https://dx.doi.org/10.1016/j.trb.2024.102887

A data-driven optimization approach to improving maritime transport efficiency

Ran Yan[a], Yan Liu[b], Shuaian Wang[c,*]

[a] *School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, email: ran.yan@ntu.edu.sg*

[b] *Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China, email: yan.y.liu@polyu.edu.hk*

[c] *Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China, email: wangshuaian@gmail.com*

Abstract

Ship inspections conducted by port state control (PSC) can effectively reduce maritime risks and protect the marine environment. The effectiveness of PSC depends on accurately selecting ships with higher risk for inspection. Ship risk profile (SRP) is currently the most common method of quantifying ship risk, but the thresholds of the factors that determine a ship's risk and classification in the SRP framework are subjective and can make the ship selection process less efficient. In this study we propose a data-driven bi-objective nonlinear programming model, referred to as the SRP+ model, to optimize the thresholds in the original SRP framework. To solve the model, we first linearize the nonlinear optimization model using the big-M method, and then develop an augmented epsilon-constraint method to transform the bi-objective model to a single-objective model and obtain all Pareto optimal solutions. We also conduct a case study using real PSC inspection records at the Hong Kong port to construct and validate the SRP+ model. The results suggest that the threshold of the total weighting points to classify a ship as high-risk ship should be slightly increased, the thresholds of ship age should be significantly increased, the threshold of historical deficiency number should be increased, while the threshold of historical ship detention times should be decreased. The proposed SRP+ model can benefit both conservative and open-minded port authority decision makers by identifying ships with more deficiencies and/or higher detention probability more efficiently. The model can also be applied to other risk management problems in transportation and supply chain management, in addition to the maritime transport domain.

Keywords: Maritime transport, data-driven optimization, bi-objective optimization model, predictive analytics, threshold optimization

1. Introduction

Ship inspection by port state control[1] (PSC) is an important method of reducing risk and improving the sustainability of maritime transport. However, due to limited resources and the high costs of inspection, only a small proportion of foreign visiting ships can be inspected. Thus, effectively selecting the highest-risk visiting ships flying foreign flags is a central operational problem for PSC. One common approach is referred to as ship risk profile (SRP), which is a risk-based scheme where the ships labeled as higher risk are subject to more frequent and in-depth inspections, while those with lower risks bear a smaller inspection burden (Tokyo MoU 2014). The SRP framework represents a significant change as it transforms and modernizes the port state control system (Paris MoU 2012, Yang et al. 2018, Zhen et al. 2022).

The SRP information sheet adopted by the Tokyo MoU is shown in Table 1. A ship's generic features (ship type and age), operating and maintenance status (flag performance[2], recognized organization (RO) performance[3], and company performance[4]), and historical inspection performance (deficiency and detention conditions in the last 36 months) are considered. Weighting points are allocated to the risk parameters. In terms of type, ships categorized as chemical tankers, gas carriers, oil tankers, bulk carriers, passenger ships, and container ships are assigned two points, as they suffer from higher navigation risks and the accident losses can be significant. Ship flag performance deteriorates from white to grey and then to black, and a ship with a flag on the black list is assigned one point. Similarly, if the RO performance (company performance) is evaluated as low/very low (low/very low/no inspection) within the previous 36 months, the ship is assigned 1 (2) point(s). According to the total weighting points and specific criteria, a ship is classified as a high risk ship (HRS), standard risk ship (SRS), or low risk ship (LRS); see Table 2.

The aforementioned criteria used in the SRP information sheet are objective and justifiable, but other parameters in the SRP framework are overreliant on expert knowledge and are therefore subjective. For example, the threshold of ship age is 12 years, and thus a ship older than 12 is assigned 1 weighting point and 0 otherwise. However, it is unclear why this threshold instead of 11, 13, or some other value is set (such as 21, which is the average age of the world merchant fleet as of 2020 (UNCTAD 2022)). Whether a

[1]Port state control refers to the inspection of foreign ships in national ports to verify that their conditions and operations comply with international and regional regulations.

[2]A flag is the nationality of a ship, and flag performance is determined by the inspection and detention conditions of the flag's fleet in the previous three calendar years.

[3]An RO is a ship management organization that is recognized by at least one member authority of the Tokyo MoU, and RO performance is determined by the inspection and detention history of all ships belonging to that RO in the previous three calendar years.

[4]Company here refers to a ship's international safety management (ISM) company, and company performance is determined by the detention and deficiency history of all ships belonging to that company on the basis of a running 36 months.

Table 1: SRP information sheet adopted by the Tokyo MoU (Tokyo MoU 2014)

Parameters	Values[5]	Weighting points	Criteria for LRS
Ship type	Chemical tanker, gas carrier, oil tanker, bulk carrier, passenger ship, container ship	2	
Ship age	$>$ 12 years	1	
Flag performance in Black-Grey-White list of Tokyo MoU	Black	1	White, and should be audited by the International Maritime Organization
RO performance evaluated by Tokyo MoU	Low/very low	1	High, and should be an RO recognized by the Tokyo MoU
Company performance evaluated by Tokyo MoU	Low/very low/no inspection within previous 36 months	2	High
Deficiencies within previous 36 months	Inspections that recorded over 5 deficiencies	The number of inspections that recorded over 5 deficiencies	All inspections have recorded 5 or fewer deficiencies and the ship has had at least one inspection within previous 36 months
Detentions within previous 36 months	More than 2 detentions	1	No detention

Table 2: Ship risk category

Ship risk profile	Criteria	Inspection time window
High-risk ship (HRS)	When the sum of weighting points is above 3	2 to 4 months
Standard-risk ship (SRS)	Neither HRS nor LRS	5 to 8 months
Low-risk ship (LRS)	All the criteria for LRS are met	9 to 18 months

significant decline in conditions will always be observed when the ship reaches 12 years old is questionable, at least from a statistical perspective. Similarly, in terms of deficiency and detention, only the inspections that recorded over 5 deficiencies in the previous 36 months are considered in the total weighting points; if there have been more than 2 detentions in the previous 36 months, 1 weighting point is assigned to the

[5]It should be noted that only the parameter values that are attached with positive weighting points are listed in this table. The other parameter values with zero weighting point are not listed.

ship. It is unclear why 5 (instead of other values such as 4, 6, or 10) and 2 (instead of other values such as 1 or 3) are deemed to be the thresholds for the total deficiency number and detention times, respectively. Again, whether recording more than 5 deficiencies in an inspection or having more than 2 detentions in the last three years indicates a higher risk remains a question. Finally, a ship is currently classified as HRS if it is allocated more than 3 total weighting points, but again this value appears arbitrary. The thresholds used in the SRP framework seem to lack justification.

We further illustrate the irrationality and rationality of the thresholds of the current SRP model based on statistics on the actual inspection records at the Hong Kong port from 2014 to 2021. The statistics suggest that the current threshold of HRS at 3 is on the low side, and should be increased to 5 or 6 to better distinguish ships with a higher risk. The threshold of ship age at 12 is much lower than the age values that can distinguish ship risk levels, and should be better if it is increased from 16 to 18. Similarly, the threshold of the number of deficiencies in the last 36 months should be increased from the current 5 to a larger number such as 7. The threshold of the detention times in the last 36 months at 2 seems to be reasonable; however, there are other candidates such as 1 and 5, and which one is the most suitable should be more closely explored. Meanwhile, we also find that the thresholds set for ship flag performance and company performance are reasonable. Due to the limited inspection records, the rationality of the threshold of ship RO performance cannot be validated. The detailed analysis is presented in Appendix A.

It can thus be seen that the current approach is overreliant on domain knowledge and may lead to inconsistencies between the evaluation generated by the SRP model and the ship's real situation, thereby reducing the SRP model's effectiveness. In this paper, we optimize the thresholds in the SRP model using real port state control inspection records from the Hong Kong port. The justification of only optimizing the thresholds are as follows. As shown in Table 1, there are three major components of the SRP framework: parameters, values (thresholds), and weighting points attached. The parameters considered form the skeleton of the SRP, as considering more or fewer parameters would change the SRP's basic framework and the data collection and processing steps, which conflicts the goal of retaining the SRP as one objective of this study. The values (thresholds) and weighting points attached to the parameters, and the total weighting points leading to an HRS are highly correlated to each other: a ship's parameter would first be compared with the threshold attached, beyond which the weighting point(s) will be assigned, and whether a ship is classified as an HRS is determined by the total weighing points for all parameters. Therefore, optimizing either values (thresholds), or weighting points, is equivalent to optimizing both simultaneously. In this study, to simplify the model structure, we only optimize the values (thresholds). Finally, to determine the thresholds of which parameters to optimize, we analyze the underlying of each of the parameters. The values considered for ship type are objective and justifiable, as the ship types of concern, namely chemical tankers, gas carriers, oil tankers, bulk carriers, passenger ships, and container ships, are indeed associated with higher navigation risks and accident losses. The values of the management and operation parties of a ship, i.e., flag, RO, and company, are based on these parties' calculated performances by the Tokyo MoU, which are determined by the actual inspection performances of the fleet under them. Therefore, they can be regarded as justifiable. For the other parameters, including ship age and deficiency and detention conditions in the past 36 months, we optimize their thresholds, as these thresholds are subjective and lack of solid underlying. Moreover, we optimize the total weighting points to classify one ship to HRS as the overall and final result, as this threshold can be viewed as a coordinator that considers all parameters together and can address the issues brought about by not optimizing the threshold of some parameters. We refer to this improved version of the SRP model as the SRP+ model. We address the following research questions:

1. What is the optimal threshold of each parameter in the SRP+ model? What is the optimal threshold for classifying a ship as HRS?

2. How does the SRP+ model perform compared with the original SRP model?

3. How does the decision makers' attitudes toward the SRP+ model's deviation from the original framework influence its performance?

To answer these questions, we construct a bi-objective mixed-integer nonlinear model to optimize the threshold of ship age, the thresholds of deficiencies and detention times within the last 36 months, and the threshold of the total weighting points to classify a ship as HRS, while keeping other settings fixed. The bi-objective functions aim to reach a balance between two contradictory goals. The first is to minimize the variation in the proportions of ships classified as HRS before and after the threshold optimization, because an HRS is likely to involve more deficiencies and a higher probability of detention and thus deserves the most frequent inspection. The second is to maximize inspection efficiency by maximizing the average number of deficiencies of the ships classified as HRS. Deficiency/deficiencies detected onboard a ship is the major outcome of a PSC inspection. A larger number of deficiencies shows more non-compliances onboard a ship, and thus indicates a worse ship condition. The two objectives of the optimization model are contradicting to each other, and we use the following extreme cases for explanation. If we only consider the objective of maximizing the average number of deficiencies on ships classified as HRS, then only one ship will be classified as HRS, which is the ship that has the largest number of deficiencies. In this case, the second objective is maximized, while the first objective is also maximized, contradicting to "minimizing" this objective. If we only keep the first objective to retain the original HRS proportion while ignoring the second objective, there is no need to construct the bi-objective model as the current practice has perfectly followed the first objective. Or in other words, randomly classifying a set of ships as HRS following the HRS ratio of the training set is an optimal solution, but obviously doing so is meaningless. Therefore, we need to consider both objectives to reflect the requirements of the SRP+ model. In other words, our aim is on the one hand to retain the proportion of ships classified as HRS while on the other hand to optimize the decision on which ships should be classified as HRS, so as to increase the efficiency of the identification of ships with higher risk. To solve the model, we first linearize the bi-objective mixed-integer nonlinear optimization model by using a big-M method. We then use an augmented ϵ-constraint method to transform one of the objectives (to minimize the variation) to a constraint by imposing an upper bound ϵ on the variation. Here, ϵ can be interpreted as the port decision makers' tolerance of the variation in the proportions of the ships classified as HRS before and after threshold optimization. If the port decision makers are conservative, meaning that they have little tolerance to the deviation of HRS proportion before and after the optimization of thresholds in the SRP framework, a smaller ϵ is used. In contrast, if the port decision makers are open-minded and seek high efficiency, a larger ϵ is used. After the transformation, we face a single-objective integer linear optimization problem where the deviation in the HRS proportion is turned to a constraint. The transformed model can be easily solved to obtain all Pareto-optimal solutions.

