Transportation Research Part D 159 (2026) 105491

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Data-driven investigation of ship fuel consumption integrating causal inference and hierarchical analysis☆

Wenjie Cao [a,b], Xinjian Wang [a,b,c,*], Wei Zhang [d], Huanhuan Li [e], Siming Fang [f], Zaili Yang [b,*]

[a] Navigation College, Dalian Maritime University, Dalian 116026, PR China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK
[c] State Key Laboratory of Maritime Technology and Safety, Dalian 116026, PR China
[d] Centre for Maritime and Logistics Management, Australian Maritime College, University of Tasmania, Maritime Way, Newnham, TAS 7248, Australia
[e] School of Engineering, University of Southampton, Southampton SO17 1, UK
[f] Navigation College, Jimei University, Xiamen 361021, PR China

A R T I C L E I N F O

Keywords:
Maritime transport
Maritime decarbonisation
Ship fuel consumption
Causal inference
Machine learning
Explainable artificial intelligence

A B S T R A C T

This study develops a data-driven analytical framework that integrates causal inference with hierarchical analysis to reveal the underlying mechanisms influencing ship fuel consumption (SFC). The framework combines multi-source data fusion, direct linear non-gaussian acyclic model-based causal discovery, and double machine learning with causal forests model to estimate Average Treatment Effects (ATEs), followed by interpretive structural modelling for hierarchical decomposition. Experimental results under the expanded DAG-derived adjustment specification indicate that Daily sailing hours has the strongest positive conditional effect on SFC (ATE = 2.058), followed by Main engine RPM with a positive but more uncertain effect estimate (ATE = 0.268). The hierarchical analysis further organises the directional-dependence network into interpretable structural levels, thereby illustrating possible multi-level transmission patterns among operational and environmental factors. This framework provides an exploratory and graph-informed analytical basis for interpreting directional dependencies and observed-covariate conditional effect patterns in SFC, offering cautious decision support for data-driven energy management. The source code is publicly available at: https://github.com/AdvMarTech/ship_fuel_consum_causalinfer.

1. Introduction

With the deepening of global trade integration, maritime transport, as the backbone of international trade, carries approximately 90% of global cargo volume and plays an essential role in driving economic development worldwide (Ma et al., 2026; Wang et al., 2023b). However, the continued expansion of maritime trade has brought increasing attention to ship fuel consumption (SFC) and its associated environmental impacts. According to the Fourth Greenhouse Gas Study published by the International Maritime

☆ This article is part of a special issue entitled: 'Machine Ln. Maritime' published in Transportation Research Part D.

* Corresponding authors.
E-mail addresses: cwj0060@dlmu.edu.cn (W. Cao), wangxinjian@dlmu.edu.cn (X. Wang), vera.zhang@utas.edu.au (W. Zhang), huanhuan.li@soton.ac.uk (H. Li), fangsiming@jmu.edu.cn (S. Fang), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.trd.2026.105491
Received 30 October 2025; Received in revised form 15 June 2026; Accepted 22 June 2026
Available online 29 June 2026

Organisation (IMO), carbon dioxide emissions from international shipping rose from 962 million tonnes in 2012 to 1,056 million tonnes in 2018, marking an increase of 9.3% and accounting for 2.89% of total global greenhouse gas emissions (International Maritime Organization, 2020). In response to the growing challenge of climate change, the IMO adopted a revised greenhouse gas reduction strategy in 2023, targeting a peak in emissions from international shipping as soon as possible and aiming to achieve net zero around 2050 (International Maritime Organization, 2023). To achieve this goal, the global shipping industry is facing increasing pressure from energy transition demands and environmental regulations (Li et al., 2025). The IMO has introduced a series of mandatory measures, including the Ship Energy Efficiency Management Plan (SEEMP, an onboard operational management plan aimed at improving a ship's energy efficiency), the Energy Efficiency Design Index (EEDI, a mandatory design standard that benchmarks the CO_2 efficiency of new ships), and the Carbon Intensity Indicator (CII, an operational metric used to rate a ship's annual carbon-intensity performance), which provide a regulatory framework and technical standards to support the green transformation of the shipping industry (Luo et al., 2024; Wang et al., 2022). In this context, shipping companies, port authorities, and policymakers are actively exploring various strategies for energy conservation and emissions reduction individually and collectively. Among these efforts, improving operational efficiency to control SFC has become a central focus (Li et al., 2024c; Wang et al., 2025b). Compared with ship design improvements, operational measures typically involve lower investment, shorter implementation times, and faster outcomes, making them particularly effective and applicable in the practice of maritime emissions reduction (Cao et al., 2025c; Coraddu et al., 2017; Yang et al., 2019).

