https://dx.doi.org/10.1016/j.marpolbul.2023.115068

A port statistics-based generic ballast water estimation and risk assessment approach and its application to Chinese ports

Dong Chen[a], Xiwen Bai[b], Zhaojun Wang[c,d], Dong Yang[a]*

[a] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong, China

[b] Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

[c] University of Delaware, 305 Robinson Hall, Newark, DE 19716, USA

[d] Smithsonian Environmental Research Center, 647 Contees Wharf Rd, Edgewater, MD 21037, USA

* Corresponding author. Email: dong.yang@polyu.edu.hk

Abstract

The Ballast Water Management Convention now enters the experience-building stage, but developing countries have no adequate ballast water information yet. Against this background, considering the richness and generality of common port statistics, we develop a new generic approach to estimate discharge volumes and assess associated risks. This is one of the few efficient and feasible ways for port authorities to manage real discharged ballast water. Discharge volumes during 2017-2020 and risks in 2017 are analyzed for bulker and tanker. Results show that: (1) Ports in Yangtze River Delta and Bohai Rim received most ballast water, especially Ningbo-Zhoushan port as high-risk region receiving about 65 million tons per year; (2)With a higher ratio of discharge volumes to cargo throughputs, the tanker tends to release more ballast water than the bulker; (3) Chinese ports suffer more than 0.9 of the probability of non-indigenous species introduction. All these findings help implement convention globally.

Keywords: Ballast water estimation; Risk assessment; Port statistic; Lasso regression; Chinese port

1. Introduction

Vessel ballast water as the important vector of the harmful non-indigenous species (NIS) brings a significant risk to marine environment and a great challenge to management (Carlton, 1985; Molnar et al., 2008). In total, the average 2-3 billion metric tons of ballast water are shipped across the globe every year (Veldhuis et al.,2010). Harmful aquatic organisms and pathogens (HAOP) are main NIS in ballast water, such as Phytoplankton (harmful algae) and pathogens bacteria, and recently microplastic as the source of NIS is also detect in ballast water (Ruiz et al., 2001; Naik et al., 2019). These harmful and invasive NIS have a great negative impact on the aquatic ecosystem and local economy around the world. More than 1000 non-indigenous aquatic species have been transferred to coastal Europe and produced a great harm like the anoxic water bloom in Bilbao Harbor (Gollasch, 2006; Butrón et al., 2011). The NIS invasion events in North America have increased exponentially and in U.S. cause annual economic loss over $100 billion (Ruiz et al., 2001; Miller et al., 2007). New Zealand receives nearly 3 million tons of ballast water with NIS from foreign ports and Australian Tasmania spends huge cost about AU $23 million to remove the invasion marine pest like the paralytic shellfish (Hewitt et al., 2005; Campbell et al., 2013). As a result, the hazards of ballast water attract increased attentions from the global academic institutions and international organizations.

The International Maritime Organization (IMO) has acknowledged the significance to manage ballast water and to control the risk of NIS spread in 1973 (IMCO, 1973) and enacted the International Convention on the Control and Management of Ship's Ballast Water and Sediments (BWM Convention) in 2004 (IMO, 2004). The BWM Convention has entered into force globally in 2017, covering 75.11% of world merchant shipping tonnages (Batista et al., 2017). Nowadays, to emphasize on effective monitoring and enforcement of BWM Convention, the IMO moves to the experience-building phase (EBP) and emphasizes the evidence-based submission and analysis of ballast water data (MEPC.290 (71); MEPC.297 (72)). Unlike management authorities of U.S, Canada, and Australia collect and compile ballast water data (Chan et al., 2013; Cope et al., 2015), most developing countries have not developed an official organization for data collection of their ports, where China is a typical example. Although China has implemented the BWM Convention since 2019, China’s vessel ballast water discharge and source information are not yet systematically gathered and published in a unified platform. The volumes of ballast water released into the China sea area are usually estimated by scholars using the ship size of DWT (deadweight tonnages) or its ratio to the ballast capacity as predictor (Zhang et al., 2017; Wan et al., 2021). However, these estimated discharge data are incomplete and only updated to 2017 (Zhang et al., 2017). In addition, the explanatory variable for estimation is single and produces the bias omitting other meaningful factors.

To fill the research gap, we propose a widely applicable and effective ballast water estimation model by lasso regression. This model has the simplicity advantage that only requires common port statistics. Every port authority can adopt the model to estimate the ballast water volumes, no matter whether the D1 exchange standard and the D2 treatment standard of BWM Convention are performed. Besides, building upon the work of Seebens et al. (2013), we also provide a modified risk assessment model. Combining both models as an innovative approach, we apply it to Chinese ports to study a real-world case. Various ballast water scenarios in China from 2017 to 2020 can be estimated and associated risks in 2017 are assessed. And thus, the port authority can obtain valuable discharge and risk information to develop management strategies, which is important for the initial implementation of BWM Convention.

The contribution of our work is three-fold. First, common port statistics of monthly total imports, exports, and deadweight tonnages (DWTs) ensure the data accessibility and generality for many undeveloped countries. These statistics are rich but never well used for ballast water estimation and risk assessment at port level. Second, we develop a generic approach for ballast water management with the above statistics. Third, we apply our model to Chinese major ports' ballast water estimations within the period 2017-2020, which help better perform the BWM convention and submit evidence-based data under experience-building phase. Some interesting findings are beneficial to make policy suggestions.

The rest of this paper is organized as follows. In Section 2, the related studies of ballast water management are reviewed and summarized. In section 3, the approach of ballast water estimation model and associated risk assessment model is developed. In section 4, the data and their properties are described. In section 5, the results and findings are presented. Lastly, Section 6 concludes and discusses this study.

2. Literature Review

Since the nineteenth century, solid ballast material had been replaced by water that bring a huge change for related maritime research. The latest research on ballast water can be mainly divided into three categories: first, the approaches for ballast water estimation and management; second, the risk assessment of possible harmful species invasion within ballast water; and third, the policies and regulations under the BWM.

The majority of studies on ballast water involve ballast water estimation as the indispensable part that is the cornerstone for their whole research. A series of approaches and variables for estimation are proposed by scholars. For example, considering the nature of ballast water in shipping, the GEF-UNDP-IMO GloBallast Partnerships Programme (2000-2017) gives the ballast water (BW) capacities and associated ratios of DWT for different ship types, intended to help developing countries estimate and implement the BWM Convention (Matheickal et al., 2017). It lays a foundation for ballast water estimation. As the application of this method, Enshaei and Mesbahi (2011) develop a ballast water export/import framework for UK ports. They establish a primary correlation between port and ballast water operation and identify top 20 ballast water recipient ports in UK. Maglić et al. (2015) estimate the ballast water volumes according to gross tonnage or deadweight. Using the estimated results for the port of Rijeka, they conduct the discrete event simulation of the ballast water management and show barge-based ballast treatment facilities are not under heavily utilization. As the further exploration of ballast water theory, combined with decision supporting system (DSS), David et al. (2012) firstly introduce the generic ballast water discharge assessment model, which considers different cargo loading/unloading conditions. Their model is well validated by real ballast water discharge data for the Port of Koper, Slovenia. Meanwhile, based on the statistical knowledge, Seebens et al. (2013) consider the possible zero release and ship stopovers to propose a double-logarithmic regression to represent the volumes of ballast water. They compare the difference of invasion risks between coastal ecosystems. Holzer et al. (2017) and Verna et al. (2021) conduct the linear regression between the commodity such as LNG export and the ballast water discharges to reflect the trade shift. They find a positive relation between both factors. Recently, with the development of machine learning technique, Yoong and Enshaei (2016) develop the artificial neural networks (ANNs) with DWT as the indicator according to the ship type to represent the ballast water volumes. They estimate the ballast water for 31 Australian ports to identify high risk regions. Wang, Z. et al. (2021) propose a random forest model with the indicator of gross weight tonnage (GWT) for ballast water estimation. They analyse species spread risks and patterns using a higher-order network and find stricter standards may reduce risk globally. These studies bring us the insight into the array of approaches and variables that related to ballast water estimation. However, all these approaches only use single variable, such as DWT, export volume, loading/unloading operations. It lacks a global application and explanation of variables, and the associations of variables have not been well quantitative analysis yet. Moreover, most approaches require the individual vessel information, which is hard to access for many ports. All these reasons limit the ballast water estimation with a general and easy way, especially from the level of port. This paper attempts to develop a generic model of ballast water estimation with full utilization of common port statistics and help facilitate the ballast water management of the port authorities in developing countries.