We then perform a case study using real PSC inspection records collected from the Hong Kong port. We compute the optimal thresholds in the SRP+ model and examine the structural differences between the SRP+ model and the SRP scheme. We find that the threshold of the total weighting points for classifying a ship as HRS should be slightly increased from the current 3 to a higher value such as 4. The threshold of ship age should be raised significantly, for example to 21 years, which is the average age of the world merchant fleet. In addition, the threshold of the historical deficiency number should be increased, while that of the historical detention times should be decreased. These new thresholds substantially improve the efficiency of the SRP+ framework by effectively identifying the ships with more deficiencies and higher detention probability. The SRP+ model has the flexibility to adjust the threshold of each parameter in a smart way, and thus distinguish ships with higher risk according to the input data.

Our main contributions to the literature are as follows. Overall, we propose an innovative ship selection framework called SRP+ based on the currently used SRP. The SRP+ is based on a bi-objective optimization problem, aiming to on the one hand retain the ratio of ships classified to HRS, and on the other hand classify ships with a higher number of deficiencies to HRS. Our numerical experiments confirm the effectiveness of the SRP+ model in identifying HRS, and thus enhance the role of port state control in reducing maritime risks and protecting the marine environment. From the perspective of methodology, we propose a new data-driven method for ship risk management, which retains the components in the original framework while improving its efficiency by optimizing the parameters via practical data. The proposed method is applicable to risk management scenarios beyond port state control, if sufficient real data are available. From the perspective of practice, unlike research that proposes brand new methods for ship risk prediction, we retain the SRP framework but innovatively optimize the thresholds. Maritime authority decision makers are likely to more readily accept our SRP+ model as it is an extension of the current SRP framework. Considering the fact that decision makers in the maritime industry are relatively conservative and stakeholders' attitudes to innovative solutions vary widely (Sanchez-Gonzalez et al. 2019), we also demonstrate the superiority of the SRP+ model under different scenarios. Hence, decision makers with different attitudes toward deviations from the original framework can effectively use the model.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 describes the SRP+ framework. In Section 4, we conduct a case study and assess the effectiveness of the SRP+ framework in improving port state control operations. Section 5 concludes.

2. Literature review

Our work optimizes the ship risk prediction model and can assist port authority decision makers by effectively identifying the ships with more deficiencies and higher probabilities of detention. Our study is therefore relevant to research into both ship risk prediction and (maritime) risk management.

The extensive literature on ship risk prediction is aimed at improving the ship selection efficiency of port state control. The level of risk is mainly evaluated from the three perspectives of general ship risk (e.g., classifying ships into risk categories), ship deficiency (e.g., predicting the number and types of deficiencies), and ship detention (e.g., predicting the likelihood of ship detention in port state control inspections) (Sun et al. 2022b, Agra and Rodrigues 2022). Most studies apply empirical analyses for general ship risk prediction (Degré 2007, 2008, Heij and Knapp 2019, Dinis et al. 2020), while machine learning models are widely used to predict the level of ship deficiency and the detention probability (Wang et al. 2019, Yan et al. 2021, Yang et al. 2018). However, these ship risk prediction models are seldom used by port authorities, even if the empirical evidence indicates that they perform is better than the current SRP framework. This is mainly because most models deviate significantly from the SRP framework in terms of input parameters, output (the ship risk level), and mechanism. The SRP model essentially considers limited factors and combines them in a weighted-sum manner, and ships are roughly classified as HRS, SRS, or LRS. In contrast, the alternative models consider a much broader range of factors, and ship risk levels generated by sophisticated prediction models (e.g., machine learning models, in a black-box manner) can be more specific and tailored. These significant differences prevent port authority decision makers from using alternative models, either because their results are not explainable or the ship risk predictions are substantially different from those of the SRP model.

Some recent studies develop ship risk prediction models that are either in line with the current SRP framework or able to explain the prediction results. Yan et al. (2022) develop two ship risk prediction models based on gradient boosting regression trees. The first model uses features identical to those in the SRP model, while the second model uses the same types of features but with a different encoding method. They both use the Shapley additive explanation method to explain the predicted ship risk level, by quantifying the contribution of each feature. Dinis et al. (2020) develop a Bayesian network model to quantify the ship risk level, in which most inputs are identical to the SRP model but feature pre-processing is more fine-grained. The output is extended to a total of 14 risk categories, rather than the 3 categories (HRS, SRS, LRS) in the SRP framework. This stream of research generally improves the accuracy of ship risk prediction. However, no studies consider risk prediction within the framework of the current SRP strategy or quantify the impact of the deviation from the current strategy on the effectiveness of the new model. We develop a unified framework that balances these two objectives through a data-driven approach. This unified framework can help port authority decision makers improve their ship selection decisions while retaining the original selection framework.

Operational risk management and control is a significant issue in operations management and supply chain management (Kleindorfer and Saad 2005, Starr and Van Wassenhove 2014). Big data has led to a shift in operations risk management decision-making from expert judgment to data-driven prescriptive analytics (Feng and Shanthikumar 2018, Choi et al. 2018, Guha and Kumar 2018, Liu et al. 2023, Adler et al. 2022, Huber et al. 2022). Operational risk management issues have been analyzed in various problem settings. For example, Bansal et al. (2017) quantify the production risks of new products by estimating the mean and standard deviations of probability distributions of production-related uncertainties from noisy quantile judgments provided by experts. Kettunen and Salo (2017) demonstrate that uncertainties in project values can lead to inaccurate risk estimates for the aggregate value of a selected project portfolio. Markou and Corsten (2021) empirically assess the effects of risk management on inventory in the gold mining industry, while taking into account volatile gold prices. Sun et al. (2022a) analyze at-risk capacity management for vaccine candidates before regulatory approval by comparing two vaccine production modes. Similarly, we quantify the risk level of foreign visiting ships using a data-driven model and our work improves the accuracy of risk evaluation under a given policy framework by optimizing its thresholds. No other studies apply such a framework.

Maritime transport risk management is also a popular operations management research topic. The numerous studies can be roughly classified into ship risk assessment and prediction, ship risk detection and alarm, and maritime accident reviews. We focus on the category of ship risk assessment and prediction, which is typically carried out before maritime incidents or accidents happen which our study falls into. For example, Baksh et al. (2015) assess the likelihood of occurrences of near misses, mishaps, incidents, accidents, and catastrophes on a ship, based on seven safety barriers. Ulusçu et al. (2009) develop a probabilistic model to predict the occurrence of collision, grounding, ramming, fire/explosion, and sinking for ships sailing in the Strait of Istanbul. Silveira et al. (2013) use ship automatic identification system data to predict the number of ship collisions over a given period along the coast of Portugal. Ship deficiency in port state control covers a wide range of issues from ship operation and manning safety to environmental considerations to crew management. Unlike the literature that considers limited risk indicators, our work evaluates the ship risk level from a much broader perspective. In addition, as port state control inspects visiting ships flying foreign flags, we consider a broader geographical coverage than other studies (which often focus on a specific region). In summary, our work contributes to the literature on maritime risk management by developing a more efficient model for predicting ship risk and facilitating the inspection of ships with higher risks. This study is essentially different from the existing studies on ship risk prediction in that it makes the first attempts to reach a trade-off between model efficiency and acceptance by on the one hand improving the efficiency of SRP and on the other hand improving its acceptance level by the port authorities. The major reason of considering decision makers' acceptance is that the maritime industry is relatively conservative and is reluctant to revolutions, although some decision makers may be open-minded and pursue higher efficiency. The bi-objective feature of the optimization model can benefit decision makers holding different altitudes to emerging models, and thus is more flexible and has a wider market.

3. The SRP+ framework

In this section, we develop the SRP+ framework used to optimize the thresholds of parameters and ship risk classification based on real port state control inspection records. Our aim is to optimize the thresholds of ship age (denoted by x_1), the number of deficiencies over which the inspection will be counted in the total weighting points (denoted by x_2), the number of detentions within the last 36 months (denoted by x_3), and the total weighting points to classify a ship as HRS (denoted by $\hat{x}$). All of the other settings in the SRP+ framework, including the parameters and the weighting points, are identical to the original SRP model, as shown in Table 1. Let $\mathbf{x} = (x_1, x_2, x_3, \hat{x})$. Therefore, $\mathbf{x} = (12, 5, 2, 3)$ in the original SRP model. We consider two objectives in the SRP+ framework. First, we consider minimizing the difference in the proportions of HRS before and after optimization, because we expect that a similar proportion of ships with higher risk can be classified as HRS and thus inspected as in the original SRP model. Second, we consider maximizing the average number of deficiencies among the ships classified as HRS, because we expect the SRP+ framework to be more effective in identifying ships with higher risks. The constraints of the SRP+ framework specify the ranges of the decision variables. The notation used to develop the SRP+ framework is listed in Table 3.

Table 3: Notation used in SRP+

Sets and indices	
Z_+	The set of nonnegative integers.
S	The set of ships in the historical inspection data set.
s	The index of a ship in the historical inspection data set, $s \in S$.
$\overline{def}_s$	The multiset[6] of the number of deficiencies of ship $s \in S$ in each PSC inspection in the previous 36 months.
def'_s	The number of deficiencies in an inspection in the previous 36 months of ship s, $def'_s \in \overline{def}_s$
$\overline{det}_s$	The set of inspections of ship $s \in S$ in the previous 36 months in which it is detained. The total number of detentions of ship s can thus be represented by $\lvert\overline{det}_s\rvert$.
Parameters	
$type_s$	The type of ship $s \in S$, which is binary and takes a value of 1 if ship s is of a type of concern in the SRP (i.e., chemical tanker, gas carrier, oil tanker, bulk carrier, passenger ship, or container ship) and 0 otherwise.
age_s	The age of ship $s \in S$, which is an integer calculated by the time interval in years between the ship's keel laid date and the PSC inspection date, according to the inspection record in the historical data set.
$flag_s$	The flag state performance of ship $s \in S$, which is binary and takes a value of 1 if the flag state of ship s is on the black list published by the Tokyo MoU and 0 otherwise.
RO_s	The RO performance of ship $s \in S$, which is binary and takes a value of 1 if the RO performance of ship s is low or very low and 0 otherwise.
$company_s$	The company performance of ship $s \in S$, which is binary and takes a value of 1 if the company performance of ship s is low or very low, or there is no inspection on any ship belonging to the company (i.e., unknown) in the last 36 months and 0 otherwise.
R_{HRS}	The proportion of HRS in the whole historical inspection data set.
def_s	The real number of deficiencies of ship $s \in S$ in the current inspection.
max_def	The maximum number of deficiencies of all ships in S in the current inspection.

[6] A multiset is a modification of set that allows for multiple instances for each of its elements, e.g., $\{1, 1, 2, 3\}$.

M	A sufficiently large number.
Decision variables	
x_1	The threshold of ship age above which the ship will be given 1 weighting point. This is a nonnegative integer.
x_2	The threshold of the number of deficiencies recorded in each inspection within the previous 36 months, above which the ship will be given 1 weighting point. This is a nonnegative integer.
x_3	The threshold of the total number of detentions within the last 36 months above which the ship will be given 1 weighting point. This is a nonnegative integer.
$\hat{x}$	The threshold of the total weighting points above which the ship will be classified as HRS. This is a nonnegative integer.
W_s	The total weighting points of ship $s \in S$ according to the SRP+ framework.

