DATA-DRIVEN APPROACHES FOR INTELLIGENT SHIP FUEL CONSUMPTION MANAGEMENT: MACHINE LEARNING MODELS, OPTIMIZATION TECHNIQUES, AND DOMAIN KNOWLEDGE

HAOQING WANG

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Data-Driven Approaches for Intelligent Ship Fuel Consumption Management: Machine Learning Models, Optimization Techniques, and Domain Knowledge

Haoqing Wang

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

February 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

Haoqing Wang (Name of student)

Abstract

Maritime transport serves as the backbone of the global supply chain. Ship fuel consumption constitutes a significant portion of maritime transport costs, while its emissions pose substantial environmental risks. Predicting and optimizing ship fuel consumption under varying scenarios is a pivotal procedure in enhancing ship fuel efficiency and sustainable maritime transport. This thesis leverages real-world data, machine learning models, optimization techniques, and domain knowledge in the shipping industry to prompt intelligent fuel consumption management.

The first study designs an innovative post hoc algorithm to correct the predictions of fuel consumption obtained by the off-of-shelf machine learning model—classification and regression trees (CART). We call this algorithm "PH-CART", which serves as a semi-domain knowledge-aware decision tree to combine domain knowledge in maritime transport with CART using a linear optimization model. Based on a real-world dataset, the experiment demonstrates that PH-CART outperforms CART in terms of both accuracy and robustness. Furthermore, the PH-CART model exhibits a high level of interpretability because it generates a univariate function representing the relationship between ship sailing speed and fuel consumption. This function is developed by incorporating all contextual information from other variables. The study contributes to sustainable maritime transport

by offering more accurate and robust predictions of ship fuel consumption. Moreover, this study provides a new perspective by applying domain knowledge to an industry-specific issue in the transportation domain.

The second study uses domain knowledge to develop two innovative methods for predicting ship fuel consumption—the first is a physics-informed neural network (PI-NN) model that improves the interpretability of the black-box model while maintaining accuracy and the second is a mixed-integer quadratic optimization (MIQO) model that considers more forms of feature variable expressions in an additive white-box model. The proposed approaches address the tradeoff between model interpretability and model accuracy in ship fuel consumption prediction. The experiment results demonstrate that the PI-NN model improves the interpretability of the black-box model while preserving accuracy. The MIQO model considers alternative variable expressions, leading to the flexibility of the white-box model. Finally, SHapley Additive exPlanations (SHAP) is used to explain how each feature value contributes to the predictions of the black-box model, thereby providing insights into how each value of feature variables affects fuel consumption. This study provides a solution to the tradeoff between model interpretability and model accuracy and can promote the application of data-driven models in ship fuel consumption prediction. Moreover, this study gives implications for the application of explainable machine learning models in practice.

Many shipping companies are unwilling to share their raw data because of data privacy concerns. However, certain problems in the maritime industry become much more solvable or manageable if data are shared. In the third study, we develop a two-stage method based on federated learning (FL) and optimization techniques to predict ship fuel consumption and optimize ship sailing speed. Be-

cause FL only requires parameters rather than raw data to be shared during model training, it can achieve both information sharing and data privacy protection. Our experiments show that FL develops a more accurate ship fuel consumption prediction model in the first stage and thus helps obtain the optimal ship sailing speed setting in the second stage. The proposed two-stage method can reduce ship fuel consumption by 2.5%–7.5% compared to models using the initial individual data. Moreover, the proposed FL framework protects the data privacy of shipping companies while facilitating the sharing of information among shipping companies.

Keywords: maritime transport; ship fuel consumption management; green shipping; sailing speed optimization; machine learning; data-driven optimization.

Publications and Working Papers

Arising from the thesis

- Wang, H., Yan, R., Wang, S., Zhen, L. (2023). Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. *Transportation Research Part C: Emerging Technologies*, 157, 104361.

- Wang, H., Yan, R., Au, M. H., Wang, S., Jin, Y. J. (2023). Federated learning for green shipping optimization and management. *Advanced Engineering Informatics*, 56, 101994.

- Wang, H., Liu, Y., Wang, S., Zhen, L. (2025). A machine learning-enhanced algorithm incorporating domain knowledge for sustainable maritime transport. Under revision at *Transportation Research Part C: Emerging Technologies*.

Others

- Wang, H., Sun, Q., Wang, S. (2024). Data-driven models for optimizing second-hand ship trading strategies under contextual information. *Naval Research Logistics*, 1–17.

- Wang, H., Liu, Y., Du, Y., Wang, S. (2024). Optimal refund policy design for ship berthing appointment mechanism. *Annals of Operations Research*, 1–28.

- Wang, H., Wang, W., Jin, Z. (2024). Mechanism for allocating delay to constituent activities in project management. *Computers & Industrial Engineering*, 197, 110603.

- Yao, J., Yi, W., Wang, H., Zhen, L., Liu, Y. (2022). Stackelberg game model for construction waste disposal network design. *Automation in Construction*, 144, 104573.

- Wang, H., Liu, Y., Wang, S., Zhen, L. A debiasing cost-sensitive learning approach for optimal decisions in ship management. Submitted.

- Wang, H., Liu, Y., Wang, S., Zhen, L. Shipping network under emissions trading system. Submitted.

- Wang, H., Chen, M., Liu, S., Wang, S. Value of dispatching flexibility in on-demand delivery. Working Paper.

Acknowledgments

While organizing this thesis from the end of 2024 to the beginning of 2025, I became acutely aware of how swiftly time passes. I am grateful to many people who have supported me throughout this journey of exploration and growth. I feel truly fortunate to have their companionship along the way.

First of all, I would like to express my deepest gratitude to my advisor Professor Shuaian Wang. In the beginning, my academic foundation was limited. Prof. Wang guided me step by step. His creativity, passion, patience, and technical skills have helped me understand what rigorous academic research truly entails. It was Prof. Wang's continuous support that carried me to this milestone. Prof. Wang is also one of the hardest-working people I have ever met. He never pushes us, yet he is our best role model. Whenever I felt complacent or craved leisure, thinking of how tirelessly my advisor continued to excel despite already being outstanding, I knew I had no excuse to slack off or halt my progress. It is such an honor for me to have Professor Shuaian Wang as my advisor. I am also grateful to Prof. Yan Liu, who is my co-supervisor, for his support and guidance.

I would like to express my sincere gratitude to Dr. Sheng Liu, who served as my supervisor during my one-year visiting study at the University of Toronto, Rotman School of Management. Sheng is incredibly creative and possesses strong

expertise. After each meeting with him, I always walk away with valuable insights. I feel truly fortunate to have had the opportunity to conduct research under his guidance. Additionally, I would like to express my gratitude to Dr. Mingliu Chen from the University of Texas at Dallas for his guidance on the research I conducted during my visiting study. He is precise, kind, and patient. Both Sheng and Mingliu are not much older than me; they are rising stars, and their example has inspired me to keep moving forward.

I am also indebted to all of my co-authors: Prof. Lu Zhen, Prof. Qinghe Sun, Dr. Yuquan Du, Dr. Ran Yan, and among others. I am grateful for their guidance and advice.

There are many friends who have greatly accompanied and supported me. I thank all the members of Prof. Wang's group, especially Simon and Xi. We shared meals and spent enjoyable moments together. I am also very fortunate to have studied at MN037, where I met many wonderful people, including Silong, Xiaoqian, and others. Outside the department, I would like to thank all the friends I met at the badminton club, both in Hong Kong and Toronto. Thank you for bringing joy into my life.

Finally, I would like to express my heartfelt gratitude to my parents for their unwavering support and encouragement. They are my pillar of strength. I am fortunate to be their daughter.

Thanks to all the amazing people in my life, including those mentioned here and many others who are not.

Contents

List of Figures

List of Tables

LIST OF TABLES

Chapter 1

Introduction

1.1 Background

Maritime transport is the foundation of the global supply chain (Bläser et al., 2024; Zhou et al., 2019). It is responsible for transporting more than 80% of global trade in terms of volume (UNCTAD, 2022). Ships mainly rely on heavy oil as their source of power, which results in emissions (e.g., carbon dioxide and sulfur dioxide) being released into the environment. Therefore, there is an urgent need to reduce emissions. A possible solution is to use clean energy. However, implementing such measures can be costly and requires government support and advancement. The more cost-effective option is to address shipping emissions at the operational level (Psaraftis & Kontovas, 2014). This process typically consists of two main steps. The first step is to predict a ship's fuel consumption; the second step is then to optimize its sailing speed based on this prediction (Du et al., 2019; Yan et al., 2020).

Data-driven research has become a prominent topic in the maritime domain. Exploring how to utilize data to extract more information and guide maritime operations is a valuable subject of study. In the field of ship fuel consumption management, some literature, e.g., Du et al. (2019) and Yan et al. (2020), uses ship noon reports to train machine learning models. Following the two-stage paradigm, they first employ off-the-shelf machine learning models to predict ship fuel consumption and then use the predicted results to optimize the ship's sailing speed. Instead of relying on off-the-shelf machine learning models, this thesis explores how to tailor machine learning models to align with domain knowledge in the maritime industry or directly incorporate domain knowledge to modify the internal structure of the machine learning models. We also employ the two-stage paradigm to

examine the critical impact of more accurate fuel consumption predictions on fuel management.

1.2 Thesis Outline

This thesis comprises three studies that use data-driven approaches to achieve intelligent fuel consumption management. The thesis is structured around three interrelated studies as shown in Figure 1.1. Study I (i.e., Chapter 2) focuses on integrating domain knowledge with machine learning to enhance both predictive accuracy and model interpretability. This chapter investigates how we can incorporate domain knowledge and constraints to guide machine learning models toward more meaningful and actionable insights. Study II (i.e., Chapter 3) dives deeper into the trade-off between interpretability and accuracy in ship fuel consumption modeling. This chapter analyzes when and how simpler models can perform competitively with complex black-box models. Study III (i.e., Chapter 4) combines prediction and optimization, extends the research to federated learning, and proposes a privacy-preserving and distributed framework for fuel consumption prediction and sailing speed optimization. Chapter 4 addresses data sensitivity and decentralization issues in the maritime industry and enables collaborative model training across multiple data holders without compromising data privacy.

Chapter 2: Integrating Domain Knowledge with Machine Learning: Enhancing Predictive Accuracy and Interpretability.[1] Chapter 2 introduces a post-hoc algorithm designed to refine fuel consumption predictions generated by classification and regression trees (CART), a widely used method in ship fuel con-

[1] This study is under revision at *Transportation Research Part C: Emerging Technologies*.

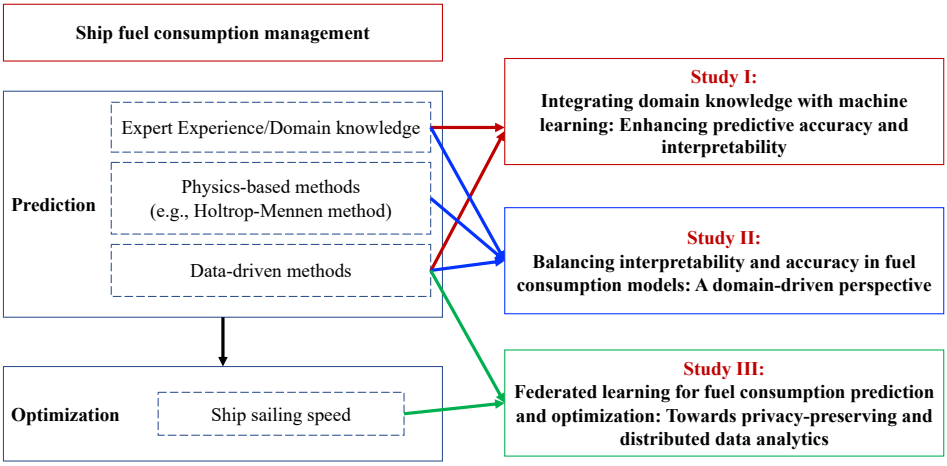


Figure 1.1: The framework the thesis

sumption prediction (Yan et al., 2020). Given that all other feature variables remain fixed, the relationship between the specified feature variable and the target predicted by CART follows a piecewise linear function. This piecewise linear function is entirely data-driven without any constraints, allowing it to be either monotonic or non-monotonic, which may contradict the domain knowledge in the shipping industry. Chapter 2 focuses on a well-known domain knowledge: *the relationship between ship sailing speed and fuel consumption is characterized as a convex non-decreasing function*. By solving a linear optimization model to obtain a piecewise linear surrogate function, the proposed post-hoc algorithm adjusts the CART results to align with domain knowledge, thereby enhancing prediction accuracy and robustness. Moreover, the piecewise linear surrogate function enhances the interpretability of machine learning models.

Chapter 3: Balancing Interpretability and Accuracy in Fuel Consumption

Models: A Domain-Driven Perspective.[2] Chapter 3 develops two models for ship fuel consumption prediction: one is a tailored neural network model, and the other is a mixed-integer quadratic optimization approach. The two proposed models strike a balance between model interpretability and model accuracy. We analyze and compare the performance of these two models, providing explanations and insights based on maritime domain knowledge. Finally, we employ SHapley Additive exPlanations (SHAP) to explain how each feature value contributes to the predictions of the final ship fuel consumption.

Chapter 4: Federated Learning for Fuel Consumption Prediction and Optimization: Towards Privacy-Preserving and Distributed Data Analytics.[3] Chapter 4 utilizes a new learning approach: federated learning, which enables decentralized training of machine learning models across multiple devices or data sources while keeping the data localized. As the maritime industry is a relatively traditional sector and is highly monopolized, data privacy and security pose significant barriers to the widespread adoption of data-driven methods. Federated learning provides a solution for data sharing and privacy protection. Chapter 4 validates the feasibility of federated learning in fuel consumption management. Moreover, we employ a two-stage paradigm, demonstrating that the more accurate prediction results achieved through federated learning can assist in optimizing ship sailing speed in the second stage.

At last, Chapter 5 concludes the thesis by summarizing the key findings and

[2]This study has been published: Wang, H., Yan, R., Wang, S., Zhen, L. (2023). Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. *Transportation Research Part C: Emerging Technologies*, 157, 104361.

[3]This study has been published: Wang, H., Yan, R., Au, M. H., Wang, S., Jin, Y. J. (2023). Federated learning for green shipping optimization and management. *Advanced Engineering Informatics*, 56, 101994.

discussing the emerging trends in fuel consumption prediction.

Chapter 2

Integrating Domain Knowledge with Machine Learning: Enhancing Predictive Accuracy and Interpretability

2.1 Introduction

Ship fuel consumption is affected by many factors, e.g., ship sailing speed, ship size, cargo weight, draft, and weather conditions (Lo & McCord, 1995). The most widely studied influencing factor is ship sailing speed because the physical characteristics of a ship are fixed. Both academia and industry recognize that there is a non-decreasing convex functional relationship between ship sailing speed and fuel consumption (Adland et al., 2020). Ronen (1982) mentioned that "the bunker fuel consumption of the main engines is directly related to the third power of the speed". In real-world scenarios, complex external conditions (e.g., waves and currents), the condition of the hull (e.g., fouling and wear), variations in engine efficiency, as well as ship design and operational practices, collectively contribute to deviations of the fuel consumption curve from the ideal cubic relationship (Edyvean, 2010; Fan et al., 2022). For example, using real-world data from a global liner shipping company, Wang and Meng (2012) find that there is a power function relationship between ship sailing speed and fuel consumption, with the exponent value ranging from 2.7 to 3.3 based on different ship types and voyage legs. Their results demonstrate that the third power relationship is indeed a good approximation. The roughly cubic relationship between a ship's sailing speed and fuel consumption, as well as the critical role of sailing speed in predicting fuel consumption, is well-established domain knowledge in maritime studies.

Recently, many studies have acknowledged the superior performance of machine learning models, as these advanced models are capable of incorporating a greater number of factors and capturing more complex relationships (Uyanık et al., 2020). Machine learning models that analyze the importance of different factors

further confirm that ship sailing speed is the most important variable in determining fuel consumption (Yan et al., 2020). However, most machine learning models are complex and off-the-shelf machine learning models cannot impose constraints on the relationship between two variables during the learning process based on data. In our context, if machine learning methods are applied to predict fuel consumption, we cannot guarantee the non-decreasing convex relationship between fuel consumption and ship sailing speed. Note that, in practical situations, ship sailing speed and fuel consumption do not follow a perfect cubic relationship. Therefore, the domain knowledge emphasized in our study is the non-decreasing convex relationship between ship sailing speed and fuel consumption. This relationship cannot be guaranteed by machine learning models, which motivates us to develop a semi-domain knowledge-aware machine learning model specifically designed for predicting ship fuel consumption. To explain further, since the machine learning model can ensure the important role of ship sailing speed in principle, the domain knowledge that we use in our study is that *the relationship between ship sailing speed and fuel consumption is characterized as a convex non-decreasing function*.

To incorporate domain knowledge in predicting ship fuel consumption into the machine learning model, we propose a post hoc correction algorithm that leverages the output results of the machine learning model. Specifically, we use a classification and regression tree (CART) model as the underlying machine learning model and introduce an infinite-dimensional optimization model to apply domain knowledge to correct the results. However, due to the inherent complexity of solving infinite-dimensional optimization problems, we transform the problem into a linear optimization model, supported by theoretical evidence. The proposed pro-

cess, including the CART model and post hoc correction using a linear optimization model, is called "PH-CART" and is proven to be more accurate and robust than the CART model. Furthermore, the PH-CART model ultimately produces a univariate non-decreasing linear piecewise convex function between ship sailing speed and fuel consumption. This function effectively incorporates contextual information from other variables to produce more accurate predictions and has high interpretability. Overall, the PH-CART model provides a fresh perspective, combining domain knowledge and the machine learning model to produce a model with both superior performance and high interpretability.

The remainder of this chapter is organized as follows. Section 2.2 summarizes the relevant literature and discusses our research contributions to theory and practice. Section 2.3 describes the research problem in detail, develops models, and proves theoretical evidence. Section 2.4 conducts experiments based on a public dataset and analyses the results. Finally, conclusions are drawn in Section 2.5.

2.2 Literature Review

In this section, we review two streams of studies that are closely related to our study: 1) ship fuel consumption prediction, and 2) interpretable models. We then summarize the contributions of our work in Section 2.2.3.

2.2.1 Research on Ship Fuel Consumption Prediction

The prediction of ship fuel consumption has attracted considerable attention in maritime studies due to the dominant role of fuel costs in ship operations (Meng et al., 2016). Fuel consumption is also recognized as a key factor in ensuring the sus-

tainability of shipping practices. Studies explore various approaches to predicting ship fuel consumption, aiming to provide valuable operational guidance to reduce fuel consumption and thus contribute to enhancing the efficiency and sustainability of ship operations (Ozsari, 2023; Yan et al., 2020). Early studies in the field of ship fuel consumption forecasting predominantly rely on hydromechanics principles (Holtrop & Mennen, 1982) and statistical models (Meng et al., 2016; Wang & Meng, 2012). In recent years, there has been a notable shift toward using ship operating data and advanced machine learning or deep learning models to predict fuel consumption. Our study also falls into this category. Data collection technologies have made it increasingly feasible to gather detailed information about ship operations, such as ship sailing speed, draft, and weather conditions. Taking advantage of such data, the development of data-driven models enables more precise and reliable predictions of fuel consumption. Wang et al. (2018a) propose to use LASSO regression to address the multicollinearity problem when applying variables to predict ship fuel consumption. Their approach is proven to be effective. Du et al. (2019) develop a neural network model to predict ship fuel consumption based on a ship's noon report—a report prepared by the captain at noon each day to describe the sailing profile to onshore officers. They find that the neural network model outperforms statistical models. Yan et al. (2020) use the random forest algorithm as a predictive tool for ship fuel consumption. The predicted value of ship fuel consumption also provides an efficient input to the downstream optimization model. Uyanık et al. (2020) conduct a comprehensive study of machine learning models used in ship fuel consumption. They compare the performance of various machine learning models, including ridge and LASSO regression, support vector regression, and tree-based algorithms. Le et al. (2020b) use a multilayer

perceptron artificial neural network to predict ship fuel consumption and validate the performance of the model using real-world data from container ships of different sizes. Readers are referred to Fan et al. (2022) for a comprehensive summary of all models adopted in predicting ship fuel consumption, as we only summarize studies using machine learning methods here.

Based on the above literature review on the use of machine learning models to predict ship fuel consumption, current research primarily focuses on applying machine learning models using ship operating data (Yan et al., 2020). However, there is a lack of research on combining domain knowledge in maritime transport with machine learning models. Integrating domain knowledge can enhance the accuracy and interpretability of such models. The best-known domain knowledge in ship fuel consumption is that the relationship between ship fuel consumption and sailing speed is a non-decreasing convex function (Wang & Meng, 2012). However, this relationship cannot be guaranteed in off-the-shelf machine learning models, which are typically designed to capture data patterns without explicitly considering domain-specific constraints or principles. Therefore, this study proposes the innovative PH-CART model to conduct a post hoc correction on the results of the CART model. PH-CART is a machine learning method that considers the relationship between ship fuel consumption and sailing speed, and is thus highly interpretable.

2.2.2 Research on Interpretable Models

Loyola-Gonzalez (2019) divide models into white-box models and black-box models. White-box models require no additional models to explain the results, while

black-box models are challenging to interpret, often requiring additional models to help explain the outcomes. Mainstream machine learning models (tree-based algorithms and neural network models) are all black-box models (Loyola-Gonzalez, 2019). Model interpretability has thus attracted attention in recent years. Lundberg (2017) provide milestone work in improving the interpretability of machine learning models. They propose the innovative SHapley Additive explanation (SHAP) method, which uses Shapley values from game theory to assign a SHAP value to each feature in each data sample. Their work is widely recognized and has over 14,000 citations. Many studies adopt SHAP to interpret the results of machine learning models based on their specific research questions (Wang et al., 2022b; Yan et al., 2022). For example, Yan et al. (2022) apply SHAP to interpret the results of gradient boosting regression trees in the port state control problem. Wang et al. (2022b) adopt SHAP to explain the prediction results of venue popularity and to further quantify the contribution of each variable. Chen et al. (2024) use SHAP to explain how prediction results are affected by class imbalance data. In general, interpretable machine learning models aim to explain the results of machine learning models, thereby making the principles of black-box models clearer and more convincing to industry experts. Our study also adopts a post hoc approach similar to SHAP, thus improving the interpretability of the machine learning model. In the domain of management science/operations management (MS/OM), some studies hold the opinion that model interpretability refers to the ability of a model to directly generate decision rules (Sun et al., 2022). For example, Ban et al. (2019) propose to use a residual tree method to prescribe the decision rules in a dynamic procurement problem. Sun et al. (2022) develop a risk-based port efficiency evaluation model, which provides an interpretable framework to quantify

the efficiency of a port. Elmachtoub and Grigas (2022) propose to revise the loss functions of decision trees to directly minimize decision loss and thus obtain optimal decision rules. From this point of view, our study also contributes to the interpretability of decision rules because we can obtain a highly interpretable univariate non-decreasing convex function that fits the domain-specific requirement and simultaneously takes into account information from other dimensions (e.g., weather conditions and draft).

Little research focuses on model interpretability in the context of ship fuel consumption prediction models. To the best of our knowledge, the study of Wang et al., 2023 is the only study that considers interpretability in predicting ship fuel consumption. However, their study focuses on developing two models—a highly interpretable regression model and a neural network model with high accuracy—to discuss the trade-off between interpretability and accuracy. In contrast, our study aims to correct the results of the CART model using domain knowledge and thus directly improve model interpretability. Moreover, our PH-CART model produces a function between ship fuel consumption and sailing speed, which can provide a reference for ship operators in various weather conditions and can thus be viewed as a decision rule under different conditions.

2.2.3 Research Contributions

The theoretical and practical contributions of our research are summarized as follows.

Theoretical contributions. The theoretical contributions of our study are highlighted throughout the literature review. First, our study tackles an important

issue in maritime studies—the prediction of ship fuel consumption. The proposed PH-CART model outperforms the off-the-shelf CART model in terms of accuracy and robustness. This improved performance can be attributed to the incorporation of domain knowledge pertaining to ship fuel consumption and sailing speed. Thus, in contrast to previous studies that solely apply machine learning models to predict ship fuel consumption, our research offers a fresh perspective by integrating domain knowledge with machine learning models. Second, our PH-CART model integrates machine learning techniques with optimization methods. We also provide theoretical evidence supporting the development of a linear optimization model to obtain a piecewise linear surrogate function. Third, the PH-CART model exhibits a high level of interpretability, as it provides a univariate function representing the relationship between ship fuel consumption and sailing speed. Although previous studies explore this relationship, they do not consider the incorporation of other contextual information. The PH-CART model demonstrates that incorporating contextual information contributes to more accurate predictions. Overall, the PH-CART model successfully leverages contextual information while simultaneously enhancing model interpretability.

Practical contributions. This study has two main practical applications. First, shipping is a traditional industry and expert opinion plays a pivotal role in the decision-making process of ship navigation. Advanced machine learning models are used infrequently in shipping practice due to the difficulty of explaining their results, which experts often struggle to interpret. Our study addresses this limitation by conducting a post hoc correction to align the outcomes of the machine learning model with domain-specific requirements. As a result, our research offers a novel perspective that can foster the practical application of machine learn-

ing models in the shipping industry. Second, as the PH-CART model generates more accurate and robust predictions, these results can be utilized in ship operations to support decision making during voyages. Consequently, our study also contributes to energy conservation and environmental protection.

2.3 Problem Description and Methodology

2.3.1 Preliminaries

State-of-the-art decision tree methods, e.g., CART (Breiman et al., 1984), C4.5 (Quinlan, 1993), and ID3 (Quinlan, 1986), are widely applied in classification and regression problems. They all take a top-down approach to determine the split criteria by solving an optimization problem. Amongst the various decision tree methods, CART stands out as a prominent technique (Bertsimas & Dunn, 2017), serving as the foundation for the widely recognized random forest method. However, relying solely on the principles of decision trees may not necessarily yield results that align with practical considerations or domain knowledge. Nanfack et al. (2023) emphasize that machine learning models need to comply with domain-specific requirements, which could also improve the interpretability of machine learning models.

To facilitate clear expression, we here represent the ship sailing speed (denoted by s) and other feature variables (denoted by $\boldsymbol{x}$) separately, despite their shared role as auxiliary data when predicting ship fuel consumption rate, denoted by y. The collected data is denoted by $S_N = \{(\boldsymbol{x}_i, s_i, y_i)\}_{i=1}^N$, where $i = 1, ..., N$ specifies a certain data record.