The risk assessment of ballast water is another popular research direction. Despite of the D-2 standard of BWM convention, the risk assessment of ballast water has no uniform criterion (David et al., 2015) and can be classified into three categories around the G7 guidelines (MEPC.289(71)). The first category mainly focuses on the environmental similarity that compares both environments between the ballast water donor port and recipient port, such as differences of salinity and temperature. The probabilities of organism transfer and survival are two fundamental conditions. Keller et al. (2011) explore the risk of NIS introduced into the Great Lakes by global shipping based on the port environmental conditions. Wonham et al. (2013) simulate the relationship between the invasion risk of NIS and the propagule pressure in new environment to solve the predicting population problem. The second category adopts the biogeographic concept that the overlapping regions have the direct risk of sharing species. For a donor port, the number of invaded biogeographic regions decides the likelihood to bring risk to a new recipient port, even the high-risk species do not invade this new recipient port yet. A good example is that Minchin (2007) discusses the link between the alien biota and the habitat alterations, which presents a scheme to classify the levels of certainty for different NIS. The third category considers the historical information, physiological tolerance, reproduction rates of target harmful species to monitor and predict the risk of these species within their life circles. Hewitt (2003) discusses the potential invasion of species with high environmental tolerances and reproduction rates. MacIsaac et al. (2004) develop a spatially explicit "gravity" model to predict the spread of a crustacean waterflea-Bythotrephes. These studies bring us different views and categories of the risk assessment. But methods of these categories still have some shortages: (1) The method in first category is limited when the differences of environments are small; (2) The method in second category cannot predict the risk for unknown invasive species; (3) The method in third category is only suitable to the harmful species selected subjectively by managers under special circumstances. Consequently, it requires a more sophisticated risk assessment model. Seebens et al. (2013) proposes a probability model of NIS spread, covering all three above categories. It integrates the successive spread stages of NIS. These stages include: (1) the identification stage for NIS; (2) the introduction stage for NIS; (3) the establishment stage for NIS. This paper extends the application of Seebens' study from individual vessel to individual port for risk assessment of above three spread stages.

Many policies and regulations of ballast water are attached great attentions by researchers. The development of ballast water policy is different across the countries, but some common grounds under the context of BWM bring us the enlightenments. The first common ground is how to build the prevention mechanism of marine species invasion. Hewitt and Campbell (2007) discuss the biosecurity concerns by fisheries and conservation management agencies. They use Australia and New Zealand as examples to explain how to form a mechanism with the principles Ecologically Sustainable Development. David and Gollasch (2008) review the regional BWM approaches in European and provide the suggestions to help establish a common legal mechanism in EU wide. The second common ground is the shift from ballast water exchange toward numeric concentration-based discharge limits. Albert et al. (2013) review the regulatory framework and standards issued by IMO, U.S. Coast Guard and U.S. Environmental Protection Agency and individual states. They find numeric concentration-based ballast water discharge limits vary from region to region and the requirement of more precise risk-assessment methodology and universal standard become important. Verna and Harris (2016) review the ballast water regulations in U.S and compare them with those in Alaska. The results show the BWM exemptions for crude oil tankers in Alaska may cause the unnecessary risk, especially from the intra-coastal vessel traffic. These studies contribute to the legislation of ballast water in developing countries, who perform as members of the BWM. Combined with the discharge estimation and risk assessment, this paper gives some policy suggestions and managerial implications to China as example.

To summarize, previous research suffers from data availability and limited variable for estimation. In this study, we take the richness and generality of port statistics into consideration. We estimate ballast water by common port statistics with the measurement significance based on an advanced machine learning approach. Our port statistics contain monthly total export tonnages, total import tonnages and total DWTs. Here, we use total export tonnages to represent vessels arriving at port in ballast status; total DWTs to represent the total capacity of ballast tanks; total import tonnages to represent possible zero releases, which usually mean no ballast water on board after importing cargo. Combined these statistics, we adopt a biased regression model-Lasso regression for estimation to reduce the potential impact of collinearity among port statistics. Meanwhile, we extend the risk assessment model from Seebens et al. (2013) to analyse the ballast water risk for ports. Our risk assessment approach with the estimation model can form a general approach for ballast water decision. We finally employ the Chinese ports as the case study to show the application of this approach and discover some implications.

3. Methodology

From the holistic perspective, port estimations for ballast water are the fundamental element to predict corresponding risk. In this section, the ballast water discharge estimation model will be introduced before the risk assessment model. For the discharge estimation, common port statistics as trade indicators can reflect the operation process of ballast water at port, which is analogous in worldwide different ports. To avoid the possible overfitting caused by multicollinearity, Lasso regression technique is finally adopted. In addition, U.S ballast water data provides rich discharge records for model setup. Then for risk assessment, the extended model of Seebens is developed with the port environment information and AIS (Automatic Identification System) data of ships. It can assess the monthly NIS risk for port. The ship types of bulk carrier and tanker ship are chosen as research objects. Because both ship types contribute more than 76% of global ballast water and are often under heavy ballast status for departure while under full loaded status for return (Endresen et al., 2004).

3.1 Ballast water estimation model

In this study, we select some common port statistics as independent variables, including total import tonnages, total export tonnages and total DWTs that haven't been well used in ballast water estimation before. Common port statistics have a nature of generality and almost follow the same rule to census. No matter whether the port is in developed or developing countries, the common port statistics are also easy to be accessed and verified. However, the multiple variables of port statistics may bring the potential multicollinearity because they are all produced by trade. And the ballast water behaviour can be regarded as homogeneous for port in different regions and is just determined by ship types, due to the homogeneity of global shipping transportation service (Stopford 2008). Therefore, two hypotheses are proposed for ballast water estimation model:
Hypothesis 1: Exist multicollinearity among common port statistics as explaining variables.
Hypothesis 2: Ballast water discharge behaviour is homogeneous among different ports.

Based on these hypotheses, considering multicollinearity has a negative impact on the fitting explicability by causing variance increase, Lasso (least absolute shrinkage and selection operator) regression model becomes a better choice as mentioned above. We develop a ballast water estimation model using Lasso technique. The lasso regression uses the L1 regularization term in parameter estimation. This feature allows to control the coefficient of explanatory variable, which avoids the overfitting of multicollinearity and highlights the significant variable.

In lasso regression model, let y denote the monthly ballast water discharge at port, and x_1, x_2, x_3 denote monthly total exports, imports, DWTs, respectively. β_1, β_2, β_3 are unknown coefficients of the above three explanatory variables, and β_0 is the unknown intercept. The objective function to be minimized is shown in Equations. (1) and (2):

$$Min_{\beta}\left\{\sum_{i=1}^{N}\left(y^{(i)}-\beta_0-\sum_{j=1}^{3}x_j^{(i)}\beta_j\right)^2+\lambda\sum_{j=1}^{3}\left|\beta_j\right|\right\}$$
$$=\left\{\sum_{i=1}^{N}\left(y^{(i)}-\beta_0-\beta_1x_1^{(i)}-\beta_2x_2^{(i)}-\beta_3x_3^{(i)}\right)^2+\lambda\left(\left|\beta_1\right|+\left|\beta_2+\right|\beta_3\right|\right)\right\} \quad (1)$$

$$\textit{Subject to: } \sum_{j=1}^{3}\left|\beta_j\right|, \; S. \quad (2)$$

Here, N represents the number of monthly ballast water discharges. $\lambda\sum_{j=1}^{3}|\beta_j|$ represents the penalty term (L1 regularization term). λ is a regularization parameter that obtained when the mean-squared error (MSE) is minimized. The sum of all coefficients is limited by $S=1$.

3.2 Risk assessment model

Associated with the ballast water estimation model, we develop a risk assessment model. As shown in Figure 1, we use the port environment information such as salinity and temperature to measure the geographical similarity for potential invasive species. Meanwhile, AIS (Automatic Identification System) data are used to extract the ship movement and help measure the sailing time for above species in ballast tank. Finally, combined with the volumes of ballast water discharge at port, the risks of different spread stage can be assessed.

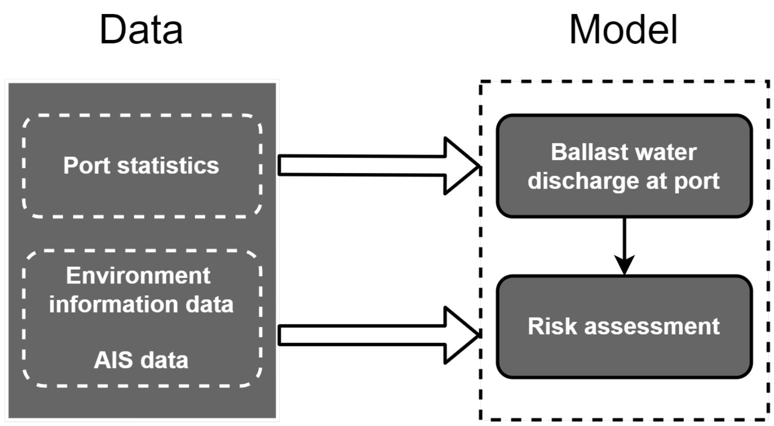


Figure 1. The risk assessment model with ballast water estimation.

