# Prediction and Optimization for Efficient Ship Inspection

*Ran Yan*[a], *Shuaian Wang*[a*], *Kjetil Fagerholt*[b]

[a]*Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

[b]*Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Norway*

## Abstract

Efficient inspection of ships at ports to ensure their compliance with safety and environmental regulations is of vital significance to maritime transportation. Given that maritime authorities often have limited inspection resources, we proposed three two-step approaches that match the inspection resources with the ships, aimed at identifying the most deficiencies (non-compliances with regulations) of the ships. The first step of each approach is a tailored tree-based prediction model that leverages historical data to predict the ships’ deficiencies, which serve as the input for the second step. The second step is an integer optimization model that matches the inspection resources with the ships to be inspected. We prove that the integer optimization models of the three approaches can be solved in polynomial time. Numerical experiments show that the proposed approaches improve the current inspection efficiency by over 4% regarding the total number of detected deficiencies. Through comprehensive sensitivity analysis, several managerial insights are generated and the robustness of the proposed approaches is validated.

## Key words

Maritime transportation, tree-based prediction models, polynomial-time algorithm, ship inspection

[*] Corresponding author. Email addresses: angel-ran.yan@connect.polyu.hk (R Yan), wangshuaian@gmail.com (S Wang), kjetil.fagerholt@ntnu.no (K Fagerholt)

## 1. Introduction

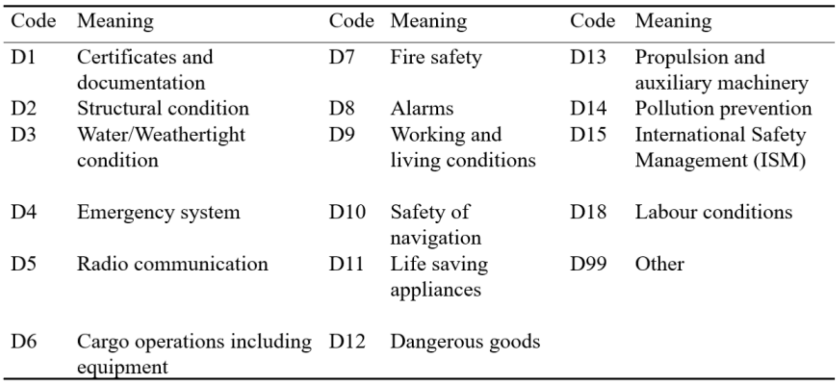
Maritime transportation forms the backbone of international trade (Ng, 2015; Dong et al., 2015; Chen et al., 2016; Lee and Song, 2017; Yu et al., 2018). According to the report of European Maritime Safety Agency, there were 3,174 ship casualties and incidents that caused 941 people injured and 53 fatalities in 2018 (European Maritime Safety Agency, 2019a). Due to the increasing volumes of maritime traffic and high losses brought by accidents at sea, growing awareness has been given to the safety of maritime traffic (Jabari et al., 2014; Ng and Lo, 2016; Zheng et al., 2017; Teye et al., 2018; Bell et al., 2020). In recent years, sustainable shipping, which requires environmental protection, reduction impact on ecosystems, and improving energy efficiency, has become another goal of the shipping industry (Angeloudis et al., 2016; European Maritime Safety Agency, 2019b). Various regulations and international conventions are proposed and implemented to guarantee maritime safety, protect the marine environment, and provide decent working and living conditions to the seafarers, including, but not limited to, the International Convention for the Safety of Life at Sea (SOLAS), the International Convention for the Prevention of Pollution from Ships (MARPOL), the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW), and the Maritime Labour Convention (MLC) proposed by the International Maritime Organization (IMO) and International Labour Organization (ILO) (UNCTAD, 2019).

The flag states of ships, under whose laws the vessels are registered or licensed, play an important role in enforcing these regulations and international conventions, and thus they are regarded as the first line of defense against substandard vessels. However, it is believed that flag states cannot perform their duties efficiently (Li and Zheng, 2008; Wang et al., 2019). As a complement to flag state control, port state control (PSC) inspection, which is an inspection regime for ports to inspect foreign visiting ships, was first implemented in 1982 and since then it has been viewed as the second line of defense against substandard vessels (Cariou et al., 2009; Heij et al., 2011).

To allow inspection information exchange, avoid duplicate ship inspections within a limited period of time in a certain region, and standardize inspection criteria and processes, the regional Memorandum of Understandings on port state control (i.e. MoUs on PSC) are signed and established. Currently, there are nine MoUs on PSC all over the world. For example, Hong Kong belongs to Tokyo MoU which is responsible for the Asia-Pacific Region. When the foreign ships come to the port state, the port state authority first needs to select the ships to be inspected. Then, available PSC officers (PSCOs) are assigned to these ships for PSC inspections. There are two types of PSC inspection: initial inspection and follow-up inspection. During the inspection process, a ship condition found not to be in compliance with the requirements of the relevant convention is denoted as a deficiency (IMO, 2017). After finishing an inspection, the inspection results, including ship deficiencies and detention, together with ship information are recorded in the corresponding database.

17 deficiency codes are required by Tokyo MoU as presented in Table 1. Except for D99, the remaining 16 deficiency codes can be grouped into four deficiency categories as follows: C1: ship safety (D4 Emergency system, D7 Fire safety, D11 Life saving appliances, and D12 Dangerous goods), C2: ship management (D1 Certificates and documentation, D9 Working and living conditions, D14 Pollution prevention, D15 International Safety Management (ISM), and D18 Labour conditions), C3: ship condition and structure (D2 Structural condition, D3 Water/Weathertight condition, D6 Cargo operations including equipment, and D13 Propulsion and auxiliary machinery), and C4: communication and navigation (D5 Radio communication, D8 Alarms, and D10 Safety of navigation). It should be noted that the deficiencies and deficiency categories are of equal importance as they are all derived from major international maritime regulations and conventions.

Table 1. Description of deficiency codes



| Code | Meaning | Code | Meaning | Code | Meaning |
|---|---|---|---|---|---|
| D1 | Certificates and documentation | D7 | Fire safety | D13 | Propulsion and auxiliary machinery |
| D2 | Structural condition | D8 | Alarms | D14 | Pollution prevention |
| D3 | Water/Weathertight condition | D9 | Working and living conditions | D15 | International Safety Management (ISM) |
| D4 | Emergency system | D10 | Safety of navigation | D18 | Labour conditions |
| D5 | Radio communication | D11 | Life saving appliances | D99 | Other |
| D6 | Cargo operations including equipment | D12 | Dangerous goods | | |

The overall inspection process suggests that the PSCOs play the key role in PSC inspection as they are responsible for conducting the inspection and deciding the inspection results (Ravira and Piniella, 2016; Graziano et al., 2017, 2018a). A PSCO should be an experienced person with both theoretical knowledge and seagoing experience. Common backgrounds of PSCOs can be naval architects, merchant marine captains, chief engineers, and ratio officers (Intercargo, 2000; Ravira and Piniella, 2016). As required, during an inspection, a PSCO will use his/her professional judgment to decide whether and in what aspects the ship should be further inspected. The PSCO will also use his/her expertise to decide what deficiencies should be recorded and whether to detain a ship. However, it is indicated that due to discretion, subjectivity, individuality, professional judgement, different backgrounds and expertise, PSCOs at the same port may have different expertise in identifying different categories of deficiencies (Ravira and Piniella, 2016; Graziano et al., 2017; Graziano et al., 2018a). For instance, there are two PSCOs on duty for one day, and PSCO 1 used to be a captain who is good at dealing with deficiencies related to communication and navigation, while PSCO 2 has naval background and is good at addressing deficiencies on the ship condition and structure. Assume now that two ships visiting the port are selected to be inspected: ship 1 has main deficiencies in structure and ship 2 has many deficiencies in radio communication. Ideally, we should assign PSCO 1 to inspect ship 2 and assign PSCO 2 to inspect ship 1; otherwise, deficiencies might be missing due to the lack of professional backgrounds and knowledge. This example shows that the inspection efficiency and effectiveness can be improved if ship deficiency conditions and the expertise of PSCOs are matched. To achieve this objective, the deficiency conditions of the ships, which can be represented by the number of deficiencies in each category (the number can be zero) need to be first predicted. Then, the expertise of PSCOs should be considered when assigning them to the ships to be inspected.

Considering the PSCOs' different expertise, this study proposes three approaches for ship deficiency condition prediction and PSCO assignment to improve the inspection efficiency. Our key contributions from a theoretical and practical point of view are summarized as follows.

First, from a theoretical point of view, we develop three sequential prediction and optimization approaches for the PSCO assignment problem. The first approach predicts the number of deficiencies in each deficiency category for each ship in a way that minimizes the mean squared error (MSE). The numbers of deficiencies are a natural choice of target to predict. The predicted values are subsequently used in a PSCO assignment model (model M1 in Section 4.1). The second approach predicts, instead of the number of deficiencies in each category for each ship, the number of deficiencies each PSCO can identify for each ship (also in a way that minimizes the MSE). The predicted values are subsequently used in a slightly different PSCO assignment formulation (model M2 in Section 4.2). The third approach also predicts the number of deficiencies each PSCO can identify for each ship. However, instead of minimizing the MSE as in the second approach, this approach adopts a loss function motivated by the structure of the optimization problem. It aims to minimize the mean squared difference regarding the overestimates (i.e., predicted value minus actual value) in the numbers of deficiencies that can be detected among the PSCOs for each ship (denoted by MSO for short). The prediction results are then applied to a PSCO assignment formulation (model M2 in Section 4.2). We demonstrate, on the basis of the three approaches, that (i) there may be different choices of targets to predict in the prediction model and then feed the targets into an optimization model and (ii) the structure of the optimization model may provide useful information to guide the training of the prediction model, even if the overall prediction and optimization procedure is sequential. Therefore, prediction models that show worse performance regarding classical regression metric (e.g. MSE) would not necessarily generate worse decisions in the following optimization models. Besides, we have rigorously proved that the optimization models can be solved in polynomial time of the length of its input parameters.

Second, from a practical point of view, we address a meaningful problem in maritime transportation. Improving inspection efficiency and effectiveness is a critical measure for PSC MoUs to guarantee maritime safety and protect the marine environment. One key point is realizing accurate identification of the deficiencies of the coming ships, which benefits from accurate prediction. Based on the three prediction models and the optimization model proposed in this study, the expertise of PSCOs can be fully utilized in dealing with various deficiency conditions of the ships. Particularly, compared with random assignment of PSCOs, the proposed three models can help to detect 4.70%, 4.55%, and 4.86% more deficiencies, respectively, after inspecting the same groups of ships by using the same PSCO resources. Comprehensive robustness analysis shows that even if there may be some uncertainties in measuring the expertise of PSCOs, the PSCO assignment scheme generated by the third proposed model can still identify more than 90% of the real deficiencies and significantly outperforms random PSCO assignment. From the perspective of application, as reported by Tokyo MoU, there were totally 31,589 PSC inspections and the total number of deficiencies detected was 73,441 in 2017 (Tokyo MoU, 2018a). This indicates that the average number of deficiencies of one ship in one PSC inspection is about 2.32. If our models are applied, about 3,569 more deficiencies can be detected (as 4.86% more deficiencies can be identified compared with random PSCO assignment), which can be viewed as inspecting about 1,538 more ships with the same inspection resources. Therefore, human, material and financial resources could be saved if the inspection efficiency is improved.

## 2. Literature review

As PSC inspection plays a vital role in guaranteeing maritime safety and protecting the marine environment, a large and growing body of literature has investigated different aspects of PSC inspection (Yan and Wang, 2019). These studies can be classified into three main categories: studies of the influence of port state control, studies of MoU management, and studies of improving the efficiency of PSC inspection. As our research topic lies in the areas of comments on MoU management and improving PSC inspection efficiency, recent and related studies of the two areas are reviewed.

### 2.1 Comments on MoU management

Although the effectiveness of PSC inspections in improving the safety level of maritime transport has been widely recognized by industry and academia, there are still critical challenges faced by port state authorities. One of the biggest challenges is the discrepancy in the inspection process and criteria among different PSC MoUs, port states of the same MoU, and even PSCOs at the same port. More specifically, variations in the treatment of vessels across the MoUs were identified and reported by Sampson and Bloor (2007), Knapp and Franses (2007), Knapp and van de Velden (2009), and Kara (2016), and the differences in inspections within the same MoU were found by Bateman (2012), Graziano et al. (2018b), and Şanlıer (2020), while the different treatment caused by different backgrounds and expertise of the PSCOs was investigated by Ravira and Piniella (2016) and Graziano et al. (2017, 2018a). It is of vital importance to achieve harmonization in PSC inspections, or the ship operators will recognize that they no longer necessarily gain a great deal from efforts to comply with regulations and thus substandard ships will "port shop", i.e., choose to call the ports with looser PSC inspection criteria.

The models proposed in this paper could help to address the problems brought about by the diverse backgrounds and expertise of PSCOs at the same port by matching the ship deficiency conditions with PSCOs' expertise. Besides, the phenomenon of "port shop" can also be alleviated by improving inspection efficiency.

### 2.2 Improving the efficiency of PSC inspection

A general PSC inspection contains four main steps: ship selection when foreign ships come to the port state, assignment of PSCOs to inspect the selected ships, conduction of ship inspection, and making and recording decisions on the inspection results. A great deal of previous research into the PSC inspection process has focused on improving the efficiency of ship selection for port state authorities. A risk-based Bayesian network (BN) model was developed by Yang et al. (2018a) to predict the detention probabilities of the coming ships while another data-driven BN model was proposed by Wang et al. (2019) to predict the total deficiency number of the coming ships based on various influencing factors. Based on the outcomes of the BN model proposed by Yang et al. (2018a), Yang et al. (2018b) developed a strategic game model to figure out the optimal inspection rate for the port state authorities. By examining several databases, a four-step protocol which considered ship detention risk and the incident risk was presented to rank the ships according to their risk level (Heij and Knapp, 2019).

To improve the efficiency of ship inspection process, several studies have investigated the correlations among the detected deficiencies and the correlations among those deficiencies and the external factors, such as ship age, ship type, and ship gross tonnage (GT). The major measure adopted is the association rule mining technology. Tsou (2018) used association rule mining methods to examine the relationship among the deficiencies as well as the relationship of the external factors and the detected deficiencies in the detained ship database of Tokyo MoU. Yan et al. (2019) developed two inspection schemes for PSC inspection based on the probabilities of the occurrence of deficiencies in the database and the application of the association rule mining algorithm to those deficiencies.

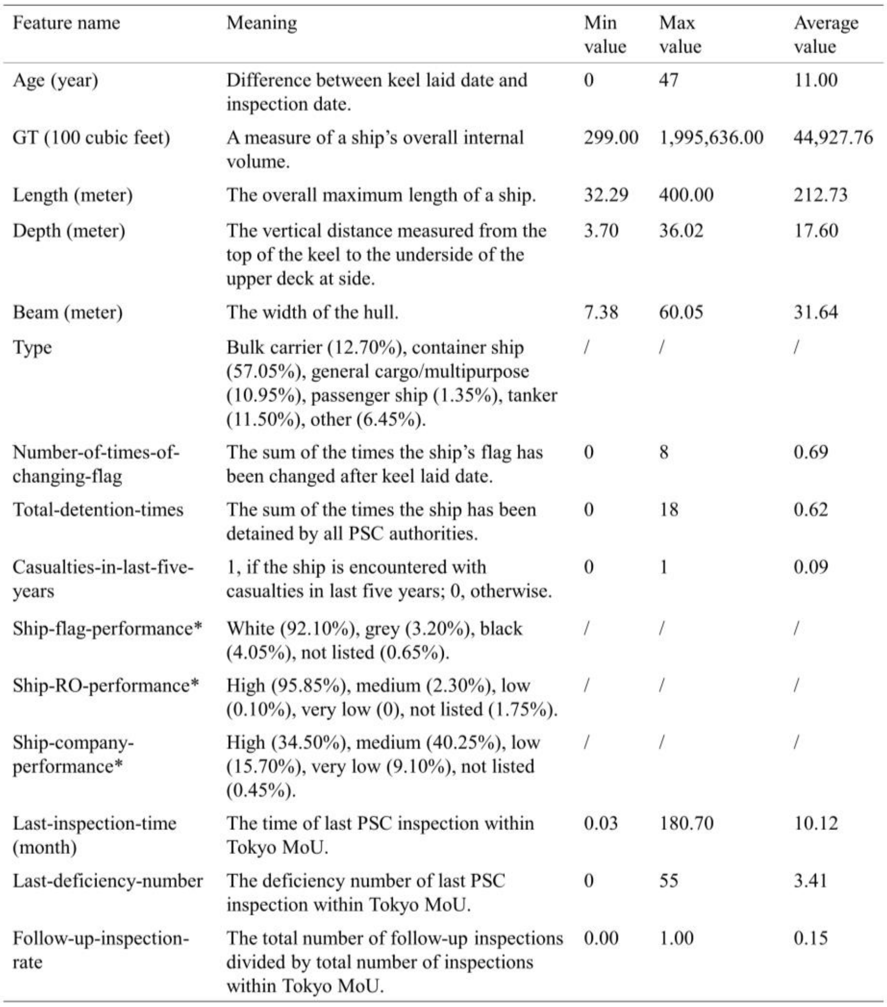
Several studies have discussed and analyzed the factors that would influence the results of PSC inspection, including detected deficiencies and ship detention. By adopting econometric models, it was shown that ship generic factors, such as ship age, flag, and type would have determinant impact on the reported deficiency number (Cariou et al., 2007). Factors leading to a large number of deficiencies and ship detention were also analyzed and identified (Cariou and Wolff, 2015; Chen et al., 2019). Apart from ship-related factors, differences in MoUs, port states, and PSCOs could also lead to diversities in deficiency identification and detention decisions (Knapp and Franses, 2007; Ravira and Piniella, 2016; Şanlıer, 2020).