3.1. Development of a bi-objective model for SRP+

From the notation in Table 3, the total weighting points of ship $s \in S$, i.e., W_s, can be calculated as follows

$$\begin{aligned} W_s =& type_s \times 2 + \mathbf{1}(age_s > x_1) + flag_s \times 1 + RO_s \times 1 + company_s \times 2 \\ &+ count(\overline{def}_s, x_2) + \mathbf{1}(\left|\overline{det}_s\right| > x_3) \times 1, \quad s \in S, \end{aligned} \tag{1}$$

where $\mathbf{1}(\omega)$ is an indicator function that takes a value of 1 if event ω is true/occurs and 0 otherwise. In the above, $count(\overline{def}_s, x_2)$ is a newly defined function that returns the number of inspections in set $\overline{def}_s$ with more than x_2 deficiencies detected. It can be calculated as follows

$$count(\overline{def}_s, x_2) = \sum_{def'_s \in \overline{def}_s} \mathbf{1}(def'_s > x_2), \quad s \in S. \tag{2}$$

The bi-objective mixed-integer nonlinear model, which is denoted as the threshold optimization problem (TOP), can be formulated as follows

$$[\text{TOP}] \qquad \min \left| \frac{\sum_{s\in S} \mathbf{1}(W_s > \hat{x})}{|S|} - R_{HRS} \right| \tag{3a}$$

$$\max \frac{\sum_{s\in S} \mathbf{1}(W_s > \hat{x}) \times def_s}{\sum_{s\in S} \mathbf{1}(W_s > \hat{x})} \tag{3b}$$

$$\text{s.t. } (1) \tag{3c}$$

$$\sum_{s\in S} \mathbf{1}(W_s > \hat{x}) \geq 1, \tag{3d}$$

$$x_1, x_2, x_3, \hat{x} \in N. \tag{3e}$$

Objective function (3a) minimizes the difference in the proportions of HRS in the whole historical inspection data set before and after threshold optimization. Objective function (3b) maximizes the average number of deficiencies among the ships that are classified as HRS. Constraint (3c) defines the total weighting points of each ship $s \in S$. Constraint (3d) requires at least one ship to be classified as HRS. The left-hand side of constraint (3d) is the total number of ships classified as HRS, which is expressed as an indicator function. The right-hand side of constraint (3d) is the threshold 1. Constraints (3e) specify the domains of the decision variables.

3.2. Linearization of the TOP model

We must linearize the TOP model to solve it using off-the-shelf mixed-integer linear optimization software. First, to remove the absolute value operation in (3a), we introduce a continuous decision variable y_1 where $y_1 = \left|\frac{\sum_{s\in S}\mathbf{1}(W_s>\hat{x})}{|S|} - R_{HRS}\right|$ in the optimal solution. Therefore, the first objective function becomes

$$\min y_1 \tag{4a}$$

$$\text{s.t. } y_1 \geq \frac{\sum_{s\in S}\mathbf{1}(W_s > \hat{x})}{|S|} - R_{HRS}, \tag{4b}$$

$$y_1 \geq R_{HRS} - \frac{\sum_{s\in S}\mathbf{1}(W_s > \hat{x})}{|S|}. \tag{4c}$$

Then, to linearize the indicator functions in (1), we introduce the binary decision variables z_s^1 and z_s^3. We define that decision variable z_s^1 takes the value of 1 if the age of ship s is greater than the threshold x_1 and 0 otherwise; decision variable z_s^3 takes the value of 1 if the total detention times of ship s in the previous 36 months is greater than x_3 and 0 otherwise. Then, z_s^1 and z_s^3 are used to linearize (1) as follows:

$$\begin{aligned} W_s =& type_s \times 2 + z_s^1 \times 1 + flag_s \times 1 + RO_s \times 1 + company_s \times 2 \\ &+ count(\overline{def}_s, x_2) + z_s^3 \times 1, \quad s \in S, \end{aligned} \tag{5a}$$

$$(age_s - x_1) - M \times z_s^1 \leq 0, \quad s \in S, \tag{5b}$$

$$(age_s - x_1) + M \times (1 - z_s^1) \geq 1, \quad s \in S, \tag{5c}$$

$$(|\overline{det}_s| - x_3) - M \times z_s^3 \leq 0, \quad s \in S, \tag{5d}$$

$$(|\overline{det}_s| - x_3) + M \times (1 - z_s^3) \geq 1, \quad s \in S, \tag{5e}$$

where M is a sufficiently large number. The value of M in constraints (5b)–(5e) can be set to age_s, $1 + \max_{s\in S} age_s$, $|\overline{det}_s|$, and $1 + \max_{s\in S} |\overline{det}_s|$, respectively.

Similarly, we introduce a binary decision variable $\hat{z}_s$ to linearize constraints (4b) and (4c), where $\hat{z}_s$ takes the value of 1 if the total number of weighting points of ship s is larger than $\hat{x}$ or 0 otherwise.

Therefore, (4b) reduces to

$$y_1 \geq \frac{\sum_{s\in S} \hat{z}_s}{|S|} - R_{HRS}, \tag{6}$$

and (4c) reduces to

$$y_1 \geq R_{HRS} - \frac{\sum_{s\in S} \hat{z}_s}{|S|}. \tag{7}$$

The following constraints are also introduced to represent the indicator function that assesses whether the total number of weighting points of ship s is greater than $\hat{x}$:

$$(W_s - \hat{x}) - M \times \hat{z}_s \leq 0, s \in S, \tag{8a}$$

$$(W_s - \hat{x}) + M \times (1 - \hat{z}_s) \geq 1, s \in S. \tag{8b}$$

The highest possible total weighting points that a ship can have according to Table 1 is $\overline{W}$, which is

$$\overline{W} = 2 + 1 + 1 + 1 + 2 + \max_{s\in S} \left|\overline{def}_s\right| + 1 = 8 + \max_{s\in S} \left|\overline{def}_s\right|. \tag{9}$$

The value of M is set to $\overline{W}$ in (8a) and the value of M is set to $\overline{W} + 1$ in (8b) because its right-hand side is 1. In the above, $\left|\overline{def}_s\right|$ is the total number of inspections conducted on ship $s \in S$. $\hat{z}_s$ is also used to linearize objective function (3b), which is reduced to

$$\max \frac{\sum_{s\in S} \hat{z}_s \times def_s}{\sum_{s\in S} \hat{z}_s}. \tag{10}$$

Nevertheless, (10) is still nonlinear, so we let $y_2 = \frac{\sum_{s\in S} \hat{z}_s \times def_s}{\sum_{s\in S} \hat{z}_s}$, and thus $y_2 \times \sum_{s\in S} \hat{z}_s = \sum_{s\in S} \hat{z}_s \times def_s$, which needs to be linearized and added into the constraints. Note that y_2 is a continuous decision variable between 0 and the maximum number of ship deficiencies max_def in the current inspection. In addition, $\sum_{s\in S} \hat{z}_s$ is the total number of ships that are classified as HRS according to the SRP+ model in the historical data set, and $\sum_{s\in S} \hat{z}_s$ lies in $\{1, ..., |S|\}$ as we require at least one ship to be classified as HRS, according to constraint (3d). For each ship $s \in S$, let $v_s = y_2 \times \hat{z}_s$, where v_s is a non-negative continuous variable as it is the product of a binary variable $\hat{z}_s$ and a positive continuous variable y_2. Then, $v_{\hat{s}}$ can be linearized by the following inequalities, which should be added to the model:

$$v_s \leq y_2, \tag{11a}$$

$$v_s \leq M \times \hat{z}_s, \tag{11b}$$

$$v_s \geq y_2 - M(1 - \hat{z}_s), \tag{11c}$$

$$v_s \geq 0, \tag{11d}$$

where M is the upper bound of continuous variable y_2, i.e., max_def.

Finally, we introduce binary decision variables $z_{def'_s}$ to linearize the function $count(\overline{def}_s, x_2)$, which take a value of 1 if $def'_s > x_2, s \in S, def'_s \in \overline{def}_s$ and 0 otherwise. Then, $count(\overline{def}_s, x_2)$ in (1) can be reduced to

$$\sum_{def'_s \in \overline{def}_s} z_{def'_s} \tag{12a}$$

$$\text{s.t. } (def'_s - x_2) - M \times z_{def'_s} \leq 0,\ s \in S,\ def'_s \in \overline{def}_s, \tag{12b}$$

$$(def'_s - x_2) + M(1 - z_{def'_s}) \geq 1,\ s \in S,\ def'_s \in \overline{def}_s, \tag{12c}$$

where the values of M in (12b) and (12c) can be set as def'_s for ship $s \in S$ and $def'_s \in \overline{def}_s$ and $1 + \max_{s \in S, def'_s \in \overline{def}_s} def'_s$ (i.e., one plus the largest historical number of deficiencies among all ships in S because the right-hand side of (12c) is 1), respectively.

After linearization, the model TOP is transformed into a bi-objective mixed-integer linear optimization model, which is denoted by TOP$'$.

[TOP$'$]

$$\min y_1 \tag{13}$$

$$\max y_2 \tag{14}$$

$$s.t.\ \ W_s = type_s \times 2 + z_s^1 \times 1 + flag_s \times 1 + RO_s \times 1 + company_s \times 2 + \sum_{def'_s \in \overline{def}_s} z_{def'_s} + z_s^3 \times 1, s \in S, \tag{15}$$

$$y_1 \geq \frac{\sum\limits_{s \in S} \hat{z}_s}{|S|} - R_{HRS}, \tag{16}$$

$$y_1 \geq R_{HRS} - \frac{\sum\limits_{s \in S} \hat{z}_s}{|S|}, \tag{17}$$

$$\sum_{s \in S} \hat{z}_s \geq 1, \tag{18}$$

$$\sum_{s \in S} v_s = \sum_{s \in S} \hat{z}_s \times def_s, \tag{19}$$

$$v_s \leq y_2, s \in S, \tag{20}$$

$$v_s \leq max\_def \times \hat{z}_s, s \in S, \tag{21}$$

$$v_s \geq y_2 - max\_def(1 - \hat{z}_s), s \in S, \tag{22}$$

$$(W_s - \hat{x}) - \overline{W} \times \hat{z}_s \leq 0, s \in S, \tag{23}$$

$$(W_s - \hat{x}) + (1 + \overline{W}) \times (1 - \hat{z}_s) \geq 1, s \in S, \tag{24}$$

$$(age_s - x_1) - age_s \times z_s^1 \leq 0, s \in S, \tag{25}$$

$$(age_s - x_1) + (1 + \max_{s \in S} age_s) \times (1 - z_s^1) \geq 1, s \in S, \tag{26}$$

$$(\left|\overline{det}_s\right| - x_3) - \left|\overline{det}_s\right| \times z_s^3 \leq 0, s \in S, \tag{27}$$

$$(\left|\overline{det}_s\right| - x_3) + (1 + \max_{s \in S} \left|\overline{det}_s\right|) \times (1 - z_s^3) \geq 1, s \in S, \tag{28}$$

$$(def'_s - x_2) - def'_s \times z_{def'_s} \leq 0, s \in S, def'_s \in \overline{def}_s, \tag{29}$$

$$(def'_s - x_2) - (1 + \max_{s \in S, def'_s \in \overline{def}_s} def'_s) \times (1 - z_{def'_s}) \geq 1, s \in S, def'_s \in \overline{def}_s, \tag{30}$$

$$x_1, x_2, x_3, \hat{x} \in N, \tag{31}$$

$$z_s^1, z_s^3, \hat{z}_s \in \{0, 1\}, s \in S, \tag{32}$$

$$z_{def'_s} \in \{0, 1\}, s \in S, def'_s \in \overline{def}_s, \tag{33}$$

$$y_1 \geq 0, \tag{34}$$

$$y_2 \geq 0, \tag{35}$$

$$v_s \geq 0, s \in S. \tag{36}$$

3.3. Solution approach to the TOP′ model

Several methods have been used to solve multi-objective optimization models, such as the weighted sum method, generic algorithms, the ϵ-constraint method (Chankong and Haimes 2008), the augmented ϵ-constraint method (Mavrotas 2009), and the improved augmented ϵ-constraint method (Mavrotas and Florios 2013). The drawback of the weighted sum method is that it cannot generate non-supported Pareto optimal solutions (which are solutions that are not on the boundary of the convex hull formed by the Pareto optimal solutions). In addition, scaling the objective is usually needed in the weighted sum method (Mrabti et al. 2022). The drawbacks of generic algorithms include that maybe not all the Pareto optimal solutions can be found, there can be huge computational burdens, and problem-specific parameter tuning is needed (Konak et al. 2006). In our problem, we can safely adopt solution approaches based on ϵ-constraint method and its revisions as the upper and lower bounds of objective function (13) and the granularity of the possible values are determined. To be specific, the range is between 0, which indicates that the proportion of ships classified as HRS in the SRP+ framework is the same as that in the original SRP model, and $\max(R_{HRS}, 1 - R_{HRS})$, which indicates that none or all of the ships are classified as HRS in the SRP+ framework. Thus, we can safely convert objective function (13) into a constraint. Furthermore, if we use the data set containing a total of $|S|$ ships to construct the bi-objective model, the granularity of the right-hand side of objective function (13) as a constraint is $\frac{1}{|S|}$. In other words, the candidate value set of the right-hand side is $\{y_1^{\min}, y_1^{\min} + \frac{1}{|S|}, ..., y_1^{\max} - \frac{1}{|S|}, y_1^{\max}\}$, and increasing the right-hand side by $\frac{1}{|S|}$ means classifying one more ship to HRS. As the number of ships belonging to HRS is discrete and bounded, which is the major feature of the problem on hand, after traversing all the values in the candidate set, all Pareto optimal solutions and the Pareto frontier can be obtained. Moreover, it can be expected that all the possible right-hand side values would lead to a feasible solution to the optimization model.