We consider three independent and dynamic invasion risk probability that the probability of NIS identification between the donor and recipient port water, the probability of NIS introduction from donor to recipient port water and the probability of NIS population establishment in recipient port water (Seebens et al., 2013; Saebi et al., 2020). The monthly comprehensive NIS spread risk probability is defined by these three probabilities:

$$P(NIS\ spread)_m = P(Non\text{-}indigenous)_m \times P(Intro)_m \times P(Estab)_m \,. \quad (3)$$

For detail, our formulas of three kinds of risk probabilities are presented as follows:

i. The likelihood of being non-indigenous

$$P(Non\text{-}indigenous)_m = \sum_{i}^{N} p_{ij}^{NI} / N_{port} \,. \quad (4)$$

We calculate the average chance that a species is non-native for the recipient port j. For a ship from the donor port i to the recipient port j, p_{ij}^{NI} is a binary variable: It equals to 0 for native disperse between the same or neighboring ecoregions and 1 for NIS disperse between the different ecoregions. N_{port} is the monthly frequency of port calls.

ii. The likelihood of introduction

$$P(Intro)_m = \rho\left(1 - e^{-\lambda D_{port}}\right)e^{-\mu \Delta t_{avg}} \quad . \tag{5}$$

We calculate the introduction risk with monthly ballast water estimation D_{port}. Factor ρ represents the ballast water treatment efficiency with the default value of 1. Δt_{avg} is the monthly average traveling time of all ships. Set $\mu=0.02$ as the mortality rate and $\lambda=3.22\times10^{-6}$ as the species introduction potential (Saebi et al., 2020).

iii. The likelihood of establishment

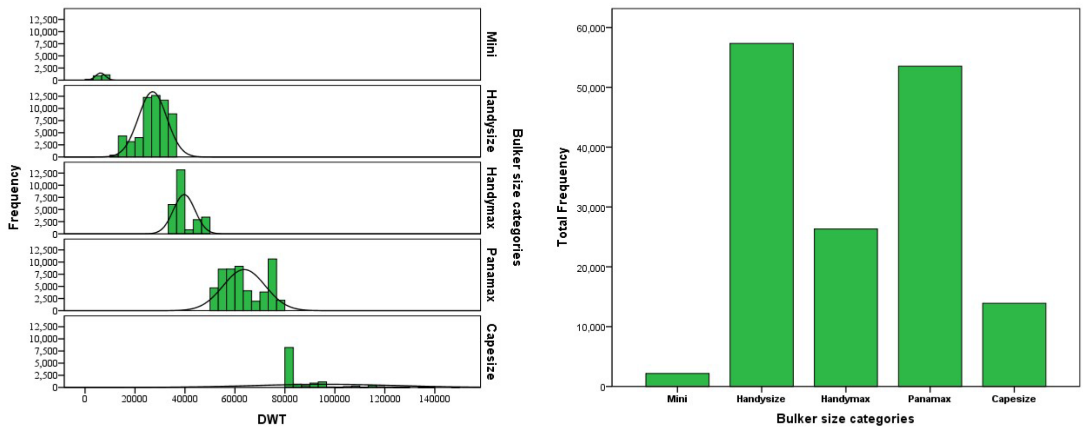
$$P(Estab)_m = \alpha_{init} e^{-\frac{1}{2}\left[\left(\frac{\Delta T_{avg}}{\delta T}\right)^2 + \left(\frac{\Delta S_{avg}}{\delta S}\right)^2\right]} \quad . \tag{6}$$

We considered the Gauss-distributions of temperature and salinity to evaluate environmental similarities. ΔT_{avg} and ΔS_{avg} are the average temperature and salty differences between the donor port and the recipient port. α_{init} is the survival probability. Set $\delta T=2^{\circ}C$, $\delta S=10\,ppt$ and $\alpha_{init}=1$ for maximum survival rate (Seebens et al., 2013).

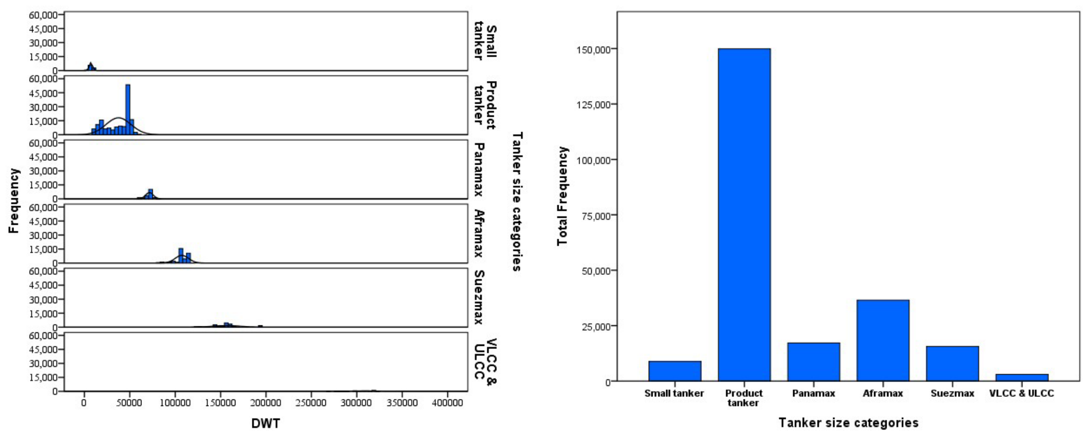
4. Data

Firstly, considering the general homogeneity of ballast water discharge mode, we construct the estimation model by the ballast water information platforms of U.S. National Ballast Information Clearinghouse (NBIC), which has received ballast reports from all vessels arriving at ports in the United States since 2004 and provides the rich ballast water release reports and samples. We use the ballast water release data from NBIC within the period 2010-2020 to build up our estimation model. In the meantime, to match the ballast water releases, the associated port statistics are derived from the USA Trade Online. Detailed information is described as follows.

Ballast water release data: NBIC can give access to online data for every port call, including IMO number, arrival date, discharge volume and gross tonnage etc. We extract 159,216 releases for bulker and 263,637 releases for tanker, from 68 major ports in the USA. As shown in Figure 2, these data cover the most size categories of two ship types. Handysize and Panamax bulk carriers contribute more than 75% of releases. Product and Panamax tanker ships contribute more than 73% of releases. Notably, Product tankers account for about 57% of releases.



(a) Bulk carrier size categories



(b) Tanker ship size categories

Figure 2. 2010-2019 ballast water discharge frequencies of different ship size categories.

Port statistics data: For imports and exports statistics of the above 68 major ports, we obtain data from the USA Trade Online according to the Harmonized System (HS) codes of commodities transported by bulkers and tankers (see Table 1). For DWTs statistics, Lloyd's Maritime Intelligence provides the information for vessels in service. We match the DWT data with the ballast water releases, according to the ship names and the IMO numbers.

Table 1. Main categories of HS codes for both ship types.

Ship types	**Two- or four-digit HS codes**
Bulk Carrier	10,11,12,26, 47,2701,2702,2703,2704,7204, etc.
Tanker Ship	2706,2707,2708,2709,2710,2712,2713,2714,2715, etc.

In order to ensure the data uniformity in time dimension, we merge the daily ballast water releases data from different sources-coastwise, overseas and unknown. Then we sum the daily data to produce the monthly data for each ballast water recipient port. The frequency distribution of monthly ballast water data among different ports presents a normal distribution (see Table S1 in Supplement materials). Finally, we obtain the panel dataset of monthly discharge volumes and related statistics from 2010 to 2020, from combining three kinds of port statistics. Table 2 shows the descriptive statistics of the variables in our dataset (unit: thousands of tons).

Table 2. Descriptive statistics of the variables.

Bulk Carrier						
Variable	Description	Observations	Mean	Std. Dev.	Min	Max
BWD	Port monthly ballast water discharge (unit: thousands of tons)	4,387	142.675	304.163	0	2686.224
Exports	Port monthly total exports (unit: thousands of tons)	4,387	400.765	686.256	0.000407	5499.777
Imports	Port monthly total imports (unit: thousands of tons)	4,387	56.325	85.160	0.000103	857.136
DWTs	Port monthly total DWTs (unit: thousands of tons)	4,387	805.383	1378.950	2.926	12360.770
Tanker Ship						
Variable	Description	Observations	Mean	Std. Dev.	Min	Max
BWD	Port monthly ballast water discharge (unit: thousands of tons)	4,401	199.026	457.293	0	4220.372
Exports	Port monthly total exports (unit: thousands of tons)	4,401	194.984	428.591	0.002501	3596.973
Imports	Port monthly total imports (unit: thousands of tons)	4,401	661.960	861.605	0.000159	6229.908
DWTs	Port monthly total DWTs (unit: thousands of tons)	4,401	1850.144	2503.859	5.887	18339.580

Using the dataset, we also validated the hypotheses and identify whether our model is reasonable and feasible. Table 3 reports the variance inflation factor (VIF) test for explanatory variables. It shows multicollinearity exists between the port statistics of exports and DWTs (*Hypothesis 1*). The multicollinearity is moderate for bulker and minor for tanker. Besides, we conduct the F-test at 95% confidence level to diagnose the port-fixed effect. The *p-value* for bulker is 0.089 and for tanker is 0.074, which means the ballast water behaviour can be assumed as homogeneous for different ports (*Hypothesis 2*). Therefore, the pooled regression of lasso form is enough accurate for generic estimation. And according to Belsley et al. (2005), when the VIF is below the allowed maximum of 10, the lasso technique can eliminate the effects of multicollinearity without excluding any variables.