Although there are a large number of studies on improving the PSC inspection efficiency, to the best of our knowledge, there is no literature on developing PSCO assignment schemes to improve inspection efficiency by considering the expertise and backgrounds of PSCOs and the deficiency conditions of the ships.

## 3. Data description and the PSCO assignment problem

The Asia Pacific Computerized Information System (APCIS) provided by the Tokyo MoU and World Register of Ships (WRS) database are used in this study. APCIS is a public website based database of PSC inspections conducted by the member authorities of the Tokyo MoU. It contains ship generic information and historical PSC inspection records within the Tokyo MoU (including the specific deficiencies detected for each inspected ship). WRS is a comprehensive database providing hundreds of features on ship construction, engine, dimension, registration, ownership, fixtures, and class, etc. We select the most relevant features of PSC inspection from WRS based on the literature. The features selected from APCIS and WRS are combined by ship IMO number, and there are 15 input features in total. The description of the features and their statistical information used in this study are provided in Table 2. For ships that have never had any inspection within Tokyo MoU, the values for “last inspection time”, “last deficiency number” and “follow-up inspection rate” are set to be “none” (not included in Table 2).

Table 2. Description of input features



| Feature name | Meaning | Min value | Max value | Average value |
|---|---|---|---|---|
| Age (year) | Difference between keel laid date and inspection date. | 0 | 47 | 11.00 |
| GT (100 cubic feet) | A measure of a ship’s overall internal volume. | 299.00 | 1,995,636.00 | 44,927.76 |
| Length (meter) | The overall maximum length of a ship. | 32.29 | 400.00 | 212.73 |
| Depth (meter) | The vertical distance measured from the top of the keel to the underside of the upper deck at side. | 3.70 | 36.02 | 17.60 |
| Beam (meter) | The width of the hull. | 7.38 | 60.05 | 31.64 |
| Type | Bulk carrier (12.70%), container ship (57.05%), general cargo/multipurpose (10.95%), passenger ship (1.35%), tanker (11.50%), other (6.45%). | / | / | / |
| Number-of-times-of-changing-flag | The sum of the times the ship’s flag has been changed after keel laid date. | 0 | 8 | 0.69 |
| Total-detention-times | The sum of the times the ship has been detained by all PSC authorities. | 0 | 18 | 0.62 |
| Casualties-in-last-five-years | 1, if the ship is encountered with casualties in last five years; 0, otherwise. | 0 | 1 | 0.09 |
| Ship-flag-performance* | White (92.10%), grey (3.20%), black (4.05%), not listed (0.65%). | / | / | / |
| Ship-RO-performance* | High (95.85%), medium (2.30%), low (0.10%), very low (0), not listed (1.75%). | / | / | / |
| Ship-company-performance* | High (34.50%), medium (40.25%), low (15.70%), very low (9.10%), not listed (0.45%). | / | / | / |
| Last-inspection-time (month) | The time of last PSC inspection within Tokyo MoU. | 0.03 | 180.70 | 10.12 |
| Last-deficiency-number | The deficiency number of last PSC inspection within Tokyo MoU. | 0 | 55 | 3.41 |
| Follow-up-inspection-rate | The total number of follow-up inspections divided by total number of inspections within Tokyo MoU. | 0.00 | 1.00 | 0.15 |

* Note: Ship flag performance, recognized organization (RO) performance, and company performance are calculated based on flag Black-Grey-White list, RO performance list, and company performance list provided by Tokyo MoU, respectively. The performance of the flags on white-list is better than those on grey-list, and much better than those on black-list. For RO and company, the performance gets worse in the sequence of “high”, “medium”, “low”, and “very low”. If the performance of the ROs and companies is not shown on the lists, the performance state is recorded as “not listed”.

We use a total of 2,000 inspection records at the Hong Kong port in 2016 (638 records), 2017 (641 records) and 2018 (721 records) in our study. One thing that needs to be mentioned is that collecting data from the two databases is a time-consuming task: due to the various data fields and the structure of the database, “copy-paste” is needed for data collection. We use the PSC inspection records at the Hong Kong port because we have visited the Marine Department of Hong Kong Special Administrative Region (HKSAR) and discussed with the PSCOs here for several times. We learned that the PSC Section of Hong Kong Marine Department has four PSCOs who are all experienced experts in all aspects of PSC. Besides, it is required that a PSCO should participate in strict trainings and assessments before becoming a qualified PSCO according to the requirements of the Hong Kong Marine Department, and the PSCOs also need to attend regular training programs and seminars. Therefore, we suppose that the PSCOs at the Hong Kong port can identify all the deficiencies in each category for each inspected ship. Nevertheless, it should be noted the PSCOs at some ports may not be that experienced, and thus the Tokyo MoU has developed several co-operation programs to enhance consulting, cooperating and exchanging information among the authorities (Tokyo MoU, 2018a). The models proposed in our study aiming to match the ship conditions with the PSCOs' expertise can also be viewed as a type of cooperation and thus are more suitable for those ports with PSCOs of divergent expertise. We randomize the whole dataset and divide it into training set, validation set and test set with each containing 70%, 15% and 15% of all data entries, i.e.,1400, 300 and 300 data entries, respectively.

According to the working process of the PSC authorities, in the morning of each day, a set of ships (denoted by $S$ ) to be inspected will be selected among all the ships coming to the port state on that day. A total of $P$ PSCOs will then be assigned for ship inspection. It is not uncommon that some PSCOs have limited expertise in some aspects of PSC because of limited work experience and training. It is, therefore, valuable to leverage historical inspection data and predict the number of deficiencies in each category for each ship, and based on the predicted number, to assign PSCOs with the relevant expertise to inspect the ships. Let $C=4$ be the number of categories of deficiencies (i.e. ship safety, ship management, ship condition and structure, and communication and navigation mentioned in Section 1). The expertise of PSCO $p$ for inspecting deficiency category $c$ is denoted by $u_{pc}$, $p=1,...,P$, $c=1,2,3,C$. $u_{pc}$ is actually the percentage of deficiencies of category $c$ that can be detected by PSCO $p$, and $0 \le u_{pc} \le 1$. The smaller $u_{pc}$ is, the more deficiencies in $c$ are likely to be ignored by PSCO $p$. The expertise (which is represented by percentage) can be evaluated by tests, questionnaires, and interviews. Considering the workload for the PSCOs, we further require that the maximum number of ships that a PSCO can inspect for each day is $\Theta$. We try to assign the available PSCOs to the selected ships in a way that maximizes the total number of deficiencies in all the $C$ categories of all the ships that can identified.

The prediction and optimization models proposed in this study work in the following way: deficiency prediction models with three different targets/model structures are first developed. Based on the prediction results, optimization models for PSCO assignment to maximize the inspection efficiency are then proposed. Several comparisons are made and comprehensive sensitivity analyses is conducted to generate managerial insights and validate the robustness of the models.

## 4. Prediction and optimization approaches

In our prediction and optimization approaches, a prediction model is first developed to predict the key unknown parameters in the optimization model. Based on the predicted values, an optimization model is then constructed to generate decisions. Particularly, we propose three prediction models denoted by MTR-RF1, MTR-RF2, and MTR-RF3 and two assignment models denoted by M1 and M2 with details provided in Table 3.

Table 3. Prediction and optimization models

| Model | Prediction targets | Splitting criteria | Decision trees | Assignment model | Assignment decision |
|---|---|---|---|---|---|
| MTR-RF1 | Number of deficiencies under each deficiency category | MSE | $f^{MTR}(\mathbf{x})$ | M1 | A1 |
| MTR-RF2 | Number of deficiencies identified by each PSCO | MSE | $f'^{MTR}(\mathbf{x})$ | M2 | A2 |
| MTR-RF3 | Number of deficiencies identified by each PSCO | MSO | $f''^{MTR}(\mathbf{x})$ | M2 | A3 |

### 4.1 Prediction of natural targets and optimization

It is natural to predict the number of deficiencies in each category for each ship based on historical records. Therefore, we first develop random forest regression model (denoted by MTR-RF1) to predict the number of deficiencies in each category for each ship based on the features in Table 2.

### 4.1.1 Prediction model

We use random forest (RF) as the prediction model. RF is a state-of-the-art machine learning model with high accuracy and is widely used (Friedman et al., 2001; Liaw and Wiener, 2002; Breiman, 2017). Since constructing a decision tree is a sub-procedure, we present decision tree in Section 4.1.1.1. and RF in Section 4.1.1.2.

#### 4.1.1.1 Multi-target regression (MTR) tree

Decision tree (denoted by DT for short) is a popular supervised machine learning model. At the beginning, all the training examples are stored in the root node. Then, the root node is recursively split into successive nodes which contains subsets of the training set until coming to the preset stopping criterion or the current node cannot be further split (i.e. all the examples are of the same output value). Each split of the nodes in the decision tree aims to reduce the variance among the records in the successive nodes. According to the target, decision trees that predict categorical target are called classification trees while decision trees that predict numerical target are called regression trees. The target is one-dimensional in traditional decision tree while the targets can be multi-dimensional in multi-target regression (MTR) tree (Blockeel and De Raedt, 1998). In this study, the outputs are four-dimensional (either the number of deficiencies under the four categories or the number of deficiencies detected by the four PSCOs), and thus the MTR trees are constructed by using classification and regression tree (CART) algorithm (Friedman, 2001; Harrington, 2012; Breiman, 2017). The procedure is as follows (Blockeel, 1998; Friedman et al., 2001).

The input information for decision tree construction contains the training dataset and termination conditions. We denote the set of $J$ input features as $(x_1, x_2, ..., x_J)$ and the set of $K$ targets as $(y_1, y_2, ..., y_K)$. An input feature is denoted by $x_j$, $j = 1, ..., J$, and the value set of this feature is denoted by $\Omega_j$. A specific value of this feature is denoted by $w_j$, $w_j \in \Omega_j$. For example, for the variable ship-flag-performance which has four states: white, grey, black, and not listed, the states are first changed to numbers, with 1 representing white, 2 representing grey, 3 representing black, and 4 representing not listed. Then, we can have $\Omega_j = \{1, 2, 3, 4\}$. A target is denoted by $y_k$, $k = 1, ..., K$ and $K = 4$. In addition, we denote the training dataset containing $N$ data entries as $D = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), ..., (\mathbf{x}^N, \mathbf{y}^N)\}$. We use $e = 1, ..., N$ to refer to both an inspection record and the ship in the current record. Notably, if a ship is inspected several times, its inspection records are treated independently. A data entry is denoted by $(\mathbf{x}^e, \mathbf{y}^e)$ with $e = 1, ..., N$, where $\mathbf{x}^e = (x^{e1}, x^{e2}, ..., x^{ej}, ..., x^{eJ})$ contains $J$ features and $\mathbf{y}^e = (y^{e1}, y^{e2}, ..., y^{ek}, ..., y^{eK})$ contains $K$ targets. The construction process of a CART-based MTR tree requires finding the *best split* pair $(j^*, w^*_{j*})$, $w^*_{j*} \in \Omega_{j*}$ of the nodes that minimizes the total within-subset variation in the two successive nodes when splitting. Denote the set of $I$ termination conditions as $\Gamma = (\gamma_1, \gamma_2, \gamma_3, ..., \gamma_I)$. The main steps to construct an MTR tree are presented as shown in Procedure 1 in Appendix A.

In our problem, we choose the 15 features in Table 2 as $\mathbf{x}$ ($J = 15$) and the number of deficiencies in each category as $\mathbf{y}$ ($K = 4$). For each ship in record $e = 1, ..., N$, the input features are $\mathbf{x}^e = (x^{e1}, x^{e2}, ..., x^{ej}, ..., x^{eJ})$. Because we have several machine learning models, in this model we represent the targets by $\boldsymbol{\alpha}^e = (\alpha^{e1}, \alpha^{e2}, \alpha^{e3}, \alpha^{eC})$ instead of using

$\mathbf{y}$, where $\alpha^{ec}$ is the number of deficiencies in category $c$, $c=1,\ldots,C$ (we use $C$ to represent the number of deficiency categories instead of using $K$) of ship $e$, and then

$$(j^*, w^*_{j*}) \in \underset{\substack{j\in(x_1,\ldots,x_J)\\ w_j\in\Omega_j}}{\arg\min}\left[\sum_{e\in R_1(j,w_j)}\sum_{c=1}^{C}\left(\alpha^{ec}-\frac{1}{|R_1(j,w_j)|}\sum_{e_1\in R_1(j,w_j)}\alpha^{e_1c}\right)^2+\sum_{e\in R_2(j,w_j)}\sum_{c=1}^{C}\left(\alpha^{ec}-\frac{1}{|R_2(j,w_j)|}\sum_{e_2\in R_2(j,w_j)}\alpha^{e_2c}\right)^2\right] \quad (1)$$

where $R_1(j,w_j)=\{e\in R_0 \mid x^{ej}\le w_j\}$ and $R_2(j,w_j)=\{e\in R_0 \mid x^{ej}> w_j\}$.

#### 4.1.1.2 Random forest

Like traditional DTs, the MTR trees can also be ensembled by using bagging (Breiman, 1996) and bootstrapping (Breiman, 2001) to reduce overfitting and increase prediction accuracy. In this study, we adopt random forest (which is based on bagging) to ensemble MTR trees proposed in Section 4.1.1.1. Compared to a single decision tree, the decision trees contained in the RF have two layers of randomness: a new training set generated by bootstrapping (i.e. randomly selecting a certain number of samples from the whole dataset with replacement) in the original training set is used to construct each decision tree, and a subset (with a preset fixed size) of all features is used to split each node in each decision tree (Friedman et al., 2001). A detailed construction procedure of MTR tree based random forest (MTR-RF) model is provided in Appendix B (Breiman, 2001; Kocev et al., 2007).

### 4.1.2 Optimization model

Among all the foreign ships visiting the port, the ships to be inspected are selected based on guidelines provided by the Tokyo MoU (2018). For each ship $s\in S$ selected to be inspected, we can only obtain its input features, while the number of deficiencies under deficiency category $c$ is unknown. With a little abuse of notation, we denote the unknown number of deficiencies in category $c$ of ship $s$ by $\alpha^{sc}$, $s\in S$, $c=1,2,3,C$ ($\alpha^{ec}$ is the known number of deficiencies in category $c$ of ship $e$ in the training set, $e=1,\ldots,N$). The predicted values of $\alpha^{sc}$, denoted by $\hat{\alpha}^{sc}$, $s\in S$, $c=1,2,3,C$, can be obtained by using the RF model proposed in Section 4.1.1.2. To achieve the maximum inspection efficiency, the sum of the product of the estimated deficiency number of each deficiency category and the corresponding inspection expertise of that deficiency category of the assigned PSCO (denoted by “inspection expertise” for short) should be as large as possible. The justification for matching the deficiency categories with the expertise of PSCOs is as follows. The decision (outcome) of a PSC inspection contains ship deficiency (specific deficiency types and total deficiency number) and ship detention. During a PSC inspection, the PSCO gets onboard and inspect the condition of the ship. For any condition that is not in compliance with the related regulations and conventions, it will be recorded as a deficiency. On the contrary, ship detention is not directly observed; instead, it is determined by the detected deficiencies and the PSCOs' judgement. Therefore, if the deficiency condition of the ships can be matched with the expertise of the PSCOs, the most proper PSCO (who can identify the existing deficiencies as many as possible and make rational detention decision) can be assigned to inspect the ship for better ship deficiency identification and detention decision making. Following this idea, we define binary decision variable $z_{ps}$ that is set to 1 if PSCO $p$ is assigned to inspect ship $s$ and 0, otherwise, and the PSCO assignment model can be expressed by mathematical model M1.