Here, we utilize the augmented ϵ-constraint method to solve model TOP′, because the original ϵ-constraint method may generate weakly Pareto optimal solutions. Weakly Pareto optimal solutions are not dominated by any other feasible solutions in terms of all objectives, but they may be worse than another solution in terms of some objectives and equally good in terms of the other objectives. Therefore, these solutions are in fact not Pareto optimal and thus the ϵ-constraint method is computationally inefficient. Note that the augmented ϵ-constraint method and the improved augmented ϵ-constraint method are the same for bi-objective optimization models, which is the case for our model TOP′. The results of model TOP′ can help us better understand and analyze the trade-offs between the two contradictory objectives. Therefore, port authority decision makers can benefit from the Pareto frontier and make informed decisions. The procedure for applying the augmented ϵ-constraint method to solve model TOP′ is given in Algorithm 1.

4. Case study of the Hong Kong port

4.1. Data description

In this section, we conduct a case study of the Hong Kong port using PSC inspection records from January 2014 to December 2021 to validate the performance of TOP′ for the SRP+ framework developed in Section 3. The data set contains a total of 4,432 inspection records. After deleting the inspection records without valid SRP information, 4,211 inspection records are left and they form the data set used for model construction and validation. The encoding methods and descriptions of the features are presented in Table 4. The whole data set is split randomly into a training set, which contains 80% of the inspection records (i.e., 3,369), and a test set, which contains the remaining 20% (i.e., 842). The training set is used to construct the optimization models, while both sets are used to validate model performance. Although we only maximize the average number of deficiencies of HRS in the SRP+ model, we evaluate the performance of the SRP+ framework using the average number of deficiencies of HRS and the detention ratio of HRS according to the inspection records, as both are major outcomes of a port state control inspection. The training set includes a total of 1,171 ships in HRS, the average number of deficiencies of which is 6.1033 and 101 of which are detained. The test set includes a total of 842 ships in HRS, the average number of deficiencies of which is 6.6054 and 23 of which are detained.

Algorithm 1 Generation of Pareto frontier for model TOP′

Input: Model TOP′.

Output: The Pareto frontier.

Step 1: Treat model TOP′ as an optimization model with only one objective: objective function (13) subject to (15) to (36), and then solve the single-objective model to obtain the lower bound of objective function (13), as denoted by $y_1^{\min}$.

Step 2: Treat model TOP′ as an optimization model with only one objective: objective function (14) subject to (15) to (36), and then solve the single-objective model to obtain the upper bound of objective function (14) denoted by $y_2^{\max}$.

Step 3: Transform objective function (14) into a constraint of model TOP′: $y_2 = y_2^{\max}$, and then solve TOP′ as an optimization model with only one objective function (13) to obtain the upper bound of objective function (13) denoted by $y_1^{\max}$. That is, $y_1^{\max}$ is the optimal value of the following model:

$$\begin{gathered} \min y_1 \\ s.t. \;\; y_2 = y_2^{\max}, \\ (15) \; to \; (36). \end{gathered}$$

Step 4: Given the range of objective function (13) as $[y_1^{\min}, y_1^{\max}]$, we set $\delta = \frac{1}{|S|}$ to discretize $[y_1^{\min}, y_1^{\max}]$ so as to obtain the Pareto frontier, denoted by set P. We initialize $P = \{(y_1^{\max}, y_2^{\max})\}$ and a temporary threshold $y_1^{\text{threshold}} = y_1^{\max} - \delta$.

Step 5:

while $y_1^{\text{threshold}} \geq y_1^{\min}$ **do**

Solve the following model

$$\begin{gathered} \max y_2 + \frac{y_1^{\text{threshold}} - y_1}{y_1^{\max} - y_1^{\min}} \\ s.t. \;\; y_1 \leq y_1^{\text{threshold}}, \\ (15) \; to \; (36). \end{gathered}$$

Denote by y_1^* and y_2^* the optimal values of y_1 and y_2, respectively. (y_1^*, y_2^*) is a new point on the Pareto frontier. Let $P = P \bigcup \{(y_1^*, y_2^*)\}$. Let $y_1^{\text{threshold}} = y_1^* - \delta$.

Return: P.

Table 4: Feature processing method and description for the whole data set

Feature	Processing method and description
Ship type	Chemical tanker, gas carrier, oil tanker, bulk carrier, passenger ship, and container ship are encoded as 1 and others as 0. In the data set as a whole, this feature is encoded to "1" for 3,584 (85.11%) ships and "0" for 627 (14.89%) ships.
Ship age	The difference in years between ship keel laid date and the inspection time recorded. The average ship age in the data set is 10.64 with a standard deviation of 6.84. The minimum value is 0 and the maximum value 48.
Flag performance	Ships with flag state performance on the black list are encoded as 1 and others as 0. In the data set, this feature is encoded as 1 for 121 (2.87%) ships and 0 for 4,090 (97.13%) ships.
RO performance	Ships with RO performance low or very low are encoded as 1 and others as 0. No ships in the data set have low or very low RO performance, and thus for all ships this feature is encoded as 0.
Company performance	Ships with ISM company performance unknown, low, or very low are encoded as 1 and others as 0. In the data set, this feature is encoded as 1 for 847 (20.11%) ships and 0 for 3,364 (79.89%) ships.
Deficiencies within previous 36 months	The set of the number of deficiencies identified in PSC inspections on a ship in the previous 36 months within the Tokyo MoU. For example, for a ship inspected on December 31, 2019 with an IMO number of 9044138, its historical deficiency multiset for the previous 36 months is $\{4, 0, 4, 0, 1, 1, 5, 1, 1, 6, 2\}$.
Detention within previous 36 months	The total number of detentions in PSC inspections of a ship in the previous 36 months within the Tokyo MoU. The average number of ship detentions in the data set as a whole is 0.20 with a standard deviation of 0.68. The minimum value is 0 and the maximum value 12.

4.2. Computational experiments

4.2.1. SRP+ performance on the training set

Based on Algorithm 1, it is easy to compute that the range of y_1 is [0, 0.347581] on the training set. We set $\delta = \frac{1}{|S|} = \frac{1}{3369}$ to obtain all Pareto optimal solutions, and we identify 348 unique solutions. The Pareto frontier is depicted in Figure 1, and the corresponding values of decision variables $\mathbf{x} = (x_1, x_2, x_3, \hat{x})$ are depicted in Figures 2 to 5.

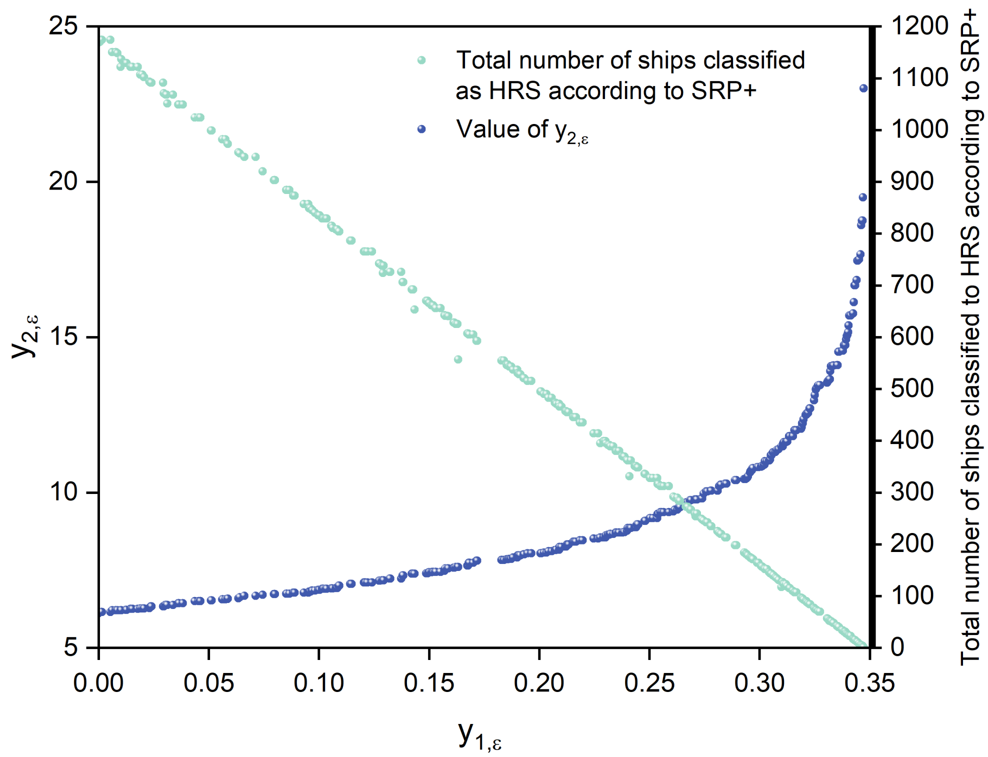

Figure 1: Pareto frontier of model TOP′

Figure 1 shows the trade-off between $y_{1,\epsilon}$ and $y_{2,\epsilon}$: As the difference in the proportions increases, the average number of deficiencies of the ships classified as HRS exhibits an increasing pattern. This indicates that as the difference between the proportions of ships classified as HRS in the SRP and SRP+ models becomes more apparent (the SRP+ framework is allowed to deviate more from the original SRP model), the SRP+ framework becomes more efficient in identifying ships with higher risks. Figures 2 shows that as ϵ increases, the threshold of the total weighting points $\hat{x}$ exhibits a general stepped shape. Meanwhile, there are a number of jump points (which are highly likely to be caused by the variations and noises in the training data) interspersed within the steps. Such trend explains why the number of ships classified as HRS in Figure 1 generally decreases. Intuitively, as the number of ships classified as HRS gradually decreases, the threshold of the total weighting points becomes more stringent, and the average number of deficiencies of the selected HRS also increases accordingly. Therefore, the basic idea and the effectiveness of the SRP framework are retained after threshold optimization in the SRP+ framework. Figure 2 also indicates that $\hat{x}$ takes the value of 4 in nearly half of the cases, which is slightly higher than that of the original SRP framework at 3.