CART is proposed by Breiman et al. (1984). It provides a top-down mechanism to address both classification and regression problems by employing different split criteria. As ship fuel consumption rate is a continuous variable, we here briefly introduce the process of realizing regression by CART. Suppose $\boldsymbol{x}$ is a n-dimensional vector. In our case, there are a total of $n+1$ feature variables, including ship sailing speed as one of the variables. We use $x^{(j)}$ $(j = 1, ..., n+1)$ to denote the j-th feature variable. V_j represents the set of all feature values of $x^{(j)}$ and $v_j \in V_j$ specifies a certain feature value. From the root node, CART splits nodes by choosing the optimal pair $(x^{(j)}, v_j)$ that minimizes the sum of squared errors for both the left and right nodes. We denote the left space and right space by $R_1(j, v_j) = \{(\boldsymbol{x}, s)|x^{(j)} \leq v_j\}$ and $R_2(j, v_j) = \{(\boldsymbol{x}, s)|x^{(j)} > v_j\}$, respectively. The optimal pair $(x^{(j)}, v_j)$ is selected by minimizing the following problem (Breiman et al., 1984):

$$\begin{aligned} \min_{x^{(j)}, v_j \in V_j, j=1,...,n+1} & \left(\sum_{\{i|(\boldsymbol{x}_i, s_i) \in R_1(j,v_j)\}} \left(y_i - \frac{\sum_{\{i|(\boldsymbol{x}_i, s_i) \in R_1(j,v_j)\}} y_i}{|R_1|} \right)^2 \right. \\ & \left. + \sum_{\{i|(\boldsymbol{x}_i, s_i) \in R_2(j,v_j)\}} \left(y_i - \frac{\sum_{\{i|(\boldsymbol{x}_i, s_i) \in R_2(j,v_j)\}} y_i}{|R_2|} \right)^2 \right), \end{aligned} \tag{2.1}$$

where $|R_1|$ and $|R_2|$ denotes the number of samples in set R_1 and R_2, respectively. That is, the CART traverses all the possible pairs $(x^{(j)}, v_j)$ and then selects the optimal one that minimizes Problem (2.1). CART adopts two parameters to restrict the complexity of the tree: the maximum depth of a tree and the minimum sample number per leaf (Breiman et al., 1984). That is, after reaching the maximum depth of a tree or the minimum sample number per leaf, the tree stop splitting and we

get a constructed tree.

2.3.2 Problem Description

By collecting historical data $S_N = \{(\boldsymbol{x}_i, s_i, y_i)\}_{i=1}^N$, we first train a CART model, denoted by $T_\theta(\boldsymbol{x}, s)$, where θ represents the parameters of the constructed tree. For the newly observed data $(\boldsymbol{x}_0, s_0)$, we can obtain the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$ by inputting the vector of the feature variables $\boldsymbol{x}_0$, i.e., by reducing the space dimension to a two-dimensional plane with only fuel consumption rate and ship sailing speed through $\boldsymbol{x}_0$. Based on the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$, this study develops a non-decreasing piecewise linear surrogate function to calculate the final estimated fuel consumption rate. This process utilizes all information provided by feature variables and considers the domain-specific requirement. By and large, we develop a semi-domain knowledge-aware machine learning model in ship fuel consumption prediction, which could improve the prediction accuracy and model interpretability at the same time.

To simplify notation, we use $T_\theta(\boldsymbol{x}, s)$ to refer to the trained tree in the subsequent discussion. For the newly observed data $(\boldsymbol{x}_0, s_0)$, the prediction value of ship fuel consumption rate is obtained by averaging all samples belonging to the same leaf node as $(\boldsymbol{x}_0, s_0)$:

$$T_\theta(\boldsymbol{x}_0, s_0) = \sum_{i=1}^{N} \frac{1_{(\boldsymbol{x}_i, s_i) \in R(\boldsymbol{x}_0, s_0)}}{|\{(\boldsymbol{x}_j, s_j) \in R(\boldsymbol{x}_0, s_0)\}|} y_i, \tag{2.2}$$

where $\{(\boldsymbol{x}_i, s_i) \in R(\boldsymbol{x}_0, s_0)\}$ represents the set of samples belonging to the same leaf node as the new observation $(\boldsymbol{x}_0, s_0)$, and $|\{(\boldsymbol{x}_j, s_j) \in R(\boldsymbol{x}_0, s_0)\}|$ calculates the number of elements in the set. Additionally, by providing only $\boldsymbol{x}_0$ as input, we

can obtain the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$.

In maritime transport, predicting fuel consumption is a key process for guiding ship operations. Decision tree methods and their advanced counterparts (e.g., random forest) are widely used to predict ship fuel consumption (Li et al., 2022; Uyanık et al., 2020; Yan et al., 2020). However, no study has taken domain knowledge into account when developing tree-based models to predict ship fuel consumption. In the context of ship fuel consumption, the crucial domain knowledge is that the correlation between ship sailing speed and fuel consumption follows a non-decreasing convex function. However, no off-the-shelf machine learning model can guarantee this relationship.

In a tree-based model, the relationship between ship fuel consumption and ship sailing speed is a piecewise constant function, which is not non-decreasing convex. This is because the principle of the CART model is to divide the input space into multiple non-overlapping regions based on conditional rules, with the output in each region being either a fixed value or the result of a simple linear relationship. We illustrate this point by Example 1.

Example 1. Using the public dataset provided by Petersen (2011) (we will describe the details about this dataset in Section 2.4.1), we randomly select a voyage and randomly select 80% records of this voyage as the training dataset to train the CART model to predict the ship fuel consumption rate (unit: kg/s). Since our focus is primarily on highlighting the issue of piecewise constant functions, we here only select two feature variables: ship sailing speed through water (unit: knot) and the crosswind speed (unit: m/s), for the purpose of visualization. We present the statistical characteristics of the three selected variables for the voyage in Table 2.1. As Figure 2.1(a) shows, the two feature variables split the space into rectangles.

Given a new observation with the values of ship sailing speed and crosswind, the corresponding predicted fuel consumption rate can be found by mapping the point in the 3D space. Inspecting Figure 2.1(b), it shows that when the value of crosswind is given, the relationship between ship sailing speed and fuel consumption rate can be obtained by projecting the XOZ plane, which results in a piecewise constant function. Figure 2.2 shows the piecewise constant function when the value of crosswind is 5.60 m/s. Obviously, the result in Figure 2.2 is not in line with the domain knowledge in maritime transport—ship sailing speed and fuel consumption follow a non-decreasing convex function because it is neither non-decreasing nor convex.

Table 2.1: Statistical characteristics of the three selected variables

Variables	Mean	Standard deviation
Fuel consumption rate (kg/s)	0.570	0.117
Ship sailing speed (knot)	17.715	5.238
Crosswind speed (m/s)	6.242	4.298

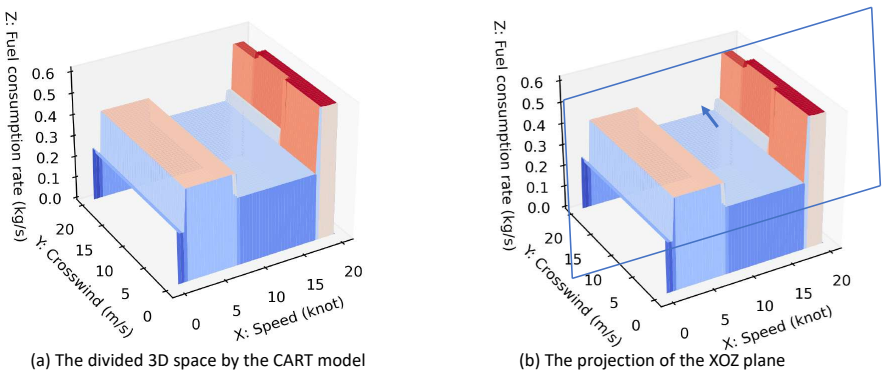


Figure 2.1: The result of the CART model

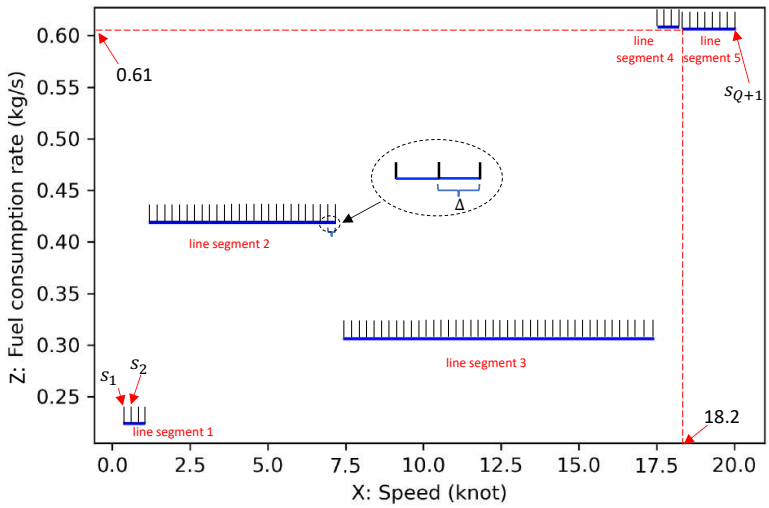


Figure 2.2: The piecewise constant function

2.3.3 PH-CART Model

In the PH-CART model, we need to conduct a post hoc correction on the relationship between ship sailing speed and fuel consumption rate. Given any new observation $\boldsymbol{x}_0$, we can obtain the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$. We use $P_{\boldsymbol{x}_0}$ to denote the number of line segments in the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$. For example, $P_{\boldsymbol{x}_0} = 5$ in Figure 2.2. For different $\boldsymbol{x}_0$, the value of $P_{\boldsymbol{x}_0}$ may be different. Moreover, the value of $P_{\boldsymbol{x}_0}$ is less than or equal to the number of leaf nodes in $T_\theta(\boldsymbol{x}, s)$ because we first utilize auxiliary data $\boldsymbol{x}_0$ to prescribe $T_\theta(\boldsymbol{x}_0, s)$. Still using Figure 2.1 as an example, there might be two leaf nodes with distinct split conditions like this: the split condition of one leaf node is that if ship sailing speed is larger than 18.20 knots and crosswind is larger than 7.11 m/s, fuel consumption rate is 0.70 kg/s; the split condition of another node is that if ship sailing speed is larger than 18.20 knots and crosswind is smaller than 6.88 m/s, fuel con-

sumption rate is 0.61 kg/s. Considering a specific sample where the crosswind is 5.60 m/s, the first leaf node would be eliminated after providing the crosswind input to derive the piecewise constant function between ship sailing speed and fuel consumption rate. Consequently, $P_{\boldsymbol{x}_0}$ remains less than or equal to the number of leaf nodes in $T_\theta(\boldsymbol{x}, s)$. In this section, we develop methods for achieving post hoc correction to make the results of CART in line with the domain knowledge in maritime transport—the correlation between ship sailing speed and fuel consumption follows a non-decreasing convex function.

Given the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$, we aim to find the optimal non-decreasing convex function $g^*_{\boldsymbol{x}_0}(s)$:

[M1]

$$g^*_{\boldsymbol{x}_0}(s) \in \underset{g_{\boldsymbol{x}_0}(s)\in\mathcal{G}}{\arg\min} \int_{s_{\min}}^{s_{\max}} |g_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds, \tag{2.3}$$

where $\mathcal{G} = \{g_{\boldsymbol{x}_0}(s) | g_{\boldsymbol{x}_0}(\lambda s_1 + (1-\lambda)s_2) \leq \lambda g_{\boldsymbol{x}_0}(s_1) + (1-\lambda) g_{\boldsymbol{x}_0}(s_2),\ 0 < \lambda < 1,\ s_{\min} \leq s_1 \leq s_{\max},\ s_{\min} \leq s_2 \leq s_{\max};\ g_{\boldsymbol{x}_0}(s_3) \leq g_{\boldsymbol{x}_0}(s_4),\ s_{\min} \leq s_3 < s_4 \leq s_{\max};\ g_{\boldsymbol{x}_0}(s) \geq y_{\min},\ g_{\boldsymbol{x}_0}(s) \leq y_{\max}\}$.

That is, $\mathcal{G}$ denotes the family of non-decreasing convex functions that satisfy $g_{\boldsymbol{x}_0}(s) \geq y_{\min}$ and $g_{\boldsymbol{x}_0}(s) \leq y_{\max}$. The $s_{\min}$ and $s_{\max}$ here denote the minimum and maximum ship sailing speeds and $y_{\min}$ and $y_{\max}$ are values of the minimum and maximum fuel consumption rate. For example, $s_{\min}$ and $s_{\max}$ could be the minimum and maximum values of the in-sample ship sailing speed[1], respectively; $y_{\min}$ and $y_{\max}$ could be the minimum and maximum values of the in-sample ship fuel

[1]Note that we use the minimum value and the maximum value of the in-sample data as the left and the right endpoint, respectively. If a smaller value or a larger value than the left endpoint or the right endpoint occurs in the test dataset, the left endpoint or the right endpoint should be that smaller or larger value in the test dataset.

consumption rate, respectively. Obviously, the optimal $g^*_{\boldsymbol{x}_0}(s)$ should be as close as possible to $T_\theta(\boldsymbol{x}_0, s)$ as it reflects the contextual information provided by $\boldsymbol{x}_0$. However, finding such an optimal $g^*_{\boldsymbol{x}_0}(s)$ is an infinite-dimensional optimization problem, which is difficult to handle (Warwicker & Rebennack, 2022). Therefore, we propose to develop a piecewise linear surrogate function for post hoc correction.

We denote the piecewise linear surrogate function by $f_{\boldsymbol{x}_0}(s)$. A small enough interval Δ is defined to discretize the line segments in $T_\theta(\boldsymbol{x}_0, s)$. For example, $\Delta = 0.001$ knot. Then, the total number of intervals is $Q = \frac{s_{\text{max}} - s_{\text{min}}}{\Delta}$. We assume that Q is an integer because we can always find a small enough interval Δ to split these line segments into integer sub-segments. As a result, we obtain $Q+1$ points, including the left and right endpoints (see Figure 2.2 for an example). We denote each point after discretization as $s_q, q = 1, ..., Q+1$. The approximated value of ship fuel consumption for each s_q is denoted by $\hat{y}_q$, which should be as close as possible to $T_\theta(\boldsymbol{x}_0, s_q)$. Connecting all $(s_q, \hat{y}_q), q = 1, ..., Q+1$, we obtain the piecewise linear surrogate function $f_{\boldsymbol{x}_0}(s)$:

$$\frac{f_{\boldsymbol{x}_0}(s) - \hat{y}_q}{\hat{y}_{q+1} - \hat{y}_q} = \frac{s - s_q}{s_{q+1} - s_q}, \; q = 1, ..., Q. \tag{2.4}$$

To approximate all the values of $\hat{y}_q$ and thus obtain the piecewise linear surrogate function $f_{\boldsymbol{x}_0}(s)$, four criteria are employed:

i) $f_{\boldsymbol{x}_0}(s)$ is a non-decreasing function;

ii) $f_{\boldsymbol{x}_0}(s)$ is a convex function;

iii) $f_{\boldsymbol{x}_0}(s) \geq y_{\text{min}}$ and $f_{\boldsymbol{x}_0}(s) \leq y_{\text{max}}$;

iv) $f_{\boldsymbol{x}_0}(s)$ is as close as possible to $T_\theta(\boldsymbol{x}_0, s)$.

Our objective is to identify the optimal piecewise linear surrogate function $f^*_{\boldsymbol{x}_0}(\cdot)$ from the family of all possible piecewise linear surrogate functions $\mathcal{F}$: [M2]

$$f^*_{\boldsymbol{x}_0}(\cdot) \in \underset{f_{\boldsymbol{x}_0}(\cdot)\in\mathcal{F}}{\arg\min} \int_{s_{\min}}^{s_{\max}} |f_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds, \tag{2.5}$$

where

$$\mathcal{F} = \left\{ f_{\boldsymbol{x}_0}(s) \;\middle|\; \begin{array}{c} f_{\boldsymbol{x}_0}(s) = \frac{(s-s_q)(\hat{y}_{q+1}-\hat{y}_1)}{\Delta} + \hat{y}_q, \\ s \in [s_q, s_{q+1}), s_1 = s_{\min}, s_{Q+1} = s_{\max}, \hat{y}_q \in \mathbb{R}, q = 1, \ldots, Q; \\ f_{\boldsymbol{x}_0}(\lambda s^{'} + (1-\lambda)s^{''}) \leq \lambda f_{\boldsymbol{x}_0}(s^{'}) + (1-\lambda) f_{\boldsymbol{x}_0}(s^{''}), 0 < \lambda < 1; \\ s_{\min} \leq s^{'} \leq s_{\max}, s_{\min} \leq s^{''} \leq s_{\max}; \\ f_{\boldsymbol{x}_0}(s^{'}) \leq f_{\boldsymbol{x}_0}(s^{''}),\ s_{\min} \leq s^{'} < s^{''} \leq s_{\max}; \\ f_{\boldsymbol{x}_0}(s) \geq y_{\min},\ f_{\boldsymbol{x}_0}(s) \leq y_{\max} \end{array} \right\}. \tag{2.6}$$

Proposition 1 $\lim_{\Delta\to 0}(\int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds - \int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds) = 0.$

Proof. We originally aim to find the optimal $g^*_{\boldsymbol{x}_0}(s)$ by solving Problem (2.3), i.e., [M1]. Suppose that we split $g^*_{\boldsymbol{x}_0}(s)$ into Q sub-segments of equal length by Δ. By connecting all the points $(s_q, g^*_{\boldsymbol{x}_0}(s_q)),\ q = 1, ..., Q+1$, we can construct a piecewise linear function, denoted by $f^{\#}_{\boldsymbol{x}_0}(s)$. Obviously, $f^{\#}_{\boldsymbol{x}_0}(s) \in \mathcal{F}$. As shown in Figure 2.3, $f^{\#}_{\boldsymbol{x}_0}(s)$ must be on or above $g^*_{\boldsymbol{x}_0}(s)$ because $g^*_{\boldsymbol{x}_0}(s)$ is a convex function.

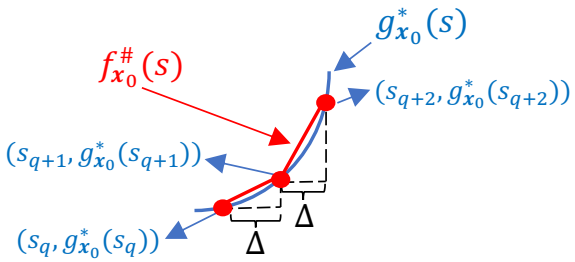


Figure 2.3: The illustration of Proposition 1

Therefore, the gap between $g^*_{\boldsymbol{x}_0}(s)$ and $f^{\#}_{\boldsymbol{x}_0}(s)$ is:

$$\begin{aligned} gap_1(\Delta) &= \int_{s_{\min}}^{s_{\max}} |f^{\#}_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds - \int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds \\ &= \sum_{q=1}^{Q} \left(\int_{s_q}^{s_{q+1}} |f^{\#}_{\boldsymbol{x}_0}(s_q) - T_\theta(\boldsymbol{x}_0, s)| ds - \int_{s_q}^{s_{q+1}} |g^*_{\boldsymbol{x}_0}(s_q) - T_\theta(\boldsymbol{x}_0, s)| ds \right). \end{aligned} \tag{2.7}$$

Moreover, $gap_1(\Delta)$ satisfies the following condition:

$$gap_1(\Delta) \leq \sum_{q=1}^{Q} \frac{\Delta}{2} \times (g^*_{\boldsymbol{x}_0}(s_{q+1}) - g^*_{\boldsymbol{x}_0}(s_q)) \leq \frac{\Delta}{2} \times (g^*_{\boldsymbol{x}_0}(s_{\max}) - g^*_{\boldsymbol{x}_0}(s_{\min})). \tag{2.8}$$

As $y_{\min} \leq g^*_{\boldsymbol{x}_0}(s) \leq y_{\max}$, we have:

$$gap_1(\Delta) \leq \frac{\Delta}{2} \times (y_{\max} - y_{\min}). \tag{2.9}$$

When $\Delta \to 0$, $\frac{\Delta}{2} \times (y_{\max} - y_{\min}) \to 0$. Therefore, when $\Delta \to 0$, $gap_1(\Delta)$ can

approach 0.

Because $f^*_{\boldsymbol{x}_0}(s)$ is the optimal solution of [M2], we have:

$$\int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds \leq \int_{s_{\min}}^{s_{\max}} |f^{\#}_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds. \tag{2.10}$$

Additionally, because $g^*_{\boldsymbol{x}_0}(s)$ is the optimal solution of [M1] and $\mathcal{F} \subseteq \mathcal{G}$, we have:

$$\int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds \leq \int_{s_{\min}}^{s_{\max}} |f^{\#}_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds. \tag{2.11}$$

and

$$\int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds \leq \int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds. \tag{2.12}$$

Therefore, the gap between $g^*_{\boldsymbol{x}_0}(s)$ and $f^*_{\boldsymbol{x}_0}(s)$, denoted by $gap_2(\Delta)$, must be no more than gap_1 because:

$$\begin{aligned} gap_2(\Delta) &= \int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds \\ &\leq \int_{s_{\min}}^{s_{\max}} |f^{\#}_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds \\ &= gap_1(\Delta) \end{aligned} \tag{2.13}$$

Thus, $0 \leq gap_2(\Delta) \leq gap_1(\Delta)$. According to Squeeze theorem, $gap_2(\Delta) \to 0$ when $\Delta \to 0$ because we prove $gap_1(\Delta) \to 0$ when $\Delta \to 0$ in Formula (2.9). We can conclude that $\lim_{\Delta \to 0}(\int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \int_{s_{\min}}^{s_{\max}} |g^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds) = 0$. □

Through Proposition 1, we establish the proof that the optimal piecewise lin-

ear surrogate function $f^*_{\boldsymbol{x}_0}(s)$ can serve as a highly accurate approximation to $g^*_{\boldsymbol{x}_0}(s)$. Additionally, the gap between $g^*_{\boldsymbol{x}_0}(s)$ and $f^*_{\boldsymbol{x}_0}(s)$ could approach 0 when Δ approach 0. Proposition 1 underscores the rationale behind the development and utilization of a piecewise linear surrogate function. However, calculating $\int_{s_{\min}}^{s_{\max}} |f_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)| ds$ in [M2] poses challenges due to the complex relationship between the values of $f_{\boldsymbol{x}_0}(s)$ and $T_\theta(\boldsymbol{x}_0, s)$. To be more specific, it is hard to analyze the magnitude relationship between the values of these two functions. This intricate relationship makes it difficult to determine and quantify the enclosed area between $f_{\boldsymbol{x}_0}(s)$ and $T_\theta(\boldsymbol{x}_0, s)$, thereby hindering the access to deriving $f^*_{\boldsymbol{x}_0}(s)$ by [M2]. Therefore, we next develop [M3] to minimize the absolute errors between $\hat{y}_q$ and $T_\theta(\boldsymbol{x}_0, s_q)$ multiplying by Δ.

[M3]

$$\min_{\hat{y}_q} \Delta \sum_{q=1}^{Q+1} |\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)| \tag{2.14}$$

subject to

$$\hat{y}_{q+1} \geq \hat{y}_q, \ q = 1, ..., Q \tag{2.15}$$

$$\frac{\hat{y}_{q+2} - \hat{y}_{q+1}}{s_{q+2} - s_{q+1}} \geq \frac{\hat{y}_{q+1} - \hat{y}_q}{s_{q+1} - s_q}, \ q = 1, ..., Q-1 \tag{2.16}$$

$$\hat{y}_q \geq 0, \ q = 1, ..., Q+1 \tag{2.17}$$

$$\hat{y}_1 \geq y_{\min} \tag{2.18}$$

$$\hat{y}_{Q+1} \leq y_{\max}, \tag{2.19}$$

where $\hat{y}_q, \ q = 1, ...Q+1$ denote the decision variables. Based on the optimal solutions of [M3], we can construct the piecewise linear surrogate function ac-

cording to Equation (2.4). We denote the piecewise linear surrogate function obtained by [M3] as $f_{\boldsymbol{x}_0}^{\mathrm{AE}*}(s)$. Constraints (2.15) ensure that the satisfaction of the first criterion—$f_{\boldsymbol{x}_0}^{\mathrm{AE}*}(s)$ is a non-decreasing function—by restricting the magnitude relationship between the adjacent two estimates. In Constraints (2.16), we restrict the slopes among three adjacent points to realize criterion ii), i.e., the convexity of the piecewise linear surrogate function. Note that $s_{q+2} - s_{q+1}$ is equal to $s_{q+1} - s_q$ in Constraints (2.16). We keep these terms in Constraints (2.16) for easy understanding. Constraints (2.17)–(2.19) guarantees criterion iii) by giving the domain of decision variables. Objective function (2.14) achieves criterion iv) by minimizing the sum of absolute errors between $\hat{y}_q$ and $T_\theta(\boldsymbol{x}_0, s_q)$. We next prove that minimizing the sum of absolute errors is equivalent to solving problem (2.5), i.e., [M2], when $\Delta \to 0$. That is, the optimal piecewise linear surrogate function $f_{\boldsymbol{x}_0}^{\mathrm{AE}*}(s)$ prescribed by [M3] is equivalent to the optimal piecewise linear surrogate function $f_{\boldsymbol{x}_0}^{*}(s)$ obtained by [M2] when $\Delta \to 0$.

Proposition 2 $\lim_{\Delta\to 0}(\int_{s_{\min}}^{s_{\max}} |f_{\boldsymbol{x}_0}^{*}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \int_{s_{\min}}^{s_{\max}} |f_{\boldsymbol{x}_0}^{AE*}(s) - T_\theta(\boldsymbol{x}_0, s)|ds) = 0$. *That is,* $\lim_{\Delta\to 0}(\int_{s_{\min}}^{s_{\max}} |f_{\boldsymbol{x}_0}^{*}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \Delta \sum_{q=1}^{Q+1} |\hat{y}_q^{*} - T_\theta(\boldsymbol{x}_0, s_q)|) = 0$, *where* $\hat{y}_q^{*}$, $q = 1, ..., Q$ *denotes the optimal solutions of [M3].*

Proof. We now prove that $f_{\boldsymbol{x}_0}^{\mathrm{AE}*}(s)$ highly approximates $f_{\boldsymbol{x}_0}^{*}(s)$. Without loss of generality, we use Figure 2.4 as an example to prove Proposition 2. The black line represents the piecewise constant function $T_\theta(\boldsymbol{x}_0, s)$, and the red line represents the piecewise linear surrogate function obtained by solving [M3]. There are four cases in terms of the enclosed area.

Firstly, for these piecewise lines that connect adjacent points within the same line segments but do not cross the line segment, e.g., $l_{6,7}$ in Figure 2.4, we should

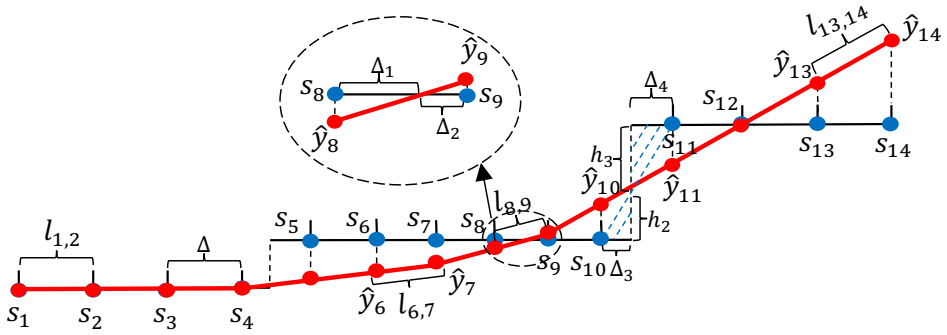


Figure 2.4: The illustration of Proposition 2

minimize the enclosed area $(|\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)| + |\hat{y}_{q+1} - T_\theta(\boldsymbol{x}_0, s_{q+1})|) \times \frac{\Delta}{2}$, which is equivalent to minimizing the corresponding objective function $\Delta(|\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)| + |\hat{y}_{q+1} - T_\theta(\boldsymbol{x}_0, s_{q+1})|)$ as Δ is a fixed constant. Actually, we can find that the error between $\hat{y}_q$ and $T_\theta(\boldsymbol{x}_0, s_q)$ contributes one-half to the area of the two trapezoids on each side. Therefore, minimizing the sum of absolute errors can directly lead to the minimum enclosed area with any value of Δ. Note that the situation where the piecewise line overlaps with the piecewise constant function (e.g., $l_{1,2}$ in Figure 2.4) can also be classified into this case.