[M1]

$$\max \sum_{p=1}^{P} \sum_{s \in S} \sum_{c=1}^{C} \hat{\alpha}^{sc} u_{pc} z_{ps} \tag{2}$$

subject to

$$\sum_{s \in S} z_{ps} \leq \Theta, \ p = 1, \ldots, P \tag{3}$$

$$\sum_{p=1}^{P} z_{ps} = 1, \ s \in S \tag{4}$$

$$z_{ps} \in \{0,1\}, \ p = 1, \ldots P, \ s \in S. \tag{5}$$

Objective (2) maximizes the inspection expertise of the PSCOs by maximizing the sum of the product of the estimated deficiency number under each deficiency category and the expertise of the selected PSCO for that corresponding deficiency category for all inspected ships. Constraints (3) limit the maximum number of ships that can be inspected by a PSCO for one day. Constraints (4) guarantee that each ship is inspected by one PSCO. Constraints (5) ensure the domain of the decision variable.

Although model M1 is an integer program, it has the following nice property, whose proof is in Appendix C.

**Proposition 1**: Model [M1] can be solved in polynomial time of the length of the input parameters.

Proposition 1 implies that the PSCO assignment model [M1] is an easy problem: even if there are hundreds of ships and tens of PSCOs, [M1] can be solved efficiently (e.g., in less than 1 second).

## 4.2 Prediction of coefficients in the objective function of optimization model

### 4.2.1 Prediction model

In model M1, the coefficients of the decision variables in the objective function are

$\sum_{c=1}^{C}\hat{\alpha}^{sc}u_{pc}$ , $s \in S$ and $p = 1,...,P$ . Therefore, instead of predicting $\alpha^{sc}$ (i.e. the number of deficiencies in category $c$ for ship $s$), we can directly predict $\sum_{c=1}^{C}\alpha^{sc}u_{pc}$ (i.e. the total number of deficiencies of ship $s$ that can be detected by PSCO $p$ ). Define $\beta^{sp} = \sum_{c=1}^{C}\alpha^{sc}u_{pc}$ , $s \in S$ and $p = 1,...,P$ . For ship $s$ , $\boldsymbol{\beta}^{s} = (\beta^{s1},...,\beta^{sp},...,\beta^{sP})$ denotes the number of deficiencies that can be detected by assigning PSCO $p = 1,...,P$ . The values for $\beta^{sp}$ (and thus $\boldsymbol{\beta}^{s}$ ) can be predicted by using the RF models developed in Section 4.1.1.2, and the prediction model is denoted by MTR-RF2. The predicted values generated by MTR-RF2 are denoted by $\hat{\beta}^{sp}$ . The procedure of constructing the MTR trees $f'^{MTR}(\mathbf{x})$ in MTR-RF2 is slightly different from the $f^{MTR}(\mathbf{x})$ in MTR-RF1: a data entry $(\mathbf{x}^{e}, \boldsymbol{\beta}^{e})$ represents ship $s$, where the input features are $\mathbf{x}^{e} = (x^{e1}, x^{e2},..., x^{ej},..., x^{eJ})$ , $J = 15$ , and the targets are $\boldsymbol{\beta}^{e} = (\beta^{e1},...,\beta^{ep},...,\beta^{eP})$ . Both MTR-RF1 and MTR-RF2 generate multi-dimensional targets: MTR-RF1 has $C$ targets for the deficiency numbers under $C$ deficiency categories while MTR-RF2 has $P$ targets for the deficiency numbers identified by the $P$ PSCOs. In particular, the choice of the best split in Step 1 in Procedure 1 for constructing an MTR tree should be revised as

$$(j^{*}, w^{*}_{j*}) \in \underset{\substack{j \in (x_1,...,x_J) \\ w_j \in \Omega_j}}{\arg\min} \left[ \sum_{e \in R_1(j,w_j)} \sum_{p=1}^{P} (\beta^{ep} - \frac{1}{|R_1(j,w_j)|} \sum_{e_1 \in R_1(j,w_j)} \beta^{e_1 p})^2 + \sum_{e \in R_2(j,w_j)} \sum_{p=1}^{P} (\beta^{ep} - \frac{1}{|R_2(j,w_j)|} \sum_{e_2 \in R_2(j,w_j)} \beta^{e_2 p})^2 \right] \quad (6)$$

where $R_1(j, w_j) = \{e \in R_0 \mid x^{ej} \le w_j\}$ and $R_2(j, w_j) = \{e \in R_0 \mid x^{ej} > w_j\}$ .

### 4.2.2 Optimization model

Based on the predicted values $\hat{\beta}^{sp}$ , optimization model M1 can be reformulated as

[M2]

$$\max \sum_{p=1}^{P}\sum_{s \in S}\hat{\beta}^{sp}z_{ps} \quad (7)$$

subject to constraints (3) to (5). The structure of [M2] is the same as that of [M1] and hence [M2] can also be solved as a linear program.

### 4.3 Prediction of fundamental parameters that are fed into the optimization model

It is common that to predict the values $\hat{\beta}^{sp}$ in [M2], we try to minimize the sum of squared errors between the predicted value and the actual value, as shown in Eq. (6). However, a closer examination into the structure of the optimization model [M2] reveals that if the predicted number of deficiencies the $P$ PSCOs can identify for a ship are overestimated or underestimated by the same value, the final optimal assignment decision will not be influenced. We use the following example to illustrate this finding:

**Example**: For any ship that is selected to be inspected, if the actual numbers of deficiencies four PSCOs can identify are 6, 7, 8, and 9, but the predicted numbers are 8, 9, 10, and 11 (i.e. all four outputs are overestimated by "2"), then the optimal assignment is not changed and we should assign PSCO 4 to inspect the ship. If the predicted number are 5, 6, 7, 8 (i.e. all four outputs are underestimated by "1"), then the optimal assignment is also not changed and we should assign PSCO 4 to inspect the ship.

Generally, if the actual numbers of deficiencies the $P$ PSCOs can identify for a ship are $n_1$, $n_2$, …, $n_P$, but the predicted numbers are $n_1+\varepsilon$, $n_2+\varepsilon$,…, $n_P+\varepsilon$ ($\varepsilon \in R$; if $\varepsilon < 0$, $|\varepsilon| \le \min(n_1, n_2, ..., n_P)$), then the resulting prediction errors do *not* adversely affect the PSCO assignment decision, because it is the *difference* in the predicted numbers among the PSCOs, rather than the absolute prediction values, that affects the assignment decision. Based on this observation, the third approach (denoted by MTR-RF3) minimizes the squared difference regarding the overestimates (i.e., predicted value minus actual value) in the predicted numbers of deficiencies among the PSCOs and then uses the prediction in a PSCO assignment formulation (model M2 in Section 4). The prediction model is revised as follows.

Decision trees contained in MTR-RF3 is denoted by $f''^{MTR}(\mathbf{x})$. Splitting criterion of $f''^{MTR}(\mathbf{x})$ is changed to minimize the sum of variance of the predicted deficiencies that can be detected by each PSCO for each ship. More specifically, in Procedure 1, the best split pair $(j^*, w^*_{j*})$ of the current splitting node is calculated by

$$(j^*, w^*_{j*}) \in \mathop{\arg\min}_{\substack{j \in (x_1,\ldots,x_J) \\ w_j \in \Omega_j}} \left[ \begin{array}{l} \sum\limits_{e \in R_1(j,s_j)} \sum\limits_{p=1}^{P-1} \sum\limits_{p'=p+1}^{P} ((\beta^{ep} - \frac{1}{|R_1(j,w_j)|} \sum\limits_{e_1 \in R_1(j,w_j)} \beta^{e_1 p}) - (\beta^{ep'} - \frac{1}{|R_1(j,w_j)|} \sum\limits_{e_1 \in R_1(j,w_j)} \beta^{e_1 p'}))^2 + \\ \sum\limits_{e \in R_2(j,s_j)} \sum\limits_{p=1}^{P-1} \sum\limits_{p'=p+1}^{P} ((\beta^{ep} - \frac{1}{|R_2(j,w_j)|} \sum\limits_{e_2 \in R_2(j,w_j)} \beta^{e_2 p}) - (\beta^{ep'} - \frac{1}{|R_2(j,w_j)|} \sum\limits_{e_2 \in R_2(j,w_j)} \beta^{e_2 p'}))^2 \end{array} \right]. \quad (8)$$

The predicted numbers of deficiencies that can be detected by each PSCO given by

MTR-RF3 based on Eq. (8) are then input to optimization model M2 to generate PSCO assignment decisions.

# 5. Computational experiments

## 5.1 Construction of MTR-RF

### 5.1.1 Introduction of hyperparameters in RF

A hyperparameter in machine learning is a parameter used to control the learning process and whose value is set before the learning process begins. As RF is an ensembled machine learning model which contains DTs as weak learners, an RF model has hyperparameters to control the overall structure and properties of the RF as well as those for its DTs. Hyperparameters for DTs are mainly used to control the complexity and serve as the regularization of the model. The hyperparameters for RF are summarized below.

(a) n_estimators: the total number of DTs contained in an RF model. As the main principle underlying bagging is that more trees are better while too few trees can lead to unstable performance, this hyperparameter should be set to the largest computationally manageable value and do not need to be tuned (Breiman, 2001; Probst and Boulesteix, 2017).

(b) max_features: the number of features considered for each split. The value range of this hyperparameter is from 1 to the total number of features in the dataset and it is an integer. Too small value will negatively affect the average performance of the trees, while too large value will reduce the randomness of each tree and thus badly influence the overall performance. Denote the total number of features as n_features, $\text{n\_features} = J = 15$. It is suggested setting $\text{max\_features} = \lfloor \text{n\_features}/3 \rfloor$ for regression trees (Friedman et al., 2001; Probst et al., 2019).

(c) max_depth: the maximum depth of each DT in the RF model. The depth of a leaf is the number of splits taken from the root node to that leaf node (Elmachtoub et al., 2020). The value range of this hyperparameter can be set from one to unlimited and it is an integer. Larger value of max_depth leads to more complex single trees.

(d) min_samples_leaf: the minimum number of examples required to be at a leaf node. The minimum value for this hyperparameter is 1 and it is an integer. Smaller value of min_samples_leaf leads to more complex single trees. It is recommended to set the value of min_samples_leaf to be 5 for regression models by default (Friedman et al., 2001).

Hyperparameters (a) and (b) control the overall structure and the property of randomness of RF, while hyperparameters (c) and (d) are related to each DT. It should also be noted that in practice the best values for these parameters will depend on the problem, and should be treated as tuning parameter (Friedman et al., 2001).

### 5.1.2 Hyperparameter tuning in RF

Hyperparameters can have a large impact on model performance and generalization ability. Although it has been proved that RF models will not overfit, several studies have shown that tuning the hyperparameters in RF would yield slightly better performance and generalization ability (Biau and Scornet, 2016; Probst et al., 2018). In this study, we aim to tune three hyperparameters: max_features, max_depth, and min_samples_leaf which can only take integer values by using a training set and a validation set. We choose MSE as the performance evaluation measure for MTR-RF1 and MTR-RF2 and MSO in the predicted ship deficiency number that can be identified among the PSCOs as the performance evaluation measure for MTR-RF3. To tune the three hyperparameters, we propose a revised grid search method. Denote the pre-defined set containing all the possible values for a hyperparameter as its constrained value space. Unlike the classical grid search which exhaustively considers all hyperparameter combinations in the constrained value spaces to form the grid, the revised grid search method could gradually reduce the search space by iteration. The procedure to tune the hyperparameters by the revised grid search is presented in Appendix D. In this study, the default value for max_features should be 5 (recall that we have 15 input features) and min_samples_leaf should be 5. To form the constrained value space, we extend the value spaces of the two hyperparameters by increasing/decreasing the default value to the same extent, i.e. we set the constrained value space for max_features as $\{3,4,5,6,7\}$ and for min_samples_leaf as $\{2,3,4,5,6,7,8\}$ . For the constrained value space of max_depth, as there is no recommended default value, we set it to be a moderate range as $\{4,5,6,7,8\}$.

## 5.2 Performance of the MTR-RF models and PSCO assignment schemes

### 5.2.1 Experiment settings and hyperparameters in MTR-RF

The settings in the numerical experiments are in accordance with the real situation at the Hong Kong port: there are 4 available PSCOs, and about 10 ships are selected for inspection every day with 2 to 3 ships assigned to one PSCO. We further assume that PSCO 1 is good at dealing with deficiency category C1, PSCO 2 is good at dealing with deficiency category C2, PSCO 3 is good at dealing with deficiency categories C3 and C4, and PSCO 4 is good at dealing with deficiency categories C4. The assumed expertise of each PSCO to inspect each deficiency category is presented in Table 4. After applying the revised grid search method to the three hyperparameters under the given constrained value spaces in MTR-RF1, MTR-RF2, and MTR-RF3, the best hyperparameter tuples for the three models are shown in Table 5.

Table 4. Expertise of each PSCO in each deficiency category

| PSCO/deficiency category | C1 | C2 | C3 | C4 |
|---|---|---|---|---|
| PSCO 1 | 0.8 | 0.5 | 0.7 | 0.6 |
| PSCO 2 | 0.7 | 0.9 | 0.4 | 0.5 |
| PSCO 3 | 0.7 | 0.6 | 0.8 | 0.7 |
| PSCO 4 | 0.4 | 0.7 | 0.6 | 0.7 |

Table 5. Best hyperparameter tuples for MTR-RF1, MTR-RF2, and MTR-RF3

| Model | max_features | max_depth | min_samples_leaf |
|---|---|---|---|
| MTR-RF1 | 4 | 8 | 5 |
| MTR-RF2 | 4 | 7 | 3 |
| MTR-RF3 | 6 | 8 | 4 |

After finding the optimal hyperparameter tuple for each model by using the training set and the validation set, in the following experiments we form a new training set by combining the current training and validation sets, and thus it contains 1,700 inspection records at the Hong Kong port. The test set contains another 300 inspection records at the Hong Kong port. We randomly and evenly divide them into 30 groups where each group contains 10 ships. We assume that the 10 ships in a group come to the port on one day and the totally 300 ships come to the port on 30 days. We also require that the maximum number of ships that can be inspected by one PSCO is three.

### 5.2.2 Performance of the three MTR-RF models

We set n_estimators = 200 for the proposed three MTR-RF models. Each MTR-RF model is trained by using the new training set and the hyperparameter tuple tuned by the revised grid search. Run each of the MTR-RF model 10 times, and the min, max, mean, and variance of MSE/MSO on the test set in the 10 runs for the three models are shown in Table 6. It can be seen that the min, mean, and max values of MSE of MTR-RF1 are all much smaller than those in MTR-RF2. The differences are caused by the values of the prediction targets in the MTR-RF models: in MTR-RF1, the prediction targets are the deficiency number under each deficiency category; while in MTR-RF2, the prediction targets are the *total* number of deficiencies a PSCO can detect if she/he is assigned to inspect the ship. Besides, it is shown that the min, mean, and max values of MSE of MTR-RF2 are all smaller than those of MTR-RF3, which indicates that MTR-RF2 performs better than MTR-RF3 as a regression model evaluated by MSE. The differences in MSE between MTR-RF2 and MTR-RF3 are caused by the property of the MTR-RF models: the splitting criteria in MTR-RF2 is to reduce the MSE in successive nodes while those in MTR-RF3 is to reduce MSO. In addition, the variance in each model is small, which implies that the performance of MTR-RF containing 200 MTR trees is stable.