SFC prediction, as a fundamental technological component of energy efficiency management in the shipping industry, holds strategic importance for achieving operational cost control and carbon emissions reduction. However, SFC is influenced by a wide range of factors, including sailing speed, loading condition, ship age, maintenance status, marine environment, and meteorological variables. These factors are often interrelated in a complex nonlinear manner, which poses substantial challenges for traditional modelling approaches to accurately quantifying the underlying mechanisms (Wang et al., 2023a; Yan et al., 2021). In recent years, Machine Learning (ML) techniques have been increasingly applied to SFC prediction, owing to their strong capabilities in processing high dimensional data and capturing nonlinear relationships (Cao et al., 2026b; Wang et al., 2024). Compared to conventional physics-based models, ML approaches offer improved effectiveness in identifying intricate patterns and interactions among influencing factors (Cao et al., 2025b; Luo et al., 2023).

Although existing ML models have significantly improved prediction accuracy, most studies primarily focus on identifying variables that are strongly correlated with SFC, while failing to uncover the actual causal relationships between these variables and fuel usage (Huang et al., 2024) due to the complexity of the issue in nature. In other words, while such models can indicate which factors are associated with SFC, they do not address whether and how the changes in these factors genuinely cause variation in consumption. This absence of causal understanding limits the development and implementation of targeted energy saving strategies. Moreover, the variables influencing SFC are often interdependent and interact in a coordinated manner, forming a complex causal network. Traditional correlation analysis overlooks the structural causality among variables, which may introduce bias in assessing the true overall contribution of each factor at a system level (Cao et al., 2025a; Li et al., 2024b; Zhang et al., 2025a).

To address the challenges outlined above, this study proposes a systematic holistic research framework that integrates data driven modelling with causal inference analysis to systematically uncover the drivers of SFC and their underlying mechanisms. The framework incorporates a range of advanced methods, including multi-source heterogeneous data fusion, causal structure identification and quantification. This integrated approach not only enables accurate prediction of SFC, but also facilitates in depth analysis of structural causality and transmission pathways among the influencing factors.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature, identifies the limitations and key challenges in existing studies, and defines the state of the art in the field and establishes the theoretical foundation. Section 3 provides a detailed description of the proposed multi-phase analytical framework, which includes multi-source heterogeneous data fusion, feature selection, and a causal inference system that integrates Direct Linear Non-Gaussian Acyclic Model (DirectLiNGAM), Causal Forest-based Double Machine Learning (DML-CF), and Interpretive Structural Modelling (ISM). Section 4 presents the empirical findings, offering an in-depth analysis of causal effect quantification, and the hierarchical structure and propagation pathways within the influence network. Section 5 outlines the theoretical contributions and practical implications of this study, offering decision support for enhancing energy efficiency and promoting green transformation in the shipping industry. Section 6 concludes this study by summarising the main findings, discussing the limitations, and proposing directions for future research.

2. Literature review

This section systematically reviews the academic background and recent advancements relevant to SFC. The review is structured around two progressive levels. First, it focuses on the traditional identification and quantitative analysis of the key drivers influencing SFC, and provides a comprehensive overview of established findings in areas such as ship design, operational practices, and external environmental factors. Building on this foundation, this section then examines the development of causal inference methods, which have emerged to address the limitations of conventional correlation-based approaches. The theoretical significance and practical applications of these methods are discussed within the context of complex system analysis, offering a foundation for uncovering the deeper and more structured mechanisms underlying the influence of these factors.

2.1. Identification and quantification of factors affecting ship fuel consumption

SFC is a nonlinear system process influenced by multiple complex factors. Accurately identifying and quantifying its primary drivers is essential for enhancing energy efficiency in shipping and developing effective energy saving strategies. Existing research indicates that the factors influencing SFC can generally be classified into three major categories: ship design, navigational operations, and external environmental conditions (Hu et al., 2025; Meng et al., 2016; Yan et al., 2021).