Secondly, if the adjacent points cross the line segment, e.g., $l_{8,9}$ in Figure 2.4, the enclosed area equals $\frac{\Delta_1}{2} \times |\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)| + \frac{\Delta_2}{2} \times |\hat{y}_{q+1} - T_\theta(\boldsymbol{x}_0, s_{q+1})|$, which is $(\frac{\Delta_1}{2} + \frac{\Delta_2}{2} \times \frac{\Delta_2}{\Delta_1}) \times |\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)|$ (see Figure 2.4 for the meaning of new notations). Regarding the Objective function (2.14), the sum of absolute errors can be re-expressed as $\frac{\Delta}{2}(|\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)| + \frac{\Delta_2}{\Delta_1}|\hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q)|)$. Multiplying by $\frac{\Delta_1}{2} + \frac{\Delta_2}{2} \times \frac{\Delta_2}{\Delta_1}$ and multiplying by $\frac{\Delta}{2}(1 + \frac{\Delta_2}{\Delta_1})$ are not equivalent. And the

difference is $(\frac{\Delta}{2}(1+\frac{\Delta_2}{\Delta_1})-(\frac{\Delta_1}{2}+\frac{\Delta_2}{2}\times\frac{\Delta_2}{\Delta_1}))\times|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|$, which is $\Delta_2\times|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|$. Recall that we have $f^{\text{AE}*}_{\boldsymbol{x}_0}(s)\geq y_{\text{min}}$ and $f^{\text{AE}*}_{\boldsymbol{x}_0}(s)\leq y_{\text{max}}$. Therefore, $\Delta_2\times|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|\leq\Delta_2(y_{\text{max}}-y_{\text{min}})$. Additionally, there are at most $P_{\boldsymbol{x}_0}$ such cases because $T_\theta(\boldsymbol{x}_0,s)$ is a piecewise constant function and $f^{\text{AE}*}_{\boldsymbol{x}_0}(s)$ is a non-decreasing function. Therefore, in this case, the difference is no more than $P_{\boldsymbol{x}_0}\times\Delta_2(y_{\text{max}}-y_{\text{min}})$. As $\Delta\to 0$ and $\Delta_2\leq\Delta$, we have $P_{\boldsymbol{x}_0}\times\Delta_2(y_{\text{max}}-y_{\text{min}})\to 0$, which indicates that such $P_{\boldsymbol{x}_0}$ cases only have a minuscule impact.

Thirdly, we need to analyze the case that piecewise lines connecting adjacent points cross two line segments. As shown in Figure 2.4, the enclosed area between the second and third line segments is $\frac{|\hat{y}_{10}-T_\theta(\boldsymbol{x}_0,s_{10})|+h_2}{2}\times\Delta_3+\frac{h_3+|\hat{y}_{11}-T_\theta(\boldsymbol{x}_0,s_{11})|}{2}\times\Delta_4$. Additionally, the area of the two shaded triangles in Figure 2.4 is fixed because h_2+h_3 is pre-determined by $T_\theta(\boldsymbol{x}_0,s_q)$ and $\Delta_3+\Delta_4=\Delta$. Therefore, in this case, we need to minimize $\frac{|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|\times\Delta_3}{2}+\frac{|\hat{y}_{q+1}-T_\theta(\boldsymbol{x}_0,s_{q+1})|\times\Delta_4}{2}$. The objective function minimizes $\frac{\Delta}{2}(|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|+|\hat{y}_{q+1}-T_\theta(\boldsymbol{x}_0,s_{q+1})|)$. The difference is $\frac{\Delta-\Delta_3}{2}\times|\hat{y}_q-T_\theta(\boldsymbol{x}_0,s_q)|+\frac{\Delta-\Delta_4}{2}\times|\hat{y}_{q+1}-T_\theta(\boldsymbol{x}_0,s_{q+1})|$. There are at most $(P_{\boldsymbol{x}_0}-1)$ such cases. Same logic as the second case, the difference is no more than $(P_{\boldsymbol{x}_0}-1)\times(\frac{\Delta_4}{2}\times(y_{\text{max}}-y_{\text{min}})+\frac{\Delta_3}{2}\times(y_{\text{max}}-y_{\text{min}}))$, which is $(P_{\boldsymbol{x}_0}-1)\times\frac{\Delta}{2}\times(y_{\text{max}}-y_{\text{min}})$. When $\Delta\to 0$, the difference approaches 0.

Finally, we analyze the situation at the endpoints of the beginning and end of the whole piecewise constant function. The absolute error of the right endpoint, e.g., s_{14} in Figure 2.4, actually contributes $\frac{1}{2}$ to the enclosed area, which equals $\frac{\Delta}{2}\times|\hat{y}_{13}-T_\theta(\boldsymbol{x}_0,s_{13})|+\frac{\Delta}{2}\times|\hat{y}_{14}-T_\theta(\boldsymbol{x}_0,s_{14})|$. This case is different from the first case because s_{14} is the endpoint and thus only contributes $\frac{1}{2}$ to the enclosed area. Therefore, there is a difference between the objective function and the enclosed area. And the difference equals $\frac{\Delta}{2}\times|\hat{y}_{Q+1}-T_\theta(\boldsymbol{x}_0,s_{Q+1})|$ for any right endpoint

s_{Q+1}. Obviously, $\frac{\Delta}{2} \times |\hat{y}_{Q+1} - T_\theta(\boldsymbol{x}_0, s_{Q+1})| \leq \frac{\Delta}{2} \times (y_{\max} - y_{\min})$. By the same logic, the difference for the left endpoint is $\frac{\Delta}{2} \times |\hat{y}_1 - T_\theta(\boldsymbol{x}_0, s_1)|$, which is also no more than $\frac{\Delta}{2} \times (y_{\max} - y_{\min})$. Therefore, the total difference at the endpoints is no more than $\Delta \times (y_{\max} - y_{\min})$. As $\Delta \to 0$, this difference approaches 0.

Considering the aforementioned four cases, the value of $(\int_{s_{\min}}^{s_{\max}} |f^*_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds - \int_{s_{\min}}^{s_{\max}} |f^{\text{AE}*}_{\boldsymbol{x}_0}(s) - T_\theta(\boldsymbol{x}_0, s)|ds)$ is no more than $P_{\boldsymbol{x}_0} \times \Delta \times (y_{\max} - y_{\min}) + (P_{\boldsymbol{x}_0} - 1) \times \frac{\Delta}{2} \times (y_{\max} - y_{\min}) + \Delta \times (y_{\max} - y_{\min})$, which approaches 0 when $\Delta \to 0$. Therefore, we can conclude that the piecewise linear surrogate function $f^{\text{AE}*}_{\boldsymbol{x}_0}(s)$ serves a high approximation to the piecewise linear surrogate function $f^*_{\boldsymbol{x}_0}(s)$. □

Model [M3] is a nonlinear optimization problem and we further linearize it to the following model:

[M4]

$$\min_{\hat{y}, z_q} \sum_{q=1}^{Q+1} z_q. \tag{2.20}$$

subject to

(2.15)–(2.19)

$$z_q \geq \hat{y}_q - T_\theta(\boldsymbol{x}_0, s_q),\ q = 1, ..., Q+1 \tag{2.21}$$

$$z_q \geq T_\theta(\boldsymbol{x}_0, s_q) - \hat{y}_q,\ q = 1, ..., Q+1, \tag{2.22}$$

where z_q is the auxiliary decision variable that replaces the absolute in the objective function. Model [M4] is a linear optimization model that can be solved by off-the-shelf optimization solvers, such as CPLEX and Gurobi. Note that the last line in

$T_\theta(\boldsymbol{x}_0, s)$ has no right endpoint. For example, the fuel consumption rate predicted by the CART model in Figure 2.2 is 0.61 kg/s for all ship sailing speed larger than 18.2 knots. As discussed before, we view the maximum value of the in-sample data (s_{max}) as the right endpoint for the last line (Arora et al., 2023). And we adopt the minimum value of the in-sample data (s_{min}) as the left endpoint (Arora et al., 2023).

In summary, Proposition 1 guarantees the rationale behind developing a piecewise linear function as an approximation for the optimal convex function. Proposition 2, on the other hand, provides proof that the method used to find the piecewise surrogate linear function $f_{\boldsymbol{x}_0}^{\text{AE}*}(s)$ ensures optimality as $\Delta \to 0$. The algorithm of the PH-CART model is summarized below.

Algorithm 1: PH-CART

Input: Training dataset $S_N = \{(\boldsymbol{x}_i, s_i, y_i)\}_{i=1}^N$; testing dataset $S_M = \{(\boldsymbol{x}_i, s_i, y_i)\}_{i=1}^M$.

Output: The estimated values of the testing dataset by CART and the piecewise linear surrogate function.

begin

Step 1. $T_\theta(\boldsymbol{x}, s) \leftarrow$ Train the CART model by dataset $S_N = \{(\boldsymbol{x}_i, s_i, y_i)\}_{i=1}^N$ based on Problem (2.1)

Step 2. $s_{\max} = \max\{s_i, i = 1, ..., N\}$, $s_{\min} = \min\{s_i, i = 1, ..., N\}$, $Q = \frac{s_{\max} - s_{\min}}{\Delta}$

for $i = 1 : M$ **do**

$T_\theta(\boldsymbol{x}_i, s) \leftarrow$ Inputting $\boldsymbol{x}_i$ to $T_\theta(\boldsymbol{x}, s)$ to get the piecewise constant function

for $q = 1 : Q + 1$ **do**

$T_\theta(\boldsymbol{x}_i, s_q) \leftarrow$ Inputting s_q to $T_\theta(\boldsymbol{x}_i, s)$ to get the estimated value for s_q

end

// Note that the q mentioned below represents the subscript $q = 1, ..., Q + 1$, instead of inheriting the final result of the above loop

$\hat{y}_q, q = 1, ..., Q + 1 \leftarrow$ Solve the optimization model [M4]

Develop $f_{\boldsymbol{x}_i}^{\mathrm{AE}*}(s)$ based on $(s_q, \hat{y}_q^*), q = 1, ..., Q + 1$ and Equation (2.4)

Output the estimated value $f_{\boldsymbol{x}_i}^{\mathrm{AE}*}(s_i)$ and $T_\theta(\boldsymbol{x}_i, s_i)$

end

end

2.4 Experiment

2.4.1 Data

We use a public dataset[2] of ship fuel consumption provided by Petersen (2011) to conduct our experiment. Petersen (2011) uses sensors to collect data from a ferry sailing between Tórshavn and Suðuroy in the Faroe Islands, Denmark. They collected many variables related to ship navigation, e.g., fuel density, fuel volume flow rate, latitude, longitude, wind speed, and wind direction. For the reader's convenience, we show the detailed raw data provided by Petersen (2011) in Table 2.2. From the comprehensive set of variables examined in Petersen (2011), we perform further calculations on the data to derive the target variable (i.e., ship fuel consumption rate) and feature variables that are widely acknowledged in both industry and academia for their significant influence on ship fuel consumption.

[2] http://cogsys.imm.dtu.dk/propulsionmodelling/data.html

Table 2.2: Details of raw data provided by Petersen (2011)

No.	Collected items	Units
1	Fuel density	kg/L
2	Fuel volume flow rate	L/s
3	Inclinometer trim angle	degrees
4	Latitude	\
5	Longitude	\
6	Port level measurements	m
7	Starboard level measurements	m
8	Speed through water	knot
9	Port propeller pitch	−10–10 V
10	Port rudder angle	−10–10 V
11	Speed over ground	knot
12	Starboard propeller pitch	−10–10 V
13	Starboard rudder angle	−10–10 V
14	Track degree magnetic	degrees
15	Track degree true	degrees
16	True heading	degrees
17	Wind angle	degrees
18	Wind speed	m/s

Firstly, we calculate the fuel consumption rate (denoted by y_i) using the first two items in Table 2.2: y_i (kg/s) = Fuel density (kg/L) $\times$ Fuel volume flow rate (L/s).

Secondly, The inclinometer trim angle (i.e., the third item in Table 2.2) refers to the trim angle of the ship measured by an inclinometer. Specifically, it describes the longitudinal balance of the ship, i.e., the angle caused by the height difference between the bow and the stern. In practical operation, the range of variation in the

trim angle is usually small (typically within a few degrees). Therefore, its impact on the overall resistance of the ship is relatively limited. Intuitively, the latitude and longitude of the ship are not related to the ship's fuel consumption.

Thirdly, items No. 6 and No. 7 in Table 2.2 are important variables because they can be used to calculate the ship's draft, which significantly impacts the ship's hydrodynamic resistance and, consequently, affects fuel consumption. In Figure 2.5, we illustrate how the sensors are installed on the ship to measure the distance to the sea surface. There are two sensors on the port side and the starboard side, resulting in the two measurements of port level and starboard level. The draft on both sides can be calculated by subtracting the detected distance by the sensor multiplied by $\cos\alpha$ from the vertical distance between the sensor and the bottom of the hull. Petersen, 2011 provides the value of α and the vertical distance between the sensor and the bottom of the hull on both the port and starboard sides. Specifically, $\alpha = 19.0°$ with a distance of 19.3m on the port side, and $\alpha = 12.6°$ with a distance of 22.1m on the starboard side. Then, we can take the average of the draft on both the port and starboard sides to obtain the final overall draft.

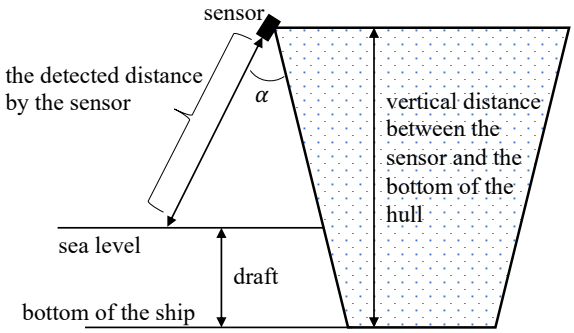


Figure 2.5: The detected distance by sensor

Fourthly, item No. 8 provides the required sailing speed.

Fifthly, items No. 9 to No. 15 do not contribute to our prediction of ship fuel consumption. Port propeller pitch (item No. 9), port rudder angle (item No. 10), starboard propeller pitch (item No. 12), and starboard rudder angle (item No. 13) are primarily responsible for the ship's navigation and do not have a direct impact on fuel consumption Since we already obtain speed through water, we do not need speed over ground (item No. 11). Track degree magnetic and track degree (items No. 14 and No. 15) true measure the ship's course relative to magnetic north and true north, respectively, and have no impact on the ship's fuel consumption.

Finally, by combining items No. 16 to No. 18, we can calculate the crosswind and headwind relative to the ship's sailing direction, which are external factors that influence the ship's fuel consumption.

By and large, the primary factor influencing fuel consumption is sailing speed. The ship's draft also has a significant impact on fuel consumption, as it directly determines the hydrodynamic resistance of the hull in the water. Additionally, there are other external environmental factors, e.g., hull fouling, wind speed, and water current. From the data, we can only obtain wind speed. To facilitate our analysis and model performance, we conduct calculations on these selected indicators, resulting in the generation of the feature vector $\boldsymbol{x}_i$. The final variables used and their corresponding calculation procedures are presented in Table 2.3.

We remove data records with null values and instances where the ship's sailing speed is below 8 knots, which indicates that the vessel has not entered a sailing state. Finally, our experimental dataset consists of 150,831 records, encompassing 244 voyages. From this dataset, we randomly allocate 200 voyages (comprising 123,243 records) for training purposes, while the remaining 44 voyages (compris-

Table 2.3: Data description

Variable	Description	Units	Calculation
y_i	The ship fuel consumption per second	kg/s	Fuel density times fuel volume flow rate
s_i	Ship sailing speed through water	knot	Directly using the raw data
$\boldsymbol{x}_i$			
draft	The average draft of the ship	m	Taking an average of the port level measurement and starboard level measurement by considering the angle between the sensor device and the hull.
headwind	Headwind speed	m/s	The relative wind speed times the cosine of the angle of the wind.
crosswind	Crosswind speed	m/s	The relative wind speed times the sine of the angle of the wind.

ing 27,588 records) are used for testing. In essence, our training phase involves using 200 voyages to train the CART model, followed by evaluating the performance of the trained CART model and the PH-CART model using the 44 test voyages. As the data change little during one voyage, we randomly select five samples from each voyage to fit the linear surrogate function. We thus ultimately fit 220 linear surrogate functions.

2.4.2 Experimental Settings

First, all of our experiments are performed on a MacBook Pro computer with an Apple M2 processor (3.5 GHz), 8 cores, and 16 GB of RAM. The optimization model [M2] is solved using IBM ILOG CPLEX Optimizer 20.1.0 via Python API.

Second, the hyperparameters (the maximum depth of a tree and the minimum number of samples per leaf) of CART are tuned using GridSearchCV and 5-fold

cross-validation in the training dataset. We summarize the search space and the optimal hyperparameters in Table 2.4.

Table 2.4: Searching space and the optimal hyperparameters of CART

Hyperparameter	Searching space	Optimal setting
The maximum depth of a tree	[4,5,6,7,8,9,10]	6
The minimum sample number per leaf	[1,2,3,4,5,6]	4

Third, the minimum and maximum values of the in-sample sailing speed of ships are 8.030 knots and 21.975 knots, respectively. The minimum and maximum values of the fuel consumption rate per second are 0.045 kg/s and 0.773 kg/s, respectively. Therefore, $s_{\text{min}} = 8.030$ knots, $s_{\text{max}} = 21.975$ knots, $y_{\text{min}} = 0.045$ kg/s, and $y_{\text{max}} = 0.773$ kg/s. To facilitate the selection of an appropriate value for Δ, we simplify the process by setting the minimum value of s (s_{min}) to 8.000 knots and the maximum value of s (s_{max}) to 22.000 knots during the experiment. Moreover, to avoid encountering test samples that exceed the y_{min} and y_{max} thresholds, we take a cautious approach during our experiment by expanding y_{min} to 0 kg/s and y_{max} to 1.000 kg/s. We conduct a thorough analysis and confirm that no samples in the test dataset exceed these thresholds established during our experiment. This reinforces the robustness and reliability of our approach.

Fourth, it is important to note that Problem (2.1) uses a quadratic loss function, in which the mean serves as the optimal estimate. In contrast, the optimization model [M3] is based on minimizing the sum of the mean absolute error (MAE) values, making the median the optimal estimate in this context. Considering this distinction, we select both MAE and the root mean squared error (RMSE) as evaluation metrics to assess the performance of CART and PH-CART. Moreover, we

calculate the variance of the errors between the real fuel consumption rate and the predicted fuel consumption rate obtained by CART and PH-CART. The variance of the errors serves as a metric to measure the robustness of the models.

2.4.3 Result

The results are presented in Table 2.5. As the unit of the fuel consumption rate is "kg/s," which leads to very small values for the error metrics, we report the error metrics by scaling them 100 times. Our results indicate that the accuracy of the PH-CART model surpasses that of the CART model when $\Delta = 0.1$, 0.01, and 0.001. As the value of Δ decreases, the accuracy of the PH-CART model improves. For instance, when Δ changes from 1 to 0.1, MAE decreases from 3.304 to 3.193, reflecting a notable improvement of 3.360% attributable to the PH-CART model. Additionally, RMSE decreases from 4.617 to 4.444, indicating an improvement of 3.747% using the PH-CART model. This is because as Δ decreases, the fitting accuracy of the piecewise linear function improves, leading to more precise fuel consumption predictions. Based on the results in Table 2.5, we can conclude that PH-CART achieves higher accuracy compared to CART. Upon examining the existing literature, we find that the accuracy performance of PH-CART is outstanding. For example, Ozsari (2023) adopts neural network models to predict the main engine power of container ships and achieves an accuracy level of 10^{-1}. Our results indicate an accuracy level of 10^{-2}. Although our study focuses on ferries, whose fuel consumption is typically lower than that of container ships, we believe that achieving an accuracy level of 10^{-2} is already highly commendable.

Moreover, our experimental results show that the PH-CART model exhibits

a lower error variance than the CART model when $\Delta = 0.1$. As Δ decreases, the robustness of the PH-CART model further improves. However, for $\Delta = 1$, the discretization interval proves too large to achieve the desired effect. On careful examination of Table 2.5, we observe that the accuracy and robustness of the PH-CART model exhibit minimal changes when Δ varies from 0.01 to 0.001. Therefore, we recommend setting $\Delta = 0.01$. This yields a solution time of less than 3 seconds for [M4] when $\Delta = 0.01$, which is perfectly acceptable in practical scenarios.

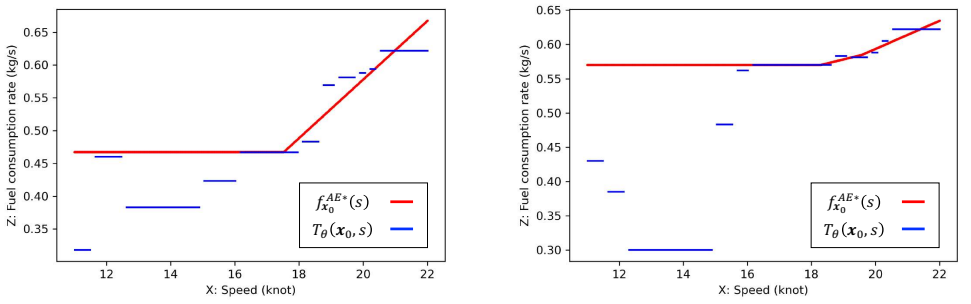
Table 2.5: Results of CART and PH-CART

Methods	CART	PH-CART $\Delta = 1$	PH-CART $\Delta = 0.1$	PH-CART $\Delta = 0.01$	PH-CART $\Delta = 0.001$
MAE ($\times$100)	3.206	3.304	3.193	3.180	3.180
RMSE ($\times$100)	4.517	4.617	4.444	4.437	4.436
Variance of errors ($\times$100)	0.202	0.213	0.197	0.197	0.196

We assume that the ship operator can make suitable operational adjustments to mitigate fuel consumption based on the predicted values, thereby eliminating the discrepancies between the two methods. By using the PH-CART model with $\Delta = 0.01$, the ship has the potential to significantly reduce fuel consumption, by 2246.4 kg per day (recall that in our dataset, the unit of ship fuel consumption is "kg/s"). This reduced fuel consumption is not only advantageous in terms of cost savings, as fuel constitutes a substantial portion of ship operating expenses, but it also contributes to the sustainable development of the shipping industry. By consuming less fuel, ships can reduce their environmental impact and help the maritime sector become more environmentally friendly and efficient.

We randomly select two examples of the optimal piecewise linear surrogate

function and visualize them in Figure 2.6. The optimal piecewise linear surrogate function captures the convex and non-decreasing relationship between ship sailing speed and fuel consumption by considering contextual information at the same time. Specifically, Figure 2.6(a) represents the case where the contextual information consists of a draft of 4.829 m, a headwind of 3.792 m/s, and a crosswind of 0.252 m/s, denoted by $\boldsymbol{x}_0 = [4.829, 3.792, 0.252]$. Based on $\boldsymbol{x}_0$, the dimensionality of the CART results is reduced to a univariate piecewise constant function (illustrated by the blue lines in Figure 2.6(a)). We then perform a post hoc correction on the piecewise constant function using PH-CART and thus obtain the optimal piecewise linear surrogate function (depicted by the red lines in Figure 2.6(a)). This optimal surrogate function enables to predict a ship's fuel consumption rate for any given value of sailing speed in the same context. Figure 2.6(b) shows another example with different contextual information. As the contextual information varies, the resulting mapped piecewise constant function also differs, yielding a distinct optimal piecewise linear surrogate function. Consequently, our PH-CART model is capable of generating various optimal piecewise linear surrogate functions based on different contextual information. It is noteworthy that while previous studies explore the relationship between ship sailing speed and fuel consumption, they focus solely on deriving a convex non-decreasing function between these two variables without considering additional contextual information. In contrast, our PH-CART model incorporates such contextual information in the post hoc correction process, resulting in the optimal $f_{\boldsymbol{x}_i}^{\mathrm{AE}*}(s)$, which effectively characterizes the relationship between ship sailing speed and fuel consumption.



(a) The optimal piecewise linear surrogate function—Example 1

(b) The optimal piecewise linear surrogate function—Example 2

Figure 2.6: The optimal piecewise linear surrogate function

2.5 Conclusion

In this study, we develop the PH-CART model to conduct a post hoc correction on the results obtained by the CART model. The PH-CART model takes into account the specific requirement of the maritime domain: the relationship between ship sailing speed and fuel consumption follows a non-decreasing convex function. To achieve this post hoc correction, we first propose an infinite-dimensional optimization model, which is difficult to solve. With the help of theoretical support, we finally convert it into a linear optimization model, which delivers a piecewise linear surrogate function that is consistent with domain knowledge. We examine the performance of the PH-CART and CART models using a publicly available shipping dataset. The experimental results show that PH-CART outperforms CART in terms of accuracy and robustness. Moreover, PH-CART can provide a univariate function between ship sailing speed and fuel consumption, thereby improving model interpretability. It is important to emphasize that this univariate function also considers the contextual information provided by other variables because PH-CART reduces the dimensionality of the CART model to two dimen-

sions based on the values of other variables. Therefore, PH-CART exhibits both high performance and high interpretability.

Chapter 3

Balancing Interpretability and Accuracy in Fuel Consumption Models: A Domain-Driven Perspective

3.1 Introduction

Ship navigation brings severe pollution to the environment as ships are mainly driven by heavy oil (Wang et al., 2022c). In order to promote the sustainable development of the shipping industry, governments and scholars all pay attention to improving ship fuel consumption efficiency and thus realizing green shipping (Fagerholt et al., 2015; IMO, 2020a; Meng et al., 2016; Wang et al., 2018b). For example, the International Maritime Organ (IMO) has promulgated a series of regulations to help achieve green shipping, such as the global sulfur content limit in fuel (IMO, 2022). Ship fuel consumption is a hot, important, and ongoing research topic in the field of maritime studies (Yang et al., 2019) because fuel cost dominates the costs of a ship (Meng et al., 2016) and generates emissions, which in turn affect sustainability (Wang et al., 2022c). Academic studies on ship fuel consumption abound and developing models to predict ship fuel consumption is a key research topic (Fan et al., 2022; Yan et al., 2021b). Some literature proposes advanced models to deliver accurate ship fuel consumption prediction. However, these advanced models in literature may be hard to be implemented in practice because they are difficult to interpret and thus experts are wary of relying on these models since shipping is a traditional industry (Yan et al., 2022) and domain knowledge plays an important role in decision-making. Therefore, developing interpretable ship fuel consumption prediction models using domain knowledge is urgently needed.

Ship fuel consumption prediction models in the literature are typically categorized as either black-box or white-box models, following the classification provided in Loyola-Gonzalez (2019). Black-box models are models based on hyper-

planes (e.g., support vector machines), biological neural networks such as those of animal brains (e.g., artificial neural networks), and probabilistic and combinatory logic (e.g., probabilistic logic networks) or models with local approximation functions (e.g., k-nearest neighbors) (Loyola-Gonzalez, 2019). White-box models are self-explanatory and do not require an additional model to explain the results (Loyola-Gonzalez, 2019). For example, linear regression is a typical white-box model. Using historical data, linear regression can yield an exact expression of the relationships between the feature variables and the estimated values. Moreover, the coefficients of the feature variables in the expressions show how these feature variables affect the outcome. White-box models are favored in practice because of their interpretability; black-box models usually have better prediction performance than white-box models (Parkes et al., 2018) but are less interpretable. Many recent studies explore the use of explainable artificial intelligence (XAI) for improving the explainability of black-box models (Gunning et al., 2019; Lundberg, 2017; Ribeiro et al., 2016). One way of achieving XAI is to develop an additional white-box model to interpret the results of a black-box model (Lundberg, 2017; Ribeiro et al., 2016; Sundararajan & Najmi, 2020). That is, a black-box model is first presented and then a white-box model is used to explain it; thus, the process involves ex-post explainability. This study is not limited to ex-post explainability. Instead, explainability or interpretability in this study indicates that the model is interpretable by itself or can be explained using a white-box model. Note that this study does not distinguish between the specific definitions of explainability and interpretability as there is no universal consensus on the definitions of either term (Doshi-Velez & Kim, 2017). The scope of explainability or interpretability in this study is in line with the definition in Doshi-Velez and Kim (2017)—"to explain or

to present in understandable terms".