Table 6. Prediction performance of MTR-RF1, MTR-RF2, and MTR-RF3

| Model | Metric | Min | Mean | Max | Variance |
|---|---|---|---|---|---|
| MTR-RF1 | MSE | 3.9756 | 4.0173 | 4.0762 | 0.0009 |
| MTR-RF2 | MSE | 15.4953 | 15.8342 | 16.1237 | 0.0437 |
| MTR-RF3 | MSE | 16.7775 | 17.1684 | 17.5571 | 0.0443 |
| MTR-RF3 | MSO | 3.0242 | 3.0513 | 3.0863 | 0.0002 |

Table 6 shows that compared to the prediction outputs of MTR-RF3, the outputs of MTR-RF2 have less variability. Meanwhile, even if the differences in the prediction targets of MTR-RF1 and MTR-RF2 are considered, the variability of MTR-RF1 is less than MTR-RF2. The reasons are as follows. For the difference between the variance of MTR-RF2 and MTR-RF3, the splitting criterion of the DTs is to minimize the MSE of ship deficiency number detected by each PSCO in MTR-RF2, whereas the splitting criterion of the DTs in MTR-RF3 is to minimize the MSO of ship deficiency number detected by each PSCO. Therefore, the target of the prediction generated by MTR-RF2 is to make the outputs as close as to their real values, while the target of the prediction generated by MTR-RF3 is to make the differences of the overestimates of each two of the outputs as small as possible. As a result, MTR-RF3 generates more flexible prediction results and when evaluating the variance of the outputs, the variance of the outputs of MTR-RF3 is larger than MTR-RF2. For the difference between the variance of MTR-RF1 and MTR-RF2, recall that the prediction targets of MTR-RF1 only consider the deficiency number under each deficiency category while both the deficiency number and the PSCOs' expertise in each deficiency category are considered in the prediction targets of MTR-RF2. Due to the nonlinearity of MTR-RF models, the impacts of the PSCOs' expertise on the deficiency number in the outputs variability can be magnified. As a result, the uncertainties are propagated to the output predictions, which leads to higher variance in MTR-RF2 compared to MTR-RF1.

### 5.2.3 Performance of the combined prediction and optimization model

We assign PSCOs based on the prediction results (10 runs) in Section 5.2.1 to the 30 groups of ships in accordance with the settings. The assignment decisions generated by assignment models based on MTR-RF1, MTR-RF2, and MTR-RF3 are denoted by A1, A2, and A3, respectively. Apart from making comparisons among the three models themselves, we also compare them with random assignment scheme and best assignment scheme in theory. The performance of random assignment scheme is the mean inspection expertise of 10,000 times of random PSCO assignment. The best assignment scheme in theory is making PSCO assignment decisions under the assumption that there is a perfect machine learning model that could predict the parameters for the optimization model totally accurate. However, this is an ideal situation that never exists because the generalization error cannot be zero. The comparison results are shown in Table 7. We further analyze the randomness of A1, A2, and A3 by calculating the min and max values of the inspection expertise and the variance of inspection expertise among the 30 groups of PSCO assignment. The results are presented in Table 8.

Table 7. Mean inspection expertise of the three models

| Group | Inspection expertise of random PSCOs assignment | Mean inspection expertise of A1 | Mean inspection expertise of A2 | Mean inspection expertise of A3 | Best inspection expertise in theory |
|---|---|---|---|---|---|
| 1 | 80.56 | 84.45 | 84.00 | 86.55 | 89.40 |
| 2 | 44.03 | 45.67 | 46.42 | 46.18 | 48.30 |
| 3 | 49.10 | 51.20 | 51.17 | 51.08 | 52.40 |
| 4 | 39.24 | 39.27 | 39.65 | 39.00 | 43.00 |
| 5 | 34.11 | 37.10 | 37.14 | 36.72 | 38.20 |
| 6 | 25.36 | 26.86 | 26.72 | 27.13 | 28.30 |
| 7 | 48.15 | 50.83 | 50.92 | 50.81 | 51.90 |
| 8 | 61.70 | 64.10 | 63.97 | 64.21 | 67.10 |
| 9 | 32.41 | 34.35 | 34.12 | 34.42 | 35.40 |
| 10 | 20.22 | 20.80 | 20.74 | 21.23 | 23.10 |
| 11 | 60.19 | 64.00 | 64.20 | 64.19 | 65.30 |
| 12 | 33.69 | 35.11 | 35.11 | 34.84 | 36.80 |
| 13 | 33.84 | 34.48 | 34.43 | 34.88 | 37.90 |
| 14 | 37.98 | 38.60 | 38.14 | 38.20 | 41.60 |
| 15 | 22.63 | 24.43 | 24.71 | 24.52 | 25.70 |
| 16 | 63.36 | 66.45 | 66.22 | 66.88 | 69.50 |
| 17 | 27.67 | 29.08 | 28.19 | 29.14 | 30.40 |
| 18 | 38.16 | 38.82 | 38.56 | 38.60 | 40.80 |
| 19 | 31.75 | 34.76 | 34.95 | 33.69 | 36.00 |
| 20 | 44.36 | 47.58 | 47.85 | 47.67 | 49.50 |
| 21 | 32.82 | 34.26 | 34.33 | 34.55 | 36.20 |
| 22 | 31.22 | 34.19 | 33.80 | 34.27 | 35.10 |
| 23 | 29.67 | 31.37 | 31.40 | 31.49 | 34.30 |
| 24 | 33.42 | 34.09 | 33.99 | 34.36 | 37.00 |
| 25 | 51.16 | 52.93 | 53.00 | 53.24 | 55.70 |
| 26 | 23.23 | 25.18 | 24.99 | 24.07 | 26.60 |
| 27 | 25.88 | 26.50 | 26.65 | 26.62 | 28.70 |
| 28 | 60.52 | 61.99 | 62.02 | 62.28 | 65.60 |
| 29 | 22.75 | 24.97 | 24.55 | 25.22 | 26.00 |
| 30 | 27.88 | 28.50 | 28.15 | 27.57 | 31.10 |
| **Average** | **38.90** | **40.73** | **40.67** | **40.79** | **42.90** |
| **Ratio*** | **(90.68%)** | **(94.94%)** | **(94.80%)** | **(95.08%)** | **(100%)** |

Note*: calculated by $\frac{\textit{Average of mean inspection expertise}}{\textit{The best inspection expertise in theory}} \times 100\%$ .

Table 8. Randomness of model performance

| | Min inspection expertise | | | Max inspection expertise | | | Variance of inspection expertise | | |
|---|---|---|---|---|---|---|---|---|---|
| Group/ inspection scheme | A1 | A2 | A3 | A1 | A2 | A3 | A1 | A2 | A3 |
| 1 | 83.4 | 83.7 | 83.4 | 86.3 | 85.7 | 87.1 | 1.4745 | 0.3600 | 1.1745 |
| 2 | 44.3 | 45.2 | 45.6 | 46.5 | 47.2 | 46.5 | 0.6261 | 0.2456 | 0.0876 |
| 3 | 51.2 | 50.9 | 50.3 | 51.2 | 51.2 | 51.2 | 0.0000 | 0.0081 | 0.0756 |
| 4 | 38.7 | 39.0 | 38.9 | 40.2 | 40.7 | 39.3 | 0.2361 | 0.1925 | 0.0120 |
| 5 | 37.1 | 37.1 | 35.1 | 37.1 | 37.5 | 37.1 | 0.0000 | 0.0144 | 0.3636 |
| 6 | 26.3 | 25.7 | 26.7 | 27.7 | 27.7 | 28.0 | 0.2844 | 0.4556 | 0.1641 |
| 7 | 50.7 | 50.8 | 50.8 | 50.9 | 51.2 | 50.9 | 0.0081 | 0.0096 | 0.0009 |
| 8 | 63.1 | 63.1 | 63.1 | 65.0 | 65.0 | 65.0 | 0.4400 | 0.2181 | 0.4089 |
| 9 | 33.6 | 33.8 | 34.0 | 34.8 | 34.8 | 34.8 | 0.2005 | 0.1216 | 0.0456 |
| 10 | 20.3 | 20.0 | 20.7 | 20.9 | 21.3 | 21.4 | 0.0300 | 0.1544 | 0.0621 |
| 11 | 64.0 | 64.0 | 64.0 | 64.0 | 64.7 | 64.7 | 0.0000 | 0.0940 | 0.0849 |
| 12 | 35.1 | 33.9 | 33.8 | 35.2 | 35.7 | 35.8 | 0.0009 | 0.1989 | 0.5864 |
| 13 | 33.2 | 33.2 | 34.2 | 35.7 | 35.5 | 35.3 | 0.5896 | 0.8121 | 0.0876 |
| 14 | 38.0 | 37.6 | 37.6 | 39.6 | 38.3 | 39.6 | 0.4440 | 0.0504 | 0.5740 |
| 15 | 23.9 | 23.9 | 24.2 | 24.8 | 25.2 | 25.4 | 0.0561 | 0.1049 | 0.1196 |
| 16 | 65.4 | 65.5 | 66.7 | 67.2 | 66.6 | 66.9 | 0.4145 | 0.1416 | 0.0036 |
| 17 | 28.1 | 27.8 | 28.1 | 29.4 | 28.8 | 29.4 | 0.2716 | 0.1029 | 0.2704 |
| 18 | 38.7 | 38.3 | 37.9 | 38.9 | 38.9 | 38.9 | 0.0096 | 0.0404 | 0.2100 |
| 19 | 33.6 | 33.6 | 32.6 | 35.6 | 35.6 | 34.3 | 0.5244 | 0.6585 | 0.2909 |
| 20 | 47.2 | 47.5 | 47.2 | 48.0 | 48.0 | 48.0 | 0.0996 | 0.0525 | 0.0801 |
| 21 | 33.8 | 33.8 | 33.8 | 34.7 | 34.8 | 35.1 | 0.1304 | 0.1161 | 0.1885 |
| 22 | 33.4 | 33.4 | 33.4 | 34.6 | 34.6 | 35.1 | 0.2189 | 0.1800 | 0.2261 |
| 23 | 30.5 | 30.3 | 31.4 | 31.6 | 31.6 | 31.6 | 0.0921 | 0.1420 | 0.0029 |
| 24 | 33.8 | 33.8 | 33.9 | 35.4 | 34.3 | 35.7 | 0.2069 | 0.0289 | 0.3444 |
| 25 | 52.7 | 53.0 | 52.7 | 53.2 | 53.0 | 53.5 | 0.0261 | 0.0000 | 0.0444 |
| 26 | 24.6 | 24.3 | 23.6 | 25.3 | 25.3 | 24.8 | 0.0596 | 0.1509 | 0.1821 |
| 27 | 25.7 | 26.2 | 25.9 | 26.7 | 26.7 | 26.8 | 0.1100 | 0.0225 | 0.0636 |
| 28 | 61.3 | 61.3 | 61.3 | 62.7 | 63.0 | 63.5 | 0.2109 | 0.2316 | 0.6176 |
| 29 | 24.5 | 24.1 | 25.0 | 25.2 | 25.0 | 25.3 | 0.0421 | 0.0565 | 0.0136 |
| 30 | 27.1 | 26.8 | 27.1 | 30.3 | 30.3 | 29.6 | 1.9640 | 1.3205 | 0.8461 |
| **Average** | **40.11** | **40.05** | **40.10** | **41.29** | **41.27** | **41.35** | **0.2924** | **0.2095** | **0.2411** |

Table 7 shows that on average, all the three models can realize about 95% of the best inspection expertise in theory on average while A3 has the best performance regarding mean inspection expertise. Table 8 indicates that the performance of A1, A2, and A3 are stable. We can draw the following conclusions:

(a) The performance of all the three newly proposed PSCO assignment schemes is stable and is much better than the performance of random PSCO assignment. This shows that the PSCO assignment schemes generated by combining MTR-RF models with PSCO assignment models are efficient compared with the currently used random PSCO assignment at the port states.

(b) The performance of A1 is better than A2, although they both use MSE as the splitting criterion for constructing the MTR-RFs. The difference between A1 and A2 is that they have different prediction targets. The prediction targets in A1 are the deficiency numbers under each deficiency category which are natural choices. Meanwhile, the prediction targets in A2 are the deficiency numbers that can be identified by each PSCO which also considers the expertise of PSCs and is determined by the parameters of the following optimization model. The difference in performance of A1 and A2 indicates that although different targets can be chosen for a combined prediction and optimization model, their performance can be divergent.

(c) The performance of A3 is better than A2, although MTR-RF3 performs much worse as a regression model than MTR-RF2 if evaluated by MSE. This indicates that when combining machine learning model (e.g. decision tree and random forest) with optimization model, the choices for prediction targets, the properties of machine learning model (e.g. splitting criteria in decision trees), and model evaluation metrics can be varying. High-quality decisions are based on either precise prediction generated by the machine learning model or combination of the structure and property of the optimization model with the machine learning model.

(d) Table 8 indicates that A2 has the least variance while A1 has the largest variance in the total inspection expertise generated by the optimal assignment among A1, A2, and A3. The possible reasons are as follows. Although the splitting criterion of MTR-RF2 and that of MTF-RF3 is different, the outputs of MTR-RF2 and MTR-RF3 can serve as the parameters of the decision variables in the optimization model of A2 and A3 directly. On the other hand, the outputs of MTR-RF1 need to be further combined with the expertise of the PSCOs to serve as the parameters of the decision variables in the optimization model of A1. The further processing might magnify the variability of the total inspection expertise in the final optimal assignment decision, which leads to highest variance of A1 compared to A2 and A3. Although MTR-RF3 predicts the deficiency number detected by each PSCO like MTR-RF2, the splitting criteria in A3 is not relevant to the values of the prediction targets directly like that in MTR-RF2. Therefore, MTR-RF3 has larger variance in the outputs compared to MTR-RF2 as shown in Table 6. When combining the prediction results as the input with the optimization models, the variability can be magnified. Therefore, A3 has the larger variance in the total inspection expertise generated by the optimal assignment decision compared to A2.

An illustration of insights of the superiority of A3 is presented in Appendix E. We also present the detailed inspection expertise under each deficiency category of A1, A2, and A3 as shown in Table 9.

Table 9. Inspection expertise under each deficiency category

| Method/ deficiency category | C1: ship safety | C2: ship management | C3: ship condition and structure | C4: communication and navigation |
|---|---|---|---|---|
| Original test set | 630 | 478 | 289 | 407 |
| Best in theory | 459.9 | 356.5 | 209.5 | 261 |
| A1 | 447.24 (97.25%) | 309.03 (86.68%) | 201.81 (96.33%) | 263.84 (101.09%) |
| A2 | 446.15 (97.01%) | 309.10 (86.70%) | 201.25 (96.06%) | 263.59 (100.99%) |
| A3 | 446.72 (97.13%) | 316.39 (88.75%) | 198.89 (94.94%) | 261.61 (100.23%) |

It can be seen from Table 9 that A1, A2, and A3 can achieve more than 85% of the inspection expertise compared to the best situation in theory. Especially, except for C2: ship management, more than 95% of the best inspection expertise in theory can be achieved by the three combined prediction and assignment models. The results further indicate that all the three models that match PSCOs' inspection expertise with ship deficiency condition are effective and accurate.

### 5.3 Comparison with other state-of-the-art prediction models

In this section, comparisons of the proposed tree-based prediction models with other state-of-the-art and popular prediction models are made. We select three machine learning models for prediction: ridge regression, the least absolute shrinkage and selection operator (LASSO) regression, and support vector regression (SVR) for comparison. Their performance of predicting the deficiency number under each deficiency category is presented in Section 5.3.1 and the total inspection expertise realized when combining with assignment models is presented in Section 5.3.2.

#### 5.3.1 Regression performance

All the three models are implemented by using scikit-learn in Python with the hyperparameter tuples tuned by grid search on the validation set. Like the experiments in Section 5.2, we also run the three models 10 times with the optimal hyperparameter tuples. Their performance is shown in Table 10.

Table 10. Prediction model performance

| Model | Metric | Min | Mean | Max | Variance |
|---|---|---|---|---|---|
| MTR-RF1 | MSE | 3.9756 | 4.0173 | 4.0762 | 0.0009 |
| MTR-RF2 | MSE | 15.4953 | 15.8342 | 16.1237 | 0.0437 |
| MTR-RF3 | MSE | 16.7775 | 17.1684 | 17.5571 | 0.0443 |
| Ridge regression | MSE | 15.9990 | 15.9990 | 15.9990 | 0 |
| LASSO regression | MSE | 25.0756 | 25.0756 | 25.0756 | 0 |
| SVR | MSE | 26.0432 | 26.0432 | 26.0432 | 0 |

Table 10 indicates that the mean MSE of the outputs of ridge regression is smaller than that of MTR-RF3, while the mean MSE of the outputs of LASSO regression and SVR is bigger than that of MTR-RF2 and MTR-RF3. Besides, the outputs of the ridge regression, LASSO regression, and SVR are determined once the hyperparameters of the three models are fixed, therefore their performance is quite stable. While in the tree-based models, randomness in the outputs can still exist even if the hyperparameters are given.