Table 3. The results of multicollinearity diagnostics.

Explanatory variable	**Bulk Carrier**		**Tanker Ship**	
	VIF	Tolerance	VIF	Tolerance
DWTs	5.86	0.170648	3.31	0.302058
Exports	5.60	0.178571	3.23	0.309292
Imports	1.11	0.902598	1.92	0.521385
Mean VIF	4.19		2.82	

Secondly, for risk assessment model, we assess the risks in 2017 of ten important Chinese coastal ports-Shanghai, Tianjin, Dalian, Qingdao, Yingkou, Rizhao, Ningbo-Zhoushan, Xiamen, Guangzhou, and Shenzhen. The imports and exports statistics of these ports are obtained from China shipping database in term of HS codes. The DWTs statistics of these ports are also extracted from Lloyd's Maritime Intelligence. In addition, the port environment and biogeographical information are obtained from Saebi et al. (2020), including temperature and salty data of 669 world ports and the Marine Ecoregion of the World (MEOW) data. The ship movement data are extracted by the AIS, e.g., port of calls, sailing time of ships etc.

Overall, the collected data is considered appropriate for applying our approach to Chinese ports for a case study. Noticeably, the U.S. is currently not a party to BWM Convention, and the United States Coast Guard (USCG) has its own BWM regulations requiring most vessels entering U.S. waters to install ballast water management systems (BWMSs) since 2014. However, these onboard BWMSs will not change the discharged volumes of ballast water. Because they are the facilities for treatment of the viable organism and bacteria in ballast water, not releasing the ballast water. Therefore, implementation time points of such treatment policy in two countries have no time-fixed effect, and thus U.S research data for model setup will not have a significant impact on the follow-up risk assessment.

5. Results

5.1 Model buildup

Our ballast water estimation model is established by the lasso regression using the Scikit-learn library of machine learning. The optimal regularization parameter-lambda λ is searched and determined by 10-fold cross validation. We can describe the port monthly ballast water discharge of both ship types by (unit: thousands of tons):

$$\textbf{\textit{Bulker:}}\ BWD = 0.098 Exports - 0.056 Imports + 0.168 DWTs - 28.950, \tag{7}$$

$$\textbf{\textit{Tanker:}}\ BWD = 0.736 Exports - 0.067 Imports + 0.071 DWTs - 30.826. \tag{8}$$

To our best knowledge, extant studies have not conducted a robustness check when estimate the ballast water volumes. In order to demonstrate the generality of lasso regression results, we conduct the robustness checks through varying the sample sizes and removing variables (Verbeek, 2008). The sample sizes are changed with different time windows. Year 2017 is the time node of BWM's entry into force, therefore, it can be used to adjust the study period. Specifically, Model 1 represents the full sample size from 2010 to 2019, as same as Equations. (7) and (8). Model 2 represents the medium sample size from 2010 to 2016. Model 3 represents the small sample size from 2017 to 2019. Besides, to compare with previous research, we set Model 4 with a single variable-*Exports* and Model 5 with a single variable-*DWTs*. We analyze both models by OLS (ordinary least squares).

Table 4 summarizes the results of five models. In general view, for fitting effects, the findings suggest that MAPE (Mean Absolute Percentage Error) of lasso models (Model 1~3) varies among 8.47%~15.88% for bulker and 8.05%~8.58% for tanker, while it varies among 15.21%~16.90% for bulker and 10.89% ~20.96% for tanker by the single variable models (Model 4~5). The fitting results are accepted within the area of 50% discrepancy (David et al., 2012). Adjusted-R^2 of models including all port statistics are more than 0.9 for bulker and 0.85 for tanker. The significance level of port statistics as explanatory variables is at 1% levels. It implies the port statistics contribute to an excellent estimation performance. For robustness effects, our models of multi-variables (Model 1~3) have smaller RMSEs (Root Mean Square Error) than those of Model 4 and 5 with a single variable. In addition, the robustness of lasso regression is better guaranteed that approaching unbiased estimation as the increase of sample size.

In view of model variable, the port statistics in lasso regression (Model 1~3) show a good interpretability for ballast water estimation. Specifically, the coefficients of *Exports* and *DWTs* indicate a positive relation with estimated discharge volumes and keep consistent with previous research. The effect of *DWTs* is larger than that of *Exports* for bulker, opposite results for tanker. One possible reason is some tankers called at port are used for offshore storage of crude oil, in which *DWTs* has no contribution to ballast water. Notably, as the sample size becomes sufficient, the coefficient sign of *Imports* becomes negative, especially in Model 1 that -0.056 for bulker and -0.067 for tanker. Although the negative coefficient induced by *Imports* is not great value, it demonstrates *Imports* should not be omitted and suggests a situation that some ships unload imported cargo at first, then they load exported cargo without no ballast water releases, which explains the negative effect of importing behaviors. These findings support the common port statistics have a favorable generality and feasibility applied in estimation model.

Table 4. The results of the robustness check.

	Bulk Carrier				
Variable	Model 1 (2010-2019)	Model 2 (2010-2016)	Model 3 (2017-2019)	Model 4 (2010-2019)	Model 5 (2010-2019)
Exports	0.099***	0.095***	0.086***	0.411***	
	(0.002)	(0.005)	(0.004)	(0.002)	
Imports	-0.056***	-0.054**	0		
	(0.014)	(0.015)	(0.021)		
DWTs	0.168***	0.174***	0.147***		0.213***

	(0.001)	(0.003)	(0.002)		(0.001)
Constant	-28.950***	-30.187***	-13.463***	-22.123***	-28.568***
	(1.473)	(1.707)	(2.716)	(1.985)	(1.415)
Observations	4387	3038	1349	4387	4387
Lambda (Control shrinkage)	0.99	0.99	0.66	-	-
RMSE	80.79	92.41	85.46	113.48	89.93
MAPE	12.14%	15.88%	8.47%	16.90%	15.21%
Adjusted-R^2	0.936	0.942	0.918	0.861	0.861
F-Test	21468.5***	16457.4***	5034.6***	27109.9***	57542.4***

	Tanker Ship				
Variable	Model 1 (2010-2019)	Model 2 (2010-2016)	Model 3 (2017-2019)	Model 4 (2010-2019)	Model 5 (2010-2019)
Exports	0.736***	0.553***	0.83***	0.985***	
	(0.003)	(0.011)	(0.009)	(0.006)	
Imports	-0.067***	-0.023***	0		
	(0.003)	(0.004)	(0.004)		
DWTs	0.071***	0.069***	0.051***		0.158***
	(0.001)	(0.002)	(0.002)		(0.001)
Constant	-30.826***	-36.356***	-29.557***	6.885**	-93.909***
	(2.951)	(3.231)	(4.594)	(2.904)	(4.273)
Observations	4401	3070	1331	4401	4401
Lambda (Control shrinkage)	0.01	0.9	0.93	-	-
RMSE	131.16	145.38	147.53	175.32	227.91
MAPE	8.12%	8.05%	8.58%	10.89%	20.96%
Adjusted-R^2	0.9	0.865	0.945	0.853	0.752
F-Test	13033.5***	6323.4***	7635.9***	25523.2***	13307.3***

Note: Standard errors are in parentheses. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively. The unit of panel datasets is thousands of tons.

Furthermore, we conduct an out of sample forecast test to validate the generalization ability with 95% Confidence Interval (CI) and Prediction Interval (PI). The testing dataset are samples in 2020 and the training dataset are ten years (2010-2019). As Figure 3 shows, the estimations fall in the range of PI with a 0.93 of Adjusted-R^2 for bulker. For tanker, when the monthly discharge volumes are larger than 3×10^3 thousands of tons, the model underestimates the ballast water discharge. But we can accept the results with a 0.90 of Adjusted-R^2. Most estimations fall within the range of three standard deviations. Compared with the results of Verna et al. (2021), our model improves the R-squared by 7% for bulker and by 3% for tanker.