3.1.1 Literature Review

We review two streams of literature that are closely related to our research: i) ship fuel consumption prediction models; and ii) interpretable models in maritime studies.

3.1.1.1 Ship Fuel Consumption Prediction Models

Fuel consumption is a critical factor in ship routing as it generates high costs and environmental pollution. There are many studies that take into account ship fuel consumption when optimizing ship sailing speed (Fagerholt et al., 2010). In recent years, with the development of informatization and the accumulation of data, an increasing number of studies have recognized the new insights that data can provide for fuel consumption prediction. For example, Du et al. (2019) study the ship sailing speed and trim optimization problem. They first predict fuel consumption rates using a neural network model based on the noon report data and then optimize the speed of the shipping route by the dynamic programming algorithm. Utilizing historical ship voyage data to develop more accurate fuel consumption prediction models, thereby optimizing vessel operations, has become a significant topic in maritime research (Yan et al., 2021b).

The literature on ship fuel consumption prediction models is divided into black-box models and white-box models, and details are shown in Table 3.1. In short, the critical issue in ship fuel consumption prediction is to obtain the influencing variables and then develop an accurate ship fuel consumption prediction model

based on these variables. The first stream of studies favors black-box models (Du et al., 2019; Le et al., 2020b; Yan et al., 2020), which usually outperform white-box models (Le et al., 2020b; Ma et al., 2023). However, the results of black-box models are hard to interpret. Sometimes, even experts in the maritime industry struggle to explain the outcomes of black-box models, and managers in shipping companies may consider it risky to apply models with low interpretability in practice. The second stream of studies presents statistic models which are white-box models with high interpretability. White-box models have advantages in exploring the explicit relationship between ship fuel consumption and its influencing factors (Adland et al., 2020; Le et al., 2020a; Meng et al., 2016; Wang & Meng, 2012). But they usually cannot capture complex interactions among feature variables and thus the prediction performance of white-box models is usually not as good as black-box models.

Therefore, there is a tradeoff in predicting ship fuel consumption: white-box models provide high interpretability but poor prediction performance, and black-box models provide low interpretability but good prediction performance. In practice, both model interpretability and accuracy are important (Carvalho et al., 2019; Loyola-Gonzalez, 2019). However, the literature does not address the tradeoff between interpretability and accuracy on the ship fuel consumption prediction problem. Studies on ship fuel consumption do not consider how to improve the interpretability of black-box models using constraints based on domain knowledge. Moreover, the literature mainly uses off-the-shelf white-box models and does not consider using domain knowledge available in the shipping field to develop more flexible white-box models by expanding the forms of feature variable expressions. Therefore, a theoretical solution to this tradeoff is urgently needed.

Table 3.1: Literature on ship fuel consumption prediction

Literature[1]	White-box model	Black-box model	Contents and findings
Wang and Meng (2012)	✓		There is a power function relationship between ship sailing speed and fuel consumption, and the power of sailing speed is between 2.7 and 3.3.
Meng et al. (2016)	✓		Analyze the relationship between fuel consumption and its influencing factors by Spearman's rank correlation coefficients.
Du et al. (2019)		✓	Develop neural network models to predict ship fuel consumption and optimize the ship sailing speed dynamically.
Yan et al. (2020)		✓	Adopt random forest to predict ship fuel consumption and optimize the ship sailing speed based on the predicted results.
Adland et al. (2020)	✓		Estimate the ship fuel consumption-speed curve and doubt the slow-steaming strategy based on empirical findings.
Le et al. (2020a)	✓		Adopt a linear regression model to predict ship fuel consumption.
Le et al. (2020b)	✓	✓	Develop a black-box multilayer perceptron artificial neural network (MLP) to predict ship fuel consumption and compare its prediction performance with two white-box multiple-regression models, showing the effectiveness of the MLP model.
Ma et al. (2023)	✓	✓	Develop both white-box model and black-box to predict ship fuel consumption and find that the white-box model has poor performance.
Uyanık et al. (2023)		✓	Develop decision tree model and neural network model to predict ship fuel consumption. The neural network model is proven to be more effective than the decision tree model.

[1] Note that studies on ship fuel consumption prediction are not limited to those listed in Table 3.1. As there are literature reviews on ship fuel consumption, this study does not go through it in a detailed way. Readers are referred to Yan et al. (2021b) and Fan et al. (2022) and the references therein.

3.1.1.2 Interpretable Models in Maritime Studies

A few recent studies focus on the interpretability of black-box models in the maritime domain. Kim and Lim (2022) propose machine learning models for predicting maritime accidents and develop additive white-box models based on SHAP

(SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to interpret the predicted values. Zhang et al. (2022a) also adopt SHAP to interpret the outcomes of tree-based machine learning models in predicting maritime accidents. Veerappa et al. (2022) use SHAP to explore the internal mechanisms for classifying the types of ships. He et al. (2021) identify the factors affecting ship detention using SHAP. Yan et al. (2022) adopt SHAP to provide explanations for the port state control problem. All of these studies develop posterior explanatory models—they first build a black-box machine learning model and then develop an additive white-box model based on SHAP to interpret the results of the black-box model. In detail, SHAP assigns an important value to each feature value for all the predictions (Kim et al., 2021). By aggregating the important values of all feature values in a sample and the mean value of the predicted target in the training dataset, SHAP develops an additive white-box model to explain the predicted values of each sample (Lundberg, 2017). Therefore, SHAP provides insights into how the predicted values are obtained in a black-box model and thus is suitable for ex-post explanations. In this study, after proposing two self-explanatory models, SHAP is adopted to interpret the results of an off-of-shelf black-box model. SHAP can interpret how the black-box model arrives at each predicted value of ship fuel consumption and how ship fuel consumption is affected by feature variables of different values. To the best of our knowledge, this study is the first to explore the issue of interpretability in predicting ship fuel consumption, with the aim of providing a comprehensive solution to the tradeoff between interpretability and accuracy.

Based on domain knowledge, this study proposes two approaches to using domain knowledge to address the tradeoff between interpretability and accuracy in

predicting ship fuel consumption. This study differs from earlier studies on ship fuel consumption prediction in several ways. First, a black-box model is developed for predicting ship fuel consumption that uses the physics constraints identified in domain knowledge to improve the model's interpretability. Second, by considering different forms of feature variable expressions from the perspective of domain knowledge, a mixed-integer quadratic optimization (MIQO) model is solved to fit a linear regression model, which is an additive white-box model with a high level of flexibility. Last, SHAP is adopted to identify how black-box models yield predicted values and how a change in a feature value affects the predicted ship fuel consumption.

3.1.2 Objectives and Research Questions

Two innovative approaches are developed for balancing model interpretability and model accuracy. The first method applies the constraints of physics (i.e., domain knowledge) to construct a neural network model. This model is named the physics-informed neural network (PI-NN) model, which is on par with the fully-connected neural network (fully-NN) model in terms of performance but has high interpretability because the PI-NN model uses domain knowledge to make the neural network model more intuitive. The second method uses an MIQO model to fit a linear regression model by selecting the best form of variable expressions that influence ship fuel consumption. This model is called the MIQO model for solving the best forms (BF) of variable expressions (MIQO-BF). The MIQO-BF model is an additive white-box model that aims to give the best linear regression formula. The two proposed models provide solutions to the tradeoff between

interpretability and flexibility (i.e., accuracy). The PI-NN model helps improve the interpretability of black-box models while preserving prediction performance, and the MIQO-BF model allows the statistical models to consider more forms of variable expressions while maintaining explainability. Moreover, by solving the MIQO-BF model, this study gains insights into the forms in which the variables affect fuel consumption and yield the optimal linear regression model at the same time. The performance of the MIQO-BF model may be slightly poorer than that of the PI-NN model, but it is highly explainable.

By building the two models, this study answers the following three research questions.

Q1: To what extent can the PI-NN model explain the fully-NN model? That is, is there a way to build a convincing neural network model to predict ship fuel consumption using domain knowledge that improves model interpretability while maintaining model accuracy?

Q2: In what forms do feature variables affect fuel consumption? That is, what relationship does the MIQO-BF model obtain between the feature variables and ship fuel consumption? Is the obtained relationship explainable in practice?

Q3: What are the differences among the MIQO-BF, the PI-NN, and other artificial intelligence (AI) models (e.g., fully-NN) in terms of prediction performance?

In addition to the two proposed self-interpretable models, this study also uses SHAP for the posterior explanation of the black-box model. SHAP directly shows how the predicted values of ship fuel consumption are obtained from different feature values and offers an additive white-box model. Unlike the MIQO-BF model, SHAP has to be developed after the machine learning model is used. Thus, the MIQO-BF model, the PI-NN model, and SHAP explore the issue of interpretabil-

ity in ship fuel consumption prediction from different perspectives.

3.1.3 Innovation and Contributions

The theoretical and practical contributions of this research are summarized as follows.

Theoretical contributions. This study presents significant theoretical contributions in the context of the ship fuel consumption prediction problem by introducing two innovative models: the PI-NN model and the MIQO-BF model. First, applying the knowledge of physics to reconstruct a neural network model is an innovation in ship fuel consumption prediction. By dissecting the neural network into two components, one addressing air resistance and the other water resistance, interpretability is notably enhanced without compromising model accuracy. Empirical experiments confirm the effectiveness of the PI-NN model, underscoring the improvement in interpretability for black-box models. Second, the MIQO-BF model yields an optimal additive model by solving the MIQO programming. Unlike other additive models, MIQO-BF accommodates a wider range of feature variable expressions, thereby enhancing the flexibility of white-box models. Lastly, the application of SHAP for interpreting machine learning models, while not novel in itself, marks the rare instance of SHAP being used to explain ship fuel consumption predictions. These results emphasize the paramount role of ship sailing speed in fuel consumption, particularly in operational routes where the impact of wind is less pronounced.

Practical contributions. This study also offers substantial practical contributions. First, it provides practical solutions to the trade-off between model in-

terpretability and accuracy. Practitioners and managers can leverage their domain knowledge to enhance the interpretability of black-box models or opt for the MIQO-BF model, which is more flexible and considers various forms of feature variable expressions. Second, the findings establish clear relationships between feature variables and ship fuel consumption, offering invaluable insights into the key determinants of fuel consumption. Lastly, by enhancing the interpretability of black-box models, this study encourages the adoption of such models in practice. These advanced models exhibit high predictive performance and have the potential to reduce vessel emissions, thus benefiting both industry and the environment.

In summary, this study delivers comprehensive insights into the crucial issue of interpretability versus accuracy in ship fuel consumption prediction, resulting in significant theoretical and practical contributions. The introduction of innovative models, the demonstration of how domain knowledge can strike a balance between interpretability and accuracy, and the practical implications for other industries underscore the far-reaching impact of this research. Given the limited focus on model interpretability in previous studies, this research has the potential to promote the application of advanced, interpretable black-box models, ultimately leading to increased industry profitability and reduced environmental impact.

The remainder of this chapter is organized as follows. Section 3.2 develops methodologies: the PI-NN model and the additive MIQO-BF model. Section 3.3 introduces the dataset used in the experiments. Section 3.4 illustrates the settings for methods and shows the results. Section 3.4 also makes a further analysis based on the results. Section 3.5 provides an ex-post way to understand the prediction results of machine learning models. Conclusions are presented in Section 3.6.

3.2 Methodologies

As shown in Figure 3.1, James et al. (2013) provide an illustration of interpretability and flexibility. Flexibility refers to the degree to which a model can capture different forms among the feature variables. For example, the linear regression model is restrictive as it can generate only a linear function between the input variables and the output. However, the linear regression model is easy to understand. In general, white-box models are highly interpretable but inflexible, whereas black-box models are flexible because they can capture complex relationships between the inputs and the output. Note that there is no clear metric in the literature for measuring model flexibility; hence, many studies use accuracy as an alternative (Gunning et al., 2019; James et al., 2013). This study argues that flexibility and accuracy are complementary—highly flexible models perform better as they can capture more complex relationships between the feature variables and the output. Therefore, this study does not strictly delineate accuracy and flexibility.

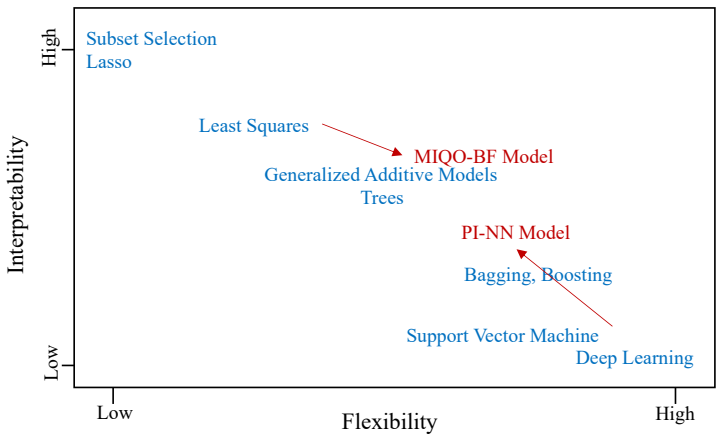


Figure 3.1: Tradeoff between interpretability and flexibility (excerpted from James et al., 2013 page 25 and adapted)

Figure 3.1 shows that the PI-NN model moves to a point of higher interpretability from the class of deep learning models. As the physics constraints are added to the neural network model, the flexibility of the PI-NN model decreases. The MIQO-BF model shifts from least squares to a point of higher flexibility because more forms of feature variable expressions are considered. However, the interpretability of the MIQO-BF model decreases slightly as it changes the original values of the feature variables. Therefore, this study provides two options for addressing the tradeoff between model interpretability and model accuracy in the ship fuel consumption prediction problem. The proposed PI-NN model makes the black-box model more interpretable while preserving accuracy. The MIQO-BF model considers more forms of the feature variable expressions while developing an explainable additive model. We next introduce the two models in detail.

3.2.1 PI-NN Model for Ship Fuel Consumption Prediction

The vector of feature variables is denoted by $\boldsymbol{x}$. When predicting ship fuel consumption, prevailing methods develop a model to solve the function:

$$y = F(\boldsymbol{x}), \tag{3.1}$$

where y kg/s represents the fuel consumption per second (the unit of ship fuel consumption can be changed according to the recorded data). Note that it may be difficult to give explicit expressions for some complex black-box models. Here Formula (3.1) is just adopted to emphasize that the current research usually directly inputs feature variables into the model without additional constraints based on domain knowledge.

From a physical point of view, oil consumed by ships is used to produce energy; specifically, burning oil is the process of converting chemical energy into internal energy and then converting internal energy into mechanical energy. The mechanical energy generated mainly overcomes the resistance of water and air and is finally converted into internal energy. According to Newton's Third law, a ship gains thrust and thus velocity. $E(y)$ denotes the energy generated by y kg/s fuel. Thus, the following formula can be formulated:

$$E(y) \propto E_A(\boldsymbol{x}_A) + E_W(\boldsymbol{x}_W), \tag{3.2}$$

where $E_A(\boldsymbol{x}_A)$ and $E_W(\boldsymbol{x}_W)$ are the energy used to overcome air resistance and water resistance, respectively, and $\boldsymbol{x}_A$ and $\boldsymbol{x}_W$ are the vectors of variables that affect air and water resistance, respectively. Therefore, from the perspective of domain knowledge or physics, different kinds of variables could be distinguished, i.e., $\boldsymbol{x}_A$ and $\boldsymbol{x}_W$, when building the fuel consumption prediction model to improve model interpretability and persuasiveness. Raissi et al. (2020) develop a physics-informed deep-learning framework that takes the Navier-Stokes equations into account. They add a hidden layer to capture the Navier-Stokes equations. Motivated by their research, this study proposes the PI-NN model that restricts the relationship shown in Formula (3.2).

The structure of the proposed PI-NN model is shown in Figure 3.2. The input layer consists of three types of variables: variables $\boldsymbol{x}_{A'}$ that only affect E_A, variables $\boldsymbol{x}_{W'}$ that only affect E_W, and variables $\boldsymbol{x}_{AW}$ that affect both E_A and E_W. For example, wind speed affects E_A (Meng et al., 2016), ocean current affects E_W (Chang et al., 2013), and draft affects both E_A and E_W (Rakke et al., 2012). The

history data of $\boldsymbol{x}_{A'}$ and $\boldsymbol{x}_{W'}$ will input to the separated two parts of hidden layers in Figure 3.2. And the history data of $\boldsymbol{x}_{AW}$ will be input to all hidden layers. The neurons of the last layer in the two split hidden layer parts will be connected to two different neurons, respectively. The values of these two neurons after the activation function f are denoted by $f(\hat{y}_A)$ and $f(\hat{y}_W)$, which represent the fuel consumed by the ship to overcome air resistance and water resistance, respectively. According to the above-mentioned physics (see Formula (3.2)), the sum of $f(\hat{y}_A)$ and $f(\hat{y}_W)$ should be the predicted value of ship fuel consumption, denoted by $\hat{y}$. The structure indicates that the PI-NN model prunes a fully connected neural network model based on domain knowledge. Specifically, by adopting domain knowledge, this study trains two neural network models in the hidden layer level and finally combines the output using an equation provided by physics constraint. The input variables of the two separated neural network models are classified by domain knowledge, which is also the pruning of the neural network model from the input layer. Moreover, if there is only one class of variables in the input layer, the PI-NN model will become an ordinary neural network model. Note that although there are two separate parts of neurons in the hidden layer, the PI-NN is an integrated neural network model as only the final fuel consumption data can be collected in practice.

According to the principle of the neural network model, the final predicted value $\hat{y}$ is obtained by multiplying input values and the weights of all connected neurons and then summing the values that are calculated by the activation function. The process of calculating the predicted value based on weights is called forward calculation. The key problem in the model training process is to get the optimal weights that connect consecutive neurons. The backpropagation method (Rumel-

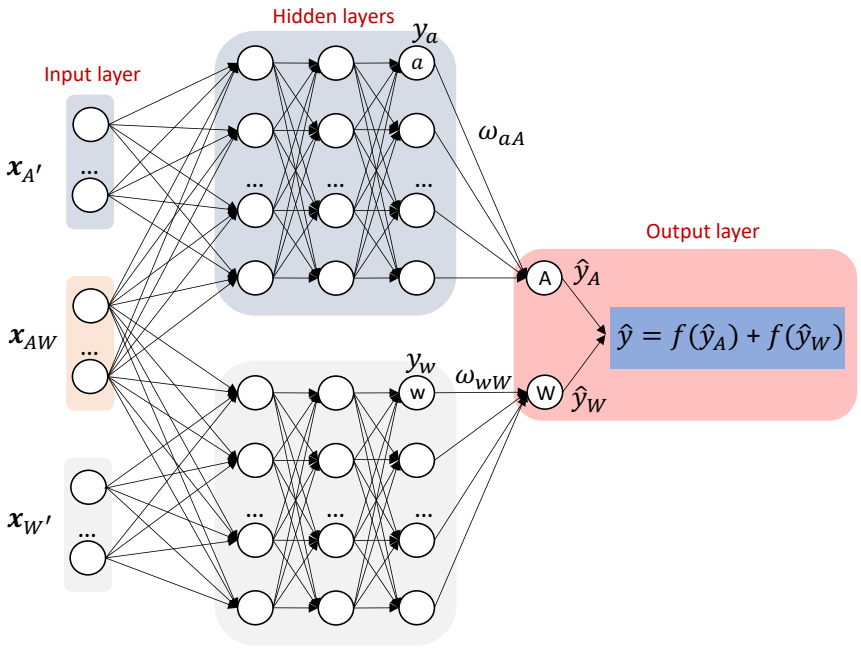


Figure 3.2: The structure of PI-NN model

hart et al., 1986) is used to obtain the optimal weights in the PI-NN model. Mean squared error (MSE) is used as the loss function since this study targets a regression problem:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2 = \frac{1}{2}(y - (f(\hat{y}_A) + f(\hat{y}_W)))^2. \tag{3.3}$$

Notations a and w denote the neurons in the last hidden layer of the two separated hidden layers, respectively. The weights of connected neurons between the last hidden layer and neurons A and W (see Figure 3.2) are denoted by ω_{aA} and ω_{wW}, respectively. According to the backpropagation method, the update increments of the weights ω_{aA} and ω_{wW} should be:

$$\begin{aligned}\Delta\omega_{aA} &= -\eta\frac{\partial L(y,\hat{y})}{\partial\omega_{aA}} \\ &= -\eta\frac{\partial L(y,\hat{y})}{\partial f(\hat{y}_A)}\frac{\partial f(\hat{y}_A)}{\partial\omega_{aA}} \\ &= \eta[y-(f(\hat{y}_A)+f(\hat{y}_W))](f^{'}(\hat{y}_A)\times y_a)\end{aligned} \tag{3.4}$$

$$\begin{aligned}\Delta\omega_{wW} &= -\eta\frac{\partial L(y,\hat{y})}{\partial\omega_{wW}} \\ &= -\eta\frac{\partial L(y,\hat{y})}{\partial f(\hat{y}_W)}\frac{\partial f(\hat{y}_W)}{\partial\omega_{wW}} \\ &= \eta[y-(f(\hat{y}_A)+f(\hat{y}_W))](f^{'}(\hat{y}_W)\times y_w),\end{aligned} \tag{3.5}$$

where y_a and y_w are the values of the neurons in the last hidden layer of the two separated hidden layers, respectively. And the values of y_a and y_w can be obtained by the forward calculation of the PI-NN model. η denotes the learning rate, which determines the convergence speed of the PI-NN model. The backpropagation process of other neurons is the same as the idea of the Formula (3.4) and Formula (3.5). In summary, the PI-NN model first randomly initializes each weight and conducts forward calculation to predict the ship fuel consumption, and then optimizes the weights according to the backpropagation method. For example, the updated values of ω_{aA} and ω_{wW} are $(\omega_{aA}+\Delta\omega_{aA})$ and $(\omega_{wW}+\Delta\omega_{wW})$, respectively. Then, the forward calculation is performed according to the updated weights and the backpropagation process is conducted again to optimize the weights. When the preset number of iterations is reached, the PI-NN model outputs the final predicted value.

The used dataset will be introduced in Section 3.3. The detailed hyperparameter settings, e.g., the number of neurons in each hidden layer, the learning rate η,

and the activation function, will be explained in Section 3.4. In summary, by introducing domain knowledge, the PI-NN model is proposed to give a solution to the tradeoff between accuracy and interpretability from the perspective of improving the interpretability of black-box models. The PI-NN model may not outperform the fully connected neural network model with the same network structure, but it can explain the black-box model at over 97% level as shown in results in Section 3.4.

3.2.2 MIQO-BF Model for Ship Fuel Consumption Prediction

Given that navigators or managers in shipping companies already have domain knowledge and have applied their knowledge and experience to make decisions for many years (Yan et al., 2021a), black-box models are not so widely used in practice in the shipping industry because even experts in the maritime industry struggle to interpret these models and thus hold the opinion that applying black-box models in practice is unreliable (Yan et al., 2022). Models with high interpretability are preferred in practice (Yan et al., 2022). In Section 3.2.1, the interpretability of black-box models is improved by domain knowledge. However, some white-box models, e.g., linear approximation, may be more prevalent in practice (Yan et al., 2021b) though they may not perform as well as black-box models (Le et al., 2020b; Parkes et al., 2018; Uyanık et al., 2020). Therefore, this section proposes a method to consider different forms of feature variable expressions and thus improve the flexibility of white-box models.

3.2.2.1 Preliminary

Linear regression is a common choice for developing highly explainable models. However, in some scenarios, there are many feature variables and informative feature variables need to be selected to build regression models that can interpret data accurately with high comprehensibility (Tan et al., 2008). The task of selecting $k, k \leq p$ out of p feature variables in a linear regression model given n observations, is the best subset selection problem (Natarajan, 1995). Given predictor matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, response vector $\boldsymbol{Y} \in \mathbb{R}^n$, and regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^p$, the best subset of feature variables can be obtained by solving the following non-convex problem:

$$\min_{\boldsymbol{\beta}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \qquad \text{subject to} \qquad \|\boldsymbol{\beta}\|_0 \leq k, \tag{3.6}$$

where $\|\boldsymbol{\beta}\|_0 = \sum_{i=1}^{p} \mathbb{I}(\beta_i \neq 0)$ and $\mathbb{I}(\cdot)$ denotes the indicator function. The best subset selection problem is an NP-hard problem (Natarajan, 1995). By solving Problem (3.6), the best k feature variables that interpret the target variable can be obtained.

Research on adopting optimization techniques to solve the best subset selection problem mainly lies in solving a convex approximation of Problem (3.6) (Bertsimas & King, 2016). Bertsimas et al. (2016) propose an MIQO approach to solve the best subset selection problem. By introducing binary decision variables that restrict the number of selected feature variables, the solution to the MIQO approach will be the solution to the best subset selection problem, i.e., Problem (3.6). The general MIQO formulation is

$$\min_{\boldsymbol{\beta},\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \tag{3.7}$$

subject to

$$\sum_{i=1}^{p} z_i \leq k \tag{3.8}$$

$$-Mz_i \leq \beta_i \leq Mz_i, i = 1, ..., p \tag{3.9}$$

$$z_i \in \{0, 1\}, i = 1, ..., p \tag{3.10}$$

$$\boldsymbol{\beta}, \boldsymbol{z} \in \mathbb{R}^p, \tag{3.11}$$

where z_i is a binary variable and $\sum_{i=1}^{p} z_i$ indicates the number of nonzeros in $\boldsymbol{\beta}$. That is, Constraint (3.8) ensures that the number of selected feature variables cannot exceed k. M is a constant that satisfies $M \geq \|\hat{\boldsymbol{\beta}}\|_\infty$, where $\hat{\boldsymbol{\beta}}$ is the vector of estimated coefficients. Constraints (3.9) guarantee that $\beta_i = 0$ if $z_i = 0$. MIQO is proven to handle small to moderate instances of the best subset selection problem (Hazimeh & Mazumder, 2020).

3.2.2.2 Method for Solving Additive Models Exactly

Referring to the idea that adopts the MIQO model to solve the best subset selection problem, an MIQO model is proposed to select the best forms of feature variable expressions. Suppose that there is a set of feature variables $F = \{1, ..., |F|\}$, where $f \in F$ indicates the index of feature variables and the total number of indexes is $|F|$. This study considers V different forms of feature variables and use $v = 1, ..., V$ to denote each form of expression. Suppose that all the feature variables are continuous variables. Therefore, the model ends up with $|F| \times V$ feature

variables. For example, there is a dataset of ship fuel consumption that contains 8 feature variables in the beginning. In addition to the 8 feature variables, the model also considers their logarithmic transformation, exponential transformation, quadratic and cubic transformation. That is, there are $|F| = 8$ feature variables in the beginning, $V = 5$ forms of expressions, and $|F| \times V = 40$ feature variables in the end. The task is to minimize the prediction error by selecting $|F|$ feature variables among $|F| \times V$ feature variables and guarantee that only one form of expression can be selected for the same index of feature variables. Although some heuristic algorithms for solving the best subset selection problem, such as forward stepwise and backward stepwise (Derksen & Keselman, 1992; Hastie et al., 2020), can be revised by adding the constraints of selecting one form of expression for the same index of feature variables, they do not guarantee to provide the optimal solutions (Derksen & Keselman, 1992). Therefore, MIQO programming is adopted to select the best forms of feature variable expression and this study abbreviates the model as MIQO-BF. Instead of original feature variables, the MIQO-BF model enables more forms of variables to be extended in a linear regression model and thus improves model flexibility. To the best of our knowledge, this is the first attempt to solve the variable expressions using MIQO. The objective of the MIQO-BF model is to minimize the MSE to select optimal forms of expressions of feature variables because the additive white-box model refers to linear regression in this study.