### 5.3.2 PSCO assignment performance

We combine the prediction results generated by the three regression models with optimization model M2 and make comparison with A3 regarding the total inspection expertise, as A3 has the best performance in PSCO assignment among the proposed models. The assignment decision generated by combining ridge regression with M2, LASSO regression with M2, and SVR with M2 are denoted by A4, A5, and A6, respectively. Comparison results over the 30 groups of ships based on 10 runs are shown in Table 11. Randomness of model performance is resented in Table 12.

Table 11. Comparison of PSCO assignment model performance

| | Random assignment | A3 (MTR-RF3+M2) | A4 (ridge+M2) | A5 (LASSO+M2) | A6 (SVR+M2) | Best in theory |
|---|---|---|---|---|---|---|
| Average | 38.90 | **40.79** | 40.59 | 40.36 | 40.48 | 42.90 |
| Ratio* | 90.68% | **95.08%** | 94.62% | 94.08% | 94.36% | 100% |

Note*: calculated by $\frac{\textit{Average of mean inspection expertise}}{\textit{The best inspection expertise in theory}}\times 100\%$

Table 11 shows that A3 achieves the highest inspection expertise among A3, A4, A5, and A6, which indicates the superiority of the combined tree-based prediction model with the structure of optimization model.

## 5.4 Model extension

In the current prediction and assignment models, the importance of the four deficiency categories is viewed as identical. Nevertheless, their importance can be different under certain situations, e.g. in the concentrated inspection campaign (CIC) where deficiencies in some categories should be paid more attention to in PSC inspections. To extend our models to deal with the situations where the importance of the deficiency categories is different, we attach each deficiency category with a relative importance score, which is denoted by $w_c, c = 1, 2, 3, C$ and is no less than 1. The larger the value is, the more important the deficiency category is. In the current model, $w_c = 1, c = 1, 2, 3, C$ . For mathematical model M1, we can combine the importance score directly in Equation (2) by revising the objective function to be $\max \sum_{p=1}^{P}\sum_{s\in S}\sum_{c=1}^{C} w_c \hat{\alpha}^{sc} u_{pc} z_{ps}$ , while the prediction model MTF-RF1 needs not be revised. Then, we denote $\sum_{c=1}^{C} w_c \hat{\alpha}^{sc} u_{pc} = \lambda^{sp}$ , which can be viewed as the weighted total deficiency number identified by PSCO $p = 1, ..., P$ of ship $s \in S$ and can be predicted by using the MTR-RF models developed in Section 4.1.1.2. The predicted values for $\lambda^{sp}$ are denoted by

$\hat{\lambda}^{sp}$, and the objective function of mathematical model M2 can be revised to $\max\sum_{p=1}^{P}\sum_{s\in S}\hat{\lambda}^{sp}z_{ps}$. Especially, the total inspection expertise generated by the three models where the differences in the deficiency category importance are considered is denoted by A1', A2', and A3' respectively.

We use an example to illustrate the working process and results of the proposed models where C1: ship safety is more important than the other deficiency categories. The relative importance score can be assigned by the ports in practice. In this example, we assume that $w_1 = 1.5$ and $w_c = 1, c = 2,3,C$. Mean inspection expertise of the three models is presented in Table 13. The performance of random assignment scheme is the mean inspection expertise of 10,000 times of random PSCO assignment. The inspection expertise under each deficiency category is shown in Table 14.

Table 13. Mean inspection expertise of the three models
(considering deficiency category importance)

| Group | Inspection expertise of random PSCOs assignment | Mean inspection expertise of A1’ | Mean inspection expertise of A2’ | Mean inspection expertise of A3’ | Best inspection expertise in theory |
|---|---|---|---|---|---|
| 1 | 80.56 | 83.68 | 82.72 | 83.77 | 89.40 |
| 2 | 44.03 | 46.60 | 46.29 | 46.54 | 48.30 |
| 3 | 49.10 | 51.47 | 51.41 | 51.31 | 52.40 |
| 4 | 39.24 | 39.44 | 39.29 | 38.86 | 43.00 |
| 5 | 34.11 | 37.24 | 37.26 | 37.17 | 38.20 |
| 6 | 25.36 | 27.23 | 27.02 | 26.66 | 28.30 |
| 7 | 48.15 | 50.93 | 51.10 | 50.96 | 51.90 |
| 8 | 61.70 | 64.45 | 64.41 | 64.23 | 67.10 |
| 9 | 32.41 | 34.60 | 34.75 | 34.68 | 35.40 |
| 10 | 20.22 | 19.79 | 19.85 | 20.09 | 23.10 |
| 11 | 60.19 | 63.87 | 63.93 | 63.81 | 65.30 |
| 12 | 33.69 | 34.56 | 34.44 | 34.76 | 36.80 |
| 13 | 33.84 | 34.31 | 34.60 | 34.11 | 37.90 |
| 14 | 37.98 | 39.56 | 39.20 | 39.70 | 41.60 |
| 15 | 22.63 | 24.23 | 25.03 | 24.72 | 25.70 |
| 16 | 63.36 | 66.24 | 66.60 | 67.34 | 69.50 |
| 17 | 27.67 | 28.65 | 28.62 | 29.01 | 30.40 |
| 18 | 38.16 | 38.90 | 39.10 | 38.47 | 40.80 |
| 19 | 31.75 | 33.84 | 33.77 | 33.06 | 36.00 |
| 20 | 44.36 | 47.10 | 47.41 | 47.00 | 49.50 |
| 21 | 32.82 | 34.13 | 34.00 | 34.06 | 36.20 |
| 22 | 31.22 | 33.73 | 33.64 | 34.45 | 35.10 |
| 23 | 29.67 | 31.33 | 31.07 | 31.02 | 34.30 |
| 24 | 33.42 | 33.87 | 34.52 | 34.29 | 37.00 |
| 25 | 51.16 | 53.21 | 53.04 | 53.39 | 55.70 |
| 26 | 23.23 | 24.16 | 24.48 | 24.63 | 26.60 |
| 27 | 25.88 | 26.70 | 26.74 | 26.70 | 28.70 |
| 28 | 60.52 | 62.15 | 62.44 | 63.17 | 65.60 |
| 29 | 22.75 | 23.84 | 23.66 | 24.02 | 26.00 |
| 30 | 27.88 | 29.73 | 29.73 | 29.41 | 31.10 |
| **Average** | **38.90** | **40.65** | **40.67** | **40.71** | **42.90** |
| **Ratio*** | **(90.68%)** | **(94.76%)** | **(94.80%)** | **(94.90%)** | **(100%)** |

Note*: calculated by $\frac{Average\ of\ \ mean\ inspection\ expertise}{The\ best\ inspection\ expertise\ in\ theory}\times 100\%$ .

Table 14. Inspection expertise under each deficiency category

| Method/ deficiency category | C1: ship safety | C2: ship management | C3: ship condition and structure | C4: communication and navigation |
|---|---|---|---|---|
| Original test set | 630 | 478 | 289 | 407 |
| Best in theory | 459.9 | 356.5 | 209.5 | 261 |
| A1 | 447.24 (97.25%) | 309.03 (86.68%) | 201.81 (96.33%) | 263.84 (101.09%) |
| A1’ | 449.58 (97.76%) | 303.20 (85.05%) | 204.27 (97.50%) | 262.49 (100.57%) |
| A2 | 446.15 (97.01%) | 309.10 (86.70%) | 201.25 (96.06%) | 263.59 (100.99%) |
| A2’ | 449.68 (97.78%) | 303.30 (85.08%) | 203.57 (97.17%) | 263.57 (100.98%) |
| A3 | 446.72 (97.13%) | 316.39 (88.75%) | 198.89 (94.94%) | 261.61 (100.23%) |
| A3’ | 451.95 (98.27%) | 307.29 (86.20%) | 201.22 (96.05%) | 260.93 (99.97%) |

Table 13 shows that if different weights of deficiency categories are taken into account, the total inspection expertise achieved by the three inspection strategies is no larger than the situation when the deficiency categories are of the same importance. Moreover, if C1 is regarded to be more important and is attached with a larger importance score, the realized inspection expertise under C1 increases in all the three inspection strategies as presented in Table 14.

## 5.5 Sensitivity analysis

In this section, we analyze how the distribution of the expertise of PSCOs would influence the performance of the proposed PSCO assignment models. To be concise, the sensitivity analysis is conducted on A3 as it achieves the maximum mean inspection expertise among A1, A2 and A3. Four groups of sensitivity analyses (SA) are performed: SA1: composition of a group of PSCOs; SA2: divergence in expertise of a PSCO; SA3: adequacy of PSCO resources; SA4: uncertainty in PSCOs’ expertise.

### 5.5.1 SA1: composition of a group of PSCOs

First, we analyze how the composition of a group of PSCOs would influence the results. Suppose there are five groups of PSCOs (denoted by SA1G1 to SA1G5, respectively) of the same total expertise and different expertise distributions while one PSCO has the same expertise to inspect the four deficiency categories. Groups SA1G1 to SA1G4 contain PSCOs with various expertise, i.e. some of them are experienced while some are newcomers. More specifically, the variations of the expertise of each PSCO are increasing from SA1G1 to SA1G4. On the contrary, the four PSCOs in SA1G5 have the same expertise. The expertise of each PSCO to each deficiency category of the five groups is shown in Table 15. The analysis results of SA1 are shown in Figure 1.

Table 15. Expertise PSCOs in SA1

| SA1G1 | C1 | C2 | C3 | C4 | SA1G 2 | C1 | C2 | C3 | C4 |
|---|---|---|---|---|---|---|---|---|---|
| PSCO 1 | 0.775 | 0.775 | 0.775 | 0.775 | PSCO 1 | 0.85 | 0.85 | 0.85 | 0.85 |
| PSCO 2 | 0.725 | 0.725 | 0.725 | 0.725 | PSCO 2 | 0.75 | 0.75 | 0.75 | 0.75 |
| PSCO 3 | 0.675 | 0.675 | 0.675 | 0.675 | PSCO 3 | 0.65 | 0.65 | 0.65 | 0.65 |
| PSCO 4 | 0.625 | 0.625 | 0.625 | 0.625 | PSCO 4 | 0.55 | 0.55 | 0.55 | 0.55 |
| **SA1G3** | C1 | C2 | C3 | C4 | **SA1G4** | C1 | C2 | C3 | C4 |
| PSCO 1 | 0.925 | 0.925 | 0.925 | 0.925 | PSCO 1 | 1.0 | 1.0 | 1.0 | 1.0 |
| PSCO 2 | 0.775 | 0.775 | 0.775 | 0.775 | PSCO 2 | 0.8 | 0.8 | 0.8 | 0.8 |
| PSCO 3 | 0.625 | 0.625 | 0.625 | 0.625 | PSCO 3 | 0.6 | 0.6 | 0.6 | 0.6 |
| PSCO 4 | 0.475 | 0.475 | 0.475 | 0.475 | PSCO 4 | 0.4 | 0.4 | 0.4 | 0.4 |
| **SA1G5** | C1 | C2 | C3 | C4 | | | | | |
| PSCO 1 | 0.7 | 0.7 | 0.7 | 0.7 | | | | | |
| PSCO 2 | 0.7 | 0.7 | 0.7 | 0.7 | | | | | |
| PSCO 3 | 0.7 | 0.7 | 0.7 | 0.7 | | | | | |
| PSCO 4 | 0.7 | 0.7 | 0.7 | 0.7 | | | | | |

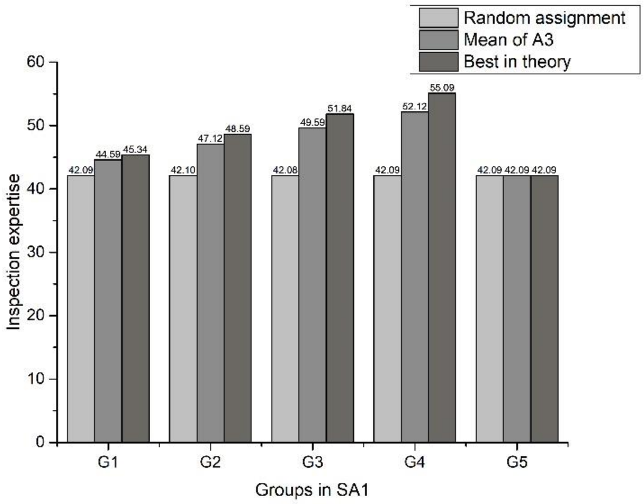


Figure 1. Analysis results of SA1

Several conclusions can be drawn from Figure 1. First, as the divergence of expertise of the group of PSCOs become larger, both the best inspection expertise in theory and the mean inspection expertise achieved by using A3 increase, as the diverse conditions of the inspected ships can be better matched with the more varied expertise of the group of PSCOs. Second, the superiority of the PSCO assignment scheme generated by A3 over random PSCO assignment becomes more obvious when the inspection expertise of the PSCOs gets more diverse. Third, the mean inspection expertise achieved by A3 is equal to the best inspection expertise in theory when all the PSCOs have the same expertise. However, as the expertise of the group of PSCOs gets more varied, the gap between mean inspection expertise and the best inspection expertise in theory gets larger. This indicates that predicting errors of the prediction model (i.e. MTR-RF3) have a larger influence on the final assignment model when the expertise of PSCOs becomes more diverse, as the assignment scheme relies more on the predicted number of deficiencies of a ship that can be identified if assigned to a PSCO.

The extreme situation is that when all the PSCOs have the same expertise, the mean inspection expertise achieved by A3 equals the best inspection expertise in theory and random PSCO assignment, as the PSCO assignment is totally random under this condition and has nothing to do with the prediction results of MTR-RF3. Nevertheless, it should also be noted that even in SA1G4, where the expertise of the PSCOs are most varied, the mean inspection expertise is approaching 95% of the best inspection expertise in theory, and the PSCO assignment performance of A3 is 24% better than random PSCO assignment. The results indicate that our model is more suitable to be applied than random PSCO assignment scheme when the expertise of the PSCOs is divergent. When the expertise of all the PSCOs is the same, our model is equal to random PSCO assignment.

### 5.5.2 SA2: divergence in expertise of a PSCO

Second, we analyze how various expertise of a PSCO in different deficiency categories would influence the results. We consider four groups of PSCOs (denoted by SA2G1 to SA2G4, respectively) where the total expertise is the same for each PSCO and the total expertise to inspect one deficiency category is the same for each group (i.e. the sum of each row and the sum of each column are the same in all groups). In SA2G1 to SA2G3, the PSCOs have different expertise to inspect different deficiency categories, while in SA2G4, all PSCOs have the same expertise in different deficiency categories. More specifically, the variations of the PSCOs are increasing from SA2G1 to SA2G3: the sum of absolute variations of all PSCOs in SA2G1, SA2G2 and SA2G3 is 1.8, 2.2 and 2.6, respectively. The expertise of each PSCO in each deficiency category of the four groups is shown in Table 16. The results of the analyses are presented in Figure 2.

Table 16. Expertise of PSCOs in SA2

| **SA2G1** | C1 | C2 | C3 | C4 | **SA2G2** | C1 | C2 | C3 | C4 |
|---|---|---|---|---|---|---|---|---|---|
| PSCO 1 | 0.9 | 0.8 | 0.6 | 0.5 | PSCO 1 | 0.8 | 0.5 | 1.0 | 0.5 |
| PSCO 2 | 0.6 | 0.8 | 0.7 | 0.7 | PSCO 2 | 0.9 | 0.8 | 0.5 | 0.6 |
| PSCO 3 | 0.5 | 0.6 | 0.9 | 0.8 | PSCO 3 | 0.6 | 0.7 | 0.6 | 0.9 |
| PSCO 4 | 0.8 | 0.6 | 0.6 | 0.8 | PSCO 4 | 0.5 | 0.8 | 0.7 | 0.8 |
| **SA2G3** | C1 | C2 | C3 | C4 | **SA1G4** | C1 | C2 | C3 | C4 |
| PSCO 1 | 0.8 | 0.9 | 0.7 | 0.4 | PSCO 1 | 0.7 | 0.7 | 0.7 | 0.7 |
| PSCO 2 | 0.8 | 0.4 | 0.8 | 0.8 | PSCO 2 | 0.7 | 0.7 | 0.7 | 0.7 |
| PSCO 3 | 0.8 | 0.5 | 0.8 | 0.7 | PSCO 3 | 0.7 | 0.7 | 0.7 | 0.7 |
| PSCO 4 | 0.4 | 1.0 | 0.5 | 0.9 | PSCO 4 | 0.7 | 0.7 | 0.7 | 0.7 |

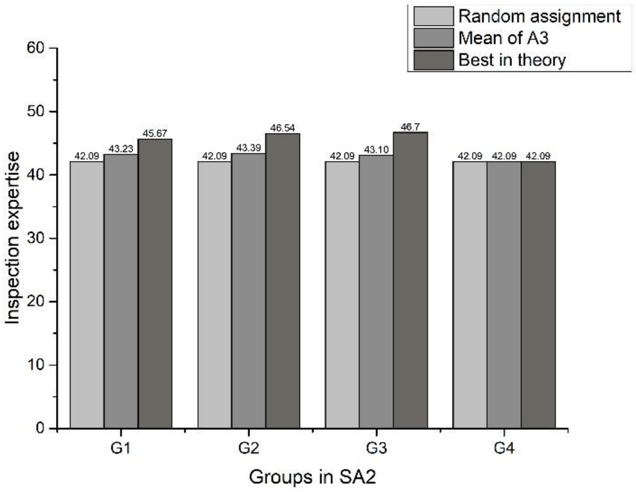


Figure 2. Analysis results of SA2

As shown in Figure 2, when the total expertise of each PSCO is the same and the total expertise to inspect one deficiency category for each group is the same, the best inspection expertise in theory shows gentle increase when the divergence of the PSCOs' expertise increases. Nevertheless, due to the randomness in the dataset and the model performance, the predicted mean inspection expertise does not show this trend: when the variations in the expertise of the PSCOs increase, the predicted mean inspection expertise can either increase or decrease modestly.