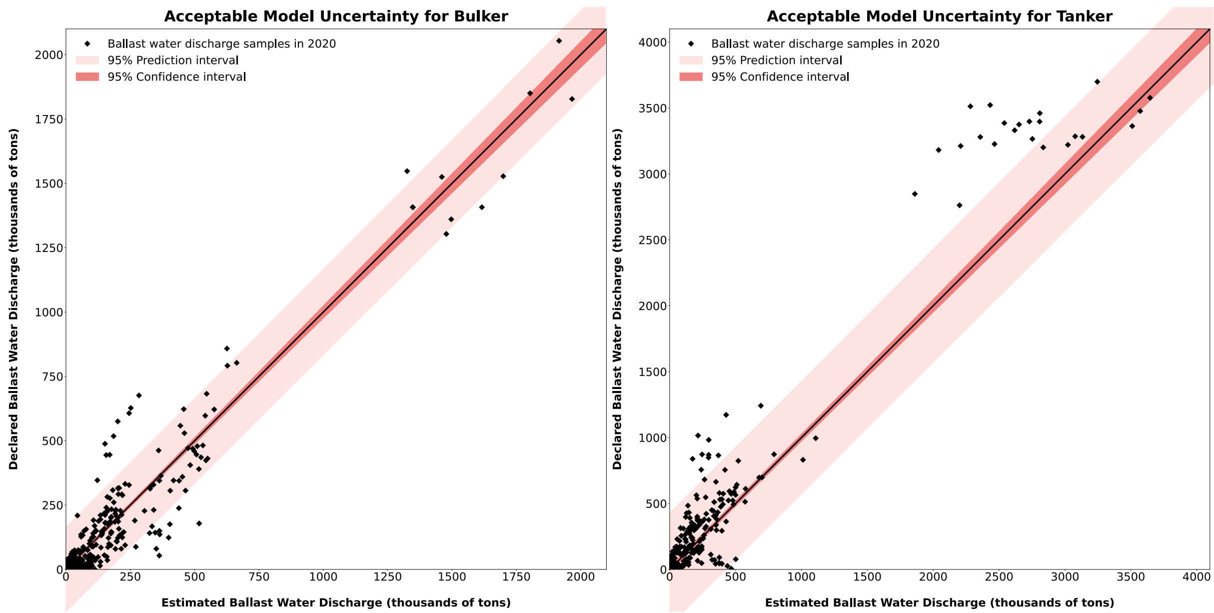


Figure 3. Declared ballast discharges versus estimated ballast discharges in 2020.

Note: Each dot represents a monthly discharge into a single port, 1:1 regression line plotted for reference in black.

5.2 Estimation for ballast water discharge into China

Based on the built lasso regression model, we estimate the annual ballast water volume (sum of 12 months) released by bulkers and tankers to Chinese ports. The results (see Table 5) show, during the years of 2017-2020, the volume contributed by bulkers has a trend of slow decrease from approximately 3.3 to 2.8 billion tons per year, while the volume contributed by tankers has increased from 1.9 to 2.2 billion tons. The top 5 recipient ports are located at Yangtze River Delta and Bohai Rim. They receive more than 80% of ballast water from both ship types. In contrast, the Guangzhou port and Shenzhen port in Pearl River Delta account only for 6%~8%. Notably, the Ningbo-Zhoushan port has received the maximum amount of ballast water with an annual average of about 65 million tons from either of the ship types. These findings are in line with the attributions and function of these ports.

Table 5. The volume of ballast water discharge into China by both ship types (unit: thousands of tons).

	Ballast water estimations for bulker				
Port	2017	2018	2019	2020	Total of years
Ningbo-Zhoushan	57,929	59,709	68,105	68,913	254,654
Rizhao	54,656	57,269	45,851	46,048	203,824
Tianjin	55,744	51,664	45,184	47,082	199,674
Qingdao	49,235	44,804	34,065	31,281	159,385
Shanghai	34,559	33,697	27,117	21,709	117,082
Guangzhou	30,696	26,295	16,210	16,951	90,151
Yingkou	21,979	24,349	22,704	20,395	89,427
Dalian	12,548	14,143	13,900	15,364	55,955
Xiamen	9,853	9,787	9,104	9,357	38,101
Shenzhen	4,939	3,792	2,068	2,387	13,186
Total of ports	332,138	325,507	284,307	279,487	

	Ballast water estimations for tanker				
Port	2017	2018	2019	2020	Total of years
Ningbo-Zhoushan	56,925	58,392	65,653	75,803	256,772
Dalian	30,286	31,006	37,740	38,045	137,076
Qingdao	26,134	29,490	31,334	31,918	118,875
Tianjin	22,969	25,837	26,365	25,747	100,917
Rizhao	11,993	14,705	14,615	17,314	58,627
Yingkou	16,047	13,045	12,895	12,667	54,654
Shanghai	9,519	8,203	8,832	8,363	34,916
Guangzhou	6,633	6,453	6,338	6,791	26,215
Shenzhen	6,020	6,310	5,410	5,737	23,478
Xiamen	659	848	2,092	2,095	5,693
Total of ports	187,184	194,288	211,274	224,479	

From another point of view, the ratios of ballast water volumes to cargo throughputs are also calculated. As Table 6 shows, for ten Chinese ports, the average ratio of 32% for tankers is higher than that of 18% for bulkers. Because, at the same DWT size, a tanker ship usually has the larger ballast tank capacity (Verling et al., 2005). As the main terminals of dry bulk cargoes and oil products, the ports in Bohai Rim (Tianjin, Qingdao, Yingkou, Rizhao, Dalian) have the maximum ratio between the ballast water volume and cargo throughput. Notably, for tanker, Shenzhen port accounts for the biggest ratio of ballast water to throughput. These results indicate tankers have more ballast water volume per arrival and the ports with many oil terminals are in high risk.

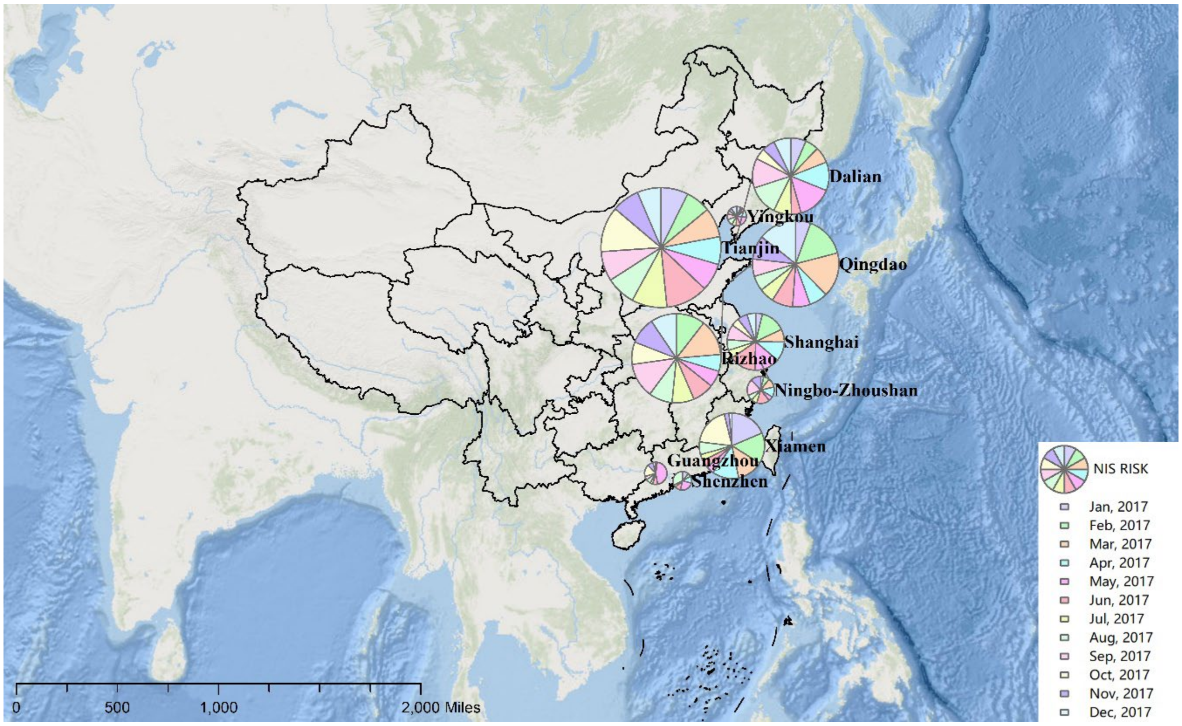
Table 6. The ratio of ballast water discharge volumes to cargo throughputs for both ship types.