[MIQO-BF]

$$\min_{\boldsymbol{\beta},\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \tag{3.12}$$

subject to

$$\sum_{v=1}^{V} z_f^v = 1, f \in F \tag{3.13}$$

$$-Mz_f^v \leq \beta_f^v \leq Mz_f^v, f \in F, v = 1, ..., V \tag{3.14}$$

$$z_f^v \in \{0, 1\}, f \in F, v = 1, ..., V \tag{3.15}$$

$$\boldsymbol{\beta}, \boldsymbol{z} \in \mathbb{R}^{\mathcal{P}}, \tag{3.16}$$

where $\mathcal{P}$ denotes the value of $|F| \times V$, and thus $\boldsymbol{X} \in \mathbb{R}^{n \times \mathcal{P}}$, $\boldsymbol{Y} \in \mathbb{R}^n$, and $\boldsymbol{\beta} \in \mathbb{R}^{\mathcal{P}}$. Continuous decision variable β_f^v represents the coefficient value of each feature variable, and binary decision variable z_f^v represents whether a certain form of expression is selected. Constraints (3.13) ensure that one feature can only be formulated by one form of expression. Thus, the MIQO-BF model is a typical MIQO model, which can be solved by off-the-shelf optimization solvers.

The MIQO-BF model is developed to take different forms of feature variable expressions into account with the aim of maintaining the interpretability of the model and improving its flexibility. The MIQO-BF model gives the exact expression between ship fuel consumption and feature variables. That is, it can be known how feature variables affect ship fuel consumption from the MIQO-BF model, and then infer how the predicted value of fuel consumption is obtained given certain values of feature variables. Domain knowledge is also helpful in determining the forms of variables. For example, the widely recognized relationship between ship sailing speed and fuel consumption is cubic. Wang and Meng (2012) exactly show that the power of sailing speed is between 2.7 and 3.3 using data from five ships. Therefore, when transforming the variable of shipping sailing speed, the MIQO-BF model considers performing the power of 2.7, the power of 2.9, the power of 3.0, the power of 3.1, and the power of 3.3 transformations. Moreover, the obtained optimal expressions of variables also provide insights into the relationship

between ship fuel consumption and feature variables in turn. The detailed information on feature variables and their transformations is discussed in Section 3.4.

In summary, two ways are proposed to trade-off between model accuracy and model interpretability: improve the interpretability of black-box models while maintaining accuracy or provide more forms of feature variable expressions for white-box models while preserving interpretability. The first approach uses the PI-NN model, which combines domain knowledge and the black-box model to improve interpretability without losing too much accuracy. The second uses the MIQO-BF model, which is an explainable white-box model obtained by solving an MIQO model. Two solutions are provided for addressing the tradeoff by comparing the performance of the PI-NN model and the MIQO-BF model. That is, the PI-NN model is more suitable for cases that require a high level of accuracy whereas the MIQO-BF model is more suitable for cases that require a high level of interpretability. To the best of our knowledge, both approaches are innovative in predicting ship fuel consumption.

3.3 Data

A public dataset[1] of ship fuel consumption provided by Petersen (2011) is used for the experiment. As shown in Figure 3.3, there is a ferry sailing between Tórshavn and Suðuroy, Faroe islands. The sailing time of one voyage is about 2 hours. Taking advantage of sensors, the dataset in Petersen (2011) records the fuel consumption data and other relevant variables (e.g., port and starboard level measurements, speed through water, and wind speed) of the ferry (Petersen et al., 2012a; Petersen

[1] http://cogsys.imm.dtu.dk/propulsionmodelling/data.html

et al., 2012b). Readers are referred to Petersen (2011) for a detailed description of the data. The variables used in this study are shown in Table 3.2. In view of the different sampling frequencies of each sensor and thus the different statistical frequencies of each variable, 10 seconds is chosen as a unit to merge data. Next, this study introduces how to calculate the needed feature variables based on the originally recorded variables in Table 3.2.

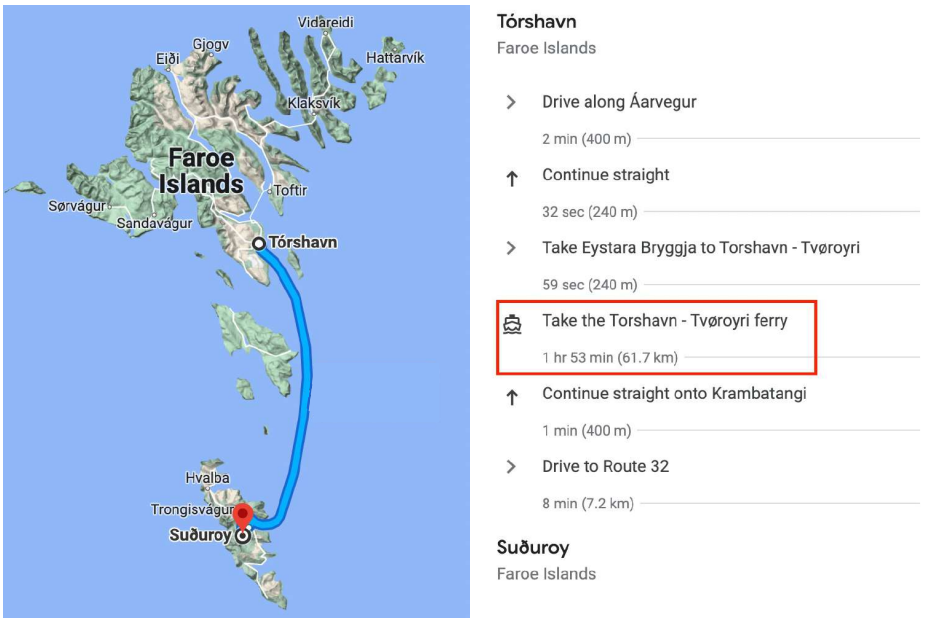


Figure 3.3: The shipping line between Tórshavn and Suðuroy

Draft. The ferry is equipped with two level measurement devices on the port side and the starboard side (Petersen, 2011). As shown in Figure 3.4, there is an angle θ between the device and the hull, and the device detects the distance to sea level (D). Given the vertical distance between the sensor and the bottom of the hull (H), the draft of the ship (d) when the ship is sailing can be calculated by the

Table 3.2: Data description

Variable	Description	Units
FD	Fuel density	kg/L
FV	Fuel volume flow rate	L/s
L_{draft}	Port level measurement	m
R_{draft}	Starboard level measurement	m
STW	Speed through water	knot
ρ	The angle of the wind relative to the heading direction of the ship	degrees
WS	Relative wind speed measured by an onboard sensor	m/s

following equation:

$$d\ (\text{m}) = H - D \times \cos\theta. \tag{3.17}$$

The installation parameters, i.e., H and θ, of the device are known. For the device on the port side, $H = 19.3$m and $\theta = 19°$ (Petersen, 2011). And for the device on the starboard side, $H = 22.1$m and $\theta = 12.6°$ (Petersen, 2011). The detected distance D, i.e., L_{draft} and R_{draft}, is recorded by the device. The average of the port and starboard drafts is taken as the ship's draft. Therefore, the final obtained draft of the ship, denoted by d_{avg}, is:

$$d_{\text{avg}}\ (\text{m}) = \frac{(19.3 - L_{\text{draft}} \times \cos 19°) + (22.1 - R_{\text{draft}} \times \cos 12.6°)}{2}. \tag{3.18}$$

Using knowledge of physics in shipping, d_{avg} is classified as a variable affecting both E_W and E_A because the draft determines the area of the ship in contact with water and air (Rakke et al., 2012), thus affecting the friction from water and air.

Headwind and crosswind. Through relative wind speed WS and the angle of

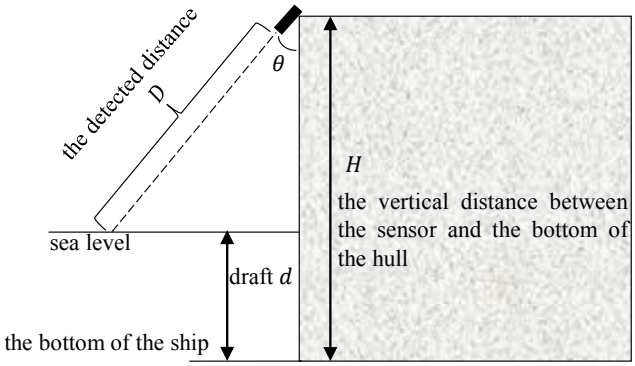


Figure 3.4: Draft measurement

the wind relative to the heading direction of the ship (see Figure 3.5), the headwind S_{head} and crosswind S_{cross} can be obtained:

$$S_{\text{head}}\ (\text{m/s}) = WS \times \cos\rho \tag{3.19}$$

$$S_{\text{cross}}\ (\text{m/s}) = WS \times \sin\rho. \tag{3.20}$$

According to Figure 3.5 and Formula (3.19), the negative value of headwind S_{head} represents tailwind and the positive value of headwind S_{head} means that the ship is sailing against the wind. According to Formula (3.20), the value of crosswind S_{cross} can be positive or negative, which indicates the different directions of the crosswind. Based on domain knowledge in shipping, the direction of the crosswind does not matter because the crosswind is in the vertical direction. Therefore, the absolute value of S_{cross} is used as the basic form of crosswind in the following experiments. Obviously, S_{head} and S_{cross} are variables that affect E_A.

Speed through the water. The variable STW indicates the ship's sailing speed through the water, which combines the ship's sailing speed over the ground

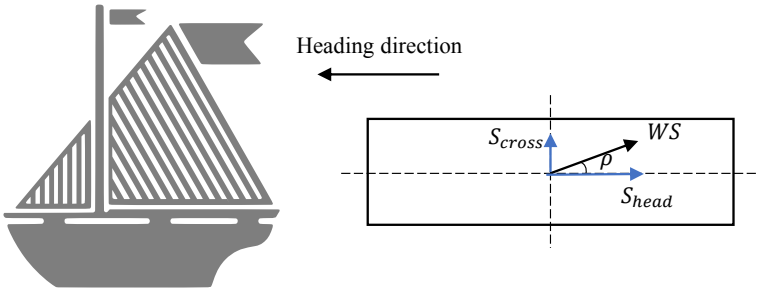


Figure 3.5: Wind measurement

and ocean currents (Petersen, 2011). Therefore, this variable can be used directly. STW is classified as a variable affecting E_W because it measures the sailing speed over the ground and ocean currents (Petersen, 2011).

Ship fuel consumption. Finally, the fuel consumption can be calculated by the following formula:

$$y \text{ (kg/s)} = FD \times FV. \tag{3.21}$$

Outlier records are deleted from the experimental dataset. First, records with null values are deleted. Second, records of ship sailing speed lower than 8 knots are deleted because these records infer that the ship is at anchor or just begins sailing. Finally, there are 150,831 records containing 245 voyages in the experimental dataset. This study randomly chooses 200 voyages (including 123,243 records) for training and 45 voyages (including 27,588 records) for testing in the experiment. The magnitude of the data used in this study is far greater than that of many other studies that use ship noon reports (one record per day) to predict ship fuel consumption (Du et al., 2019; Wang & Meng, 2012; Yan et al., 2020), which makes the results more convincing.

3.4 Experiment

All experiments are performed on a MacBook Pro computer with an Apple M2 processor (3.5 GHz), 8 cores, and 16 GB of RAM. Gurobi 10.0.0 is used as the optimization solver.

The optimal parameter settings of the PI-NN model are shown in Table 3.3. The optimal hyperparameters are obtained through o3 on the training dataset consisting of the randomly selected 200 voyages. The fully-NN model for comparison is equipped with the same network structure and the neuron number between layers is 4-10-10-10-2-1. According to the literature, the feature variables are normalized when training the neural network model (Beşikçi et al., 2016).

Table 3.3: Parameter settings of the PI-NN model

Parameter	Searching space	Optimal setting
Number of hidden layers	[1,2,3,4]	3
Number of neurons in each hidden layer[1]	[6,8,10,12]	10
Number of neurons in the input layer	\	4
Number of neurons in the output layer	\	2
Activation function	[Relu,Sigmoid]	Sigmoid
Learning rate	[0.001,0.01,0.1]	0.1
Number of epochs	[50,80,100]	100
Number of batch size	[32,64,128]	64

[1] The number of neurons in each hidden layer consists of the number of neurons in each hidden layer of the two separate parts in the PI-NN model. Therefore, 10 neurons indicate that there are 5 neurons in each layer of the separated two parts of hidden layers.

3.4.1 Settings for the MIQO-BF Model

For variables d_{avg}, S_{head}, and S_{cross}, their square root transformation, logarithmic transformation, quadratic and cubic transformation are considered. As discussed in Section 3.3, the absolute value of S_{cross} is used because the direction of the cross-wind does not matter. But the direction of the headwind will affect fuel consumption. Obviously, sailing with a tailwind will save fuel consumption. Therefore, when making transformations, the sign of S_{head} needs to be preserved. Parameter μ is defined to keep the sign

$$\mu = 2 \times \mathbb{I}(S_{\text{head}} > 0) - 1, \tag{3.22}$$

where $\mathbb{I}(\cdot)$ is an indicator function. If $S_{\text{head}} > 0$, then $\mathbb{I}(S_{\text{head}} > 0) = 1$; if $S_{\text{head}} <= 0$, then $\mathbb{I}(S_{\text{head}} > 0) = 0$. Therefore, $\mu = 1$ if the ship is sailing against the wind and -1 otherwise. It is well-recognized that there is an approximately cubic relationship between ship sailing speed and fuel consumption. Wang and Meng (2012) show that the value of the power is between 2.7 and 3.3. Thus, the expressions of $STW^{2.7}$, $STW^{2.9}$, $STW^{3.0}$, $STW^{3.1}$ and $STW^{3.3}$ is considered. The forms of expressions of feature variables are summarized in Table 3.4.

Table 3.4: Forms of expressions of variables

Variable	Expression 1	Expression 2	Expression 3	Expression 4	Expression 5
STW	$STW^{2.7}$	$STW^{2.9}$	$STW^{3.0}$	$STW^{3.1}$	$STW^{3.3}$
d_{avg}	d_{avg}	$\ln(d_{\text{avg}})$	$\sqrt{d_{\text{avg}}}$	d_{avg}^2	d_{avg}^3
S_{head}	S_{head}	$\ln(\lvert S_{\text{head}}\rvert)$	$\mu\sqrt{\lvert S_{\text{head}}\rvert}$	μS_{head}^2	S_{head}^3
S_{cross}	$\lvert S_{\text{cross}}\rvert$	$\ln(\lvert S_{\text{cross}}\rvert)$	$\sqrt{\lvert S_{\text{cross}}\rvert}$	S_{cross}^2	$\lvert S_{\text{cross}}\rvert^3$

3.4.2 Results and Discussion

The performance of the PI-NN model, the fully-NN model, and the MIQO-BF model are shown in Table 3.5. To illustrate the universality of the neural network model, a tree-based machine learning model—XGBoost model (XGB)—is also applied for comparison (Chen & Guestrin, 2016). And the hyperparameters of XGB are tuned by GridSearchCV (Yan et al., 2021a) the finally adopted hyperparameters are shown in Table 3.6. The mean absolute error (MAE), MSE, and the mean absolute percentage error (MAPE) are used to measure the accuracy of the models. The variances of absolute value difference between predicted fuel consumption and real fuel consumption are calculated, providing a measure of the stability of the model performance:

$$Var = \frac{\sum_{n=1}^{N}(|y_n - \hat{y_n}| - (\frac{\sum_{n=1}^{N}|y_n - \hat{y_n}|}{N}))^2}{N}, \tag{3.23}$$

where N is the total number of samples in the test dataset and $n = 1, .., N$.

Table 3.5: Results of three models

Metrics	PI-NN	fully-NN	XGB	MIQO-BF
MAE	0.0285	0.0278	0.0317	0.0353
MSE	0.0034	0.0029	0.0033	0.0035
MAPE	4.70%	4.63%	5.27%	5.88%
Var	0.0026	0.0021	0.0023	0.0023

The results show that all four models have a small variance in their absolute errors, which indicates that the models are stable. Table 3.5 shows that the PI-NN model is only slightly poorer than the fully-NN model. The MAE and MSE of the fully-NN model account for 97.54% and 85.29% of the MAE and MSE of the

Table 3.6: Parameter settings of XGB model

Parameter	Searching space	Selected setting
max_depth	[4,6,8,10]	6
n_estimators	[50,100,150]	100
learning_rate	[0.05,0.1,0.2,0.3]	0.3
sub_sample	[0.6,0.8,1]	1
colsample_bytree	\	1
colsample_bynode	\	1
min_child_weight	[1,2,3]	1

PI-NN model, respectively. That is, the PI-NN model can replace the fully-NN model at a level of more than 97% as measured by the MSE, which means that the domain knowledge introduced into the neural network model increases model interpretability while preserving accuracy. Moreover, the number of weights in the PI-NN model is fewer than the number of weights in the fully-NN model, which indicates that the training time of the PI-NN model is less than the training time of the fully-NN model. Specifically, there are 282 weights between connected neurons in the fully-NN model, which is more than double the number of weights in the PI-NN model (135). The PI-NN model saves almost 10% of training time compared with the fully-NN model in the experiment. The performance of the XGB model is slightly poorer than both the PI-NN and fully-NN models. The neural network model is inferred to be more suitable for predicting ship fuel consumption.

The MIQO-BF model does not perform as well as the other three AI models. And this result is consistent with existing literature (Le et al., 2020b; Parkes et al., 2018; Uyanık et al., 2020). However, the MIQO-BF model is highly explainable.

According to the values of decision variables, the corresponding formula of the MIQO-BF model is:

$$y \propto STW^{2.9} + \ln(d_{\text{avg}}) + S_{\text{head}}^3 + \sqrt{|S_{\text{cross}}|}. \tag{3.24}$$

Formula (3.24) indicates that the 2.9$^{\text{th}}$ power of ship sailing speed is proportional to fuel consumption, which is in line with previous studies (Le et al., 2020a; Meng et al., 2016; Wang & Meng, 2012). The logarithmic form of the ships' draft is proportional to fuel consumption. That is, the draft has a smooth effect on ship fuel consumption. Surprisingly, headwinds seem to have a greater effect on ship fuel consumption than crosswinds because the optimal expression of the crosswind is a root transformation but the optimal expression of the headwind is a cubic transformation. Based on common sense, the effect of the crosswind may be greater than that of the headwind when a ship sails. Such a counterintuitive result might be caused by the wind angle on the actual sailing route of the ferry. Figure 3.6 shows the histogram of the frequency distribution of the wind angle in the dataset. In most cases, the wind angle breaks up more wind force horizontally than vertically. This study argues that no shipping company wants to operate a sailing route that is subject to perennial crosswinds, as these create dangerous sailing conditions. Thus, as the ferry between Tórshavn and Suðuroy is already in operation, its sailing route should be appropriate for sailing, and the counterintuitive result (that the crosswind has a smaller effect than the headwind) based on the data generated by the ferry may be obtained. In summary, the result is reasonable because the experiment is based on a dataset generated by a ship in operation and the effect of the crosswind may already be taken into account by managers in the decision stage.

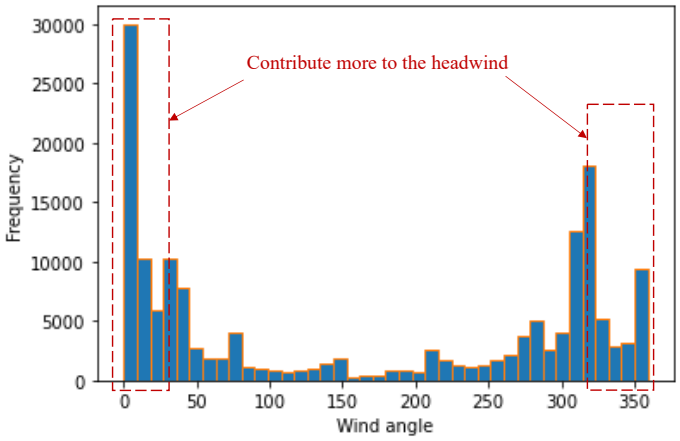


Figure 3.6: The distribution of wind angle

The findings answer the three questions put forward in Section 3.1.

R1: The proposed PI-NN model is on par with the fully-NN model to a degree of 97%. That is, the PI-NN model achieves the goal of improving model interpretability while maintaining model accuracy because it takes advantage of domain knowledge. By adding the constraints of physics, the PI-NN model becomes more acceptable to practitioners, thereby predicting ship fuel consumption more accurately.

R2: The findings indicate that the 2.9^{th} power of ship sailing speed, the logarithmic form of draft, the root transformation of the crosswinds, and the cubic transformation of the headwinds are the best formations for fitting a linear regression model to predict ship fuel consumption. The MIQO-BF model is an explainable additive model and the relationship between selected forms of variable expressions and ship fuel consumption is in line with practice. The results of the MIQO-BF model indicate that sailing speed is the most important factor in ship fuel consumption. The results of the MIQO-BF model also suggest that the head-

wind and crosswind are important variables but are not always influential because an established sailing route does not have frequent strong winds. Moreover, as wind and draft are objective, uncontrollable factors, ship captains should focus more on the effect of sailing speed on fuel consumption.

R3: The MIQO-BF model is slightly poorer than the other AI models in terms of performance, which is consistent with the findings of the literature (Le et al., 2020b; Parkes et al., 2018). Moreover, the neural network model is more suitable for predicting ship fuel consumption than the state-of-the-art tree-based models. All of the four models own good stability. In summary, the PI-NN model improves the interpretability of black-box models and the MIQO-BF model allows more variable expressions to be considered in a linear regression model. Managers can flexibly choose between the two models according to their needs for model accuracy and model interpretability.

3.5 Extension: SHAP Values

SHAP is proposed by Lundberg (2017). SHAP uses Shapley values from game theory to explain the prediction results and assigns a SHAP value to each feature value in each data sample (Lundberg, 2017; Wang et al., 2022b). SHAP provides a unified approach to interpreting model predictions and it is especially useful for explaining the prediction results of machine learning models. Different from the MIQO-BF model and the PI-NN model, SHAP is developed based on a machine learning model to interpret the already predicted value of that machine learning model. Therefore, SHAP addresses the interpretability issue from the perspective of hindsight. To make the research more comprehensive, SHAP is further adopted

to explore the feature importance for each predictor, i.e., d_{avg}, STW, S_{head}, and S_{cross} and quantifies the contribution of each feature value.

Referring to Wang et al. (2022b) and Yan et al. (2022), feature importance values (SHAP values) are calculated from two angles: global interpretability and local interpretability. Global interpretability means that the absolute SHAP values of each variable from the training data are averaged to measure the feature importance globally. Local interpretability shows how the contribution of an individual predictor varies across selected samples (Wang et al., 2022b). Readers are referred to Lundberg (2017) for more details about SHAP. All the following figures are created by the SHAP Python module (Lundberg, 2017).

Figure 3.7 shows the contributions of each predictor from the global interpretability perspective. Note that "avgLEVEL" is d_{avg}. Variables are ranked in descending order. The top variable is STW, which indicates that the ship's sailing speed is the most important factor affecting fuel consumption. And this result is in line with the consensus in the maritime field (Meng et al., 2016; Wang & Meng, 2012). The second important feature variable is d_{avg}, followd by S_{head} and S_{cross}. The ordering of S_{head} and S_{cross} is consistent with what this study has addressed in Section 3.4.2.

Four samples are randomly selected in the dataset and analyze how the four feature variables contribute to the final predictions. The expectation value (i.e., the base value) in Figure 3.8 indicates that the predicted value of ship fuel consumption per second is 0.569kg/s when any values of the feature variables are not revealed. The base value is the mean value of all the ship fuel consumption records in the training dataset. Taking the values of the feature variables into account, the final prediction is the sum of the SHAP value of each feature variable

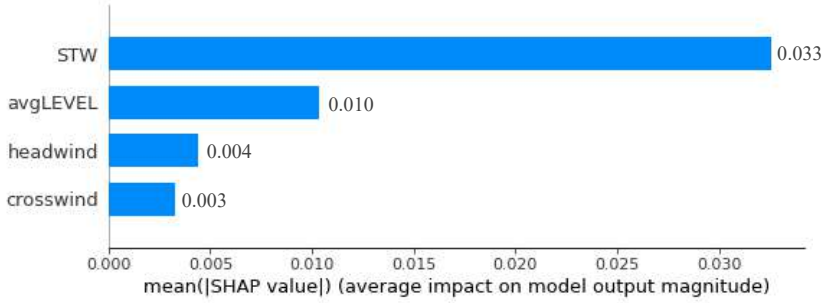
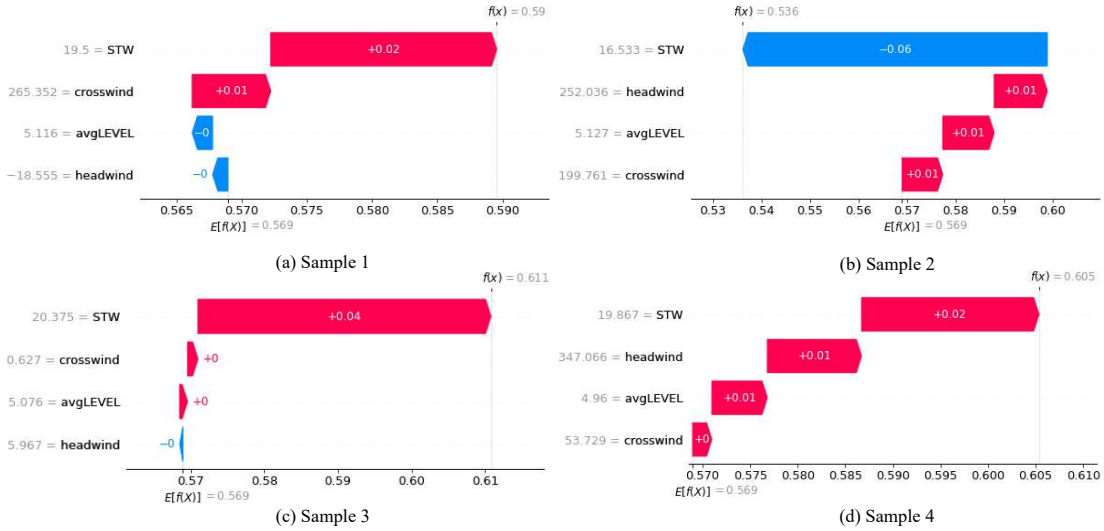


Figure 3.7: The SHAP variable importance on global interpretability

and the base value. For example, the predicted value of ship fuel consumption in Figure 3.8(b) is 0.536kg/s, which is the sum of 0.569kg/s (the base value), -0.06kg/s (the contribution of the ship sailing speed), 0.01kg/s (the contribution of headwind), 0.01kg/s (the contribution of the draft), and 0.01kg/s (the contribution of crosswind). Note that the third digit is different as the values are displayed to the last two decimal places. Figure 3.8 shows that sailing speed contributes the most to ship fuel consumption in all of the samples. In Figure 3.8(a), the crosswind value contributes positively to ship fuel consumption because the feature value of crosswind is 265.352m/s, which is quite high and increases fuel consumption. The value of the headwind variable (-18.555m/s) decreases fuel consumption as a negative headwind value indicates a tailwind. However, the feature value is low, and thus the effect is small. In Figure 3.8(b), the SHAP value of STW is negative, which means when the ship is sailing at 16.533 knots, the sailing speed is less than the average ship sailing speed, creating a negative SHAP value that is subtracted from the base value in the calculation of the predicted value. It is also found that the value of ship sailing speed in Figures 3.8(a), Figure 3.8(c), and Figure 3.8(d) all contribute positively to ship fuel consumption because the value of STW in



(a) Sample 1 (b) Sample 2

(c) Sample 3 (d) Sample 4

Figure 3.8: The SHAP value: local interpretability

these three samples is large. The SHAP values for the crosswind and headwind in Figure 3.8(b) are all positive because there are strong crosswinds and headwinds and the ship sails against the wind. In Figure 3.8(c), the SHAP values of all the variables except ship sailing speed are low. Figure 3.8(d) indicates that the ship sails against the wind and the effect of the headwind is large. The contribution of each feature variable is not independent and the SHAP value of one feature variable is influenced by other feature variables in the sample. For example, there is no big difference between the value of draft in Figures 3.8(a) and 3.8(c) but the direction of the effect is different.

