recognized by Tokyo MoU
Company performance	Low/very low/no inspection within the previous 36 months	2			High
Deficiencies within the previous 36 months	No. of inspections which recorded over 5 deficiencies	Number of inspections which recorded over 5 deficiencies			All inspections have 5 or less deficiencies and has at least one inspection within the previous 36 months
Detentions within the previous 36 months	3 or more detentions	1			No detention

Table 2. Time windows required by Tokyo MoU

Risk profile	Definition	Time window
HRS	Ships which meet criteria to a total of 4 or more weighting points.	2 to 4 months
SRS	Ships that are neither LRS nor HRS.	5 to 8 months
LRS	Ships which meet all the criteria of LRS and have had at least one inspection in the previous 36 months.	9 to 18 months

Table 3. Variables in the combined model

Variable name	Notation	States in the TAN model	Encoding method in the BRF model
Ship deficiency number	deficiency	S1:0to1, S2:2to4, S3:5+	Not included
Ship detention	detention	Not included	one-hot encoding: 1 if the ship detained and 0, otherwise
Age	age	S1: 0to7, S2: 8to12, S3: 13+	No processing
Gross tonnage (GT)	GT	S1:0to17209, S2:17210to44146, S3:44147+	No processing
Length	length	S1:0to172.0, S2:172.1to244.0, S3:244.1+	No processing
Depth	depth	S1:0to14.0, S2:14.1to20.1, S3:20.2+	No processing
Beam	beam	S1:0to27.2, S2:27.3to32.3, S3:32.4+	No processing
Type	type	S1:bulk carrier, S2:container_ship, S3:general_cargo/multipurpose, S4:passenger_ship, S5:tanker, S6:other	one-hot encoding: is_bulk_carrier: 1 for bulk carrier and 0, otherwise; is_container_ship: 1 for container ship and 0, otherwise; is_ general cargo/multipurpose: 1 for general cargo/multipurpose and 0, otherwise; is_ passenger_ship: 1 for passenger ship and 0, otherwise; is_tanker: 1 for tanker and 0, otherwise; is_other: 1 for other ship types and 0, otherwise.
Flag performance*	flag	S1:white, S2:grey, S3:black, S4:not_listed	label encoding: white->1**; grey->2; black->3; not listed->4.
Company performance*	company	S1:high, S2:medium, S3:low, S4:very_low, S5:not_listed	label encoding: high->1; medium->2; low->3; very low->4; not listed->5.
Recognized organization (RO) performance*	RO	S1:high, S2:medium, S3:low, S4:not_listed	label encoding: high->1; medium->2; low->3; not listed->4.
The number of times of changing flag	flag-changing-times	S1:zero, S2:one, S3:2+	No processing
Casualties in last five years	casualty-in-5-years	S1:zero, S2:one	one-hot encoding: casualty-in-5-years: 1 for any casualty occurs in the last 5 years and 0, otherwise.
The number of previous detentions	total-detentions	S1:zero, S2:one, S3:2+	No processing
Last PSC initial inspection time in Tokyo MoU	last-inspection-time	S1:0to5.1, S2:5.2to9.3, S3:9.4+, S4:none	No processing
The number of deficiencies in last initial inspection in Tokyo MoU	last-deficiency-no	S1:zero, S2:1to2, S3:3+, S4:none	No processing

*Note that these three features are purely ordinal data if the state of "not listed" is excluded. However, to preserve the order of states from best to worse as from "white", "grey" to "black" for ship flag performance, from "high", "medium", "low" to "very low" for ship RO performance, and from "high", "medium", "low" to "very low" for ship company performance, we treat them as ordinal data.

**Note that this indicates that the state of "white" is encoded to be "1".

Table 4. Improvement of the combined model over SRP

scheme i (considering inspection priority, within the framework of SRP)

λ_l / metric	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Deficiency	14.80%	**14.85%**	14.82%	14.65%	14.55%	14.34%	14.16%	13.93%	13.74%	13.42%	13.09%
Detention	27.85%	29.54%	29.94%	30.06%	30.18%	30.35%	30.53%	30.47%	30.53%	30.59%	**30.64%**
Average	21.33%	22.19%	**22.38%**	22.35%	22.36%	22.35%	22.35%	22.20%	22.13%	22.00%	21.87%

scheme ii (ignoring inspection priority, out of the framework of SRP)

λ_l / metric	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Deficiency	25.61%	**25.87%**	25.83%	25.61%	25.43%	25.06%	24.63%	24.25%	23.83%	23.34%	22.75%
Detention	53.13%	56.16%	57.03%	57.50%	57.79%	58.03%	58.37%	58.37%	58.55%	58.78%	**58.90%**
Average	39.37%	41.01%	41.43%	41.56%	**41.61%**	41.54%	41.50%	41.31%	41.19%	41.06%	40.83%