	The ratio for bulker				
Port	2017	2018	2019	2020	Average of years
Ningbo-Zhoushan	11%	10%	12%	12%	11%
Rizhao	26%	21%	16%	16%	20%
Tianjin	25%	24%	21%	23%	23%
Qingdao	26%	23%	17%	16%	20%
Shanghai	17%	20%	18%	19%	19%
Guangzhou	20%	17%	8%	9%	13%
Yingkou	20%	19%	27%	23%	22%
Dalian	19%	19%	17%	17%	18%
Xiamen	13%	14%	16%	20%	16%
Shenzhen	23%	20%	14%	13%	17%
Average of ports	20%	19%	17%	17%	
	The ratio for tanker				
Port	2017	2018	2019	2020	Average of years

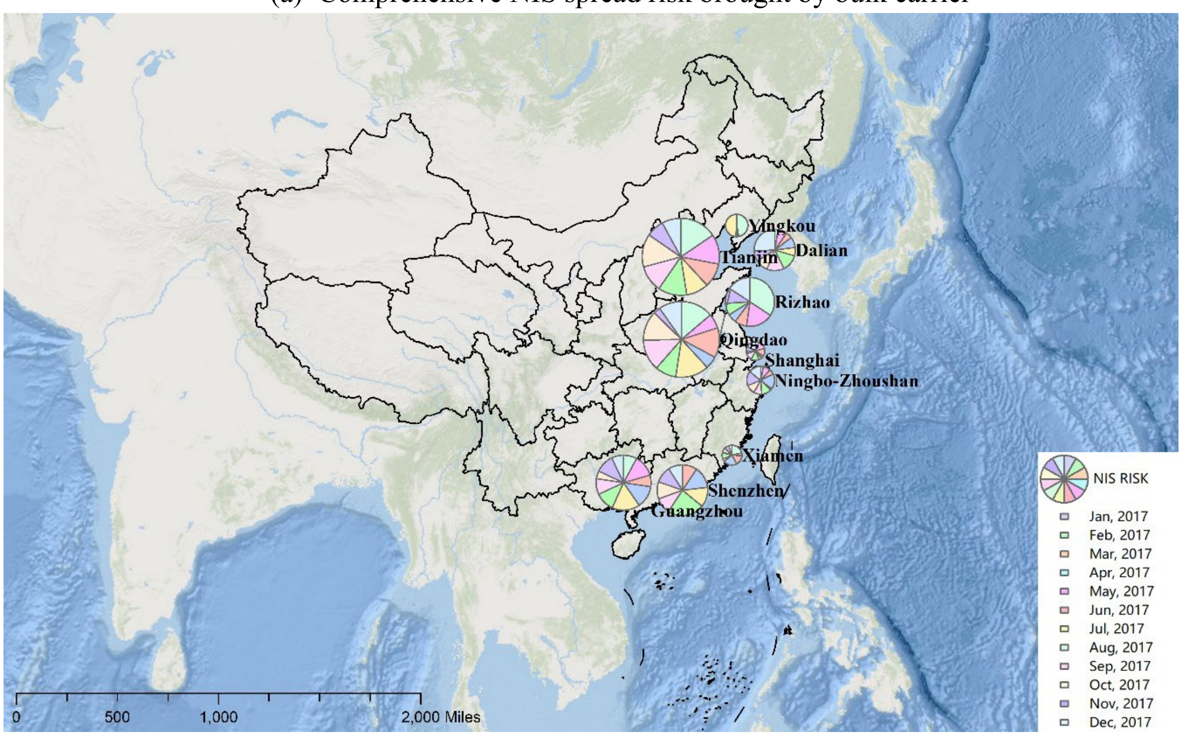
Ningbo-Zhoushan	35%	37%	37%	35%	36%
Dalian	40%	40%	39%	36%	39%
Qingdao	25%	27%	26%	25%	26%
Tianjin	40%	39%	37%	35%	38%
Rizhao	19%	21%	21%	22%	21%
Yingkou	38%	41%	50%	42%	42%
Shanghai	30%	30%	30%	30%	30%
Guangzhou	28%	28%	29%	28%	28%
Shenzhen	45%	49%	40%	37%	43%
Xiamen	16%	17%	31%	21%	21%
Average of ports	32%	33%	34%	31%	

5.3 Risk assessment for Chinese ports

We calculate the monthly NIS spread risk for Chinese ports in 2017 (see Figure 4). As shown in Table 7, for both ship types, the distinct invasion probabilities show different results. The average probability of NIS introduction from donor to recipient port waters (*average introduction risk*) is high and exceeds over 0.9, except for Shenzhen. It implies large amounts of marine species are introduced into China Seas through the vector of vessel ballast water. The average probability of NIS identification between the donor and recipient port (*average non-indigenous species risk*) is larger than 0.4 in Xiamen port and ports of Pearl River Delta (Guangzhou port and Shenzhen port), which means these ports are more possible to find NIS from different Marine Ecoregions. However, the average probability of NIS population establishment in recipient port waters (*average establishment risk*) is quite low for all ports. Because most NIS suffer from a great difference of temperature and salinity among ports. As a result, the comprehensive NIS spread risk probability (*average NIS spread risk*) is not significant for Chinese ports. But the potential negative impact of introduced species cannot be excluded. It still threatens the biodiversity and hydrologic condition of the recipient ecosystem. Because secondary transfer of aquatic NIS by coastwise usually appears, even in the same Marine Ecoregion (Lavoie et al., 1999; Verna and Harris, 2016).



(a) Comprehensive NIS spread risk brought by bulk carrier



(b) Comprehensive NIS spread risk brought by tanker ship

Figure 4. The monthly comprehensive NIS spread risk assessment for Chinese ports in 2017.

Note: The size of pie chart exhibits the rank according to the average NIS spread risk.

Table 7. The average monthly risk assessment for Chinese port in 2017.

Port	**Average NIS spread risk**		**Risk of three distinct invasion transitions**					
			Average non-indigenous species risk		Average introduction risk		Average establishment risk	
	Bulker	Tanker	Bulker	Tanker	Bulker	Tanker	Bulker	Tanker
Tianjin	0.036	0.024	0.164	0.115	0.993	0.990	0.232	0.255
Qingdao	0.019	0.023	0.089	0.141	0.992	0.993	0.217	0.187
Rizhao	0.020	0.010	0.126	0.072	0.988	0.951	0.208	0.203
Dalian	0.015	0.007	0.096	0.159	0.949	0.991	0.187	0.047
Guangzhou	0.001	0.012	0.592	0.138	0.993	0.824	0.003	0.121
Xiamen	0.011	0.002	0.417	0.224	0.918	0.917	0.038	0.053
Shenzhen	0.001	0.010	0.713	0.128	0.713	0.789	0.002	0.124
Shanghai	0.008	0.001	0.028	0.019	0.994	0.915	0.299	0.084
Ningbo-Zhoushan	0.002	0.003	0.036	0.032	0.991	0.993	0.052	0.098
Yingkou	0.001	0.002	0.105	0.133	0.990	0.978	0.011	0.035

6. Conclusion and discussion

This study develops a new and generic ballast water discharge estimation and risk assessment approach, by integrating the common port statistics with lasso regression and the theory of risk propagation probability. This first-ever approach considers the data richness and accessibility at port level to manage ballast water under the context of BWM Convention and its experience-building phase (MEPC.297 (72)). The performance and interpretability of this approach is well validated and analyzed. Specifically, to illustrate the effectiveness of the approach, China as the one of the most important developing countries is chosen for case study. The unknown ballast water discharge volumes from 2017 to 2020 in ten main Chinese ports are estimated and the different risk probabilities in 2017 are assessed. This paper is believed to contribute to previous literature both theoretically and practically.

For ballast water estimation model, for first time, more available and public predictors are considered for port management. It is particularly suitable for the countries without ballast water records and information platforms. More importantly, classical models only depend on the correlation between DWT and ballast water on board, such as Great Lakes model and European model. Their estimated results are rough and show high difference (David et al., 2012). Our estimation model can achieve a satisfactory accuracy by shrinking the variance of every variable in lasso method with a great robustness and out of sample forecast. As ballast water decision tool, our model is also a good supplement of the BWAD model because: The BWDA model is based on the perspective of individual vessel. It considers many involved factors, including the identification of vessels that will discharge ballast water, every port of call of a vessel, and loading and unloading operations. However, the data and information accessibility of an individual vessel and its associated loading and unloading operation limits application of BWDA model in different regions, especially in developing countries. In addition, our model shows the variables-*DWTs* and *Exports* have different positive effects on discharge estimation for bulker and tanker, while variable-*Imports* presents the negative effect. For risk assessment model, we extend the comprehensive probability of NIS spread model to the port level. It integrates the estimation model to act as a powerful management support framework.

In practice, we found that: (1) For bulker and tanker, Ningbo-Zhoushan port receives most about 65 million tons ballast water per year, and ports in Yangtze River Delta and Bohai Rim are heavily affected; (2) Comparing to bulk ship, tanker contributes more ballast water to the port due to the higher ratio of discharge volumes to cargo throughputs; (3) The Chinese port are at a high risk (more than 0.9) of invasive species' introduction, although most NIS cannot establish their habitats. In fact, China as the representative of developing countries, has no special legislation for ballast water regulation. BWM is included in many other legislations (e.g., Chinese Government formulated the Biosecurity Law and the Administrative Measures for Entry and Exit Inspection and Quarantine on Ships of International Sails). However, our empirical results show these legislations are not enough to address current ballast water challenges. As the fact we can see, in recent years, more than 660 NIS have been found in ballast water and about 40 alien species of phytoplankton, 16 alien species of zooplankton and 24 alien species of harmful microorganisms (Xu et al., 2012; Wu et al., 2017). Therefore, it is urgent to call for a new ballast water legislation.