Figure 3 shows a large fluctuation in the threshold of ship age x_1. The value of x_1 ranges between 3

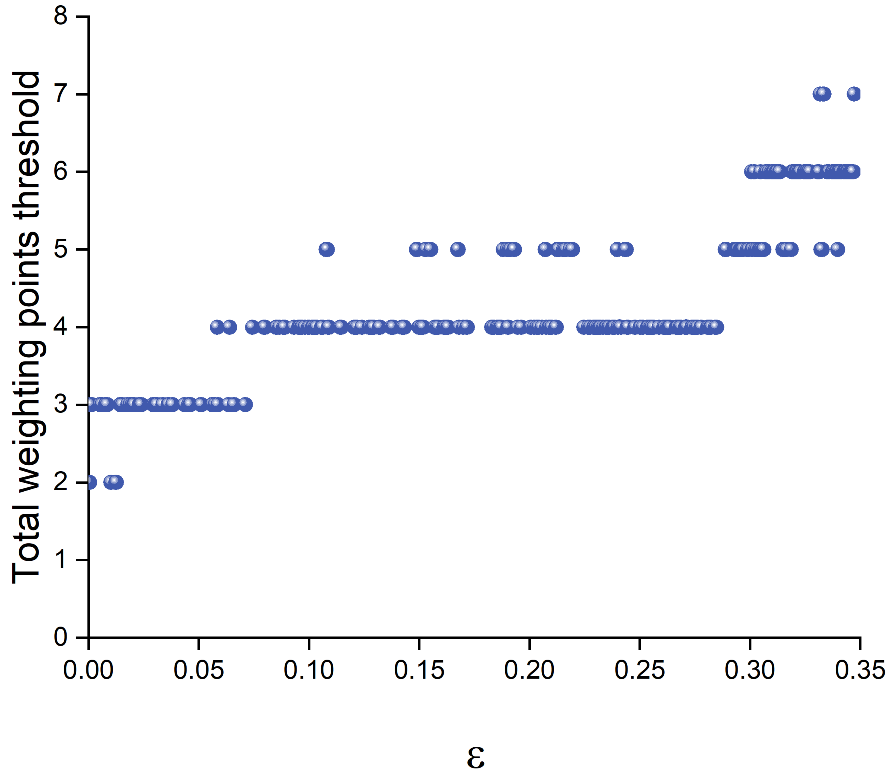

Figure 2: Threshold of total weighting points ($\hat{x}$) in model TOP′

and 43, with a large proportion of the values (about 65%) being between 16 and 24. This reflects how the average age of the global merchant fleet is approximately 21. Essentially, $x_1 = 21$ in approximately 7% of the Pareto optimal solutions, and thus we can conclude that the age threshold of 12 used in the current SRP model may be too low to identify HRS effectively. Figure 4 shows that the threshold of the historical number of ship deficiencies x_2 is almost constant at 6 when $\epsilon \in [0, 0.06]$ and drops to 4 when $\epsilon \in [0.07, 0.12]$. x_2 then increases steadily from 5 to above 10 as ϵ increases from 0.15 to 0.33. Finally, x_2 increases sharply to more than 50 when $\epsilon > 0.33$. We observe that in a large proportion of the Pareto optimal solutions (248 out of 348), x_2 is greater than the current threshold 5, showing that the current threshold may be too low and should be increased to improve the accuracy of the HRS identification. Figure 5 shows that the threshold of the historical ship detention number x_3 ranges between 0 and 4 but takes the value of 0 in about 50% of the cases and the value of 1 in about 40% of the cases. This indicates that the current threshold of 2 should be decreased, as ship detention is a rare event which is indicated by the detention rate of only 3.75% for the data set as a whole. A high threshold may make SRP less effective in distinguishing HRS from other types.

To further examine the relationship between the average number of deficiencies, the detention rate,

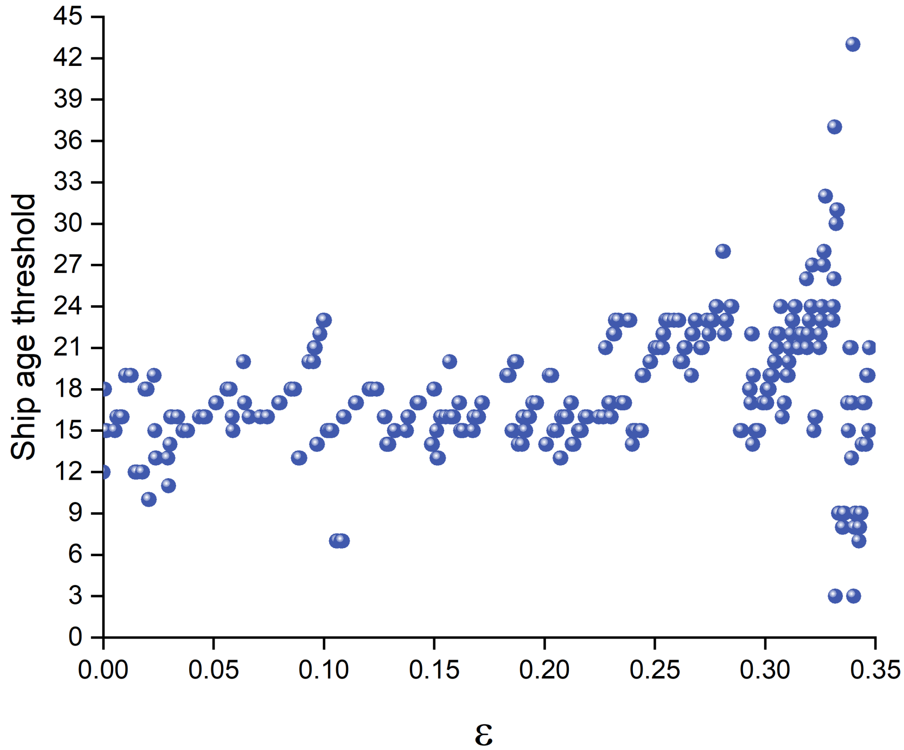

Figure 3: Threshold of ship age (x_1) in model TOP′

and the proportion of HRS, we calculate these three indicators under different intervals of ϵ. The results in Table 5 show that the SRP+ framework can improve the efficiency of HRS identification among all of the foreign visiting ships in terms of both deficiency and detention conditions under different values of ϵ, although our objective function focuses on maximizing the average number of deficiencies in HRS. In particular, when the deviation of the HRS proportion in SRP+ is only 0.0208, the mean value of the average number of deficiencies in HRS is 6.2937, which is 3.12% higher than when using the SRP model (6.1033). When the HRS proportion is 0.1220 (the corresponding deviation is 0.2256), the mean value of the average number of deficiencies is nearly 40% higher than that achieved by the SRP model. The improvement in ship detention is even more significant. When the deviation of the HRS proportion is 0.0783, the improvement of the SRP+ framework in ship detention identification improves by 18.75%. When the HRS proportion is 0.0717, the detention rate of HRS is nearly 1.5 times that of the original SRP model.

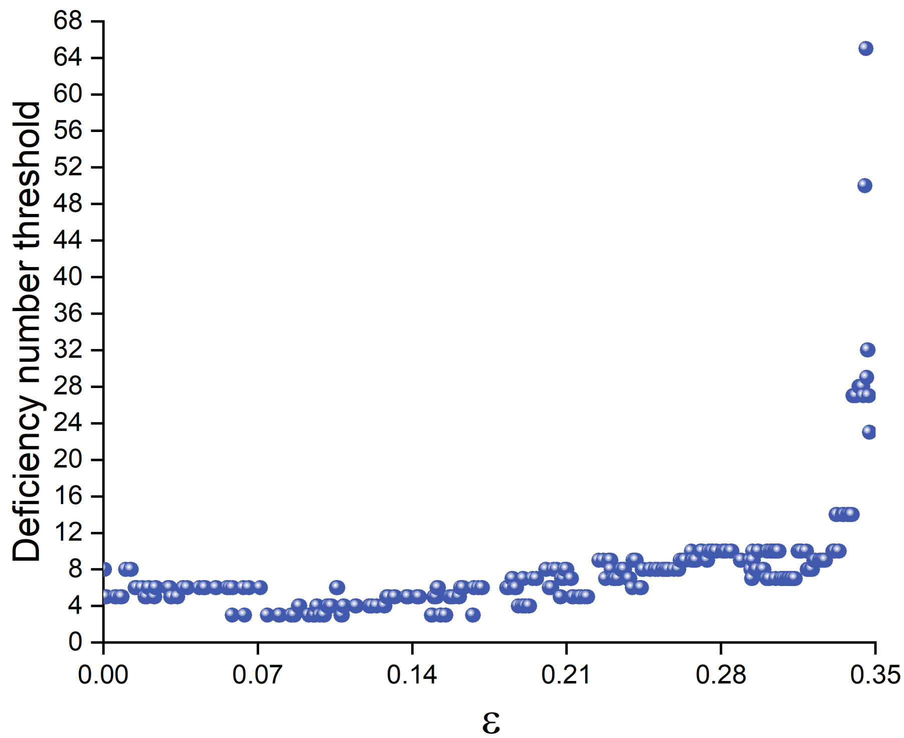

Figure 4: Threshold of number of deficiencies in the past 36 months (x_2) in model TOP$'$

Table 5: Relationship of HRS proportion and HRS conditions

Interval of ϵ	Average HRS proportion	Deviation from the original HRS proportion	Mean of the average number of deficiencies per HRS	Detention rate of HRS
$[0, 0.05)$	0.3268	0.0208	6.2937 (3.12%)[7]	0.0921 (6.83%)
$[0.05, 0.10)$	0.2693	0.0783	6.7061 (9.88%)	0.1024 (18.75%)
$[0.10, 0.15)$	0.2252	0.1223	7.1171 (16.61%)	0.1156 (33.98%)
$[0.15, 0.20)$	0.1746	0.1730	7.7265 (26.60%)	0.1366 (58.37%)
$[0.20, 0.25)$	0.1220	0.2256	8.5344 (39.83%)	0.1624 (88.32%)
$[0.25, 0.30)$	0.0717	0.2759	9.9625 (63.23%)	0.2107 (144.29%)
$[0.30, 0.35)$	0.0226	0.3249	13.5693 (122.33%)	0.3255 (277.40%)

Deficiency and detention are the most important ship risk indicators in PSC inspections. The significant improvements in both indicators show that the proposed SRP+ framework is more effective in identifying ships with higher risks than the original SRP model. This superiority remains robust across

[7]The number in parentheses shows the improvement of the SRP+ over the SRP model.

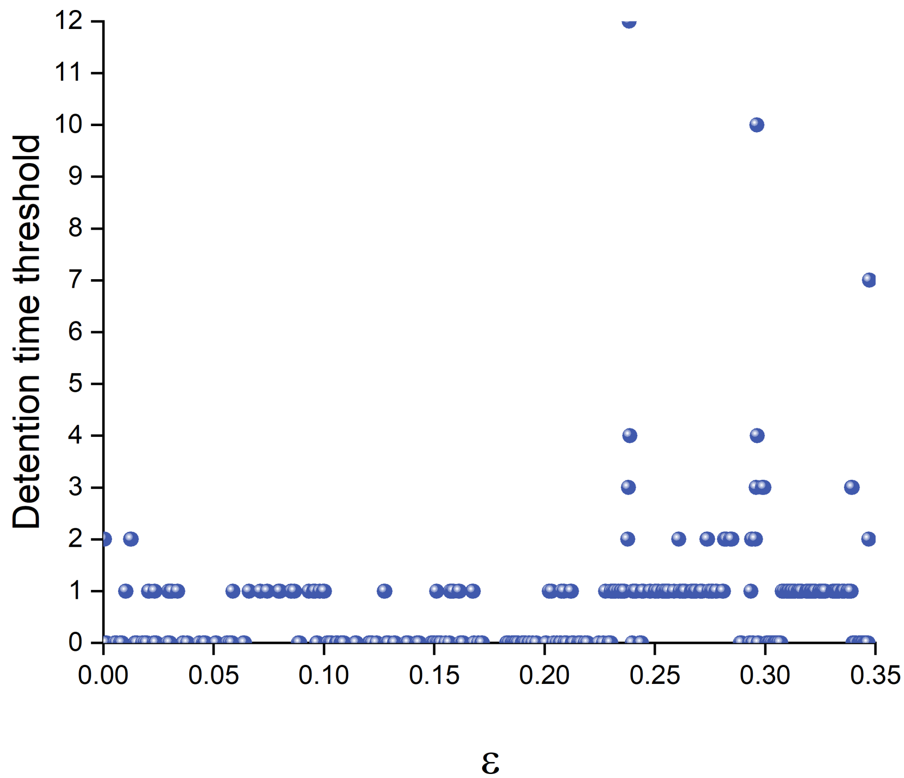

Figure 5: Threshold of detention times in the past 36 months (x_3) in model TOP′

different scenarios (under different values of ϵ) and thus can be used by port authority decision makers with different tolerances to deviations from the original framework. The SRP+ framework can thus be applied and will perform well if, for example, the HRS proportion in the original SRP policy framework is preferred (i.e., ϵ is small), if some deviations in the HRS proportion can be tolerated (i.e., ϵ is intermediate), or if limited inspection resources are focused on ships with the highest risk (i.e., ϵ is large). Even if applying the SRP+ framework directly is difficult, the SRP+ framework can help to adjust the thresholds in the SRP model, based on the values of $\mathbf{x} = (x_1, x_2, x_3, \hat{x})$ given by the Pareto optimal solutions.