Ship design factors form the foundational framework for fuel efficiency, primarily encompassing parameters such as hull shape, propeller configuration, and deadweight tonnage (Saydam et al., 2022; Tillig et al., 2015). These factors directly affect a vessel's hydrodynamic resistance during navigation and thus fundamentally determine its energy consumption levels (Kim et al., 2017). Hamed (2022) demonstrated significant differences in resistance characteristics among various hull forms: slender hulls are more advantageous in reducing wave resistance, whereas wider waterline designs help lower frictional resistance. Yan et al. (2024) systematically analysed the nonlinear relationship between the hull's principal dimensions and fuel efficiency, highlighting drag reduction and propulsion efficiency enhancement as the core goals of structural design. As a critical component of propulsion, the optimisation of propeller configuration significantly affects SFC. Inal et al. (2022) further highlighted that hybrid and electric propulsion technologies can contribute to fuel-consumption and emission reductions by improving overall system efficiency. Marques et al. (2018) achieved simultaneous reductions in SFC and emissions for a container ship by designing propellers with a minimum brake specific SFC at the engine's operating point. Maersk's Triple E class container ships feature a low speed, large diameter twin propeller design, improving SFC per unit of cargo by over 37 percent compared to their predecessors at equivalent load capacity (Halff et al., 2019). Additionally, deadweight tonnage significantly affects fuel efficiency, for example, the difference in SFC between fully loaded and ballasted conditions in ultra large crude oil carriers can reach 25 to 30 percent of total consumption (Adland et al., 2020; Nguyen et al., 2022).

Navigational operational factors, as controllable elements of ship operation, primarily include parameters such as sailing speed, voyage distance and duration, and cargo capacity (Psaraftis and Kontovas, 2014; Theocharis et al., 2019). These parameters not only directly influence SFC, but also affect overall energy efficiency through complex interactions and interdependencies (Işıklı et al., 2020). Among these, controlling sailing speed is widely regarded as one of the most effective approaches to energy saving. Studies have shown that a ship's daily fuel consumption is approximately proportional to the cube of its sailing speed (Wang and Meng, 2012). The application of slow steaming strategies has been extensively investigated as a practical method to reduce both SFC and carbon emissions. For instance, Balcombe et al. (2019) report that slow steaming can reduce fuel use and CO_2 emissions by 20 to 30 percent, and up to 60 percent under extreme conditions. Building on this, Degiuli et al. (2021) found that a 13.6 percent reduction in the speed of Panamax container ships in the Mediterranean can yield up to a 31 percent decrease in SFC and greenhouse gas emissions. In addition, Borkowski et al. (2019), through an analysis of historical operational data, revealed that sailing at 20 knots results in 130 percent higher annual SFC compared to 15 knots, while cargo transport increases by only 29 percent. Voyage duration and route selection also play a critical role in fuel efficiency by interacting with variables such as speed and weather conditions (Huang et al., 2025a).

External environmental factors, as uncontrollable variables in ship navigation, primarily include natural conditions such as sea state, ocean currents, wind speeds, and wind directions (Wang et al., 2025d). These factors significantly influence wave and wind resistance, alter additional propulsion power requirements, and affect hull stability, thereby directly impacting SFC (Wang et al., 2025a). Existing research has shown that an increase in wave height from one to two metres can raise hull resistance by 20 to 30 percent (Farag and Ölçer, 2020). Hull fouling and coating degradation increase surface roughness; cumulative fouling over six months can result in a speed loss of 1.5 to 2.0 knots (Bialystocki and Konovessis, 2016). In such cases, ship operators often increase engine power to maintain scheduled arrivals, which further raises SFC (Kontovas, 2014). Notably, there is a significant interaction between environmental and operational factors. Route optimisation strategies that incorporate environmental variables such as wind, currents, and waves can reduce emissions and SFC by up to 30 percent on medium and long distance voyages (Borén et al., 2022). Although the disruptive effects of adverse weather on ship schedules have been acknowledged by researchers such as Bell et al. (2013) and Saber et al. (2025), the quantitative relationship between fuel efficiency and weather conditions remains underexplored.