These findings show that SHAP is helpful for understanding the effect of feature values on the output. The global interpretability of SHAP makes it possible to determine the average effect of all of the feature variables on the predicted values. The local interpretability of SHAP clearly quantifies the contribution of each feature value to the final predicted value and helps in understanding the internal

mechanisms in black-box models. SHAP provides a posterior alternative for explaining machine learning models. That is, a machine learning model is trained in the first stage and then SHAP is used to interpret it in the second stage. This study comprehensively explores the issue of interpretability in ship fuel consumption prediction by improving interpretability from a model-building perspective and presenting SHAP for the posterior explanation.

3.6 Conclusion

Ship fuel consumption is an important issue in the shipping industry. In this study, two innovative approaches are developed for predicting ship fuel consumption that addresses the tradeoff between model interpretability and model accuracy. Although some black-box models are quite advanced and can deliver accurate predictions, they lack interpretability and hence are rarely applied in practice. The proposed PI-NN model incorporates the constraints of physics into a neural network model; the results of the experiment using real-world data demonstrate that the effectiveness of the PI-NN model is on par with that of the fully-NN model to a degree of 97%. An additive white-box model, the MIQO-BF model, is also developed to consider more forms of feature variable expressions based on domain knowledge. The MIQO-BF model can give an explicit expression for predicting ship fuel consumption by solving MIQO programming. Practitioners can choose between the two approaches depending on their requirements: the PI-NN model is more suitable in scenarios requiring a high level of accuracy, whereas the MIQO-BF model is more suitable in scenarios requiring a low level of accuracy but a high level of interpretability. SHAP, a popular interpretability method, is adopted

to provide explanations for the results of the machine learning model.

This study helps to promote the application of data-driven models in maritime practice as models are developed based on domain knowledge, thereby making them more acceptable to practitioners. This study argues that using data-driven models for predicting ship fuel consumption will decrease fuel consumption, which will help to reduce operating costs, protect the environment, and achieve green shipping. AI has immense potential in the shipping industry. This study provides methods for coupling AI models with domain knowledge; this study also provides alternatives for interpreting black-box AI models. This research contributes to the application of AI in the shipping industry as the findings show that domain knowledge can complement AI models. With the help of domain knowledge, AI can lead to digital transformation, energy efficiency, and predictive analytics in the maritime industry.

Chapter 4

Federated Learning for Fuel Consumption Prediction and Optimization: Towards Privacy-Preserving and Distributed Data Analytics

4.1 Introduction

The shipping industry causes harm to the environment, because ships are mainly powered by heavy fuel oil (Wang et al., 2022c). For example, the maritime industry is responsible for 15% of global carbon dioxide emissions and 13% of sulfur dioxide emissions (Faber et al., 2020; Wang et al., 2022d). The International Maritime Organization (IMO) and other official bodies have issued many regulations to reduce ship emissions. We extract some of them listed in Table 4.1. However, emissions from the maritime industry continue to increase overall (IMO, 2020b) and the resulting pollution is a matter of concern for both governments and the shipping industry. Many governments have proposed that ships use clean energy, e.g., liquefied natural gas, or be equipped with pollutant processors to reduce emissions. These two approaches are costly, and shipping companies may not be willing to implement them out of consideration for their interests, which leads to the need for government subsidies to promote their implementation (Wang et al., 2022d). In terms of ship operations, a captain can control the emission of pollutants by adjusting ship speed, as a high sailing speed usually generates more fuel consumption and thus more emissions (Du et al., 2019). Controlling sailing speed involves no additional investment, although it is necessary to arrive at port on time, and thus is a feasible method of emission control in practice. The problem of controlling sailing speed and optimizing total ship fuel consumption has attracted the attention of many scholars (Du et al., 2019; Yan et al., 2020). Our study similarly aims to reduce total fuel consumption by finding the optimal sailing speed setting. However, we take a different approach in terms of methodology and also innovatively consider the data privacy of ships.

Table 4.1: Regulations for reducing ship emissions

Organization	Regulation	Main contents
Hong Kong Marine Department (2008)	Cap. 413M	Imposing a 3.5% sulfur limit on vessel fuel.
IMO (2009)	Guidelines for voluntary use of the ship energy efficiency operational indicator (EEOI)	Define EEOI and provide guidelines for shipping operators.
IMO (2022)	MEPC.328(76)	Promulgate a series of regulations to prevent air pollution from ships, such as emission control areas and global sulfur content limit.
European Commission (2015)	EU 2015/757	Monitor, report and verify emissions caused by ships.

The emerging of the Internet of Things (IoT) has led to the paradigm of the Internet of Ships (IoS) (Zhang et al., 2021). According to Aslam et al. (2020), IoS refers to the process of intelligently connecting maritime objects, e.g., electronic devices in ships or ports, with the aim of improving efficiency, safety, and environmental sustainability. After these maritime objects are connected by sensors and communication devices, they can share data and information (Aslam et al., 2020; IMO, 2019). Three major IoS applications are smart ships, smart ports, and smart transportation (Aslam et al., 2020). The IoS paradigm can promote the accumulation of data, which facilitates data mining; in turn, this increases the use of data-driven applications in the maritime industry and improves decision-making. For example, the shipowner China Merchants Energy Shipping, mining company BHP, and classification society DNV announced a joint venture on November 17, 2022 with the goal of achieving energy efficiency and reducing greenhouse

gas emissions by sharing navigation datasets (Lloyd's List, 2022). However, not all companies are willing to share data. TradeLens, a shipping trade data and document-sharing platform jointly established by IBM and Maersk in 2018, declared at the end of November 2022 that it would cease operations after the first quarter of 2023 (TradeLens, 2022). The original vision of the platform is to achieve global information sharing and collaboration across a highly fragmented industry. Unfortunately, this high level of cooperation and support proved impossible to achieve. Concerns about the security of data sharing are a major reason why the platform failed. In this context, one key issue is how to protect private shipping-company data while still making full use of data from all parties.

Federated learning (FL) provides a solution. In short, FL replaces the direct transmission of raw data between parties with the transmission of model parameters (e.g., the weight coefficients of the neural network); that is, it utilizes the data of all parties while reducing the risk of data leakage and also the difficulties associated with the transmission of a large amount of data (McMahan et al., 2017). With the FL method, data owners can train models locally with their own data and then transfer the parameters to a central server, which will aggregate the parameters from all data owners and then distribute these aggregated parameters back to them. FL already has practical applications in some industries. For example, WeBank has successfully used FL for credit evaluation in the financial industry (WeBank, 2022). FL is also a very promising avenue in the maritime industry because the data protection mechanism should increase shipping companies' willingness to share information, which will facilitate the realization of the IoS. As discussed in many studies, shipping companies often form alliances to make shared use of ships. That is, the containers of one company can be loaded onto the ship of an-

other company within the alliance (a revenue distribution model is agreed upon in advance); in this way, it is possible to increase ships' full load rate or achieve economies of scale. Because different shipping companies can form alliances to share ships, it would seem natural for them to form alliances to share data. Under the FL mechanism, ships from different companies share information without revealing the original data, which should greatly improve the companies' willingness to take part in information-sharing. In this study, we explore the application of FL in the maritime industry using the example of ship fuel consumption and emission reduction. We show that FL can be used to reduce total ship fuel consumption by developing a more accurate fuel consumption prediction model and solving a speed optimization model.

Our research has three main objectives. First, we propose an FL framework that enables multiple ships to collaboratively train a model by sharing parameters without the risk of data leakage. Second, after adopting the trained model derived from the FL framework, we use optimization techniques to get the optimal ship sailing speed that minimizes total ship fuel consumption over the voyage while ensuring the ship arrives on time. Third, we demonstrate the feasibility of FL applications in the marine industry. The theoretical and practical contributions of our research can be summarized as follows.

Theoretical contributions. (1) To the best of our knowledge, this is the first study of the application of FL to ship fuel consumption prediction and optimization, and our study provides a theoretical reference point for practical applications of FL. (2) By sampling and generating data, we show that FL can greatly improve the performance of the ship fuel consumption prediction model, which will help reduce emissions caused by fuel oil. (3) We also prove that all parties participating

in the FL mechanism will maintain the integrity from the perspective of cooperation. (4) The FL framework also provides a theoretical solution to the data silo problem in the maritime industry.

Practical contributions. (1) Our study will encourage firms to cooperate under the FL mechanism in practice, as we show that FL can benefit all parties. (2) Our findings suggest that the companies that cooperate should have considerable scale, to avoid the issue free riders; this finding can provide guidance for the adoption of FL in practice. (3) The results of our proposed sailing speed optimization model have implications for captains' daily operations.

The remainder of this paper is organized as follows. Section 4.2 reviews the existing literature related to our study. Section 4.3 proposes the ship fuel consumption prediction method based on FL and develops models to optimize the sailing speed to reduce emissions. Section 4.4 analyzes the honest participation of shipping companies in the alliance. Section 4.5 conducts a series of numerical experiments to verify the effectiveness of our proposed method. Conclusions are presented in Section 4.6.

4.2 Literature Review

In this section, we focus on the review of the following two streams of studies that are closely related to our work: (i) ship fuel consumption prediction in the maritime industry; (ii) the principle of FL and its applications.

4.2.1 Ship Fuel Consumption Prediction in Maritime Industry

Over the past decades, ship fuel consumption prediction has been the focus in the maritime field as ship fuel consumption affects the cost of shipping companies (Yan et al., 2021b) and environmental and sustainability issues, which are a matter of great concern to the government all over the world. By and large, the research on ship fuel consumption prediction can be divided into two categories. The first category is mainly based on statistical analysis, which combines physics principles and domain knowledge to mine the influencing factors of ship fuel consumption, establishes the relationship equation between ship fuel consumption and its influencing factors, and then realizes ship fuel consumption prediction. Wang and Meng (2012) estimate the bunker consumption by sailing speed because sailing speed is a key factor affecting ship fuel consumption. They find that there is a power function relationship between ship sailing speed and ship fuel consumption and the value of the power is between 2.7 and 3.3. Meng et al. (2016) develop a statistical model to explore the decisive factors for ship fuel consumption rate. They find that ship sailing speed is the most important decisive factor, followed by weather conditions. Their model is verified by the shipping log data of four container ships in practice. Bialystocki and Konovessis (2016) fit a curve between ship fuel consumption and ship sailing speed by considering many practical factors, such as the ship's draft and displacement, weather conditions, and hull and propeller roughness. They use the noon report—a report made by the captain at noon each day to describe the sailing profile to the onshore officers (Du et al., 2019)—to test their function, which is proven to be accurate and can be applied to sister ships. Adland et al. (2020) design a dynamic data-driven framework to

estimate the ship sailing speed-fuel curve based on noon reports from 16 crude oil tankers. They present that the cubic law for ship fuel consumption does not always hold. There are also many studies using this paradigm—adopting statistic analysis to explore the relationship between ship fuel consumption and other factors (Medina et al., 2020; Yao et al., 2012). Although different studies present different statistical models and analyze different factors, in general, they all agree that the most important factor affecting ship fuel consumption is ship sailing speed and that weather conditions and physical characteristics of the ship are some significant factors (Meng et al., 2016; Wang & Meng, 2012; Wang et al., 2013).

The second category is based on machine learning and deep learning methods that have emerged in recent years, that is, training a model to obtain ship fuel consumption predictions through input variables. These models could consider more complex influence relationships and are favored by many scholars. Du et al. (2019) put forward a two-stage framework based on ANN (artificial neural network) model to optimize ship's fuel efficiency. Using data from two 9000-TEU (Twenty-foot Equivalent Unit, TEU) container ships, they prove that their models can help save over 5% bunker fuel. Yan et al. (2020) also develop a two-stage ship fuel consumption prediction model. In the first stage, they use random forest regression model to predict ship fuel consumption by inputting variables such as sailing speed, total cargo weight, and sea conditions. And then they optimize the sailing speed to reduce the ship fuel consumption in the second stage. Uyanık et al. (2020) test the performance of many machine learning models, e.g., lasso (least absolute shrinkage and selection operator) regression, support vector regression, tree-based models, and boosting algorithms, when being adopted to predict ship fuel consumption. The experiment results show that bayesian ridge regression,

kernel ridge regression, multiple linear regression, and ridge regression perform better than other machine learning models. Yan et al. (2021b) give an exhaustive literature review on principles and models being used in ship fuel consumption prediction. Readers are referred to their study for more details.

We can draw two insights from the above literature. First, the close relationship between ship sailing speed and ship fuel consumption means that research on either of these topics must consider both. Wu (2020) shows that the optimal speeds for large ships are larger than those for small ships when the goal is reducing ship fuel consumption. Most studies focus not only on predicting ship fuel consumption, but also on optimizing ship fuel consumption by changing sailing speed, based on the relationship between ship fuel consumption and sailing speed and other variables (Du et al., 2019; Wang & Meng, 2012; Wang & Wang, 2016; Wang et al., 2013; Yan et al., 2020). Therefore, in this study, we also focus on optimizing the ship sailing speed while reducing ship fuel consumption using the developed ship fuel consumption prediction model. Second, almost all studies use non-public datasets from partner companies. Indeed, the maritime industry does not have a public dataset that includes ship fuel consumption records and the variables that influence ship fuel consumption, such as sailing speed and sea conditions. Although weather condition data can be obtained, other necessary variables—ship fuel consumption and ship sailing speed—are not available to the public. Therefore, the results of previous studies are the results of very limited testing, e.g., with data from one ship or a group of sister ships. In our study, we do not use data from a specific ship or shipping company. Instead, we sample and generate data based on previous research, which is a more general approach and can be reproduced.

4.2.2 Federated Learning

FL is first developed by McMahan et al. (2017). The main idea of FL is to allow models to be trained locally. To be more specific, we usually need to collect a large amount of data and train the machine learning models or deep learning models on a central server. In the process of FL, the central server does not collect any data from clients. Instead, it distributes the model parameters to each client, and each client trains the model locally using its own data and then updates the model parameters to the central server (Li et al., 2020a; McMahan et al., 2017). According to the certain aggregation rules, the central server updates global parameters and distributes them to clients again (Yue et al., 2023). After multiple rounds of communication, a model that meets the accuracy requirements is finally obtained (Li et al., 2020b). The Federated Averaging (FedAvg) algorithm proposed by McMahan et al. (2017) is the pioneering work in FL, and the subsequent studies are almost all based on their work to make improvements or list their work as a benchmark. Our work is also based on FedAvg. The aggregation rules in FedAvg are to sum and average the parameters of all clients in each round, which is quite straightforward but effective. Many studies explore more complex aggregation ways. For example, Yue et al. (2023) add a regularization term in the aggregation function to penalize those clients who contribute less to the loss function. In other words, their work takes into account fairness in FL. Mohri et al. (2019) also put forward a framework that considers fairness when aggregating parameters. Pillutla et al. (2022) develop a robust aggregation approach to deal with the corrupted update problem. As FL needs to be implemented through constant communications between the central server and clients, the communication efficiency (Niknam et al.,

2020) and information security in transmission (Liu et al., 2020) are also key concerns. These research perspectives have little to do with our research scope, so we do not elaborate on them too much. Prior to FL, many studies develop frameworks towards distributed learning. We need to emphasize that distributed learning and FL are very different, especially in the assumption about data distribution (McMahan et al., 2017). Distributed learning aims to achieve parallel computing based on independent and identically distributed (IID) datasets (Zhang et al., 2015). FL originally aims at mining information on datasets that are not independent and identically distributed (Non-IID) (McMahan et al., 2017) and many scholars conduct research on the problem of Non-IID dataset in FL (Zhao et al., 2018; Zhu et al., 2021).

Since FL was proposed, it has attracted attention in various fields. The main advantages of FL are improving data privacy (because no personal data needs to be collected) and allowing parallel training (because the model is trained simultaneously and locally on each terminal) (Yue et al., 2023). FL has great potential to be applied in practice. The most common application is in scenarios related to mobile devices. For example, prediction tasks in the keyboard or language-related topics (Passban et al., 2022) and prediction of human trajectory or behavior (Xiao et al., 2021). And FL is very suitable for application in the medical field because patient data is usually private (Pfitzner et al., 2021). Recently, some studies have explored the potential value of FL in the maritime industry. Zhang et al. (2021) present that FL can play an important role in shipping with the development of the IoS paradigm. And they develop an FL framework for fault diagnosis in ships. Liu et al. (2022) and Ma et al. (2022) hold the opinion that FL provides a new way for avoiding serious cyber security threats to maritime transportation systems.

Research on FL in the maritime industry is in its infancy. Previous studies mainly focus on communication security when adopting FL in ships (Liu et al., 2022; Ma et al., 2022). In addition, to the best of our knowledge, the use of FL for fault diagnosis is the only specific industry application that has been discussed (Zhang et al., 2021). No study has considered using FL for ship fuel consumption prediction and optimization. As discussed in Section 4.2.1, ship fuel consumption is an important issue in the maritime industry, and it is directly affected by the operation of the ship. Some ships may consume more fuel due to poor handling or their physical characteristics (Yan et al., 2021b) and they do not want to let others know, which makes them unwilling to share raw data. FL provides a new approach that makes it possible for ships to share information without revealing raw data. Through the application of the FL mechanism, it is feasible to combine information from many more ships and thus achieve a win–win situation. Therefore, our study explores the value of FL in the shipping field, especially in ship fuel consumption prediction and optimization. We also show that ship owners have incentives to participate honestly in the FL system.

4.3 Ship Fuel Consumption Prediction and Optimization

The datasets used for studying ship fuel consumption prediction in the literature mainly come from ships' noon reports (Du et al., 2019; Yan et al., 2020), which means that there are only 365 data records in total for each ship per year. Although some studies collect data from multiple sister ships, the amount of data is still too

limited to be suitable for constructing machine learning or deep learning models. Therefore, motivated by the studies of McMahan et al. (2017), Du et al. (2019), and Yan et al. (2020), we develop an FL-based framework to predict ship fuel consumption and optimize ship fuel consumption efficiency by deciding the optimal sailing speed. Under the proposed framework, ships can share information without revealing their own raw data. In this way, more shipping companies will be willing to share information, and thus more knowledge can be mined and the problem of isolated data islands between ships can be solved. The main notations used in this chapter are shown in Table 4.2.

We first introduce the FL ideas. As shown in Figure 4.1, the central server will distribute model parameters to each ship at the beginning. And then ships train the same deep learning model based on their own data. After certain rounds of iterations, ships will update the parameters to the central server which will aggregate all parameters. Again, the central server will distribute the updated global parameters to the ships and they will train the model using the new updated global parameters. After certain rounds of communication, we can obtain a trained model. The key point in this process is that the raw data of each ship does not need to be shared with other ships or the central server, which greatly protects the privacy of ships. Note that different ships may belong to the same ship company. In our study, we take a ship as a client during the training process and we will show that companies have incentives to take part in the FL mechanism. As all ships are located in the same scenarios, i.e., sailing on the sea, they can obtain the same types of feature variables and thus the FL in our study refers to the horizontal category (Li et al., 2020a; Li et al., 2020b).

We use back propagation neural network (Rumelhart et al., 1986) as the ba-

Table 4.2: Notations of FL and [M1]

Sets	
I	Set of shipping segments, $i \in I$
K	Set of ships, $k \in K$
J	Set of ship companies, $j \in J$
Parameters	
R	Number of communication rounds
c	The fraction of ships participating in training each round
E	Number of iterations in each local training
η	Learning rate
$\boldsymbol{\theta}$	Parameters of neural network model
L_i	The length of segment i
s_{min}	The minimum sailing speed
s_{max}	The maximum sailing speed
s_i^v	Ship's discrete sailing speed on segment i
T	The maximum sailing time allowed between two ports
t_i	The cumulative sailing time to the i-th segment
n_j	Number of ships belonging to company j
Function	
$F(\boldsymbol{x}, s)$	Neural network model predicting daily ship fuel consumption (tons) with inputting feature variables $\boldsymbol{x}$ and sailing speed s
Decision Variables	
s_i	Ship's sailing speed on segment i
z_i^v	Binary decision variable that equals 1 if a ship sails with speed s_i^v on segment i and 0 otherwise

sic deep learning model in each client. The number of hidden layers is 3 and the Sigmoid function is used as the activation function. The details of our neural network model and parameter setting are shown in Appendix A. The parameters that need to be transmitted between ships are the weights between nodes in the adjacent layers of the network. Note that in addition to the feature variables, e.g., sailing speed and weather conditions, we also take into account time series variables, i.e., the ship's historical fuel consumption data. Therefore, our study actually adopts a

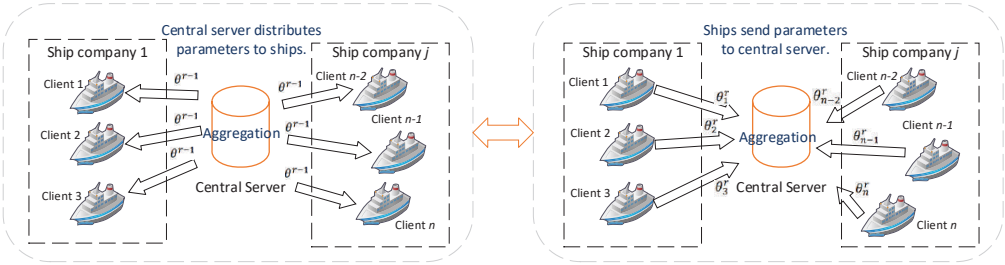


Figure 4.1: The process of FL in shipping

back propagation neural network by inputting both feature variables and time series variables, which is innovative as few studies have considered the information of ship ages. After R rounds of communication, we get the trained model. As the ship's historical fuel consumption data are also adopted as input features, we will collectively refer to them as feature variables in the following, unless otherwise specified. And we distinguish only the speed variable from the feature variables in the following.

Referring to Yan et al. (2020) and Du et al. (2019), we adopt the two-stage process to optimize ship fuel consumption efficiency. That is, we get the trained model which can give the prediction value of ship fuel consumption by inputting sailing speed and other feature variables. And then we optimize the ship's fuel consumption by deciding the optimal ship sailing speed in the case of satisfying a series of constraints for speed being the only variable that can be decided by the vessel captain and having a significant impact on ship fuel consumption. The innovation in the proposed two-stage process in our study is that instead of collecting data from ships, we use the idea of FL in the first stage, which allows each ship to train the model locally and then transfer the parameters. We use $F(\boldsymbol{x}, s)$ to denote the neural network model that predicts ship fuel consumption by inputting

feature variables $\boldsymbol{x}$ and ship's sailing speed s and $F(\boldsymbol{x}, s)$ can be obtained by the trained neural network model. Set $I = \{1, ..., |I|\}$ denotes the set of shipping segment between Port A and Port B, $i \in I$. And s_i (nautical miles per hour) denotes the sailing speed on segment i and L_i (nautical mile) represents the length of each segment. s_{min} and s_{max} are minimum and maximum sailing speed requirements of ships. Since the existing studies mainly use noon reports to conduct data analysis, we also assume that our data is obtained on a daily basis and the unit of predicted ship fuel consumption is tons per day. Subject to the constraints of ensuring that the ship arrives on time, we can formulate the following optimization model.

[M1]

$$\min \sum_{i \in I} F(\boldsymbol{x}, s_i) \frac{L_i}{24 \times s_i} \tag{4.1}$$

subject to

$$t_i = t_{i-1} + \frac{L_i}{s_i}, i \in I \tag{4.2}$$

$$t_{|I|} \leq T \tag{4.3}$$

$$t_0 = 0 \tag{4.4}$$

$$s_{\text{min}} \leq s_i \leq s_{\text{max}}, i \in I \tag{4.5}$$

where T denotes the maximum sailing time allowed between Port A and Port B, and t_i represents the cumulative voyage time until the end of the i-th segment.

The whole process of the FL-Prediction-Optimization (FLPO) process is shown in Algorithm 2. FLPO is a two-stage framework. In the first stage, a neural network model for predicting ship fuel consumption is obtained based on FL which considers the ship's data privacy. In the second stage, ship sailing speed is op-

timized to reduce fuel consumption. As $F(\boldsymbol{x}, s_i)$ represents the neural network model, which is hard to give an explicit expression, we discretize s_i in the unit of 0.1 knot for solving [M1] (Yan et al., 2020). The discretized ship sailing speed is denoted by s_i^v. That is, we can obtain the corresponding fuel consumption by inputting each s_i^v together with other feature variables. And thus $F(\boldsymbol{x}, s_i^v)$ can be viewed as a parameter in the objective. We further introduce a binary decision variable to indicate which discretized ship sailing speed is selected with the constraint that only one speed can be chosen in one segment. Therefore, model [M1] is transferred to a solvable integer optimization model, denoted by model [M2]. The details are shown in Appendix B. [M1] and [M2] are optimization models for one ship to solve. There will be many ships in the FL mechanism, and we use K to denote the total number of ships and set I_k to denote the shipping segments of each ship k.