4.2.2. SRP+ performance evaluation and policy insights

We further validate the performance of the proposed SRP+ framework by identifying HRS in the training and test sets. As there is no optimal solution to TOP' but a set of Pareto optimal solutions, we choose one Pareto solution from each interval of ϵ in Table 5 and set the value of ϵ as the median of the interval. That is, if there are a total of N_ε possible values of ϵ in an interval, the $\lfloor \frac{N_\varepsilon+1}{2} \rfloor$th largest ϵ will be chosen. The values of decision variables $\mathbf{x} = (x_1, x_2, x_3, \hat{x})$ are then used to classify ships as HRS in the training and test sets. Column "Deviation" is the value of objective (3a) which includes the absolute value sign to turn the deviation in HRS proportions before and after threshold optimization in itself, and thus the values in this column cannot be negative. Table 6 summarizes the values of ϵ, the corresponding values of $\mathbf{x} = (x_1, x_2, x_3, \hat{x})$, the number of HRS, the average number of deficiencies, and the detention rate of ships classified as HRS in both training and test sets.

Table 6: Performance of TOP$'$ under different values of ϵ

						Training set			Test set		
Interval of ϵ	Selected ϵ	x_1	x_2	x_3	$\hat{x}$	No. of HRS	Average number of deficiencies per HRS	Detention rate of HRS	No. of HRS	Average number of deficiencies per HRS	Detention rate of HRS
Original SRP model (Benchmark)	Not applicable	12	5	2	3	1171	6.1033	0.0863	261	6.6054	0.0881
$[0, 0.05)$	0.0199	18	5	0	3	1107	6.2692 (2.72%)*	0.0912 (5.78%)	251	6.7012 (1.45%)	0.0956 (8.50%)
$[0.05, 0.10)$	0.0801	17	3	1	4	903	6.7287 (10.25%)	0.0997 (15.56%)	192	7.2344 (9.52%)	0.1146 (30.03%)
$[0.10, 0.15)$	0.1211	18	4	0	4	765	7.1020 (16.36%)	0.1163 (34.89%)	144	8.1875 (23.95%)	0.1458 (65.49%)
$[0.15, 0.20)$	0.1701	16	6	0	4	605	7.7322 (26.69%)	0.1372 (59.06%)	116	9.2500 (40.04%)	0.1810 (105.43%)
$[0.20, 0.25)$	0.2294	17	9	0	4	399	8.5564 (40.19%)	0.1654 (91.78%)	82	10.0000 (51.39%)	0.2439 (176.78%)
$[0.25, 0.30)$	0.2746	22	10	1	4	246	9.9715 (63.38%)	0.2073 (140.36%)	47	12.4468 (88.43%)	0.3404 (286.31%)
$[0.30, 0.35)$	0.3250	22	9	1	6	75	13.1316 (115.15%)	0.3600 (317.39%)	14	19.7143 (198.46%)	0.5714 (548.44%)

Note*: The number in parentheses shows the improvement of the SRP+ over the SRP model.

Table 6 shows that in the training set, as the value of ϵ increases, the number of ships classified as HRS decreases dramatically and both the average number of deficiencies per HRS and the detention rate of ships classified as HRS increase faster when ϵ becomes larger. When the difference in the proportion of HRS is 8.01%, the SRP+ framework in the training set can improve the ship deficiency rate by 10.25% and ship detention by 15.56% when compared with the original SRP model. The improvement in the test set is also substantial, with the corresponding percentages of 9.52% and 30.03%, respectively. In addition, when the difference in the proportion of HRS is 12.11%, the corresponding improvements reach 16% (24%) and 35% (65%) in the training (test) set. These comparisons imply that even for conservative port authority decision makers who only aim to improve the efficiency of HRS detection within the SRP framework, the proposed SRP+ model still has great potential.

Alternatively, if the decision makers are more open-minded and have a higher tolerance, they can benefit more from the SRP+ framework by adopting larger values of ϵ, for example between 0.15 and

0.25. As shown in Table 6, the SRP+ framework with $\epsilon = 0.2294$ can improve the ship deficiency rate by over 40% and the ship detention rate by more than 90% in the training set. The improvement in both metrics is also significant in the test set, reaching over 51% and 177%, respectively. Thus, we can conclude that the SRP+ framework works better for ambitious decision makers who prefer the effectiveness of HRS identification over a small variation in the proportion.

Finally, if the inspection resources are limited and thus only a small proportion of ships can be inspected by the port authority (i.e., classified as HRS), the contribution of the SRP+ framework to HRS identification can be significant if larger values of ϵ, for example over 0.3, are applied. The number of ships classified as HRS is thus significantly reduced in both the training and test sets. The average number of deficiencies in HRS correspondingly increases by over 115% and the detention rate of HRS increases by over 3 times in the training set. The corresponding improvements in the test set are nearly 200% and 550%. In summary, the SRP+ framework performs well across different scenarios and can be adopted by port authority decision makers who have different tolerances to the level of deviation from the original framework.

In summary, by varying the values of ϵ, we can investigate how decision makers' attitudes toward deviations from the original SRP framework influence the effectiveness of the SRP+ model. We find that when a very small deviation is allowed (i.e., $\epsilon = 0.0199$), approximately 2.7% more deficiencies and 5.8% more detentions can be detected from ships classified as HRS in the training set by the SRP+, while approximately 1.5% more deficiencies and 8.5% more detentions can be detected from ships classified as HRS in the test set by the SRP+. With a relatively large deviation (i.e., $\epsilon = 0.3250$), only a small number of ships are classified as HRS, and the corresponding improvements increase to 115% and 317% in the training set and 198% and 548% in the test set. The results show that the proposed SRP+ model remains superior across different scenarios. Both conservative decision makers who only accept small deviations from their current strategies and aggressive decision makers who are open to more innovative strategies will benefit from the SRP+ framework. It is also applicable to ports with limited inspection resources where only a small proportion of ships can be inspected. Thus, decision makers of port authorities can apply it using different tolerances of the deviations from the original framework.

5. Conclusions

In this study, we develop an SRP+ framework (based on the original SRP model) to improve the selection of HRS in port state control inspections. The SRP+ framework achieves a balance between effectively identifying HRS and retaining the original SRP framework, in which the former is measured by the average number of deficiencies among the ships classified as HRS and the latter is reflected in the difference in the proportions of ships classified as HRS between the SRP and SRP+ frameworks. We first linearize the bi-objective mixed-integer nonlinear optimization model using a big-M method, and then solve the linear model using an augmented ϵ-constraint method by transforming one objective into a constraint and finding all the associated Pareto optimal solutions. We then use real PSC inspection records from the Hong Kong port to construct and validate the SRP+ framework. We derive managerial insights that can help to improve inspection operations at ports.

We find that the SRP+ framework can benefit port authority decision makers by more efficiently identifying ships with extensive deficiencies and through a higher detention probability, compared with the original SRP model. The Pareto optimal solutions also indicate that decision makers should slightly decrease the threshold of the total weighting points over which a ship is classified as HRS, increase the thresholds of ship age and the historical number of deficiencies, and decrease the threshold of historical detention times. Finally, the proposed SRP+ framework is flexible and effective with different tolerances of deviation from the original model; the higher the tolerance, the more significant the benefit. Thus, decision makers can effectively use and gain benefits from the SRP+ framework, no matter what their tolerances of the deviation from their current operations.

Although the data set used in the numerical experiments is from the Hong Kong port belonging to the Tokyo MoU, the proposed SRP+ framework can be applied to 50% of the PSC inspection regimes over the world where SRP is adopted, namely the Paris MoU (Paris MoU 2011), the Abuja MoU (Abuja MoU 2012), the Tokyo MoU (Tokyo MoU 2014), the Black Sea MoU (Black Sea MoU 2016), and the Indian Ocean MoU (Indian Ocean MoU 2020). Among them, Paris MoU and Tokyo MoU are the two most advanced PSC inspection regimes over the world, as they are believed to be of higher efficiency in terms of high-risk ship selection and inspection resources allocation. The annual reports of the PSC regimes show that there was a total of 74,574 inspections conducted in 2021 around the world, while 51,339 were conducted by MoUs adopting SRP as ship selection criteria, accounting for 68.84% of all the inspections (Tokyo MoU 2022, Paris MoU 2022, Caribbean MoU 2022, Acuerdo de Viña del Mar 2022, Abuja MoU 2022, Black Sea MoU 2022, Riyadh MoU 2022, USCG 2022, Indian Ocean MoU 2022, Mediterranean MoU 2022). Therefore, it can be anticipated that a wide range of PSC inspection regimes can benefit from the proposed SRP+ framework, no matter whether they are conservative or open-minded.

Several future research directions could be considered. First, a larger available volume of PSC inspection data could enable more parameters and thresholds in the SRP model to be simultaneously optimized, which could further improve its effectiveness. Second, the objective functions could account for a more comprehensive set of factors. For example, both ship deficiency and detention conditions could be used to measure ship risk level, and the differences in the proportions of HRS, SRS, and LRS could reflect the level of deviation from the SRP framework.

References

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

Abuja MoU, 2022. Abuja MoU annual report 2021. Accessed on 3 October.
URL https://abujamou.org/wp-content/uploads/2023/01/2021_Annual_Report.pdf.

Acuerdo de Viña del Mar, 2022. Annual report on port state control for Latin American Agreement 2021. Accessed on 10 September 2023.
URL https://www.equasis.org/Fichiers/Statistique/MOA/PSC%20Annual%20Reports%20Vina%20Del%20Mar%20MoU/PSC_ANNUAL_REPORT_VDMOU_2021.pdf.

Adler, N., Brudner, A., Gallotti, R., Privitera, F., Ramasco, J. J., 2022. Does big data help answer big questions? the case of airport catchment areas & competition. Transportation Research Part B: Methodological 166, 444–467.

Agra, A., Rodrigues, F., 2022. Distributionally robust optimization for the berth allocation problem under uncertainty. Transportation Research Part B: Methodological 164, 1–24.

Baksh, A.-A., Khan, F., Gadag, V., Ferdous, R., 2015. Network based approach for predictive accident modelling. Safety Science 80 (1), 274–287.

Bansal, S., Gutierrez, G. J., Keiser, J. R., 2017. Using experts' noisy quantile judgments to quantify risks: Theory and application to agribusiness. Operations Research 65 (5), 1115–1130.

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

Black Sea MoU, 2022. Port state control in the Black Sea Region annual report 2021. Accessed on 3 October 2023.
URL http://www.bsmou.org/downloads/annual-reports/BSMOU-AR-2021.pdf.

Caribbean MoU, 2022. Caribbean memorandum of understanding on port state control annual report 2021. Accessed on 21 September 2023.
URL https://www.equasis.org/Fichiers/Statistique/MOA/PSC%20Annual%20Reports%20Caribbean%20MoU/PSC_ANNUAL_REPORT_CMOU_2021.pdf.

Chankong, V., Haimes, Y. Y., 2008. Multiobjective Decision Making: Theory and Methodology. Courier Dover Publications, Mineola, New York.

Choi, T.-M., Wallace, S. W., Wang, Y., 2018. Big data analytics in operations management. Production and Operations Management 27 (10), 1868–1883.

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.

Degré, T., 2008. From black-grey-white detention-based lists of flags to black-grey-white casualty-based lists of categories of vessels? The Journal of Navigation 61 (3), 485–497.

Dinis, D., Teixeira, P., Soares, G., 2020. Probabilistic approach for characterising the static risk of ships using Bayesian networks. Reliability Engineering & System Safety 203 (1), 107073.

Feng, Q., Shanthikumar, J. G., 2018. How research in production and operations management may evolve in the era of big data. Production and Operations Management 27 (9), 1670–1684.