### 5.5.3 SA3: adequacy of PSCO resources

Third, we analyze the influence of the adequacy of PSCO resources on the inspection results. In our problem, four PSCOs are assigned to inspect 10 ships coming to the port state every day (benchmark, denoted by SA4G3). The maximum number of ships that can be inspected by one PSCO is three. We consider other situations where there are 8 (SA3G1), 9 (SA3G2), 11 (SA3G4), and 12 (SA3G5) ships coming to the port state every day while keeping the other settings unchanged and compare the mean inspection expertise of a single ship to identify the influence of PSCO resources. The results are shown in Figure 3.

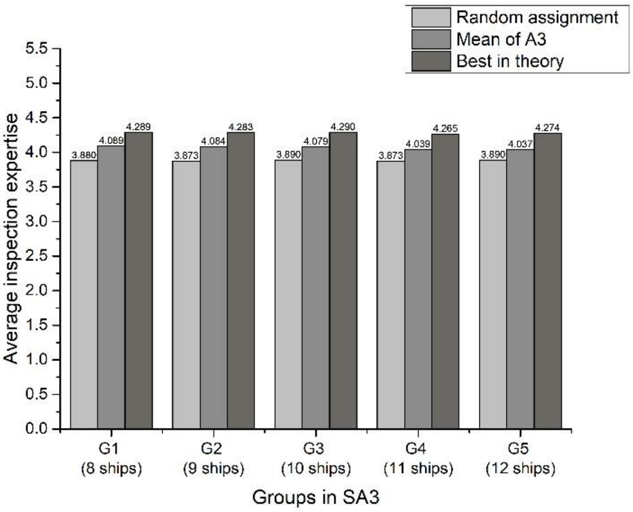


Figure 3. Analysis results of SA3

As shown in Figure 3, when the number of ships in a group grows while the number of PSCOs and the maximum number of ships can be inspected by one PSCO remain unchanged, the average inspection expertise of one ship remains stable. This indicates that the performance of the proposed models is not heavily influenced by the adequacy of the resources of PSCOs, which also shows that the model performs robustly. Besides, our model performs much better than random PSCO assignment in all situations.

### 5.5.4 SA4: uncertainty in PSCOs' expertise

Fourth, we analyze the uncertainty in PSCO expertise in each deficiency category. Although the expertise of PSCOs could be measured by tests, interviews, and questionnaires, uncertainties can exist, which means that the expertise we obtained may not be the exact expertise in reality. The expertise values presented in Table 4 are the measured inspection expertise and we suppose that the real inspection expertise is within 10% more or less than the measured inspection expertise. For example, the expertise of PSCO 1 for deficiency category C1 is 0.8, and we suppose that the real inspection expertise is uniformly distributed from 0.72 to 0.88 (accurate to two digits). We randomly select a value within this interval for each inspection expertise value and form a new expertise table of each PSCO in each deficiency category for ten times, and we can obtain ten possible real inspection expertise tables. The inspection expertise of random PSCO assignment, the best inspection in theory, and the mean inspection expertise of the ten groups are shown in Figure 4.

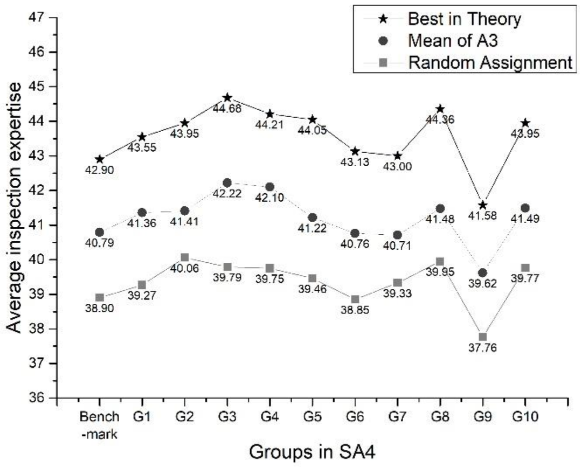


Figure 4. Analysis results of SA4

Under the assumption that the real inspection expertise of each PSCO to each deficiency category is within 10% more or less than the measured inspection expertise presented in Table 4, the variance of the best inspection expertise in theory of the 10 groups in SA4 is 0.8298. The range of the best inspection expertise in theory is 3.1 (the maximum inspection expertise of the 10 groups is 44.68 and the minimum inspection expertise is 41.58). Compared with the benchmark, which is generated by using the measured inspection expertise shown in Table 4, the differences are between −3.08% and +4.15% and are much smaller than ±10%.

As for the predicted inspection expertise, the variance of the 10 groups in SA4 is 0.4999, and the range of mean inspection expertise is 2.6 (the maximum inspection expertise of the 10 groups is 42.22 and the minimum inspection expertise is 39.62). The differences between the benchmark and the 10 groups range from −2.87% to +3.51% and are also much smaller than ±10%. We also compare the differences between predicted mean inspection expertise of the benchmark with the best inspection expertise of the 10 groups in SA4. The difference is from +1.94% to +9.54%, which indicates that the proposed models are robust even if there are some uncertainties in measuring the inspection expertise of each PSCO to each deficiency category. The average best inspection expertise of the 10 groups is 43.646 and the average predicted inspection expertise of the 10 groups is 41.237, which indicates that the proposed model can identify about 94.5% of the total deficiencies and that it always performs better than random PSCO assignment in all groups of tests as shown in the two lower lines of Figure 4.

## 6. Discussion and future research

As indicated in Section 5.2.1, the inspection expertise of each PSCO in each deficiency category shown in Table 4 is assumed by the authors as there is no such standard tests or questionnaires at the moment in the Tokyo MoU. The assumption of the inspection expertise table is that each PSCO has more expertise in one or two deficiency categories than the other PSCOs and we just use the assumed inspection expertise to illustrate the working process of the proposed models. Although massive sensitivity analysis has been conducted to evaluate the performance of the proposed models, in future research, accurate assessments would be developed to evaluate the real expertise of the PSCOs for different deficiency categories. For example, a test consisting of a theoretical part and a practical part of all the four deficiency categories, or an interview regarding the background, experience, and self-evaluation of the PSCOs, or a questionnaire for collecting the PSCOs' own preference and expertise can be held. The results of the test, interview, and questionnaire can be considered simultaneously to comprehensively evaluate the inspection expertise of the PSCOs. For the convenience of MoU management, we propose another way to evaluate the performance of the PSCOs. Suppose that there are several PSCOs at a port, and we let them to inspect a group of ships (say 10 ships or 20 ships) in a certain amount of time. Then, we compare the total number of detected deficiencies under each deficiency category of the PSCOs regarding all the ships. For the PSCO(s) who can identify the most deficiency number of a category, we denote her/his expertise to be "1". The expertise of the other PSCOs regarding this deficiency category is calculated by dividing her/his number of detected deficiencies in this category by the largest number of detected deficiencies of this category of all PSCOs. For example, the detected deficiency number for deficiency category C1 is 20, 18, 16, and 14 for PSCO1, PSCO2, PSCO3, and PSCO4, and their inspection expertise for C1 should be 1, 0.9, 0.8, and 0.7, respectively. In this way, evaluating and updating the inspection expertise of the PSCOs can be more convenient for the ports.

Another thing should be noted is that the expertise of the PSCOs can be updated over time and experience. Therefore, reevaluations should be carried out for updating the expertise of the PSCOs. We suggest that the reevaluation to be carried out once a year for the following reasons. First, the committee meeting of Tokyo MoU is held once a year. In the committee meeting, several important discussions and decisions are made, such as the application for co-operating member status, actions relating to harmonization of PSC, and approval of the final report of the concentrated inspection campaign (CIC) in the 29th committee meeting in 2018 (Tokyo MoU, 2018b). Therefore, it should be a good chance to discuss the details of reevaluations at the committee meeting. Second, setting the time interval of two reevaluations to be one year is a result of a trade-off: for one thing, the inspection expertise for the PSCOs could remain unchanged for only a period of time; for another, it can be time-consuming to prepare for the reevaluations.

Given the fact that the inspection expertise of the PSCOs can improve over time and experience and it is also a goal to improve the PSCOs' inspection expertise to be as close as possible to 1, we propose two ways to achieve the comprehensive development of the PSCO if the proposed models are applied. First, except for the assigned PSCO who is responsible to conduct the PSC inspection, other PSCOs can get onboard during the PSC inspections to learn from the PSCO with more inspection expertise in certain deficiency categories to achieve self-improvement. Second, during the regular trainings and seminars, the PSCOs can share their experience and expertise as well as discuss the difficulties they meet during the inspections with each other to achieve co-operation and progress.

As the main goal of PSC is to identify substandard ships and detain them if necessary to protect the maritime safety and protect the marine environment, ship detention probability can also be incorporated in the prediction and assignment models in the future research for better applicability and practicability. Meanwhile, PSCOs' expertise in targeting ships with high detention probability should also be evaluated and considered.

## 7. Conclusion

Maritime safety and the marine environmental protection are gaining increasing concern in recent years. PSC inspection is a widely-believed effective and efficient way to safeguard the sea. To improve the efficiency of PSC inspections, one of the key points is to identify as many deficiencies as possible using limited inspection resources. At the ports with less experienced and divergent PSCOs, this requires matching the PSCOs of different expertise and the deficiency conditions of the inspected ships, e.g. the deficiency number under each deficiency category. To achieve this goal, this paper proposes three machine learning models: MTR-RF1, MTR-RF2, and MTR-RF3 and two PSCO assignment models M1 and M2 to match the diverse ship deficiency conditions with the expertise of PSCOs. More specifically, MTR-RF1 predicts the number of deficiencies in each deficiency category for each ship in a way that minimizes the MSE between actual and predicted numbers of deficiencies; MTR-RF2 predicts the number of deficiencies each PSCO can identify for each ship by minimizing the MSE between actual and predicted deficiency numbers; MTR-RF3 predicts the number of deficiencies each PSCO can identify for each ship while adopting a loss function motivated by the structure of the optimization problem, i.e. minimizing the MSO in the numbers of deficiencies that can be detected among the PSCOs for each ship. Numerical experiments show that the performance of combination of MTR-RF3 and M2 (i.e. A3) is the best among the three proposed models, while all the three models perform much better than the currently used random PSCO assignment as they can identify about 95% of all the deficiencies compared to the best inspection expertise in theory.

By conducting sensitivity analyses, several managerial insights can be drawn. First, our model is more suitable to be applied when the expertise of the PSCOs is divergent as the superiority of the proposed models becomes more obvious when the divergence of the PSCOs increases. Second, the adequacy of the PSCO resources would not heavily influence the performance of the proposed models. Besides, even if uncertainty may exist in measuring the expertise of the PSCOs to each deficiency category, the robustness of our model is validated.

In this study, both prediction and optimization are required to generate the decision for PSCO assignment. Meanwhile, both prediction and optimization are challenging tasks, as errors cannot be avoided in the prediction problem, while the unknown parameters in the optimization model are determined by the outputs of the prediction model. For A1 and A2, the machine learning models for parameter prediction totally ignore the downstream optimization problem and only aim to minimize the prediction error which is evaluated by MSE. Although the objective is to make the predicted outputs as close as to the real outputs, inaccuracy always exists, and thus minimizing the output error cannot guarantee the best decision in theory generated by the following optimization model or generate the decision as close as to the best decision in theory. Moreover, inaccuracy in the predicted results is highly likely to be magnified when combining with the downstream optimization model and thus make the final generated decision far away from the best decision in theory. On the contrary, MSO used in MTR-RF3 (and thus in A3) is highly related to the structure and property of the downstream optimization model, as the prediction model is designed to generate outputs that make the generated decisions of the following optimization model as close as to the best decision in theory by aiming to maintain the property of the parameters in the optimization model for generating the best decision in theory.

Theoretically, the proposed MTR-RF1 and MTR-RF2 treat prediction and optimization models as sequential steps while the proposed MTR-RF3 partially combines prediction and optimization models by considering the structure and property of the optimization model when constructing the machine learning model. The numerical experiments show that although MTR-RF3 performs much worse than MTR-RF2 as a regression model evaluated by the metric of MSE, the performance of MTR-RF3 is better than MTR-RF2 when combining with the following optimization models.

Practically, the proposed models help to address a meaningful practical problem in PSC inspection. Compared with random assignment of PSCOs, the proposed three models can help to detect 4.70%, 4.55%, and 4.86% more deficiencies after inspecting the same groups of ships by using the same PSCO recourses. Meanwhile, the performance of the three models are stable and their performance would achieve 95% of the best inspection expertise in theory.

## Appendix A.

*Procedure 1: Construction of MTR tree*

*Input* Training dataset $D$ and termination conditions $\Gamma = (\gamma_1, \gamma_2, \gamma_3, ..., \gamma_I)$.

*Output* MTR tree $f^{MTR}(\mathbf{x})$: for a new example with input features $\mathbf{x}$, the predicted targets are $f^{MTR}(\mathbf{x})$.

*Step* 0 Construct a root node that contains all the examples in the training dataset (the set of indices for the examples in the root node is denoted by $\{1, ..., N\}$). The root is set as the current splitting node.

*Step* 1 Define $R_0$ as the set of indices for the examples in the current splitting node. Find the *best split* pair $(j^*, w^*_{j*})$ of the current splitting node by enumerating of all possible values of $j$ and $w_j$.

$$(j^*, w^*_{j*}) \in \mathop{\arg\min}_{\substack{j \in (x_1, ..., x_J) \\ w_j \in \Omega_j}} \left[ \sum_{e \in R_1(j, w_j)} \sum_{k=1}^{K} \left( y^{ek} - \frac{1}{|R_1(j, w_j)|} \sum_{e_1 \in R_1(j, w_j)} y^{e_1 k} \right)^2 + \sum_{e \in R_2(j, w_j)} \sum_{k=1}^{K} \left( y^{ek} - \frac{1}{|R_2(j, w_j)|} \sum_{e_2 \in R_2(j, w_j)} y^{e_2 k} \right)^2 \right]$$

where $R_1(j, w_j) = \{e \in R_0 \mid x^{ej} \le w_j\}$ and $R_2(j, w_j) = \{e \in R_0 \mid x^{ej} > w_j\}$.

*Step* 2 Use the *best split* $(j^*, w^*_{j*})$ to split the current node into two nodes that contain two sub-sets of indices of examples $R_1(j^*, w^*_{j*}) = \{e \in R_0 \mid x^{ej*} \le w^*_{j*}\}$ and $R_2(j^*, w^*_{j*}) = \{e \in R_0 \mid x^{ej*} > w^*_{j*}\}$ with output value for target $y_k$ as $\omega_1^k = \frac{1}{|R_1(j^*, w^*_{j*})|} \sum_{e_1 \in R_1(j^*, w^*_{j*})} y^{e_1 k}$ and $\omega_2^k = \frac{1}{|R_2(j^*, w^*_{j*})|} \sum_{e_2 \in R_2(j^*, w^*_{j*})} y^{e_2 k}$, respectively, $k = 1, ..., K$.

*Step* 3 Repeat *Step* 1 and *Step* 2 in a depth-first manner until coming to a node that reaches one of the preset termination conditions. Then, this node becomes a leaf node and a new node for splitting is found by backtracking.