2.2. Application of causal inference methods in complex system analysis

The driving mechanism of SFC is inherently complex and influenced by the interaction among hull structure, operational state, and external environmental conditions (Wang et al., 2025a). Achieving accurate prediction and effective control requires a thorough analysis of causal mechanisms and a quantitative assessment of the interactions among influencing factors and their specific contributions to system dynamics (Huang et al., 2024). Traditional correlation based statistical methods often overlook causal dependencies, which can result in systematic bias during attribution analysis. Consequently, the adoption of causal inference approaches has become a necessary and increasingly prominent direction in this field (Wang et al., 2025c).

Structural Causal Modelling (SCM), a classical framework for causal inference, describes causal relationships among variables and estimates causal effects by integrating structural equations with causal diagrams (Arif and MacNeil, 2023). This method has been successfully applied across various fields. Sun et al. (2023) employed SCM to estimate the Average Treatment Effect (ATE) of different groundwater extraction strategies, aiming to optimise coastal groundwater management. Liu et al. (2025a) applied SCM to investigate the causal relationships between key explanatory variables and urban water quality. In the transport domain, Singh and Borrok (2019) utilised spatio temporal Granger causality analysis to examine the causal mechanisms underlying long term groundwater patterns. Similarly, He et al. (2023) proposed a spatio temporal Granger causality based approach to capture the global dynamic transmission of traffic information and demonstrated its effectiveness in improving the accuracy of long term traffic flow prediction.

LiNGAM is a prominent causal discovery method grounded in SCM, originally proposed by Shimizu et al. (2006). The model assumes linear relationships among variables and Non-Gaussian, statistically independent noise in the data generating process, and it is solved using independent component analysis. Moneta et al. (2013) demonstrated the effectiveness of LiNGAM in identifying causal relationships among economic variables by comparing its performance with traditional methods. DirectLiNGAM, an enhanced version introduced by Shimizu et al. (2011), addresses the instability caused by random initialisation in independent component analysis and offers improved robustness and computational efficiency. More recently, Chauhan et al. (2024) integrated the causal discovery algorithm with structural equation modelling (SEM) to develop the DirectLiNGAM-SEM approach for analysing travel mode choice, validating its causal identification capability using data from the New York metropolitan area.

In recent years, innovative frameworks that integrate causal reasoning with ML have begun to emerge in the transport and energy sectors (Huang et al., 2024). These developments not only facilitate the quantification of causal dependencies among variables but also deepen the theoretical understanding of underlying mechanisms (Liu et al., 2025a). Causal Machine Learning (CML) effectively combines causally interpretable algorithms with the capacity of ML to detect statistical patterns in high dimensional, unstructured data (Dibaeinia et al., 2025; Tang et al., 2025). Causal Forest (CF), a tree based CML approach, has demonstrated strong performance in evaluating the causal effects of different interventions, offering robust statistical inference and actionable causal insights in fields such as public policy and healthcare (Wang et al., 2025c). Double Machine Learning (DML), a general framework that merges ML with causal inference, was introduced by Chernozhukov et al. (2018) and has since been widely applied. For example, Zhang et al. (2022) used DML to analyse the social impacts of London's night time underground services; Nachtigall et al. (2025) employed it to assess the limiting effects of urban planning policies on travel related CO_2 emissions; and Huang et al. (2025b) applied DML to investigate how data resources promote inclusive green growth across 301 prefectural cities.

2.3. State of the art and contributions of this study

This study presents three significant contributions to the field of SFC analysis, offering novel methodological innovations and deeper insights into maritime energy efficiency optimisation. The state-of-the-art for these contributions (C1-C3) is summarised below:

C1. A holistic analytical framework: Integrating predictive modelling and causal inference

State of the art: Current research in maritime energy efficiency often suffers from methodological fragmentation, where high-performance predictive models remain disconnected from formal causal inference methods (Liu et al., 2025b). This disciplinary divide limits the translation of predictive accuracy into actionable, evidence-based interventions, as predictive power does not equate to causal understanding (Cao et al., 2026a). The absence of a unified analytical framework that systematically connects prediction with causation restricts the development of holistic solutions for this complex system.