Grounded on our work, some managerial implications are suggested: (1) In Yangtze River Delta and Bohai Rim, set up the regional pilot ballast water information online platform for experience accumulation; (2) Fill up the history discharge data for main ports and assess NIS risk of past years to carry out hierarchical remedial measure of environment; (3) The allowed ballast water exemption rules can be reconsidered strictly in terms of ship types, which is the supplementary regulation for the current Regulation D-1 and Regulation A-3 (Hewitt and Campbell, 2007); (4) Strengthen monitor and detection of the density of the aquatic species in the high risk port. Furthermore, considering the BWM Convention D-2 treatment standard has entered into force in China, we also recommended Chinese port authorities prepare the onshore or barge-based ballast water treatment facilities and conduct biological sampling of treated ballast water during Port State Control (PSC) inspections for further reduction of NIS risk.

While the authors believe our model provides important insights for ballast water management, the lack of real ballast water records in China make the estimations hard to calibrate, which is unobservable error. The factors of port weather conditions, port ballast water storage, and port barge-based ballast water reception still have room for further investigation. These concerns will be the key points to focus on in the future.

Data availability

Automatic Identification System and Lloyd's Maritime Intelligence data were purchased and are not publicly available. The ballast water data are available from https://invasions.si.edu/nbicdb?agree=true. The import and export data of ports are available at https://usatrade.census.gov/ for U.S. and http://www.shippingdata.cn/index_en.html for China. The lists of port environment and biogeographical information are cited from https://github.com/msaebi1993/Species-Flow-Networks/tree/master/data .

References

Albert, R. J., Lishman, J. M., & Saxena, J. R. (2013). Ballast water regulations and the move toward concentration-based numeric discharge limits. Ecological Applications, 23(2), 289-300.

Batista, W. R., Fernandes, F. C., Lopes, C. C., Lopes, R. S., Miller, W., & Ruiz, G. (2017). Which ballast water management system will you put aboard? Remnant anxieties: a mini-review. Environments, 4(3), 54.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). Regression diagnostics: Identifying influential data and sources of collinearity. John Wiley & Sons.

Butrón, A., Orive, E., & Madariaga, I. (2011). Potential risk of harmful algae transport by ballast waters: The case of Bilbao Harbour. Marine pollution bulletin, 62(4), 747-757.

Campbell, A., Hudson, D., McLeod, C., Nicholls, C., & Pointon, A. (2013). The 2012–13 Paralytic Shellfish Toxin Event in Tasmania Associated with Dinoflagellate Alga, Alexandrium tamarense. Fisheries Research and Development Corporation, Canberra, Australia, 93.

Carlton, J. T. (1985). Transoceanic and interoceanic dispersal of coastal marine organisms: the biology of ballast water. Oceanography and Marine Biology, 23, 313-371.

Chan, F. T., Bailey, S. A., Wiley, C. J., & MacIsaac, H. J. (2013). Relative risk assessment for ballast-mediated invasions at Canadian Arctic ports. Biological invasions, 15(2), 295-308.

Cope, R. C., Prowse, T. A., Ross, J. V., Wittmann, T. A., & Cassey, P. (2015). Temporal modelling of ballast water discharge and ship-mediated invasion risk to Australia. Royal Society Open Science, 2(4), 150039.

David, M., & Gollasch, S. (2008). EU shipping in the dawn of managing the ballast water issue. Marine Pollution Bulletin, 56(12), 1966-1972.

David, M., Gollasch, S., Leppäkoski, E., & Hewitt, C. (2015). Risk assessment in ballast water management. In Global Maritime Transport and Ballast Water Management (pp. 133-169). Springer, Dordrecht.

David, M., Perkovič, M., Suban, V., & Gollasch, S. (2012). A generic ballast water discharge assessment model as a decision supporting tool in ballast water management. Decision support systems, 53(1), 175-185.

Endresen, Ø., Behrens, H. L., Brynestad, S., Andersen, A. B., & Skjong, R. (2004). Challenges in global ballast water management. Marine pollution bulletin, 48(7-8), 615-623.

Enshaei, H., & Mesbahi, E. (2011). Development of a methodology for estimation of ballast water operation in UK ports. Maritime Policy & Management, 38(7), 691-703.

Gollasch, S. (2006). Overview on introduced aquatic species in European navigational and adjacent waters. Helgoland Marine Research, 60(2), 84-89.

Hewitt, C. L. (2003). Marine Biosecurity Issuesin theWorldOceans: GlobalActivitiesandAustralian Directions. Ocean Yearbook Online, 17(1), 193-212.

Hewitt, C. L., & Campbell, M. L. (2007). Mechanisms for the prevention of marine bioinvasions for better biosecurity. Marine pollution bulletin, 55(7-9), 395-401.

Hewitt, C. L., Campbell, M. L., McEnnulty, F., Moore, K. M., Murfet, N. B., Robertson, B., & Schaffelke, B. (2005). Efficacy of physical removal of a marine pest: the introduced kelp Undaria pinnatifida in a Tasmanian Marine Reserve. Biological Invasions, 7(2), 251-263.

Holzer, K. K., Muirhead, J. R., Minton, M. S., Carney, K. J., Miller, A. W., & Ruiz, G. M. (2017). Potential effects of LNG trade shift on transfer of ballast water and biota by ships. Science of The Total Environment, 580, 1470-1474.

Hou, W., Chen, X., Ba, M., Yu, J., Chen, T., Zhu, Y., & Bai, J. (2022). Characteristics of Harmful Algal Species in the Coastal Waters of China from 1990 to 2017. Toxins, 14(3), 160.

IMCO, 1973. In: Proceedings of the Resolution 18, Conference for Adoption of the International Convention for Prevention of Pollution from Ships.

IMO, M. (2004). International convention for the control and management of ships' ballast water and sediments. In BWM/CONF./36.

Keller, R. P., Drake, J. M., Drew, M. B., & Lodge, D. M. (2011). Linking environmental conditions and ship movements to estimate invasive species transport across the global shipping network. Diversity and Distributions, 17(1), 93-102.

Lavoie, D. M., Smith, L. D., & Ruiz, G. M. (1999). The potential for intracoastal transfer of non-indigenous species in the ballast water of ships. Estuarine, Coastal and Shelf Science, 48(5), 551-564.

Lv, B., Jiang, T., Wei, H., Tian, W., Han, Y., Chen, L., ... & Cui, Y. (2021). Transfer of antibiotic-resistant bacteria via ballast water with a special focus on multiple antibiotic-resistant bacteria: A survey from an inland port in the Yangtze River. Marine Pollution Bulletin, 166, 112166.

MacIsaac, H. J., Borbely, J. V., Muirhead, J. R., & Graniero, P. A. (2004). Backcasting and forecasting biological invasions of inland lakes. Ecological Applications, 14(3), 773-783.

Maglić, L., Zec, D., & Frančić, V. (2015). Effectiveness of a barge-based ballast water treatment system for multi-terminal ports. Promet-Traffic&Transportation, 27(5), 429-437.

Matheickal J, Blonce A, Alonso J. 2017. The GloBallast story: reflections from a global family. Partnerships to catalyze transformational innovations in marine biosafety'. GloBallast Monograph Series No. 25. London: GEF-UNDP-IMO GloBallast Partnerships.

Miller, A. W., Lion, K., Minton, M. S., & Ruiz, G. M. (2007). Status and trends of ballast water management in the United States. Third biennial report of the National Ballast Information Clearinghouse (January 2004 to December 2005). Submitted to the United States Coast Guard, 3(3), 1-62.

Minchin, D. (2007). Aquaculture and transport in a changing environment: overlap and links in the spread of alien biota. Marine Pollution Bulletin, 55(7-9), 302-313.

Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment, 6(9), 485-492.

Naik, R. K., Naik, M. M., D'Costa, P. M., & Shaikh, F. (2019). Microplastics in ballast water as an emerging source and vector for harmful chemicals, antibiotics, metals, bacterial pathogens and HAB species: A potential risk to the marine environment and human health. Marine Pollution Bulletin, 149, 110525.

RESOLUTION MEPC.289(71) (adopted on 7 July 2017). 2017 GUIDELINES FOR RISK ASSESSMENT UNDER REGULATION A-4 OF THE BWM CONVENTION (G7).

RESOLUTION MEPC.290(71) (adopted on 7 July 2017). THE EXPERIENCE-BUILDING PHASE ASSOCIATED WITH THE BWM CONVENTION.