*Step* 4 Repeat *Step* 3 until there are no more nodes that can be split. Finally, the total training set is separated into $Q$ mutually exclusive and collectively exhaustive leaf sub-sets $R_1, R_2, ..., R_Q$ according to the input variable vector. The decision tree model can be presented by

$f^{MTR}(\mathbf{x}) = \sum_{q=1}^{Q} I(\mathbf{x} \in R_q)(\omega_q^1, \omega_q^2, ..., \omega_q^K)$, where $I(\mathbf{x} \in R_q) = \begin{cases} 1, \mathbf{x} \in R_q \\ 0, \mathbf{x} \notin R_q \end{cases}$.

In Step 1, we have a tree that may not be completed yet, denoted by $T$, and one of its leaves is the current splitting node. If the current splitting node is split at $(j, w_j)$, we will have a new tree, denoted by $T_{j,w_j}$, which has two new leaves (the left leave, or called leave 1, and the right leave, or called leave 2) with sets of indices for the examples $R_1(j,w_j)=\{e\in R_0\,|\,x^{ej}\le w_j\}$ and $R_2(j,w_j)=\{e\in R_0\,|\,x^{ej}>w_j\}$. The predicted value of the $k$th target for an example $e$ in leave 1 ( $e\in R_1(j,w_j)$ ) is $\frac{1}{|R_1(j,w_j)|}\sum_{e_1\in R_1(j,w_j)} y^{e_1k}$, which is the average value of the $k$th target among all the examples in leave 1. Therefore, it can be seen that Step 1 chooses the best split $(j^*,w^*_{j*})$ that minimizes the sum of squared error.

## Appendix B.

*Procedure 2: Construction of MTR-RF*

*Input* Training dataset $D$, termination conditions $\Gamma = (\gamma_1, \gamma_2, \gamma_3, ..., \gamma_I)$, the number of trees contained in the RF $M$, and the maximum number of features considered when splitting each node $J'$, $J' < J$.

*Output* MTR-RF $f^{MTR-RF}(\mathbf{x})$: for a new example with input feature $\mathbf{x}$, the predicted targets are $f^{MTR-RF}(\mathbf{x})$.

*Step* 1*:* Draw a bootstrap sample $D'$ of the whole training set $D$.

For $m = 1, ..., M$

*Step* 1.0 Construct a root node that contains all the examples in $D'$ (the set of indices for the examples in the root node is denoted by $\{1, ..., N\}$). The root is set as the current splitting node.

*Step* 1.1 Among all the $J$ features, randomly select $J'$ features with each selected feature denoted by $j'$. Define $R_0$ as the set of indices for the examples in the current splitting node. Find the *best split* pair $(j'^{*}, w_{j'*}^{'*})$ of the current splitting node by solving the following formula:

$$(j'^{*}, w_{j'*}^{'*}) \in \underset{\substack{j' \in (x_1, ..., x_{J'}) \\ w'_{j'} \in \Omega_{j'}}}{\arg\min} \left[ \sum_{e \in R_1(j', w'_{j'})} \sum_{k=1}^{K} \left( y^{ek} - \frac{1}{|R_1(j', w'_{j'})|} \sum_{e_1 \in R_1(j', w'_{j'})} y^{e_1 k} \right)^2 + \sum_{e \in R_2(j', w'_{j'})} \sum_{k=1}^{K} \left( y^{ek} - \frac{1}{|R_2(j', w'_{j'})|} \sum_{e_2 \in R_2(j', w'_{j'})} y^{e_2 k} \right)^2 \right]$$

where $R_1(j', w'_{j'}) = \{e \in R_0 \mid x^{ej'} \leq w'_{j'}\}$ and $R_2(j', w'_{j'}) = \{e \in R_0 \mid x^{ej'} > w'_{j'}\}$.

*Step* 1.2 Use the *best split* $(j'^{*}, w_{j'*}^{'*})$ to split the current node into two nodes that contain two sub-sets of indices of examples $R_1(j'^{*}, w_{j'*}^{'*}) = \{e \in R_0 \mid x^{ej'*} \leq w_{j'*}^{'*}\}$ and $R_2(j'^{*}, w_{j'*}^{'*}) = \{e \in R_0 \mid x^{ej'*} > w_{j'*}^{'*}\}$ with output value for target $y_k$ as $\omega_1^k = \frac{1}{|R_1(j'^{*}, w_{j'*}^{'*})|} \sum_{e_1 \in R_1(j^{*'}, w_{j'*}^{'*})} y^{e_1 k}$ and $c_2^k = \frac{1}{|R_2(j'^{*}, w_{j'*}^{'*})|} \sum_{e_2 \in R_2(j'^{*}, w'^{*}_{j'*})} y^{e_2 k}$, respectively, $k = 1, ..., K$.

*Step* 1.3 Repeat *Step* 1.1 and *Step* 1.2 in a depth-first manner until coming to a node that reaches one of the preset termination conditions. Then, this node becomes a leaf node and a new node for splitting is found by backtracking.

*Step* 1.4 Repeat *Step* 1.3 until there are no more nodes that can be split. Finally, the total training dataset is separated into $Q^m$ mutually exclusive and collectively exhaustive leaf sub-sets $R_1, R_2, ..., R_{Q^m}$ according to the input variable vector in decision tree $m$. The $m$th decision tree model can be presented by $f_m^{MTR}(\mathbf{x}) = \sum_{q=1}^{Q^m} I(\mathbf{x} \in R_q)(\omega_q^1, \omega_q^2, ..., \omega_q^K)$, where $I(\mathbf{x} \in R_q) = \begin{cases} 1, \mathbf{x} \in R_q \\ 0, \mathbf{x} \notin R_q \end{cases}$. For target $k = 1, ..., K$, the final predicted value generated for $\mathbf{x}$ by tree $m$ is represented by $\hat{y}_k^m$ for short.

*Step* 2*:* For $k = 1, ..., K$, the final predicted value generated by the RF model is the average regarding all the predicted values of the $M$ DTs, i.e. $\hat{y}_k = \frac{1}{M}\sum_{m=1}^{M} \hat{y}_k^m$. The MTR-RF model can be represented by $f^{MTR-RF}(\mathbf{x}) = (\hat{y}_1, ..., \hat{y}_k, ..., \hat{y}_K)$.

## Appendix C.

**Proof**: If $\Theta > |S|$, we can safely set $\Theta = |S|$ in model [M1] without losing optimality. Therefore, we can assume that $\Theta \leq |S|$. Since $P$ PSCOs can inspect a maximum of $\Theta P$ ships, we add $\Theta P - |S|$ dummy ships to the model, each of which has 0 deficiency in each category. Then model [M1] is equivalent to

[M1']

$$\max \sum_{p=1}^{P} \sum_{s=1}^{\Theta P} \sum_{c=1}^{C} z_{ps} \hat{\alpha}^{sc} u_{pc} \tag{9}$$

subject to

$$\sum_{s=1}^{\Theta P} z_{ps} = \Theta, \ p = 1, \ldots, P \tag{10}$$

$$\sum_{p=1}^{P} z_{ps} = 1, \ s = 1, \ldots, \Theta P \tag{11}$$

$$z_{ps} \in \{0,1\}, \ p = 1, \ldots, P, \ s = 1, \ldots, \Theta P \tag{12}$$

where parameters $\hat{\alpha}^{sc} = 0$, $s = |S| + 1, \ldots, \Theta P$, $c = 1, \ldots, C$.

Define decision vector $\mathrm{z} = (z_{ps}, \ p = 1, \ldots, P, \ s = 1, \ldots, \Theta P)$, parameter vector $\mathrm{b} = (\underbrace{\Theta, \ldots, \Theta}_{P \text{ elements}}, \underbrace{1, \ldots, 1}_{\Theta P \text{ elements}})$, and parameter matrix $A_{(P+\Theta P) \times \Theta P^2}$ that represents the coefficients in constraints (10) and (11). Defining $¢$ as the set of integers, model [M1'] can be written as

[M1"]

$$\max \sum_{p=1}^{P} \sum_{s=1}^{\Theta P} \sum_{c=1}^{C} z_{ps} \hat{\alpha}^{sc} u_{pc} \tag{13}$$

subject to

$$A\mathrm{z} = \mathrm{b} \tag{14}$$

$$0 \leq \mathrm{z} \leq 1 \tag{15}$$

$$\mathrm{z} \in ¢^{\Theta P^2}. \tag{16}$$

We can see that (i) all of the elements in $\mathrm{b}$ are integers, (ii) all of the elements in $A$ are 0 or 1, (iii) each column of matrix $A$ has exactly two elements whose values are 1, and (iv) matrix $A$ can be divided into two sub-matrices: the top $P$ rows constitute one matrix and the bottom $\Theta P$ rows constitute the other matrix, such that each sub-matrix has exactly one element of 1 in each column. Consequently, $A$ is totally unimodular and all the extreme points are optimal solutions to the linear programming relaxation of model [M1”] are integral. Hence, the integrality constraint in Eq. (16) can be dropped. In other words, model [M1”] can be easily solved as a linear programming problem.

Note that the conversion of model [M1] to model [M1”] is polynomial because $\Theta \leq |S|$. Since a linear program can be solved in polynomial time of the length of its input parameters, model [M1] can also be solved in polynomial time of the length of its input parameters. □

## Appendix D.

Denote the set of hyperparameters (i.e. max_features, max_depth, and min_samples_leaf) to be tuned as $\mathrm{K}=\{\kappa_1,\kappa_2,\kappa_3\}$ and one hyperparameter is denoted by $\kappa_i$. The minimum and maximum values each hyperparameter can take are denoted by $\kappa_i \in [\kappa_i^{\min},\kappa_i^{\max}]$, $\kappa_i \in \mathrm{K}$. The initial constrained value spaces are denoted by $R_i=\{\kappa_i^{\min},\lfloor(\kappa_i^{\min}+\kappa_i^{\max})/2\rfloor,\kappa_i^{\max}\}$, $\kappa_i \in \mathrm{K}$. The procedure to tune the hyperparameters by revised grid search is as follows:

*Procedure 3: Tuning hyperparameters by revised grid search*

```
Input    The set of hyperparameters to be tuned $\mathrm{K}=\{\kappa_1,\kappa_2,\kappa_3\}$, the minimum and
         maximum values each hyperparameter can take $\kappa_i \in [\kappa_i^{\min},\kappa_i^{\max}]$, $\kappa_i \in \mathrm{K}$, the
         initial constrained value spaces $R_i=\{\kappa_i^{\min},\lfloor(\kappa_i^{\min}+\kappa_i^{\max})/2\rfloor,\kappa_i^{\max}\}$ for all
         $\kappa_i \in \mathrm{K}$.
Output   Hyperparameter value tuple with the best performance on validation set denoted
         by $\psi^*$.
Step 1   Set the hyperparameter grid $\Psi$ to $\Psi = R_1 \times R_2 \times R_3$.
         for each $\psi \in \Psi$:
             Train MTR-RF model $f_\psi^{MTR\text{-}RF}(\mathbf{x})$ using the training set and hyperparameter
             tuple $\psi$. Measure its performance by calculating the MSE/MSO score $m_\psi$
             on the validation set.
Step 2   The hyperparameter tuple that yields minimum MSE/MSO score $m_\psi^*$ on the
         validation set is denoted by $\psi^*$, $\psi^*=\{\kappa_1^*,\kappa_2^*,\kappa_3^*\}$.
         if $\kappa_i^{\min}+2 \ge \kappa_i^{\max}$ for all $\kappa_i \in \mathrm{K}$:
             Return the optimal hyperparameter tuple $\psi^*=\{\kappa_1^*,\kappa_2^*,\kappa_3^*\}$ and terminate the
             program.
         else:
             for $\kappa_i \in \mathrm{K}$ with $\kappa_i^{\min}+2 < \kappa_i^{\max}$:
                 if $\kappa_i^* = \kappa_i^{\min}$:
                     set $\kappa_i^{\max} = \lfloor(\kappa_i^{\min}+\kappa_i^{\max})/2\rfloor - 1$, update
                     $R_i=\{\kappa_i^{\min},\lfloor(\kappa_i^{\min}+\kappa_i^{\max})/2\rfloor,\kappa_i^{\max}\}$.
```

```
        else if $\kappa_i^* = \kappa_i^{\max}$:
            set $\kappa_i^{\min} = \lfloor (\kappa_i^{\min} + \kappa_i^{\max})/2 \rfloor + 1$, update $R_i = \{\kappa_i^{\min}, \lfloor (\kappa_i^{\min} + \kappa_i^{\max})/2 \rfloor, \kappa_i^{\max}\}$.
        else:
            set $\kappa_i^{\min} = \lfloor (\kappa_i^{\min} + \kappa_i^*)/2 \rfloor$ and $\kappa_i^{\max} = \lfloor (\kappa_i^* + \kappa_i^{\max})/2 \rfloor$, update $R_i = \{\kappa_i^{\min}, \lfloor (\kappa_i^{\min} + \kappa_i^{\max})/2 \rfloor, \kappa_i^{\max}\}$.
Step 3    Repeat Step 1 and Step 2 until termination.
```

## Appendix E.

We use a randomly selected group of ships in the numerical experiment to illustrate the insights of the superiority of A3. The real inspection expertise of each PSCO to each ship (denoted by a ship-PSCO pair) in the selected group and the best PSC assignment in theory are presented in Table A.1. For simplicity, we only compare the performance of A2 (and MTR-RF2) and A3 (and MTR-RF3). The predicted inspection expertise of ship-PSCO pairs generated by MTR-RF2 and MTR-RF3 and the corresponding PSCO assignment are shown in Table A.2 and Table A.3.

Table A.1. Real inspection expertise and best PSCO assignment

| PSCO/Ship | PSCO 1 | PSCO 2 | PSCO 3 | PSCO 4 | Best PSCO assignment |
|---|---|---|---|---|---|
| **1** | 3.2 | 2.3 | 3.7 | 3.3 | 3 |
| **2** | 0.6 | 0.5 | 0.7 | 0.7 | 4 |
| **3** | 2.4 | 2.3 | 2.8 | 2.7 | 4 |
| **4** | 9.3 | 10.7 | 9.3 | 7.6 | 2 |
| **5** | 5.0 | 4.6 | 4.9 | 3.6 | 1 |
| **6** | 4.9 | 3.9 | 5.2 | 3.9 | 3 |
| **7** | 28.3 | 34.1 | 31.9 | 31.1 | 2 |
| **8** | 0.7 | 0.4 | 0.8 | 0.6 | 1 |
| **9** | 17.3 | 17.5 | 18.8 | 16.3 | 3 |
| **10** | 5.8 | 7.8 | 6.4 | 6.4 | 2 |
| Total inspection expertise | | | | | 89.4 |

Table A.2. Predicted inspection expertise and PSCO assignment of A2

| PSCO/Ship | PSCO 1 | PSCO 2 | PSCO 3 | PSCO 4 | Assigned PSCO |
|---|---|---|---|---|---|
| **1** | 2.71151 | 2.71333 | 2.83366 | 2.37304 | 2 |
| **2** | 2.19643 | 2.16080 | 2.27956 | 1.88287 | 1 |
| **3** | 1.42442 | 1.41698 | 1.47454 | 1.21828 | 4 |
| **4** | 3.16357 | 3.19042 | 3.25073 | 2.66992 | 2 |
| **5** | 2.94226 | 2.93073 | 3.06017 | 2.54735 | 1 |
| **6** | 3.46248 | 3.39896 | 3.59726 | 2.95980 | 3 |
| **7** | 21.94541 | 22.66446 | 23.59041 | 20.79491 | 3 |
| **8** | 3.30471 | 3.29913 | 3.40618 | 2.80313 | 1 |
| **9** | 6.76287 | 6.80433 | 6.97162 | 5.77351 | 3 |
| **10** | 4.58404 | 4.59511 | 4.72913 | 3.90545 | 2 |
| Total achieved inspection expertise | | | | | 85.7 |

Table A.3. Predicted inspection expertise and PSCO assignment of A3

| PSCO/Ship | PSCO 1 | PSCO 2 | PSCO 3 | PSCO 4 | Assigned PSCO |
|---|---|---|---|---|---|
| **1** | 2.28501 | 2.31229 | 2.38390 | 2.00269 | 2 |
| **2** | 2.07840 | 2.06494 | 2.16828 | 1.81122 | 1 |
| **3** | 1.73917 | 1.73968 | 1.81308 | 1.51692 | 4 |
| **4** | 3.23858 | 3.27633 | 3.33146 | 2.74851 | 2 |
| **5** | 2.63255 | 2.61521 | 2.71647 | 2.23666 | 1 |
| **6** | 3.28039 | 3.24911 | 3.40201 | 2.80859 | 3 |
| **7** | 15.68360 | 17.17298 | 16.95479 | 15.32269 | 2 |
| **8** | 3.65548 | 3.67663 | 3.77489 | 3.12439 | 1 |
| **9** | 7.47597 | 7.52486 | 7.73092 | 6.42257 | 3 |
| **10** | 4.61062 | 4.56776 | 4.74119 | 3.88015 | 3 |
| Total achieved inspection expertise | | | | | 86.5 |

Tables A.2 and A.3 show that the performance of A3 is better than A2 by 0.8 inspection expertise, while both A2 and A3 can achieve 95% of the total real inspection expertise. The main differences between A2 and A3 are that PSCO 3 is assigned to inspect ship 7 in A2 while PSCO 2 is assigned to inspect ship 7 in A3, whereas PSCO 2 is assigned to inspect ship 10 in A2 while PSCO 3 is assigned to inspect ship 10 in A3. Notably, assigning PSCO 2 to inspect ship 7 and PSCO 3 to inspect ship 10 could obtain more inspection expertise, as the difference between assigning PSCO 2 and PSCO 3 to ship 7 is 2.2 while the difference is 1.4 to ship 10. We further compare the squared error of the predicted inspection expertise of each ship-PSCO pair and the MSE score in MTR-RF2 and MTR-RF3 are shown Table A.4 and Table A.5. The squared overestimate of the predicted inspection expertise of each ship-PSCO pair and the MSO score in MTR-RF2 and MTR-RF3 are shown in Table A.6 and Table A.7.