Algorithm 2: FLPO

Input: the number of communication rounds R, the number of ships K, the fraction of ships that participate in training in each round c, the number of iterations in local training E, learning rate η, initial parameter setting $\boldsymbol{\theta}$, feature variables and ship sailing speed, and parameters needed in optimization process

Output: $\boldsymbol{\theta}^*$, $s_i^{k*}, k = 1, ..., K, i \in I_k$

// $\boldsymbol{\theta}^*$ is the finally obtained optimal parameters of the neural network model and s_i^{k*} is the optimal sailing speed of ship k in segment i

Stage 1

begin

  $\boldsymbol{\theta}^0 = \boldsymbol{\theta}$

  for $r = 0 : (R - 1)$ **do**

    $m \leftarrow \max(c \times K, 1)$

    $K_r \leftarrow$set of randomly selected m ships

    // K_r is the set of randomly selected ships by the central server, i.e., m ships will participate in training in round r. If all ships are selected, $m = K$

    for $k \in K_r$ **do**

      $\boldsymbol{\theta}_k^{r+1} \leftarrow BackPropagation(\eta, \boldsymbol{\theta}^r, E)$

      // $\boldsymbol{\theta}_k^{r+1}$ represents the parameters of ship k's model in round $r + 1$ and ships train model in parallel

    end

    // The central server will collect parameters from ship $k \in K_r$, integrate these parameters, and then distribute $\boldsymbol{\theta}^{r+1}$ to all ships

    $\boldsymbol{\theta}^{r+1} \leftarrow \frac{1}{|K_r|} \sum_{k \in K_r} \boldsymbol{\theta}_k^{r+1}$

  end

  // The central server will distribute the final $\boldsymbol{\theta}^R$ to all ships

  $\boldsymbol{\theta}^* = \boldsymbol{\theta}^R$

end

Stage 2

begin

  // $F(\boldsymbol{x}, s)$ can be obtained by $\boldsymbol{\theta}^*$. For each ship k, its feature variables and sailing speed in segment i are denoted by $\boldsymbol{x}_i^k$ and s_i^k, respectively

  for $k = 1 : K$ **do**

    solve [M2] to obtain $s_i^{k*}, k = 1, ..., K, i \in I_k$

    // The details of [M2] are shown in Appendix B

  end

end

4.4 Integrity of Alliance Parties in the FL Mechanism

Generally speaking, for machine learning models or deep learning models, as the number of samples increases, the training accuracy increases and the magnitude of the increase decreases (Wang et al., 2022a). Cui and Gong (2018) find that the exponential form is appropriate to fit the relationship between prediction accuracy and sample size. In this study, we also adopt the exponential function to characterize the curve of prediction accuracy and sample size, i.e., the curve of prediction accuracy and the number of ships. We use set J to denote the set of shipping companies, $j \in J$, and each company j has n_j ships. Suppose that if a shipping company joins the FL mechanism, all the ships belonging to this company will be used as clients. The total number of participating ships is $n = \sum_{j \in J} n_j$ and we treat n as a function of n_j. The prediction accuracy is affected by many factors, e.g., sample size, data quality, and the problem itself. As we take a dataset corresponding to the same length of the period from all ships for model training in our study, e.g., noon reports for the past 180 days, we assume that the number of ships participating in training is the variable that affects the prediction accuracy and the influence of other factors can be regarded as constant, which could be eliminated for simplifying notation. Then, the general function that describes the relationship between prediction accuracy and the number of participating ships, denoted by $H(n)$, is

$$H(n) = a \times e^{\frac{n}{b}} + c, \ a < 0, \ b < 0, \ c > 0, \tag{4.6}$$

where a, b, and c are estimated parameters.

Proposition 3 *All participating shipping companies will honestly participate in the federal organization from the perspective of cooperation.*

Proof. The derivative of $H(n)$ with respect to n_j is

$$\frac{\partial H(n)}{\partial n_j} = \frac{dH(n)}{dn}\frac{\partial n}{\partial n_j} = \frac{a \times e^{\frac{n}{b}}}{b}, j \in J, \tag{4.7}$$

which is always greater than 0, which means that the more ships serve as clients, the more accurate the model will be. Therefore, the company will let all ships be clients to maximize its own interests and the overall interests. The second derivative of $H(n)$ with respect to n_j is

$$\begin{aligned}\frac{\partial^2 H(n)}{\partial n_j^2} &= \frac{\partial}{\partial n_j}\left(\frac{dH(n)}{dn}\frac{\partial n}{\partial n_j}\right) \\ &= \frac{d^2 H(n)}{dn^2}\left(\frac{\partial n}{\partial n_j}\right)^2 + \frac{\partial^2 n}{\partial n_j^2}\frac{dH(n)}{dn} \\ &= \frac{a \times e^{\frac{n}{b}}}{b^2}, j \in J\end{aligned} \tag{4.8}$$

which is always less than 0 because $a < 0$. And thus, the marginal utility is diminishing. If company $j^{'}$ chooses to cheat, i.e., they do not use real data to train the local model, the accuracy or convergence speed of the global model will be reduced. We use α, $\alpha > 0$, to denote the influence coefficient as these cheating samples may reduce the model's accuracy. The function $\widehat{H}(n, n_{j'})$ with one cheating company $j^{'}$ can be written as follows:

$$\widehat{H}(n, n_{j'}) = a \times e^{\frac{-\alpha n_{j'} + n - n_{j'}}{b}} + c, j^{'} \in J, \tag{4.9}$$

As Equation (4.7) is strictly greater than 0, Equation (4.6) is definitely greater than Equation (4.9). Therefore, companies have no incentive to cheat from the perspective of cooperation and they will let all ships honestly participate in the federal organization. □

Proposition 4 *Although the global accuracy will be reduced if one company chooses to cheat, the company may still benefit from the FL mechanism if the number of the company's ship satisfies:*

$$-\alpha n_{j'} + n - n_{j'} > n_{j'}. \tag{4.10}$$

which means the company may cheat from the perspective of competition.

Proof. Figure 4.2 gives intuitive proof. If $-\alpha n_{j'} + n - n_{j'} > n_{j'}$, then the accuracy of the global model will still exceed the accuracy of the company $j^{'}$' own individually trained model, which means that company $j^{'}$ may cheat from the perspective of competition. Because in this way, the company's own utility will increase, but other companies' utility will decrease. □

We can find from the proof of Proposition 2 that the smaller the companies, the more likely that it chooses to cheat from the perspective of competition. But these large ship companies, e.g., company $j^{\#}$ in Figure 4.2, will have no incentive to cheat from the perspective of both cooperation and competition because cheating will reduce both their own and the global utility. Therefore, the integrity of large ship companies will guarantee the effectiveness of the FL mechanism. Moreover, it is more appropriate that the alliance is made up of companies of comparable size, which will significantly reduce the possibility of cheating and avoiding the free-riders in the FL mechanism. For example, if a small shipping company consisting

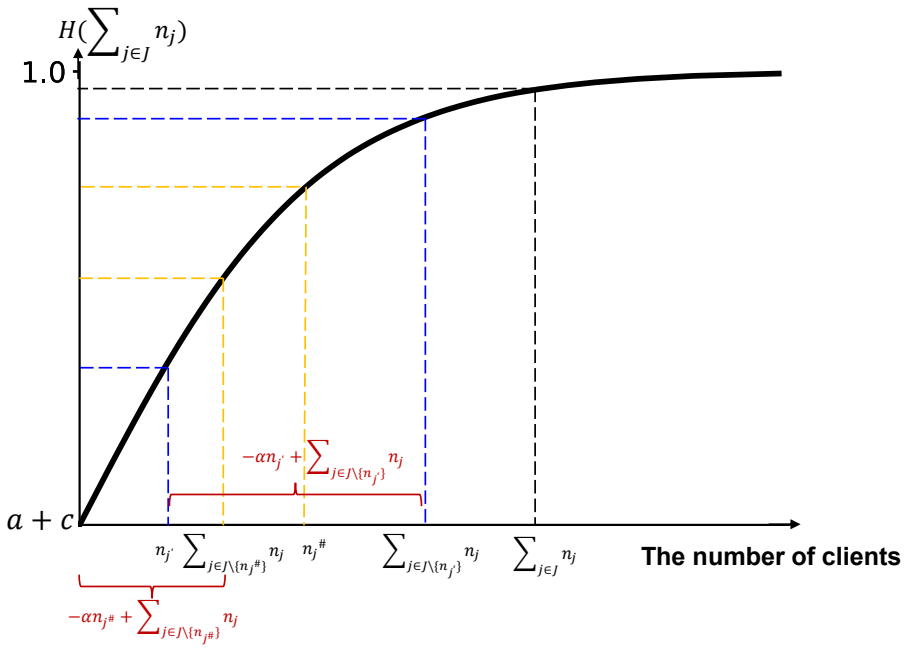


Figure 4.2: Prediction accuracy with different participating companies

of only 100 ships takes part in an FL mechanism composed of large companies with more than 1000 ships per company, he may be a free rider.

4.5 Numerical Experiments

4.5.1 Data

Recall that we need to train a neural network model $F(\boldsymbol{x}, s)$ based on feature variables $\boldsymbol{x}$ and sailing speed s to predict daily ship fuel consumption. In our study, we choose wave direction ($0 - 180°$ angle relative to the ship's heading direction), wave height, wind force, cargo weight, and the historical daily ship fuel consumption data for the past two weeks as feature variables $\boldsymbol{x}$ to train the model

together with sailing speed s. The parameter settings of our neural network model are shown in Table A1, Appendix A. We also use one month's historical vessel fuel consumption data to test the performance of FL and individual training in Appendix C. According to the existing literature, the cube of sailing speed is related to ship fuel consumption and Wang and Meng (2012) exactly show the power of sailing speed is between 2.7 and 3.3 using data from five ships. That is, we can roughly hold the opinion that there is a cubic relationship between ship sailing speed and ship fuel consumption, but the power may vary among different ships. The relationship between ship fuel consumption and other feature variables is unclear in the existing studies. We assume that ship fuel consumption, wave direction, wave height, wind force, and cargo weight jointly follow the normal distribution, i.e., $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{X}$ is the vector of five variables, $\boldsymbol{\mu}$ is the mean vector, and Σ is the covariance matrix. The following steps are adopted to sample and generate data. We finally sample and generate 180 records (i.e., the noon report for half a year) for each of the 20 ships.

Step 1. Sample from $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{X} \in \Re^5$ using Cholesky decomposition (Seeger, 2004).

Step 2. Sample from $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\varepsilon} \in \Re^5$ and $\bar{\boldsymbol{X}} = \boldsymbol{X} + \boldsymbol{\varepsilon}$. (The purpose of this step is to add the disturbance term because real-world data often have noise.)

Step 3. Randomly select the value of power within [2.85,3.15], which is the results of a 5% fluctuation of the third power. The value of sailing speed is obtained by solving the power function between sailing speed and ship fuel consumption, which is sampled in Step 2.

The above sampling and generating process require the values of the mean vector and covariance matrix, i.e., we need to know the mean of ship fuel con-

sumption, wave direction, wave height, wind force, and cargo weight, and their covariance matrix. We use the mean and variance in Yan et al. (2020) and the correlation coefficient matrix, which can be used to calculate the covariance matrix given the variance, in Wang et al. (2018a) to sample and generate variables. Details are shown in Table 4.3. We assume that the correlation coefficients between different variables are the general rules. Therefore, we can sample by adopting the correlation coefficient matrix from one study and then using the mean and variance from another study. Moreover, we normalize all the feature variables and sailing speed and map the daily ship fuel consumption data using the Sigmoid function for model training.

Table 4.3: Variable correlation coefficients and descriptive statistics

Correlation coefficients					
	Daily fuel consumption (ton/day)	Wave direction (degree)	Wave height (meter)	Wind force (beaufort force number)	Cargo weight (tons)
Daily ship fuel consumption	1				
Wave direction	0.76	1			
Wave height	0.15	0.28	1		
Wind force	0.17	0.35	0.38	1	
Cargo weight	0.62	0.68	0.06	0.05	1
The mean value and standard deviation of variables					
Mean value	16.95	87.34	2.00	4.90	28685.06
Standard deviation	9.16	98.74	0.82	1.21	2411

As mentioned in Section 4.2, the data in each client in FL is usually Non-IID (Zhao et al., 2018). Because we assume that the feature variables and fuel consumption follow a joint distribution, the above data generation process actually simulates two Non-IID situations: the same fuel consumption with different feature values and the same feature values with different fuel consumption. We argue that these two kinds of Non-IID are in line with the practice since different ships

sailing in different weather conditions may generate the same fuel consumption. We need to clarify that we do not consider forming an FL mechanism using different types of ships. For example, container ships can be served as clients in one FL mechanism and LNG carriers can be served as clients in another FL mechanism. But we do not consider forming an FL mechanism using both container ships and LNG carriers simultaneously because these two types of ships differ a lot in terms of physical characteristics. Thus, we argue that label distribution skew and feature distribution skew (Zhang et al., 2022b) do not exist within our research scope.

4.5.2 Results of Daily Ship Fuel Consumption Prediction

Using 180 records from each of the 20 ships, we carry out several experiments to validate the FL mechanism. We randomly select 60% of the records of each ship as the training dataset, 20% as the validation dataset, and 20% as the test dataset. Mean absolute error (MAE) and root mean square error (RMSE) are chosen to measure the predictive accuracy of the model. As stated in Algorithm 2, the number of communication rounds R, the fraction of ships that participate in training in each round c, and the number of iterations in local training E, i.e., the number of epochs in each round, are three key parameters in the FL mechanism (McMahan et al., 2017). We first set $R = 10$ and $c = 0.5$ to show the difference between the FL mechanism and individual training. That is, we randomly select 50% of all ships to participate in each communication round and perform parameter aggregations 10 times. The results are shown in Figure 4.3. Figure 4.3(a)$-$(c) show the results of MAE and Figure 4.3(d)$-$(f) show the results of RMSE by adopting different epochs. We can find that the FL mechanism outperforms individual training on

both MAE and RMSE for each epoch set. We further calculate the average MAE and RMSE for the 20 ships and then calculate the percentage improvements (the ratio of the absolute value of the difference between the average prediction results of the two methods to the individual training prediction accuracy) made by the FL mechanism. As shown in Table 4.4, the FL mechanism brings almost more than 50% improvement to the accuracy at each epoch set. Moreover, it is intuitive that the fluctuations of accuracy metrics in the FL mechanism are smaller. Summarily, the FL mechanism performs better than individual training in terms of accuracy and stability.

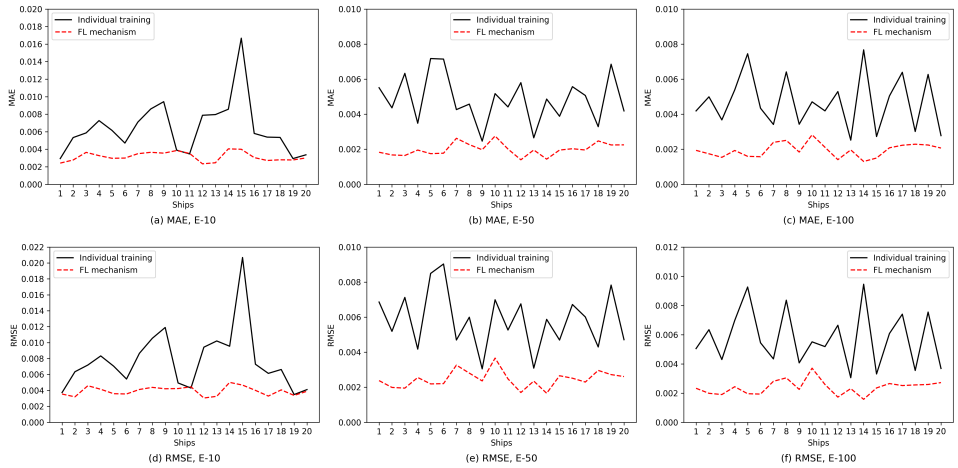


Figure 4.3: Prediction results with $\boldsymbol{R = 10}$ and $\boldsymbol{c = 0.5}$

Next, we fix $E = 10$ and $c = 0.5$ to explore the changes in accuracy with the number of communication rounds. As shown in Figure 4.4, the prediction accuracy is greatly improved when the number of communication rounds is changed from 10 to 30. After that, the two accuracy metrics do not change significantly as the number of communication rounds increases. Considering that communication is time-consuming, we suggest that there is no need to take too many times

Table 4.4: Comparision of FL mechanism and individual training

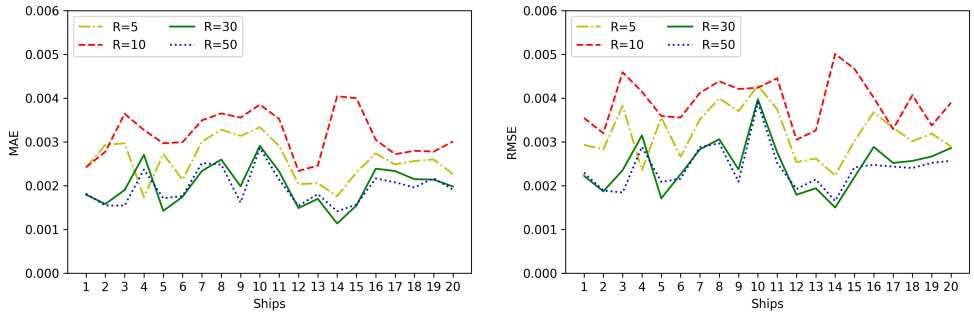
Metrics	E	% Improvements of of FL mechanism
MAE	10	50.80%
RMSE	10	49.57%
MAE	50	58.84%
RMSE	50	57.81%
MAE	100	58.40%
RMSE	100	58.53%

of communication and 30 times are enough for our dataset. Moreover, increasing local training iterations should be considered first since communication is more time-consuming than local training (McMahan et al., 2017).

Finally, we test the impact of sampling rate c by setting $E = 10$ and $R = 10$. Results are shown in Figure 4.5. We can find that the values of MAE and RMSE are greatly reduced when the fraction of ships participating in training in each communication round increases from 0.5 to 0.8. The difference between the two accuracy metrics is not obvious when $c = 0.8$ and $c = 1.0$. After several rounds of experiments, we propose that the optimal parameter configuration of the FL mechanism in our study is $R = 30$, $E = 50$, and $c = 0.8$. In practice, we recommend that increasing the number of local training times should be first considered when optimizing model parameters, followed by increasing the number of joined clients in each round and finally increasing the number of communication rounds because of the communication costs.

4.5.3 Results of Ship Sailing Speed Optimization Model

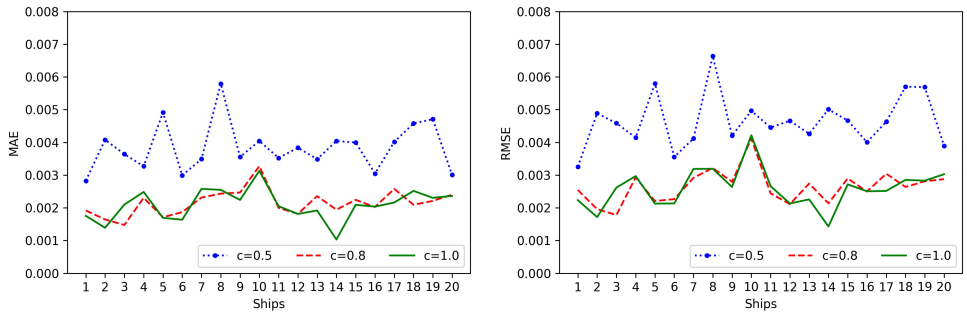
The trained model $F(\boldsymbol{x}, s)$ can be obtained by using the optimal parameter settings and the data records of the 20 ships. We randomly select two ships to validate the



(a) Changes of MAE with communication rounds

(b) Changes of RMSE with communication rounds

Figure 4.4: Prediction results with $\boldsymbol{E = 10}$ and $\boldsymbol{c = 0.5}$



(a) Changes of MAE with sampling rate

(b) Changes of RMSE with sampling rate

Figure 4.5: Prediction results with $\boldsymbol{E = 10}$ and $\boldsymbol{R = 10}$

performance of the proposed ship sailing speed optimization model using CPLEX Python API 20.1.0. And we treat the last 7 days as seven segments and optimize the daily sailing speed. The experiments are run on a laptop computer equipped with 2.6 GHz of Intel Core i7 CPU and 16 GB of RAM. According to the sampled data, we set $s_{\min} = 14$ knots and $s_{\max} = 18$ knots. As we do not have real sailing segment data, we solve model [M3] in Appendix B instead of model [M2], which can be considered equivalent because sailing speed guarantees the estimated time of arrival. The optimization model [M3] of each ship can be solved within 1 sec-

ond. Results are shown in Table 4.5 and Table 4.6 respectively. Ship 1 consumes 112.30 tons of fuel and Ship 2 consumes 113.66 tons of fuel in the last 7 days after optimization. Compared with the ship fuel consumption before optimization, ship 1 saves 6.25% and ship 2 saves 2.68%.

Table 4.5: Optimization results of Ship 1

Days	Original speed (knot)	Optimized speed (knot)
7th day to last	14.39	15.60
6th day to last	14.56	14.00
5th day to last	14.05	17.50
4th day to last	17.30	15.00
3rd day to last	17.16	15.60
2nd day to last	16.80	14.70
1st day to last	15.91	17.90
Total ship fuel consumption (tons)	119.79	112.30 (−6.25%)

Table 4.6: Optimization results of Ship 2

Days	Original speed (knot)	Optimized speed (knot)
7th day to last	17.47	18.00
6th day to last	16.27	14.00
5th day to last	13.71	14.60
4th day to last	14,34	14.60
3rd day to last	16.53	15.70
2nd day to last	15.75	17.10
1st day to last	14.27	14.60
Total ship fuel consumption (tons)	116.79	113.66 (−2.68%)

4.6 Conclusion

This study develops a two-stage FL-Prediction-Optimization framework to solve the problem of ship fuel consumption and sailing speed optimization. We first adopt FL to obtain a trained neural network model while protecting the privacy

of each ship's data. Then we solve the optimization problem to obtain the optimal sailing speed by making use of the trained model. Our numerical experiments show that our proposed two-stage method can reduce total ship fuel consumption while still ensuring that the ship arrives at the port on time; thus it can make an important contribution to emission reduction and environmental protection. Our study is the first to investigate ship fuel consumption prediction models under the FL mechanism, and provides a new way to overcome the barriers to data sharing among shipping companies. Once the FL mechanism is implemented, ship companies will be more willing to share information without fears of data leakage. Moreover, we analyze the integrity of allied parties in the FL mechanism and show that it is more appropriate for companies of comparable size to participate in one FL mechanism.

Our paper is not without limitations. First, we do not consider outlier identification as part of the process of data cleaning. However, this problem is worth addressing in further studies, as raw shipping log data are very noisy. Second, we perform our experiments in a laboratory environment, without taking into account problems that may be encountered on an actual voyage, such as the loss of the ship's signal or the inability to communicate. Third, the numerical data is generated based on assumptions because it is impractical to get data from many ships. The data sources for the existing studies using real data are only collected from a few sister ships. Moreover, to make the generated numerical data closer to reality, we refer to the mean and variance of feature variables, and the covariance between feature variables in the existing literature, which leads to limited categories of feature variables. We hope that our study can bring new insight to the industry such that shipping companies are willing to form an FL alliance, which

makes it possible for future research to test the performance of FL with real data.

Chapter 5

Concluding remarks

Chapter 2 considers the specific requirements of domain knowledge in the shipping industry, specifically that the relationship between ship sailing speed and fuel consumption follows a non-decreasing convex function. This insight is crucial for developing accurate models that can predict fuel consumption based on varying operational speeds, ensuring more effective fuel management and optimization strategies. The PH-CART method proposed in Chapter 2 improves prediction accuracy by nearly 4%, leading to a daily fuel consumption saving of 2246.4 kg.

Chapter 3 presents two innovative methods tailored to domain knowledge, aimed at addressing the trade-off between model interpretability and model accuracy in ship fuel management. These methods strike a balance, ensuring that the models remain both understandable and effective in ship fuel consumption prediction. The proposed PI-NN model achieves over 95% of the performance of the fully-NN model, while the MIQO-BF model provides an explicit expression for fuel consumption prediction.

Chapter 4 addresses the practical data privacy issues in maritime practice by proposing the use of federated learning for model training. This approach enables decentralized model development while ensuring that sensitive data remains secure and localized, overcoming challenges related to data sharing and privacy concerns in the industry. The results show that after using the model trained with federated learning to solve the optimization problem in the second stage, fuel consumption can be reduced by more than 6%. This has significant implications for the development of green shipping.

Although there are some classic hydrodynamic-based modeling approaches for predicting ship fuel consumption, e.g., the Holtrop-Mennen method (Holtrop & Mennen, 1982), these methods are mostly suited for predicting resistance under

calm water conditions and are typically applied during the ship design stage. When dealing with complex sailing environments, it becomes necessary to predict fuel consumption based on real-time factors, e.g., waves, currents, or wind. Classic hydrodynamic modeling methods struggle to achieve this level of adaptability. In recent years, with the development of the Internet of Things (IoT) technology, the concept of digital twin has started to gain prominence (Lee et al., 2022; Mauro & Kana, 2023). A digital twin ship is a virtual and digital representation of a physical entity. It can both represent the actual status of the physical ship and predict the future behavior of the ship by integrating data from various sensors, systems, and historical records (Mauro & Kana, 2023). Digital twin ships offer a novel perspective on fuel consumption: collecting real-time data from onboard sensors, simulating various scenarios, and predicting and optimizing fuel efficiency. However, implementing digital twin technology requires expensive hardware and software support, and real-time data transmission imposes higher demands on system security to prevent data breaches or attacks. Moreover, achieving data standardization and integration across different stakeholders and ship types remains a significant challenge. Therefore, before the implementation of digital twin ships becomes a reality in practice, the mainstream approach remains collecting fuel consumption and environmental data during ship operations and using models for fuel consumption prediction. Given that machine learning models can capture complex relationships among variables, developing machine learning-based fuel consumption prediction methods continues to hold both theoretical and practical significance.

References

Adland, R., Cariou, P., & Wolff, F.-C. (2020). Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet. *Transportation Research Part E: Logistics and Transportation Review*, *140*, 101972.

Arora, S., Taylor, J. W., & Mak, H.-Y. (2023). Probabilistic forecasting of patient waiting times in an emergency department. *Manufacturing & Service Operations Management*, *25*(4), 1489–1508.

Aslam, S., Michaelides, M. P., & Herodotou, H. (2020). Internet of ships: A survey on architectures, emerging applications, and challenges. *IEEE Internet of Things journal*, *7*(10), 9714–9727.

Ban, G.-Y., Gallien, J., & Mersereau, A. J. (2019). Dynamic procurement of new products with covariate information: The residual tree method. *Manufacturing & Service Operations Management*, *21*(4), 798–815.

Bertsimas, D., & Dunn, J. (2017). Optimal classification trees. *Machine Learning*, *106*, 1039–1082.

Bertsimas, D., & King, A. (2016). Or forum—an algorithmic approach to linear regression. *Operations Research*, *64*(1), 2–16.

Bertsimas, D., King, A., & Mazumder, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, *44*(2), 813–852.

Beşikçi, E. B., Arslan, O., Turan, O., & Ölçer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. *Computers & Operations Research*, *66*, 393–401.

Bialystocki, N., & Konovessis, D. (2016). On the estimation of ship's fuel consumption and speed curve: A statistical approach. *Journal of Ocean Engineering and Science*, *1*(2), 157–166.

Bläser, N., Magnussen, B. B., Fuentes, G., Lu, H., & Reinhardt, L. (2024). MATNEC: AIS data-driven environment-adaptive maritime traffic network construction for realistic route generation. *Transportation Research Part C: Emerging Technologies*, *169*, 104853.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth; Brooks.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, *8*(8), 832.

Chang, Y.-C., Tseng, R.-S., Chen, G.-Y., Chu, P. C., & Shen, Y.-T. (2013). Ship routing utilizing strong ocean currents. *The Journal of Navigation*, *66*(6), 825–835.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chen, Y., Calabrese, R., & Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. *European Journal of Operational Research*, *312*(1), 357–372.

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, *178*, 622–637.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265–282.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Du, Y., Meng, Q., Wang, S., & Kuang, H. (2019). Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. *Transportation Research Part B: Methodological*, *122*, 88–114.

Edyvean, R. (2010). Consequences of fouling on shipping. *Biofouling*, *10*, 217–225.

Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, *68*(1), 9–26.

European Commission. (2015). On the monitoring, reporting and verification of carbon dioxide emissions from maritime transport, and amending Directive 2009/16/EC. [Accessed on 1 April 2023]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02015R0757-20161216

Faber, J., Hanayama, S., Zhang, S., Pereda, P., Comer, B., Hauerhof, E., van der Loeff, W. S., Smith, T., Zhang, Y., Kosaka, H., et al. (2020). Reduction of GHG emissions from ships: Fourth IMO GHG study 2020–final report. *IMO MEPC*, *75*(7), 15.