RESOLUTION MEPC.297(72) (adopted on 13 April 2018). AMENDMENTS TO THE INTERNATIONAL CONVENTION FOR THE CONTROL AND MANAGEMENT OF SHIPS' BALLAST WATER AND SEDIMENTS, 2004.

Ruiz, G. M., Miller, A. W., Lion, K., Steves, B., Arnwine, A., Collinetti, E., & Wells, E. (2001). Status and trends of ballast water management in the United States. First Biennial Report of the National Ballast Information Clearinghouse. Submitted to United States Coast Guard, 16.

Saebi, M., Xu, J., Grey, E. K., Lodge, D. M., Corbett, J. J., & Chawla, N. (2020). Higher-order patterns of aquatic species spread through the global shipping network. Plos one, 15(7), e0220353.

Seebens, H., Gastner, M. T., Blasius, B., & Courchamp, F. (2013). The risk of marine bioinvasion caused by global shipping. Ecology letters, 16(6), 782-790.

Stopford, M. (2008). Maritime economics 3e. Routledge.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Veldhuis, M., ten Hallers, C., de la Rivière, E. B., Fuhr, F., Finke, J., Stehouwer, P. P., ... & van Slooten, C. (2010). Ballast water treatment systems:"Old" and "New" ones. WMU Journal of Maritime Affairs, 9(2), 213-222.

Verbeek, M. (2008). A guide to modern econometrics. John Wiley & Sons.

Verling, E., Ruiz, G. M., Smith, L. D., Galil, B., Miller, A. W., & Murphy, K. R. (2005). Supply-side invasion ecology: characterizing propagule pressure in coastal ecosystems. Proceedings of the Royal Society B: Biological Sciences, 272(1569), 1249-1257.

Verna, D. E., & Harris, B. P. (2016). Review of ballast water management policy and associated implications for Alaska. Marine Policy, 70, 13-21.

Verna, D. E., Minton, M. S., & Ruiz, G. M. (2021). Trade Exports Predict Regional Ballast Water Discharge by Ships in San Francisco Bay. Frontiers in Marine Science, 940.

Wan, Z., Shi, Z., Nie, A., Chen, J., & Wang, Z. (2021). Risk assessment of marine invasive species in Chinese ports introduced by the global shipping network. Marine Pollution Bulletin, 173, 112950.

Wang, H., Xiao, K., Wu, Z., Chen, J., Xiong, W., Wang, Z., ... & Bowler, P. A. (2020). Delta arrowhead (Sagittaria platyphylla) in the Yangtze River: an invasive aquatic plant and the potential ecological consequences. BioInvasions Record, 9(3).

Wang, Z., Saebi, M., Corbett, J. J., Grey, E. K., & Curasi, S. R. (2021). Integrated biological risk and cost model analysis supports a geopolitical shift in ballast water management. Environmental Science & Technology, 55(19), 12791-12800.

Wonham, M. J., Byers, J. E., Grosholz, E. D., & Leung, B. (2013). Modeling the relationship between propagule pressure and invasion risk to inform policy and management. Ecological Applications, 23(7), 1691-1706.

Wu, H., Chen, C., Wang, Q., Lin, J., & Xue, J. (2017). The biological content of ballast water in China: A review. Aquaculture and Fisheries, 2(6), 241-246.

Xu, H., Qiang, S., Genovesi, P., Ding, H., Wu, J., Meng, L., ... & Pysek, P. (2012). An inventory of invasive alien species in China. NeoBiota, 15, 1.

Yoong, J. L. X., & Enshaei, H. (2016). Development of a methodology for estimation of ballast water Imported to Australian ports. Journal of Basic and Applied Scientific Research, 6(5), 14-25.

Zhang, X., Bai, M., Tian, Y., Du, H., & Zhang, Z. (2017). The estimation for ballast water discharged to China from 2007 to 2014. Marine pollution bulletin, 124(1), 89-93.

Supplement materials

Table S1. The frequency distribution of monthly ballast water data for both ship types

Bulk Carrier				Tanker Ship			
port	Freq.	Percent	Cum.	port	Freq.	Percent	Cum.
Baltimore	120	2.74	2.74	Baltimore	120	2.73	2.73
Houston	120	2.74	5.48	BatonRouge	120	2.73	5.45
Jacksonville	120	2.74	8.22	CorpusChristi	120	2.73	8.18
LongBeach	120	2.74	10.96	Houston	120	2.73	10.91
LosAngeles	120	2.74	13.7	LakeCharles	120	2.73	13.63
Mobile	120	2.74	16.44	LongBeach	120	2.73	16.36
NewOrleans	120	2.74	19.18	LosAngeles	120	2.73	19.09
Savannah	120	2.74	21.92	Mobile	120	2.73	21.81
Tacoma	120	2.74	24.66	NewOrleans	120	2.73	24.54
Tampa	120	2.74	27.4	NewYork	120	2.73	27.27
Seattle	119	2.71	30.11	Pascagoula	120	2.73	29.99
NewYork	118	2.69	32.8	Philadelphia	120	2.73	32.72
Newark	118	2.69	35.49	PortArthur	120	2.73	35.45
Philadelphia	117	2.67	38.16	PortEverglades	120	2.73	38.17
Boston	116	2.64	40.8	Richmond	120	2.73	40.9
CorpusChristi	116	2.64	43.44	SanJuan	120	2.73	43.63
PortArthur	116	2.64	46.08	Savannah	120	2.73	46.35
PortEverglades	115	2.62	48.7	TexasCity	120	2.73	49.08
SanFrancisco	115	2.62	51.32	Beaumont	119	2.7	51.78
Wilmington (DE)	115	2.62	53.94	Boston	119	2.7	54.49
Wilmington (NC)	115	2.62	56.56	Freeport	119	2.7	57.19
Detroit	113	2.58	59.14	Jacksonville	119	2.7	59.9
BatonRouge	112	2.55	61.69	Newark	119	2.7	62.6
LakeCharles	112	2.55	64.24	Brownsville	116	2.64	65.24
Oakland	112	2.55	66.79	Galveston	116	2.64	67.87
SanJuan	108	2.46	69.25	Detroit	115	2.61	70.48
Beaumont	101	2.3	71.55	SanFrancisco	111	2.52	73.01
Chicago	99	2.26	73.81	Wilmington (NC)	109	2.48	75.48
Toledo-Sandusky	99	2.26	76.07	Wilmington (DE)	98	2.23	77.71
PanamaCity	96	2.19	78.26	Anchorage	95	2.16	79.87
Portland1	92	2.1	80.36	Tampa	89	2.02	81.89
Providence	82	1.87	82.23	Honolulu	86	1.95	83.84
Gulfport	77	1.76	83.99	Seattle	82	1.86	85.71
Stockton	70	1.6	85.59	PortCanaveral	80	1.82	87.53
Albany	68	1.55	87.14	Anacortes	76	1.73	89.25
Freeport	64	1.46	88.6	Paulsboro	72	1.64	90.89
Portland	59	1.34	89.94	Portland	71	1.61	92.5
Portsmouth	58	1.32	91.26	Tacoma	70	1.59	94.09
Vancouver	52	1.19	92.45	Bellingham	56	1.27	95.36
Cleveland	41	0.93	93.38	Portland1	43	0.98	96.34
Honolulu	41	0.93	94.31	Portsmouth	38	0.86	97.21

NewHaven	38	0.87	95.18		Ponce	33	0.75	97.96
Longview	34	0.78	95.96		WestPalmBeach	19	0.43	98.39
SanDiego	31	0.71	96.67		NewHaven	14	0.32	98.7
Ponce	30	0.68	97.35		Searsport	12	0.27	98.98
WestPalmBeach	26	0.59	97.94		Chicago	11	0.25	99.23
PortCanaveral	18	0.41	98.35		Oakland	8	0.18	99.41
Brunswick	15	0.34	98.69		Vancouver	6	0.14	99.55
Anchorage	12	0.27	98.96		Albany	4	0.09	99.64
Brownsville	8	0.18	99.14		Toledo-Sandusky	4	0.09	99.73
Galveston	8	0.18	99.32		GreenBay	3	0.07	99.8
Searsport	7	0.16	99.48		Cleveland	2	0.05	99.84
RedwoodCity	6	0.14	99.62		Longview	2	0.05	99.89
Milwaukee	4	0.09	99.71		PortAngeles	2	0.05	99.93
FallRiver	3	0.07	99.78		Providence	2	0.05	99.98
Camden	2	0.05	99.83		SanDiego	1	0.02	100
Pascagoula	2	0.05	99.88		Total	4401	100	
PortHuron	2	0.05	99.93					
TexasCity	2	0.05	99.98					
Beaufort-MoreheadCity	1	0.02	100					
Bridgeport	1	0.02	100.02					
Gary	1	0.02	100.04					
Total	4,387	100						