Table A.4. Squared error of MTR-RF2

| PSCO/Ship | PSCO 1 | PSCO 2 | PSCO 3 | PSCO 4 |
|---|---|---|---|---|
| **1** | 0.23862 | 0.17084 | 0.75054 | 0.85925 |
| **2** | 2.54859 | 2.75824 | 2.49501 | 1.39917 |
| **3** | 0.95175 | 0.77972 | 1.75684 | 2.19551 |
| **4** | 37.65583 | 56.39386 | 36.59365 | 24.30566 |
| **5** | 4.23430 | 2.78646 | 3.38496 | 1.10807 |
| **6** | 2.06646 | 0.25104 | 2.56878 | 0.88398 |
| **7** | 40.38078 | 130.77153 | 69.04925 | 106.19480 |
| **8** | 6.78453 | 8.40497 | 6.79218 | 4.85378 |
| **9** | 111.03115 | 114.39728 | 139.91054 | 110.80693 |
| 10 | 1.47855 | 10.27133 | 2.79180 | 6.22278 |
| MSE (of each pair) | | | | 26.4820 |

Table A.5. Squared error of MTR-RF3

| PSCO/Ship | PSCO 1 | PSCO 2 | PSCO 3 | PSCO 4 |
|---|---|---|---|---|
| **1** | 0.83721 | 0.00015 | 1.73213 | 1.68302 |
| **2** | 2.18567 | 2.44902 | 2.15584 | 1.23480 |
| **3** | 0.43670 | 0.31396 | 0.97402 | 1.39968 |
| **4** | 36.74080 | 55.11084 | 35.62344 | 23.53698 |
| **5** | 5.60484 | 3.93941 | 4.76782 | 1.85870 |
| **6** | 2.62314 | 0.42366 | 3.23279 | 1.19117 |
| **7** | 159.17366 | 286.52394 | 223.35935 | 248.92341 |
| **8** | 8.73485 | 10.73632 | 8.84997 | 6.37257 |
| **9** | 96.51167 | 99.50336 | 122.52453 | 97.56362 |
| 10 | 1.41462 | 10.44739 | 2.75165 | 6.34964 |
| MSE (of each pair) | | | | 39.4949 |

Table A.6. Squared overestimate of MTR-RF2

| PSCO/Ship | PSCO 1&<br>PSCO 2 | PSCO 1&<br>PSCO 3 | PSCO 1&<br>PSCO 4 | PSCO 2&<br>PSCO 3 | PSCO 2&<br>PSCO 4 | PSCO 3&<br>PSCO 4 |
|---|---|---|---|---|---|---|
| **1** | 0.81327 | 0.14277 | 0.19226 | 1.63754 | 1.79637 | 0.00367 |
| **2** | 0.00414 | 0.00028 | 0.17104 | 0.00660 | 0.22842 | 0.15737 |
| **3** | 0.00857 | 0.12242 | 0.25619 | 0.19575 | 0.35845 | 0.02442 |
| **4** | 1.88554 | 0.00760 | 1.45530 | 2.13252 | 6.65386 | 1.25259 |
| **5** | 0.15091 | 0.04749 | 1.01021 | 0.02909 | 0.38022 | 0.61965 |
| **6** | 0.87699 | 0.02730 | 0.24733 | 1.21374 | 0.19286 | 0.43896 |
| **7** | 25.81606 | 3.82203 | 15.60644 | 9.77157 | 1.27792 | 3.98201 |
| **8** | 0.08668 | 0.00000 | 0.16127 | 0.08582 | 0.48442 | 0.16245 |
| **9** | 0.02513 | 1.66732 | 0.00011 | 1.28304 | 0.02862 | 1.69492 |
| 10 | 3.95587 | 0.20695 | 1.63481 | 2.35323 | 0.50458 | 0.67845 |
| MSO (of each ship) | | | | | | 10.0031 |

Table A.7. Squared overestimate of MTR-RF3

| PSCO/Ship | PSCO 1&<br>PSCO 2 | PSCO 1&<br>PSCO 3 | PSCO 1&<br>PSCO 4 | PSCO 2&<br>PSCO 3 | PSCO 2&<br>PSCO 4 | PSCO 3&<br>PSCO 4 |
|---|---|---|---|---|---|---|
| **1** | 0.85986 | 0.16089 | 0.14617 | 1.76463 | 1.71506 | 0.00035 |
| **2** | 0.00749 | 0.00010 | 0.13482 | 0.00934 | 0.20586 | 0.12749 |
| **3** | 0.01010 | 0.10634 | 0.27274 | 0.18199 | 0.38783 | 0.03848 |
| **4** | 1.85572 | 0.00863 | 1.46392 | 2.11740 | 6.61609 | 1.24779 |
| **5** | 0.14643 | 0.03383 | 1.00824 | 0.03950 | 0.38620 | 0.67272 |
| **6** | 0.93843 | 0.03182 | 0.27900 | 1.31585 | 0.19406 | 0.49927 |
| **7** | 18.58139 | 5.42334 | 9.99131 | 3.92756 | 1.32184 | 0.69238 |
| **8** | 0.10314 | 0.00038 | 0.18583 | 0.09105 | 0.56586 | 0.20295 |
| **9** | 0.02283 | 1.55014 | 0.00285 | 1.19671 | 0.00955 | 1.42003 |
| 10 | 4.17329 | 0.22037 | 1.77015 | 2.47569 | 0.50750 | 0.74139 |
| MSO (of each ship) | | | | | | 8.0162 |

Tables A.4 and A.5 shows that the MSE of the outputs of MTR-RF3 is much larger than that of MTR-RF2, while Tables A.6 and A.7 show that the MSO of the outputs of MTR-RF3 is smaller than that of MTR-RF2. Especially, for ship 7, the MSE is 86.60 for the outputs generated by MTR-RF2 and 229.50 for the outputs generated by MTR-RF3. On the contrary, the MSO of ship 7 is 60.28 in MTR-RF2 and the MSO of ship 7 is 39.94 in MTR-RF3. The differences in the MSE and MSO of MTR-RF2 and MTR-RF3 regarding ship 7 indicate that although MTR-RF3 is less accurate in the prediction values compared to MTR-RF2, it could better predict the “relative relationship” among the four outputs. More specifically, we compare the relative relationship of the outputs in the real situation and the predicted values generated by MTR-RF2 and MTR-RF3 for ship 7 as shown in Figure A.1.

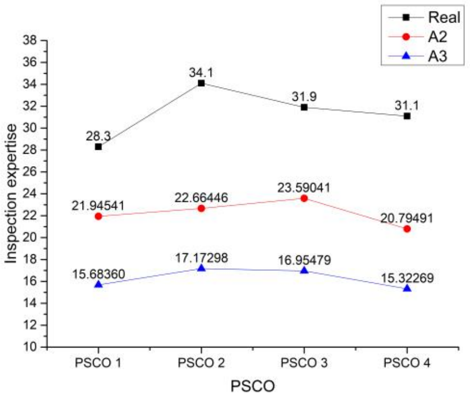


Figure A.1. Comparison of the predicted inspection expertise of MTR-RF2 and MTR-RF3

Figure A.1 shows that the relative relationship of the predicted inspection expertise of MTR-RF3 and the real situation is quite the same: PSCO 2 has the largest expertise, following by PSCO 3. Although the predicted relative inspection expertise of PSCO 4 and PSCO 1 is swapped in MTR-RF3, the gap is quite small, and is much smaller than that of MTR-RF2. However, the prediction results of MTR-RF2 suggests that PSCO 3 has the largest inspection expertise followed by PSCO 2, where there is a big gap with the real situation. Therefore, A2 assigns PSCO 3 to inspect ship 7, and A3 assigns PSCO 2 to inspect ship 7 which is the same as the optimal assignment in the real situation.

## Reference

Angeloudis, P., Greco, L., Bell, M. G., 2016. Strategic maritime container service design in oligopolistic markets. Transportation Research Part B: Methodological 90, 22–37.

Bateman, S., 2012. Maritime security and port state control in the Indian Ocean Region. Journal of the Indian Ocean Region 8(2), 188–201.

Bell, M. G., Pan, J. J., Teye, C., Cheung, K. F., Perera, S., 2020. An entropy maximizing approach to the ferry network design problem. Transportation Research Part B: Methodological 132, 15–28.

Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25(2), 197–227.

Blockeel, H., 1998. Top-down induction of first-order logical decision trees. PhD Thesis of Catholic University of Leuven.

Blockeel, H., De Raedt, L.,1998. Top-down induction of first-order logical decision trees. Artificial Intelligence 101, 285–297.

Breiman, L., 1996. Bagging predictors. Machine Learning 24(2), 123–140.

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

Breiman, L., 2017. Classification and regression trees. Routledge, Abingdon, United Kingdom.

Cariou, P., Mejia, M. Q., Wolff, F. C., 2007. An econometric analysis of deficiencies noted in port state control inspections. Maritime Policy & Management 34(3), 243–258.

Cariou, P., Wolff, F. C., 2015. Identifying substandard vessels through port state control inspections: a new methodology for concentrated inspection campaigns. Marine Policy 60, 27–39.

Chen, J., Zhang, S., Xu, L., Wan, Z., Fei, Y., Zheng, T., 2019. Identification of key factors of ship detention under port state control. Marine Policy 102, 21–27.

Chen, R., Dong, J. X., Lee, C. Y., 2016. Pricing and competition in a shipping market with waste shipments and empty container repositioning. Transportation Research Part B: Methodological 85, 32–55.

Dong, J. X., Lee, C. Y., Song, D. P., 2015. Joint service capacity planning and dynamic container routing in shipping network with uncertain demands. Transportation Research Part B: Methodological 78, 404–421.

Elmachtoub, A. N., Liang, J. C. N., McNellis, R., 2020. Decision trees for decision-making under the predict-then-optimize framework. arXiv preprint arXiv:2003.00360.

European Maritime Safety Agency, 2019a. Annual overview of marine casualties and incidents 2019. <http://www.emsa.europa.eu/news-a-press-centre/external-news/item/3734-annual-overview-of-marine-casualties-and-incidents-2019.html>.

(Accessed 18 Dec 2019).

European Maritime Safety Agency, 2019b. Sustainable shipping. <http://www.emsa.europa.eu/implementation–tasks/environment/sustainable-toolbox.html>. (Accessed 17 Dec 2019).

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer Publisher, Berlin, Germany.

Graziano, A., Cariou, P., Wolff, F. C., Mejia Jr, M. Q., Schröder–Hinrichs, J. U., 2018a. Port state control inspections in the European Union: do inspector's number and background matter? Marine Policy 88, 230–241.

Graziano, A., Schröder–Hinrichs, J. U., Ölcer, A. I., 2017. After 40 years of regional and coordinated ship safety inspections: destination reached or new point of departure? Ocean Engineering 143, 217–226.

Graziano, A., Mejia Jr, M. Q., Schröder–Hinrichs, J. U., 2018b. Achievements and challenges on the implementation of the European Directive on port state control. Transport Policy 72, 97–108.

Harrington, P., 2012. Machine learning in action. Manning Publications Co., New York, USA.

Heij, C., Bijwaard, G. E., Knapp, S., 2011. Ship inspection strategies: effects on maritime safety and environmental protection. Transportation Research Part D 16(1), 42–48.

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. < http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Assembly/Documents/A.1119%2830%29.pdf>. (Accessed 17 May 2019).

Intercargo, 2000. Port state control – a guide for ships involved in the dry bulk trades. <https://www.mardep.gov.hk/en/others/pdf/pscguide.pdf>. (Accessed 12 Oct 2018).

Jabari, S. E., Zheng, J., Liu, H. X., 2014. A probabilistic stationary speed–density relation based on Newell's simplified car-following model. Transportation Research Part B: Methodological 68, 205–223.

Knapp, S., Franses, P. H., 2007. A global view on port state control: econometric analysis of the differences across port state control regimes. Maritime Policy & Management 34(5), 453–482.

Knapp, S., Franses, P. H., 2007. Econometric analysis on the effect of port state control inspections on the probability of casualty: can targeting of substandard ships for inspections be improved? Marine Policy 31(4), 550–563.

Knapp, S., Van de Velden, M., 2009. Visualization of differences in treatment of safety inspections across port state control regimes: a case for increased harmonization efforts. Transport Reviews 29(4), 499–514.

Kocev, D., Vens, C., Struyf, J., Dzeroski, S., 2007. Ensembles of multi-objective decision trees. Proceedings of European Conference on Machine Learning 2007, 624–631.

Lee, C. Y., Song, D. P., 2017. Ocean container transport in global supply chains: overview and research opportunities. Transportation Research Part B: Methodological 95, 442–474.

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

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R news 2(3), 18-22.

Ng, M., 2015. Container vessel fleet deployment for liner shipping with stochastic dependencies in shipping demand. Transportation Research Part B: Methodological 74, 79–87.

Ng, M., Lo, H. K., 2016. Robust models for transportation service network design. Transportation Research Part B: Methodological 94, 378–386.

Probst, P., Boulesteix, A. L., 2017. To tune or not to tune the number of trees in random forest. The Journal of Machine Learning Research 18(1), 6673–6690.

Probst, P., Wright, M. N., Boulesteix, A. L., 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9(3), 1–19.

Ravira, F. J., Piniella, F., 2016. Evaluating the impact of PSC inspectors' professional profile: a case study of the Spanish Maritime Administration. WMU Journal of Maritime Affairs 15(2), 221–236.

Sampson, H., Bloor, M., 2007. When Jack gets out of the box: the problems of regulating a global industry. Sociology 41(3), 551–569.

Şanlıer, Ş., 2020. Analysis of port state control inspection data: the Black Sea Region. Marine Policy 112, 1–11.

Teye, C., Bell, M. G., Bliemer, M. C., 2018. Locating urban and regional container terminals in a competitive environment: an entropy maximising approach. Transportation Research Part B: Methodological 117, 971–985.

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

Tokyo MoU, 2018b. Tokyo MoU celebrates its 25th anniversary during its 29th committee meeting in Hangzhou, china. http://www.tokyo-mou.org/doc/PSCC29%20PRESS-f.pdf. (Accessed 3 June 2020).

Tokyo MoU, 2019. List of Tokyo MoU deficiency codes. <http://www.tokyo–

mou.org/doc/Tokyo%20MOU%20deficiency%20codes%20(December%202019).pdf>. (Accessed 12 Dec 2019).

Tsou, M. C., 2018. Big data analysis of port state control ship detention database. Journal of Marine Engineering & Technology 17, 1–9.

UNCTAD, 2019. Review of Maritime Transport 2019. <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2563>. (Accessed 16 Dec 2019).

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, 129–157.

Yan R., Zhuge D., Wang S., 2019. Development of two highly-efficient and innovative inspection schemes for PSC inspection. Asia Pacific Journal of Operations Research, in press.

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.

Yu, M., Fransoo, J. C., Lee, C. Y., 2018. Detention decisions for empty containers in the hinterland transportation system. Transportation Research Part B: Methodological 110, 188–208.

Zheng, W., Li, B., Song, D. P., 2017. Effects of risk-aversion on competing shipping lines' pricing strategies with uncertain demands. Transportation Research Part B: Methodological 104, 337–356.