Guha, S., Kumar, S., 2018. Emergence of big data research in operations management, information systems, and healthcare: Past contributions and future roadmap. Production and Operations Management 27 (9), 1724–1735.

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

Huber, M., Meier, J., Wallimann, H., 2022. Business analytics meets artificial intelligence: Assessing the demand effects of discounts on swiss train tickets. Transportation Research Part B: Methodological 163, 22–39.

Indian Ocean MoU, 2020. Indian Ocean MoU, secretariat guidance on port state control inspection. Accessed October 5, 2023, `https://iomou.nic.in/HOMEPAGE/pdf/PSC_Manual/PSCBooklet.pdf`.

Indian Ocean MoU, 2022. Indian Ocean MoU 2021 annual report. Accessed on 3 October 2023.
URL `https://www.iomou.nic.in/HOMEPAGE/pdf/Annual_Report/AnrRep2021.pdf`.

Kettunen, J., Salo, A., 2017. Estimation of downside risks in project portfolio selection. Production and Operations Management 26 (10), 1839–1853.

Kleindorfer, P. R., Saad, G. H., 2005. Managing disruption risks in supply chains. Production and Operations Management 14 (1), 53–68.

Konak, A., Coit, D. W., Smith, A. E., 2006. Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering & System Safety 91 (9), 992–1007.

Liu, B., Li, Z.-C., Wang, Y., 2023. A branch-and-price heuristic algorithm for the bunkering operation problem of a liquefied natural gas bunkering station in the inland waterways. Transportation Research Part B: Methodological 167, 145–170.

Markou, P., Corsten, D., 2021. Financial and operational risk management: Inventory effects in the gold mining industry. Production and Operations Management 30 (12), 4635–4655.

Mavrotas, G., 2009. Effective implementation of the ε-constraint method in multi-objective mathematical programming problems. Applied Mathematics and Computation 213 (2), 455–465.

Mavrotas, G., Florios, K., 2013. An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in multi-objective integer programming problems. Applied Mathematics and Computation 219 (18), 9652–9669.

Mediterranean MoU, 2022. Mediterranean memorandum of understanding on port state control annual report 2021. Accessed on 3 October 2023.
URL `http://www.medmou.org/Annual_rep/Annual%20Report%202021.pdf`.

Mrabti, N., Hamani, N., Boulaksil, Y., Gargouri, M. A., Delahoche, L., 2022. A multi-objective optimization model for the problems of sustainable collaborative hub location and cost sharing. Transportation Research Part E: Logistics and Transportation Review 164 (1), 102821.

Paris MoU, 2011. Paris MoU port state control annual report 2010. Accessed October 3, 2023, https://parismou.org/sites/default/files/2021-10/Annual%20Report%202010%20final%20%2830%20June%202011%29_0.pdf.

Paris MoU, 2012. 2011 Annual report "On Course for Safer Shipping". Accessed July 4, 2022, https://www.parismou.org/2011-annual-report-course-safer-shipping.

Paris MoU, 2022. Paris MoU port state control annual report 2021. Accessed on 26 March 2023. URL http://parismou.org/system/files/2023-06/Annual%20Report%202021%20Paris%20MoU.pdf.

Riyadh MoU, 2022. Riyadh memorandum of understanding on port state control annual report 2021. URL https://www.equasis.org/Fichiers/Statistique/MOA/PSC%20Annual%20Reports%20Riyadh%20MoU/PSC_ANNUAL_REPORT_RIYADHMOU_2021.pdf.

Sanchez-Gonzalez, P.-L., Díaz-Gutiérrez, D., Leo, T. J., Núñez-Rivas, L. R., 2019. Toward digitalization of maritime transport? Sensors 19 (4), 926.

Silveira, P. A. d. M., Ângelo Palos, Soares, G., 2013. Use of AIS data to characterise marine traffic patterns and ship collision risk off the coast of portugal. The Journal of Navigation 66 (6), 879–898.

Starr, M. K., Van Wassenhove, L. N., 2014. Introduction to the special issue on humanitarian operations and crisis management. Production and Operations Management 23 (6), 925–937.

Sun, H., Toyasaki, F., Falagara Sigala, I., 2022a. Incentivizing at-risk production capacity building for COVID-19 vaccines. Production and Operations Management early view.

Sun, Q., Chen, L., Meng, Q., 2022b. Evaluating port efficiency dynamics: A risk-based approach. Transportation Research Part B: Methodological 166, 333–347.

Tokyo MoU, 2014. New inspection regime (NIR). Accessed September 6, 2020, https://www.tokyo-mou.org/inspections_detentions/NIR.php.

Tokyo MoU, 2022. Annual report on port state control in the Asia-Pacific region. Accessed on 15 July 2023. URL https://www.tokyo-mou.org/doc/ANN21-web.pdf.

Ulusçu, Ö. S., Özbaş, B., Altıok, T., Or, İ., 2009. Risk analysis of the vessel traffic in the strait of Istanbul. Risk Analysis: An International Journal 29 (10), 1454–1472.

UNCTAD, 2022. Review of Maritime Transport 2021. Accessed December 23, 2021, https://unctad.org/webflyer/review-maritime-transport-2021.

USCG, 2022. Port state control in the United States 2021 annual report. Accessed on 3 October 2023. URL https://www.dco.uscg.mil/Portals/9/DCO%20Documents/5p/CG-5PC/CG-CVC/CVC2/psc/AnnualReports/annualrpt21a.pdf.

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: Methodological 128 (1), 129–157.

Yan, R., Wang, S., Cao, J., Sun, D., 2021. Shipping domain knowledge informed prediction and optimization in port state control. Transportation Research Part B: Methodological 149 (1), 52–78.

Yan, R., Wu, S., Jin, Y., Cao, J., Wang, S., 2022. Efficient and explainable ship selection planning in port state control. Transportation Research Part C: Emerging Technologies 145 (1), 103924.

Yang, Z., Yang, Z., Yin, J., 2018. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transportation Research Part A: Policy and Practice 110 (1), 38–56.

Zhen, L., Zhuge, D., Wang, S., Wang, K., 2022. Integrated berth and yard space allocation under uncertainty. Transportation Research Part B: Methodological 162, 1–27.

A. Analysis of the irrationality in the SRP framework

We analyze the irrationality and rationality of the SRP model in terms of the thresholds of the parameters in this appendix. It is conducted by changing the threshold of the total weighting points to classify a ship to HRS, the age threshold to give a ship one weighting point, the deficiency number threshold to give a ship one weighting point, and the detention threshold to give a ship one weighting point. The analysis is conducted on the entire data set containing a total of 4,232 samples using the average deficiency number and detention rate as the metric. In addition, we also calculate the average deficiency number and detention rate for ships in different states of ship flag performance and company performance to show that the thresholds for these parameters are rational. As there is no ship with low or very low RO performance, the rationality of RO cannot be validated. The analysis results for each parameter are presented in the following subsections.

A.1. Analysis of HRS

Over the whole data set with 4,232 inspection records, the average number of deficiencies is 4.0061, and the detention rate is 0.0373. Among the 4,232 ships, 1,435 (33.91%) are classified to HRS, and their average number of deficiencies is 6.1889 and the detention rate is 0.0864, and both metrics are much higher than those in the whole data set. Meanwhile, it is also noted that in the whole data set, the average number of deficiencies of the top 1,435 ships with the highest deficiency number is 8.5847 (which we regard as the perfect condition). Considering the fact that 158 ships are detained in the 4,232 inspection records, if all the detained ships are classified as HRS (where the total number is 1,435) under the perfect condition, the detention rate of HRS would be 0.1101. The above statistics show that overall speaking, the SRP has certain ability to capture ships with higher risk, as the average deficiency number and detention rate of HRS are much higher than those in the entire data set. Meanwhile, as reflected by the gap in the average deficiency number and detention rate between the current HRS and the perfect condition, how to classify a ship to HRS still has a lot of room for improvement.

In the current SRP, the threshold of the total weighting points to classify a ship to HRS is 3. We then change this threshold to 0, 1, 2, 4, 5, 6, and 7 and record the average deficiency number and detention rate under each threshold. A smaller threshold will result in more ships classified as HRS. The results are shown in Figure 6.

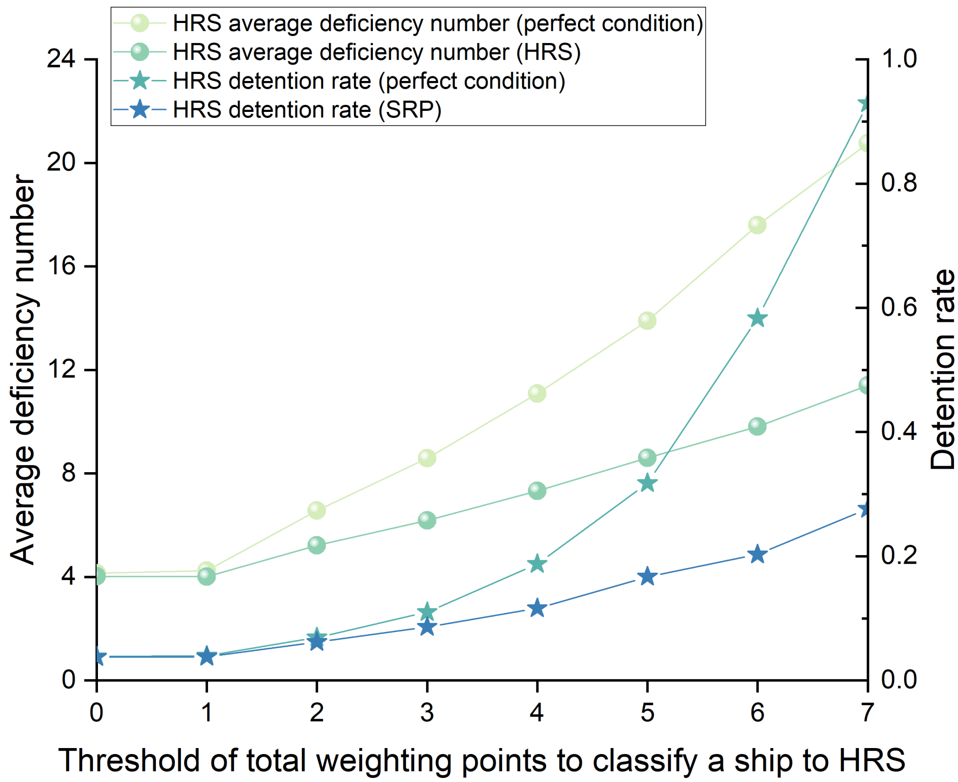

Figure 6: Deficiency and detention conditions of HRS classified by different thresholds of HRS and the performance condition

Figure 6 shows that when the threshold of HRS is small, a large number of ships will be classified as HRS, so the difference in SRP and the perfect condition is small in terms of the average number of deficiencies and detention rate. As the threshold increases, the gap between the SRP and the perfect condition increases in an increasing speed, and the increasing speeds under different thresholds are relatively stable. However, when the threshold increases from 5 to 6, there is a sharp increase in both metrics in the curves corresponding to the perfect condition. This indicates that reducing the ratio of HRS from the ratio corresponding to the HRS threshold at the current 3 to that corresponding to a higher value, such as 5 or 6, can distinguish ships with a higher risk more effectively.

A.2. Analysis of ship age

In the current SRP, the threshold of ship age is 12. We change the ship age threshold from 5 to 25 with 1 as an interval. The threshold classifies ships into two groups, where one group contains ships whose age is above the threshold, and the other group contains ships whose age is below or equal to the threshold of ship age. The results are shown in Figure 7.