Fagerholt, K., Gausel, N. T., Rakke, J. G., & Psaraftis, H. N. (2015). Maritime routing and speed optimization with emission control areas. *Transportation Research Part C: Emerging Technologies*, *52*, 57–73.

Fagerholt, K., Laporte, G., & Norstad, I. (2010). Reducing fuel emissions by optimizing speed on shipping routes. *Journal of the Operational Research Society*, *61*(3), 523–529.

Fan, A., Yang, J., Yang, L., Wu, D., & Vladimir, N. (2022). A review of ship fuel consumption models. *Ocean Engineering*, *264*, 112405.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—explainable artificial intelligence. *Science Robotics*, *4*(37), 7120.

Hastie, T., Tibshirani, R., & Tibshirani, R. (2020). Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. *Statistical Science*, *35*(4), 579–592.

Hazimeh, H., & Mazumder, R. (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Operations Research*, *68*(5), 1517–1537.

He, J., Hao, Y., & Wang, X. (2021). An interpretable aid decision-making model for flag state control ship detention based on smote and xgboost. *Journal of Marine Science and Engineering*, *9*(2), 156.

Holtrop, J., & Mennen, G. (1982). An approximate power prediction method. *International Shipbuilding Progress*, *29*(335), 166–170.

Hong Kong Marine Department. (2008). Merchant shipping (prevention of air pollution) regulation (Cap. 413M [Accessed on 1 April 2023]. https://www.legco.gov.hk/yr2022/chinese/brief/thbtpmlcr810901%5C_20220629-c.pdf

IMO. (2009). Guidelines for voluntary use of the ship energy efficiency operational indicator [Accessed on 9 January 2023]. https://gmn.imo.org/wp-content/uploads/2017/05/Circ-684-EEOI-Guidelines.pdf

IMO. (2019). Guidance on the definition and harmonization of the format and structure of maritime services in the context of e-navigation (MSC.467(101)) [Accessed on 10 January 2023]. https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/enavigation/MSC.467(101).pdf

IMO. (2020a). Fourth greenhouse gas study 2020 [Accessed on 22 January 2023]. https://www.imo.org/en/OurWork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx

IMO. (2020b). Fourth greenhouse gas study 2020. [Accessed on 9 January 2023]. https://www.imo.org/en/OurWork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx

IMO. (2022). MEPC.328(76) [Accessed on 9 January 2023]. https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Air%20pollution/Certified%20copy%20of%20MEPC.328(76).pdf

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Kim, D., Antariksa, G., Handayani, M. P., Lee, S., & Lee, J. (2021). Explainable anomaly detection framework for maritime main engine sensor data. *Sensors*, *21*(15), 5200.

Kim, G., & Lim, S. (2022). Development of an interpretable maritime accident prediction system using machine learning techniques. *IEEE Access*, *10*, 41313–41329.

Le, L. T., Lee, G., Kim, H., & Woo, S.-H. (2020a). Voyage-based statistical fuel consumption models of ocean-going container ships in Korea. *Maritime Policy & Management*, *47*(3), 304–331.

Le, L. T., Lee, G., Park, K.-S., & Kim, H. (2020b). Neural network-based fuel consumption estimation for container ships in Korea. *Maritime Policy & Management*, *47*(5), 615–632.

Lee, J.-H., Nam, Y.-S., Kim, Y., Liu, Y., Lee, J., & Yang, H. (2022). Real-time digital twin for ship operation in waves. *Ocean Engineering*, *266*, 112867.

Li, L., Fan, Y., Tse, M., & Lin, K.-Y. (2020a). A review of applications in federated learning. *Computers & Industrial Engineering*, *149*, 106854.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020b). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60.

Li, X., Du, Y., Chen, Y., Nguyen, S., Zhang, W., Schönborn, A., & Sun, Z. (2022). Data fusion and machine learning for ship fuel efficiency modeling: Part I–Voyage report data and meteorological data. *Communications in Transportation Research*, *2*, 100074.

Liu, W., Xu, X., Wu, L., Qi, L., Jolfaei, A., Ding, W., & Khosravi, M. R. (2022). Intrusion detection for maritime transportation systems with batch federated aggregation. *IEEE Transactions on Intelligent Transportation Systems*, *24*(2), 2503–2514.

Liu, Y., James, J., Kang, J., Niyato, D., & Zhang, S. (2020). Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, *7*(8), 7751–7763.

Lloyd's List. (2022). Pilot project to share voyage data aims to cut emissions. [Accessed on 17 November 2022]. https://lloydslist-maritimeintelligence-informa-com.ezproxy.lb.polyu.edu.hk/LL1143012/Pilot-project-to-share-voyage-data-aims-to-cut-emissions

Lo, H. K., & McCord, M. R. (1995). Routing through dynamic ocean currents: General heuristics and empirical results in the gulf stream region. *Transportation Research Part B: Methodological*, *29*(2), 109–124.

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, *7*, 154096–154113.

Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Ma, X., Jiang, Q., Shojafar, M., Alazab, M., Kumar, S., & Kumari, S. (2022). Disbezant: Secure and robust federated learning against byzantine attack in IoT-enabled MTS. *IEEE Transactions on Intelligent Transportation Systems*, *24*(2), 2492–2502.

Ma, Y., Zhao, Y., Yu, J., Zhou, J., & Kuang, H. (2023). An interpretable gray box model for ship fuel consumption prediction based on the shap framework. *Journal of Marine Science and Engineering*, *11*(5), 1059.

Mauro, F., & Kana, A. (2023). Digital twin for ship life-cycle: A critical systematic review. *Ocean Engineering*, *269*, 113479.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence and Statistics*, 1273–1282.

Medina, J. R., Molines, J., González-Escrivá, J. A., & Aguilar, J. (2020). Bunker consumption of containerships considering sailing speed and wind conditions. *Transportation Research Part D: Transport and Environment*, *87*, 102494.

Meng, Q., Du, Y., & Wang, Y. (2016). Shipping log data based container ship fuel efficiency modeling. *Transportation Research Part B: Methodological*, *83*, 207–229.

Mohri, M., Sivek, G., & Suresh, A. T. (2019). Agnostic federated learning. *International Conference on Machine Learning*, 4615–4625.

Nanfack, G., Temple, P., & Frénay, B. (2023). Learning customised decision trees for domain-knowledge constraints. *Pattern Recognition*, *142*, 109610.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, *24*(2), 227–234.

Niknam, S., Dhillon, H. S., & Reed, J. H. (2020). Federated learning for wireless communications: Motivation, opportunities, and challenges. *IEEE Communications Magazine*, *58*(6), 46–51.

Ozsari, I. (2023). Predicting main engine power and emissions for container, cargo, and tanker ships with artificial neural network analysis. *Brodogradnja: An International Journal of Naval Architecture and Ocean Engineering for Research and Development*, *74*(2), 77–94.

Parkes, A., Sobey, A., & Hudson, D. (2018). Physics-based shaft power prediction for large merchant ships using neural networks. *Ocean Engineering*, *166*, 92–104.

Passban, P., Roosta, T., Gupta, R., Chadha, A., & Chung, C. (2022). Training mixed-domain translation models via federated learning. *arXiv preprint arXiv:2205.01557*.

Petersen, J. P., Winther, O., & Jacobsen, D. J. (2012a). A machine-learning approach to predict main energy consumption under realistic operational conditions. *Ship Technology Research*, *59*(1), 64–72.

Petersen, J. P. (2011). Mining of ship operation data for energy conservation.

Petersen, J. P., Jacobsen, D. J., & Winther, O. (2012b). Statistical modelling for ship propulsion efficiency. *Journal of Marine Science and Technology*, *17*, 30–39.

Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: A systematic literature review. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1–31.

Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2022). Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, *70*, 1142–1154.

Psaraftis, H. N., & Kontovas, C. A. (2014). Ship speed optimization: Concepts, models and combined speed-routing scenarios. *Transportation Research Part C: Emerging Technologies*, *44*, 52–69.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*, 81–106.

Raissi, M., Yazdani, A., & Karniadakis, G. E. (2020). Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, *367*(6481), 1026–1030.

Rakke, J. G., Christiansen, M., Fagerholt, K., & Laporte, G. (2012). The traveling salesman problem with draft limits. *Computers & Operations Research*, *39*(9), 2161–2167.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Ronen, D. (1982). The effect of oil price on the optimal speed of ships. *Journal of the Operational Research Society*, *33*(11), 1035–1040.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. The MIT press.

Seeger, M. (2004). Gaussian processes for machine learning. *International Journal of Neural Systems*, *14*(02), 69–106.

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

Sundararajan, M., & Najmi, A. (2020). The many shapley values for model explanation. *International Conference on Machine Learning*, 9269–9278.

Tan, F., Fu, X., Zhang, Y., & Bourgeois, A. G. (2008). A genetic algorithm-based method for feature subset selection. *Soft Computing*, *12*, 111–120.

TradeLens. (2022). Maersk and IBM have announced the discontinuation of the TradeLens platform. [Accessed on 5 December 2022]. https://www.tradelens.com/

UNCTAD. (2022). Review of maritime transport. [Accessed on 18 July 2023]. https://unctad.org/system/files/official-document/rmt2022_en.pdf

Uyanık, T., Bakar, N. N. A., Kalenderli, Ö., Arslanoğlu, Y., Guerrero, J. M., & Lashab, A. (2023). A data-driven approach for generator load prediction in shipboard microgrid: The chemical tanker case study. *Energies*, *16*(13), 5092.

Uyanık, T., Karatuğ, Ç., & Arslanoğlu, Y. (2020). Machine learning approach to ship fuel consumption: A case of container vessel. *Transportation Research Part D: Transport and Environment*, *84*, 102389.

Veerappa, M., Anneken, M., Burkart, N., & Huber, M. F. (2022). Validation of xai explanations for multivariate time series classification in the maritime domain. *Journal of Computational Science*, *58*, 101539.

Wang, H., Yan, R., Wang, S., & Zhen, L. (2023). Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. *Transportation Research Part C: Emerging Technologies*, *157*, 104361.

Wang, H., Yi, W., Liu, Y., et al. (2022a). An innovative approach of determining the sample data size for machine learning models: A case study on health and safety management for infrastructure workers. *Electronic Research Archive*, *30*(9), 3452–3462.

Wang, L., Gopal, R., Shankar, R., & Pancras, J. (2022b). Forecasting venue popularity on location-based services using interpretable machine learning. *Production and Operations Management*, *31*(7), 2773–2788.

Wang, S., Ji, B., Zhao, J., Liu, W., & Xu, T. (2018a). Predicting ship fuel consumption based on lasso regression. *Transportation Research Part D: Transport and Environment*, *65*, 817–824.

Wang, S., & Meng, Q. (2012). Sailing speed optimization for container ships in a liner shipping network. *Transportation Research Part E: Logistics and Transportation Review*, *48*(3), 701–714.

Wang, S., Meng, Q., & Liu, Z. (2013). Bunker consumption optimization methods in shipping: A critical review and extensions. *Transportation Research Part E: Logistics and Transportation Review*, *53*, 49–62.

Wang, S., Qi, J., & Laporte, G. (2022c). Governmental subsidy plan modeling and optimization for liquefied natural gas as fuel for maritime transportation. *Transportation Research Part B: Methodological*, *155*, 304–321.

Wang, S., Qi, J., & Laporte, G. (2022d). Optimal subsidy design for shore power usage in ship berthing operations. *Naval Research Logistics (NRL)*, *69*(4), 566–580.

Wang, S., & Wang, X. (2016). A polynomial-time algorithm for sailing speed optimization with containership resource sharing. *Transportation Research Part B: Methodological*, *93*, 394–405.

Wang, Y., Meng, Q., & Kuang, H. (2018b). Jointly optimizing ship sailing speed and bunker purchase in liner shipping with distribution-free stochastic bunker prices. *Transportation Research Part C: Emerging Technologies*, *89*, 35–52.

Warwicker, J. A., & Rebennack, S. (2022). A comparison of two mixed-integer linear programs for piecewise linear function fitting. *INFORMS Journal on Computing*, *34*(2), 1042–1047.

WeBank. (2022). An industrial grade federated learning framework. [Accessed on 5 January 2023]. https://fate.fedai.org/

Wu, W.-M. (2020). The optimal speed in container shipping: Theory and empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, *136*, 101903.

Xiao, Z., Xu, X., Xing, H., Song, F., Wang, X., & Zhao, B. (2021). A federated learning system with enhanced feature extraction for human activity recognition. *Knowledge-Based Systems*, *229*, 107338.

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

Yan, R., Wang, S., & Du, Y. (2020). Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transportation Research Part E: Logistics and Transportation Review*, *138*, 101930.

Yan, R., Wang, S., & Psaraftis, H. N. (2021b). Data analytics for fuel consumption management in maritime transportation: Status and perspectives. *Transportation Research Part E: Logistics and Transportation Review*, *155*, 102489.

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

Yang, L., Chen, G., Rytter, N. G. M., Zhao, J., & Yang, D. (2019). A genetic algorithm-based grey-box model for ship fuel consumption prediction towards sustainable shipping. *Annals of Operations Research*, 1–27.

Yao, Z., Ng, S. H., & Lee, L. H. (2012). A study on bunker fuel management for the shipping liner services. *Computers & Operations Research*, *39*(5), 1160–1172.

Yue, X., Nouiehed, M., & Al Kontar, R. (2023). Gifair-fl: A framework for group and individual fairness in federated learning. *INFORMS Journal on Data Science*, *2*(1), 10–23.

Zhang, C., Zou, X., & Lin, C. (2022a). Fusing xgboost and shap models for maritime accident prediction and causality interpretability analysis. *Journal of Marine Science and Engineering*, *10*(8), 1154.

Zhang, J., Li, Z., Li, B., Xu, J., Wu, S., Ding, S., & Wu, C. (2022b). Federated learning with label distribution skew via logits calibration. *International Conference on Machine Learning*, 26311–26329.

Zhang, S., Choromanska, A. E., & LeCun, Y. (2015). Deep learning with elastic averaging SGD. *Advances in Neural Information Processing Systems*, *28*.

Zhang, Z., Guan, C., Chen, H., Yang, X., Gong, W., & Yang, A. (2021). Adaptive privacy-preserving federated learning for fault diagnosis in internet of ships. *IEEE Internet of Things Journal*, *9*(9), 6844–6854.

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*.

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2019). Review of maritime traffic models from vessel behavior modeling perspective. *Transportation Research Part C: Emerging Technologies*, *105*, 323–345.

Zhu, H., Xu, J., Liu, S., & Jin, Y. (2021). Federated learning on non-IID data: A survey. *Neurocomputing*, *465*, 371–390.

Appendices

Appendix A

We first briefly introduce the back propagation method proposed by Rumelhart et al. (1986) in a general form and then give details about the parameters in our model.

A typical neural network consists of an input layer, a hidden layer, and an output layer. We use symbols i, h, and j to denote the nodes in each layer, respectively. The weight coefficients between input layer nodes and hidden layer nodes, hidden layer nodes and output layer nodes are represented by w_{ih} and w_{hj} respectively. As shown in Figure A1, we use k to denote each sample, which has input variables x_i^k and the target t_j^k. The output layer outputs z_j^k, the closer to t_j^k the better. For each sample trained by the network, it undergoes the following calculations.

From the input layer to the hidden layer, the hidden layer first calculates the weighted sum of these input nodes:

$$net_h^k = \sum_i w_{ih} x_i^k. \tag{A1}$$

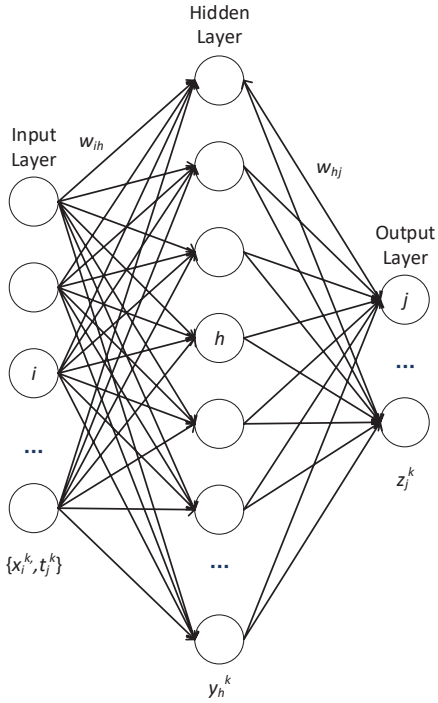


Figure A1: The structure of a typical neural network

And then activation function, denoted by f, calculates the final output of the hidden layer:

$$y_h^k = f(net_h^k) = f(\sum_i w_{ih} x_i^k). \tag{A2}$$

The output layer then calculates the weighted sum of inputs from the hidden layer:

$$net_j^k = \sum_h w_{hj} y_h^k = \sum_h w_{hj} f(\sum_i w_{ih} x_i^k). \tag{A3}$$

The final outputs will be:

$$z_j^k = f(net_j^k) = f(\sum_h w_{hj} f(\sum_i w_{ih} x_i^k)). \tag{A4}$$

To simplify notation, we use a fully connected neural network with one hidden layer as an example. For non-fully connected neural networks, we only need to exclude the corresponding weight coefficients during the derivation process. And for neural networks with multiple hidden layers, more partial derivatives of errors and weights between layers need to be calculated. The errors can be calculated by:

$$E(\boldsymbol{w})^k = \frac{1}{2}\sum_j (t_j^k - z_j^k)^2 = \frac{1}{2}\sum_j (t_j^k - f(\sum_h w_{hj} f(\sum_i w_{ih} x_i^k)))^2. \tag{A5}$$

We use η to denote the learning rate. Then, we have:

$$\begin{aligned} \Delta w_{hj} &= -\eta \frac{\partial E(\boldsymbol{w})^k}{\partial w_{hj}} \\ &= -\eta \sum_k \frac{\partial E(\boldsymbol{w})^k}{\partial net_j^k} \frac{\partial net_j^k}{\partial w_{hj}} \\ &= \eta \sum_k (t_j^k - z_j^k) f'(net_j^k) y_h^k \\ &= \eta \sum_k \delta_j^k y_h^k \end{aligned} \tag{A6}$$

where $\delta_j^k = f'(net_j^k)(t_j^k - z_j^k)$. Similarly,

$$\begin{aligned}
\Delta w_{ih} &= -\eta \frac{\partial E(\boldsymbol{w})^k}{\partial w_{ih}} \\
&= -\eta \sum_{k,j} \frac{\partial E(\boldsymbol{w})^k}{\partial z_j^k} \frac{\partial z_j^k}{\partial w_{ih}} \\
&= \eta \sum_{k,j} (t_j^k - z_j^k) \frac{\partial z_j^k}{\partial w_i^h} \\
&= \eta \sum_{k,j} (t_j^k - z_j^k) \frac{\partial z_j^k}{\partial net_j^k} \frac{\partial net_j^k}{\partial y_h^k} \frac{\partial y_h^k}{\partial w_i^h} \\
&= \eta \sum_{k,j} (t_j^k - z_j^k) f'(net_j^k) w_{hj} \frac{\partial y_h^k}{\partial net_h^k} \frac{\partial net_h^k}{\partial w_{ih}} \\
&= \eta \sum_{k,j} (t_j^k - z_j^k) f'(net_j^k) w_{hj} f'(net_h^k) x_i^k \\
&= \eta \sum_{k,j} w_{hj} f'(net_h^k) x_i^k \\
&= \eta \sum_{k} (f'(net_h^k) \sum_j \delta_j^k w_{hj}) x_i^k \\
&= \eta \sum_{k} \delta_h^k x_i^k
\end{aligned} \tag{A7}$$

where $\delta_h^k = f'(net_h^k) \sum_j w_{hj} \delta_j^k$. Therefore, the network can update weight coefficients according to Δw_{hj} and Δw_{ih}.

In our study, we adopt a three-hidden-layer neural network with 20 neurons. Based on the characteristics of our problem, there are twenty input nodes—sailing speed, wave direction, wave height, wind force, cargo weight, historical daily ship fuel consumption for the past two weeks, and one node for bias—and one output node—daily ship fuel consumption (see Figure A2). We determine the parameter settings of our model after experiments and the details are shown in Table A1.

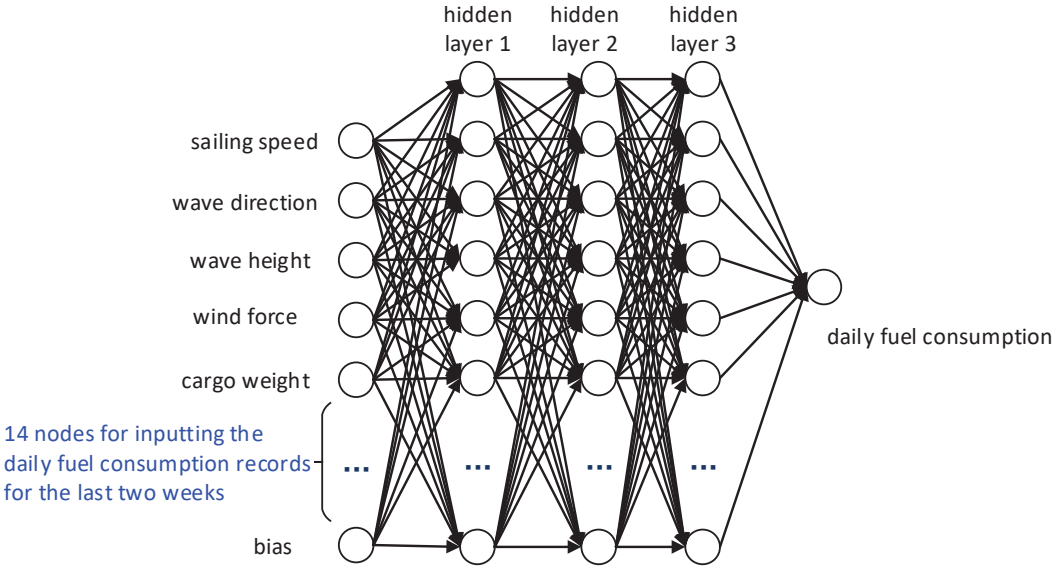


Figure A2: The structure of our neural network model

Appendix B

The ship speed s_i is discretized in units of 0.1 knot. We define

$$V = \lfloor \frac{s_{\max} - s_{\min}}{0.1} \rfloor + 1. \tag{B1}$$

We set $v = 0, 1, ..., V$. The discretized speeds are $s_i^0 = s_{\min}$, $s_i^1 = s_{\min} + 0.1$, $s_i^2 = s_{\min} + 0.1 \times 2$, $s_i^3 = s_{\min} + 0.1 \times 3$,...,$s_i^V = \min\{s_{\min} + 0.1 \times V, s_{\max}\}$. We use binary decision variable z_i^v to indicate which speed is adopted and z_i^v equals 1 if a ship sails with speed s_i^v in segment i and 0 otherwise. Then, model [M1] can be linearized as follows.

Table A1: Parameter settings

Parameter	Setting
Number of hidden layers	3
Number of neurons in each hidden layer	20
Number of neurons in the input layer	6
Number of neurons in the output layer	1
Activation function	Sigmoid
Learning rate	0.1
Number of epochs	50
Number of batch size	10

[M2]

$$\min \sum_{i \in I} \sum_{v=0}^{V} z_i^v \times F(\boldsymbol{x}, s_i^v) \times \frac{L_i}{24 \times s_i^v} \tag{B2}$$

subject to

$$t_i = t_{i-1} + \sum_{v=0}^{V} (\frac{L_i}{s_i^v} \times z_i^v), i \in I \tag{B3}$$

$$\sum_{v=0}^{V} z_i^v = 1, i \in I \tag{B4}$$

$$t_{|I|} \leq T \tag{B5}$$

$$t_0 = 0 \tag{B6}$$

$$z_i^v \in \{0, 1\}, i \in I, v = 0, ..., V. \tag{B7}$$

When conducting numerical experiments, we only have noon report data, which does not include segment data. Therefore, we develop model [M3] to replace model [M2]. We treat one day as one segment. The original sailing speed on segment i, i.e., on the day i, is denoted by $s_i^{\#}$.

[M3]

$$\min \sum_{i \in I} \sum_{v=0}^{V} z_i^v \times F(\boldsymbol{x}, s_i^v) \tag{B8}$$

subject to

$$\sum_{i \in I} \sum_{v=0}^{V} z_i^v \times s_i^v \geq \sum_{i \in I} s_i^{\#} \tag{B9}$$

$$\sum_{v=0}^{V} z_i^v = 1, i \in I \tag{B10}$$

$$z_i^v \in \{0, 1\}, i \in I, v = 0, ..., V. \tag{B11}$$

Appendix C

We test the performance of FL and individual training by changing the length of the input fuel consumption history data to one month. Results are shown in Table C1. We find that using one month of fuel consumption data does not improve the performance of both models in terms of MAE and RMSE. Considering the computing efficiency, we believe that two weeks of data is enough.

Table C1.: The performance of FL under different fuel consumption historical data inputs

Ship number	MAE		RMSE	
	Individual training	FL mechanism	Individual training	FL mechanism
Two weeks of fuel consumption historical data				
1	0.0094	0.0024	0.0109	0.0031
2	0.0085	0.0014	0.0093	0.0016
3	0.0030	0.0016	0.0034	0.0021
4	0.0040	0.0024	0.0050	0.0030
5	0.0024	0.0023	0.0028	0.0026
6	0.0018	0.0015	0.0022	0.0020
7	0.0054	0.0023	0.0065	0.0029
8	0.0070	0.0022	0.0092	0.0028
9	0.0022	0.0020	0.0028	0.0025
10	0.0021	0.0020	0.0027	0.0027
11	0.0019	0.0016	0.0024	0.0021
12	0.0015	0.0015	0.0017	0.0018
13	0.0021	0.0021	0.0025	0.0026
14	0.0021	0.0017	0.0024	0.0021
15	0.0020	0.0020	0.0027	0.0026
16	0.0060	0.0023	0.0068	0.0028
17	0.0066	0.0019	0.0072	0.0024
18	0.0072	0.0019	0.0084	0.0024
19	0.0024	0.0015	0.0030	0.0019
20	0.0039	0.0022	0.0046	0.0027
Mean value	0.0041	0.0019	0.0048	0.0024
Standard deviation	0.0025	0.0003	0.0028	0.0004
One month of fuel consumption historical data				
1	0.0094	0.0024	0.0109	0.0031
2	0.0085	0.0014	0.0093	0.0016
3	0.0030	0.0016	0.0034	0.0021
4	0.0040	0.0024	0.0050	0.0030
5	0.0024	0.0023	0.0028	0.0026
6	0.0018	0.0015	0.0022	0.0020
7	0.0054	0.0023	0.0065	0.0029
8	0.0070	0.0022	0.0092	0.0028
9	0.0022	0.0020	0.0028	0.0025
10	0.0021	0.0020	0.0027	0.0027
11	0.0019	0.0016	0.0024	0.0021
12	0.0015	0.0015	0.0017	0.0018
13	0.0021	0.0021	0.0025	0.0026
14	0.0021	0.0017	0.0024	0.0021
15	0.0020	0.0020	0.0027	0.0026
16	0.0060	0.0023	0.0068	0.0028
17	0.0066	0.0019	0.0072	0.0024
18	0.0072	0.0019	0.0084	0.0024
19	0.0024	0.0015	0.0030	0.0019
20	0.0039	0.0022	0.0046	0.0027
Mean value	0.0050	0.0023	0.0060	0.0028
Standard deviation	0.0021	0.0006	0.0025	0.0007