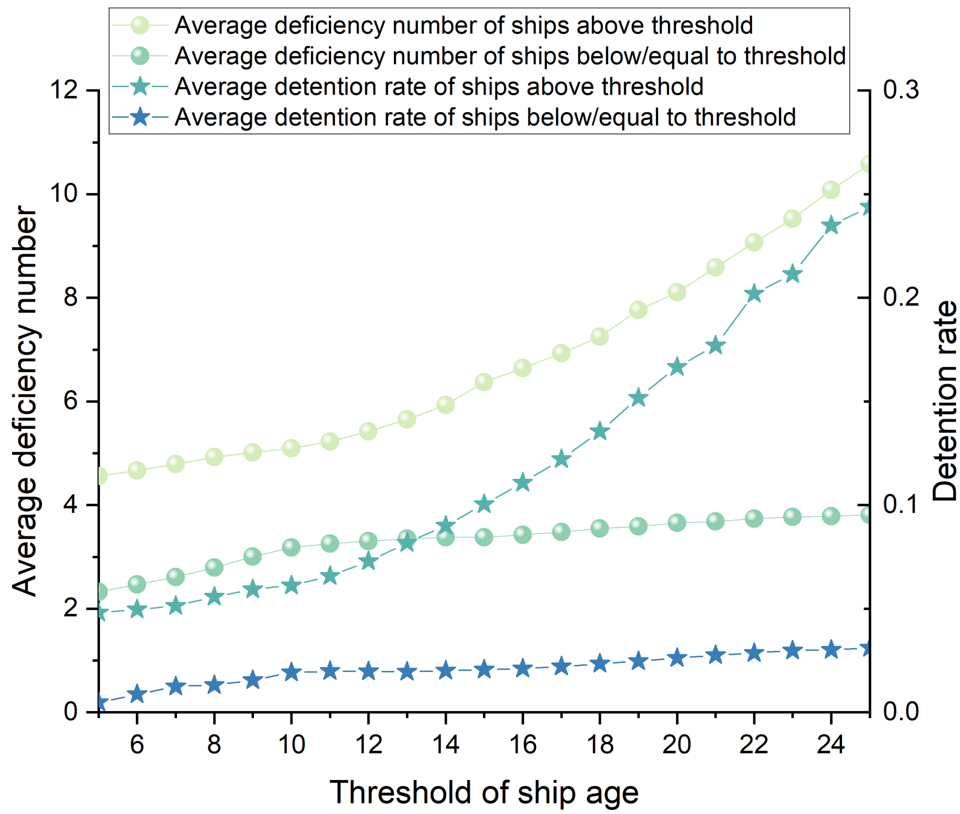

Figure 7: Deficiency and detention conditions of ships with age above and below/equal to the threshold for ship age

Figure 7 shows that the differences in the average deficiency number and detention rate of ships in the two groups are stable when ship age is no more than 11, respectively, which means that these two groups of ships do not differ too much in their inspection performance and thus the risk level. When the threshold of age is above 16 to 18, there is a surge in the difference in both metrics in the two groups, showing that it is more proper to set the age threshold to these values as they can better distinguish ships with high risk levels based on their age. Therefore, the current ship age threshold 12 may not be rational and should be further increased to about 16 to 18, or even larger.

A.3. Analysis of the number of deficiencies in the past 36 months

In the current SRP, the threshold of the number of deficiencies in the inspections of the past 36 months is 5. That is, the weighting point(s) gained from this parameter of a ship is equal to the number of inspections that have over 5 deficiencies detected in the past 36 months. To explore the inspection performances of ships with different historical deficiency conditions, we change this threshold from 1 to 10 with 1 as an interval. The threshold classifies ships into two groups, where one group contains ships that have at least one inspection with the number of detected deficiencies above the threshold, and the other group contains ships that have all inspections with less than or equal to the deficiency number threshold. We calculate the average number of deficiencies and detention rate of ships in these two groups, and the results are shown in Figure 8.

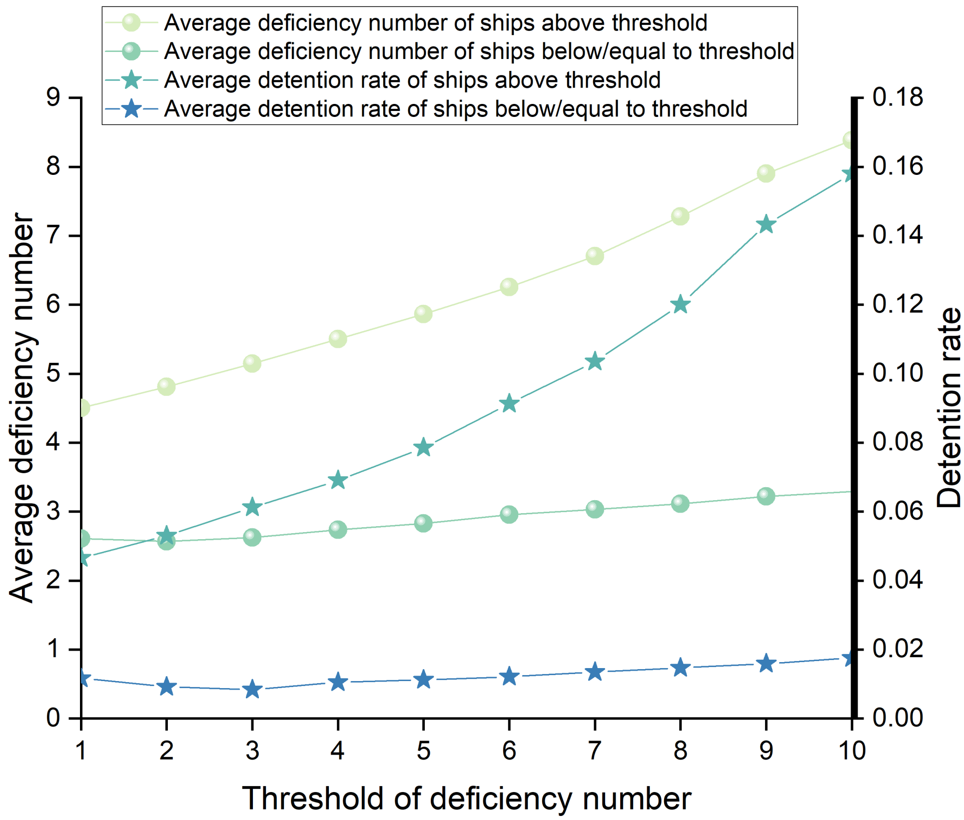

Figure 8: Deficiency and detention conditions of ships with one deficiency number in the past 36 months above and below/equal to the threshold

Figure 8 shows that the pattern of the differences in the average number of deficiencies and detention rate of ships in the two groups is similar to that of ship age. To be specific, when the threshold is small at less than 5, the differences in the average deficiency number and detention rate between the two groups are relatively stable. When the threshold is 5 or more, especially no less than 7, there is a drastic increase in the differences in the average deficiency number and detention rate between the two groups, showing that such thresholds of ship deficiency number can distinguish ships with higher risk. Therefore, the current ship historical deficiency number threshold at 5 may not be rational and should be further increased to a larger number such as 7.

A.4. Analysis of detentions in the past 36 months

In the current SRP, the threshold of the number of detentions in the past 36 months is 2. That is, if a ship has more than 2 detentions in the last 36 months, one weighting point will be given to it. We change this threshold from 0 to 5 with 1 as an interval. The threshold classifies ships into two groups, where one group contains ships that has more detentions than the threshold in the past 36 months, and the other group contains ships with fewer or equal detentions in the past 36 months compared to the threshold. We calculate the average number of deficiencies and detention rate of ships in these two groups, respectively. The results are shown in Figure 9.

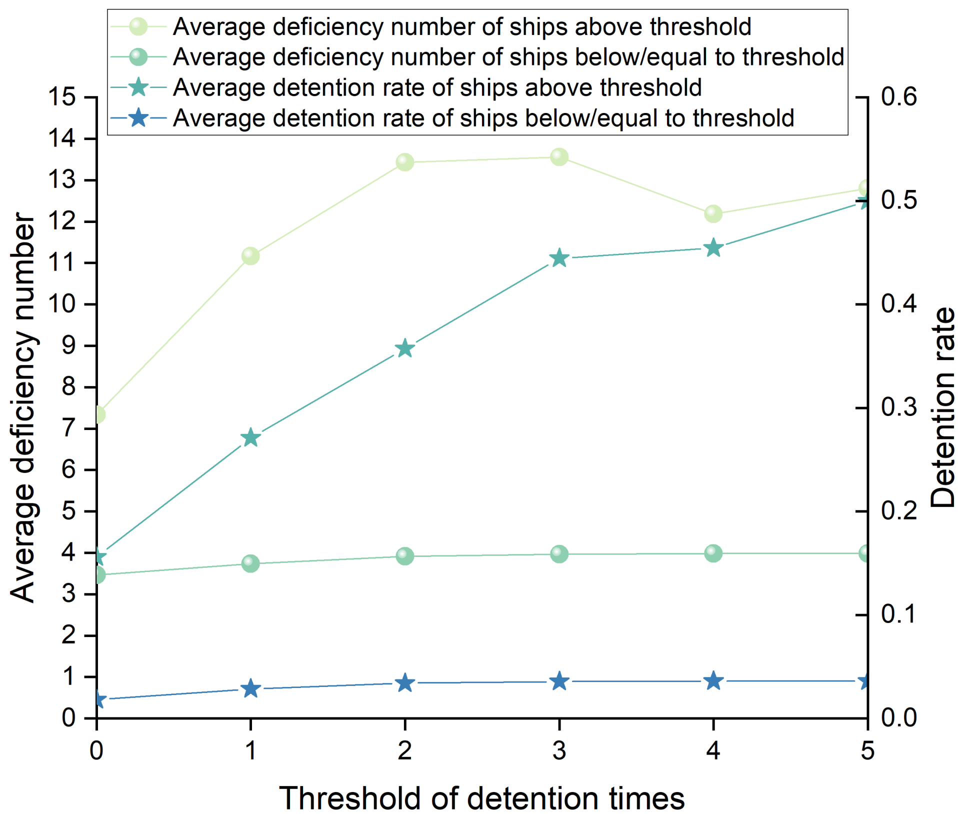

Figure 9: Deficiency and detention conditions of ships with detention times in the past 36 months above and below/equal to the threshold

Figure 9 shows that as the threshold increases from 0 to 1 and from 1 to 2, there is a sharp increase in the difference in the average number of deficiencies and detention rate between the two groups. A similar but more much moderate pattern can be found when the threshold increases from 4 to 5. It is interesting to find that when the threshold changes from 3 to 4, there is a decrease in the difference in the average deficiency number and a very slight increase in the detention rate of ships in the two groups. The above observations indicate that setting the threshold of the detention times in the past 36 months to 2 is reasonable. However, there are other candidates such as 1 and 5, and which one is the most suitable needs more explorations.

A.5. Analysis of ship flag, RO, and company performances

We calculate the average number of deficiencies and detention rate in each state of ship flag performance (i.e., white, grey, and black), RO performance (high and medium; there is no ship with low or very low RO performance), and company performance (high, medium, low, very low, and no inspection within the previous 36 months which is also denoted by state "unknown"). The results are presented in Table 7.

Table 7: Deficiency and detention conditions of ships under different states of management parties

Management party	State	Average number of deficiencies	Detention rate
Flag	White	3.69	0.026
	Grey	7.68	0.176
	Black	10.96	0.273
RO	High	3.75	0.028
	Medium	13.76	0.398
Company	High	2.61	0.012
	Medium	3.57	0.020
	Low	5.39	0.036
	Very Low	9.25	0.212
	Unknown	9.72	0.244

Table 7 shows that for ship flag performance, there is a large difference in the average number of deficiencies and detention rates of ships with flags on the white and grey lists, and on the grey and black lists. For ships with flag on the black-list, the average number of deficiencies is more than 10 and the detention rate is over 27%, which are much higher than those in the entire data set (which are 4.01 and 3.73%, respectively). This indicates that one weighting point is given to ships with flag on the black-list is reasonable. For ship RO performance, one weighting point is given to ships with low or very low performance. However, as there is no such record in our data set, it can be hard to evaluate whether the threshold for RO performance is rational. For ship company performance, ships with very low and unknown company performances have very large average number of deficiencies and very high detention rate, and ships with low company performance also have large average deficiency number and high detention rate. The deficiency number (detention rate) in all the three states are much larger (higher) than ships with high and medium company performances. Therefore, it can be seen that giving ships with low, very low or unknown company performance two weighting points is also rational.