Solution: This study develops an integrated analytical framework that can establish a systematic progression from data processing to deep causal insight. Notably, it repurposes advanced machine learning algorithms from a purely predictive capacity to serve as instrumental components within the causal inference workflow. The framework adopts a modular architecture, beginning with multi-source data fusion and rigorous feature selection to manage high-dimensional data. Subsequently, it employs ML techniques to inform the selection of candidate variables and to robustly control for confounding effects in the causal estimation stage. This integrated module is designed to first discover the causal structure, then quantify observed-covariate conditional effect patterns, and finally map the hierarchical influence of mechanisms. This systematic approach overcomes the limitations of single-method analyses and provides a standardised methodology for complex systems analysis.

C2. From correlation to causation: A two-stage causal inference module for robust policy guidance

State of the art: Research on SFC has traditionally relied on statistical correlation analysis, and to date, causal inference theory has not been systematically introduced into this field. Correlation-based approaches are widely used to examine the relationships between influencing factors and SFC, but they suffer from fundamental methodological limitations. Specifically, correlation analyses cannot distinguish true causal relationships from spurious associations, and findings derived from such methods are highly vulnerable to confounding by unobserved variables. This may result in the misattribution of causal mechanisms and systematic errors in estimating the effects of interventions (Li et al., 2024b). Consequently, policy recommendations based on correlation analysis often prove ineffective or even counterproductive. This methodological gap severely limits the transition of ship energy efficiency management from experience-driven practices to scientifically grounded and precise decision making.

Solution: This study integrates causal inference theory into SFC analysis by developing a two-stage causal inference module embedded within a broader multi-phase analytical framework. This approach moves beyond traditional correlation-based methods toward causal reasoning in ship energy efficiency research. The term "two-stage" here refers specifically to the causal inference module (Contribution C2), while the overall methodology remains multi-phase, encompassing data fusion, feature selection, causal discovery, effect estimation, and hierarchical mechanism mapping. The proposed causal inference module adopts a progressive two-stage design. In the first stage, the DirectLiNGAM algorithm is employed for data driven discovery of causal structures, enabling the automatic learning of directed acyclic graphs among variables without prior assumptions. In the second stage, DML is integrated with CF to construct the DML-CF model, which estimates observed-covariate conditional effect patterns of key operational factors. This module offers the first causally grounded basis for scientific decision making in ship energy efficiency management and marks a significant methodological advancement in this research field.

C3. From individual effects to holistic hierarchy: Mapping structural paths for strategic intervention

State of the art: While the quantification of individual causal effects represents a significant analytical advancement, this approach often yields a non-holistic understanding of the system, presenting a series of independent causal links in piecemeal without fully elucidating the underlying structure of their interdependencies (Cao et al., 2024). This perspective can obscure the critical distinction between foundational root drivers and intermediate mediating factors, treating all variables as independent intervention points. Consequently, it remains challenging to formulate strategic, multi-level interventions that target the most fundamental levers within the complex SFC system.

Solution: To address this, the final phase of the analytical framework employs ISM as a structural organisation and interpretation tool rather than as an additional causal-identification method. It takes the estimated directed graph as its input and reorganises the retained directional-dependence relationships into hierarchical levels through reachability-based decomposition. This structural decomposition helps distinguish upstream variables, intermediate transmission variables and the final response variable within the estimated network, thereby providing an interpretable map for discussing possible propagation patterns and intervention priorities. Accordingly, the contribution of ISM in this study lies in improving the readability of the estimated causal structure, offering a strategic map that highlights the most effective prioritisation and intervention points for systemic optimisation in ship energy management.

3. Materials and methodology

3.1. The proposed framework

To address the prevalent limitations of conventional correlational studies in the maritime sector, this study establishes a novel, multi-phase analytical framework. Its primary innovation lies in the systematic integration of causal discovery, robust effect quantification, and hierarchical structural analysis into a single, cohesive workflow designed to transition from statistical association to a deep, mechanistic understanding of SFC.

The proposed framework, as illustrated in Fig. 1, proceeds through three logically connected phases. In the first phase, a robust empirical foundation is established through the fusion of multi-source heterogeneous data, including ship noon reports, Automatic Identification System (AIS) data, meteorological records, and oceanographic data. Subsequently, TopoRankFS is applied to construct a parsimonious main-analysis feature subset from the high-dimensional fused dataset. This step is necessary because the interpolation, spatial–temporal matching and direction conversion of meteorological, oceanographic and operational variables inevitably generate partially overlapping, noisy and redundant descriptors, particularly among wind-, wave-, current- and resistance-related variables. In

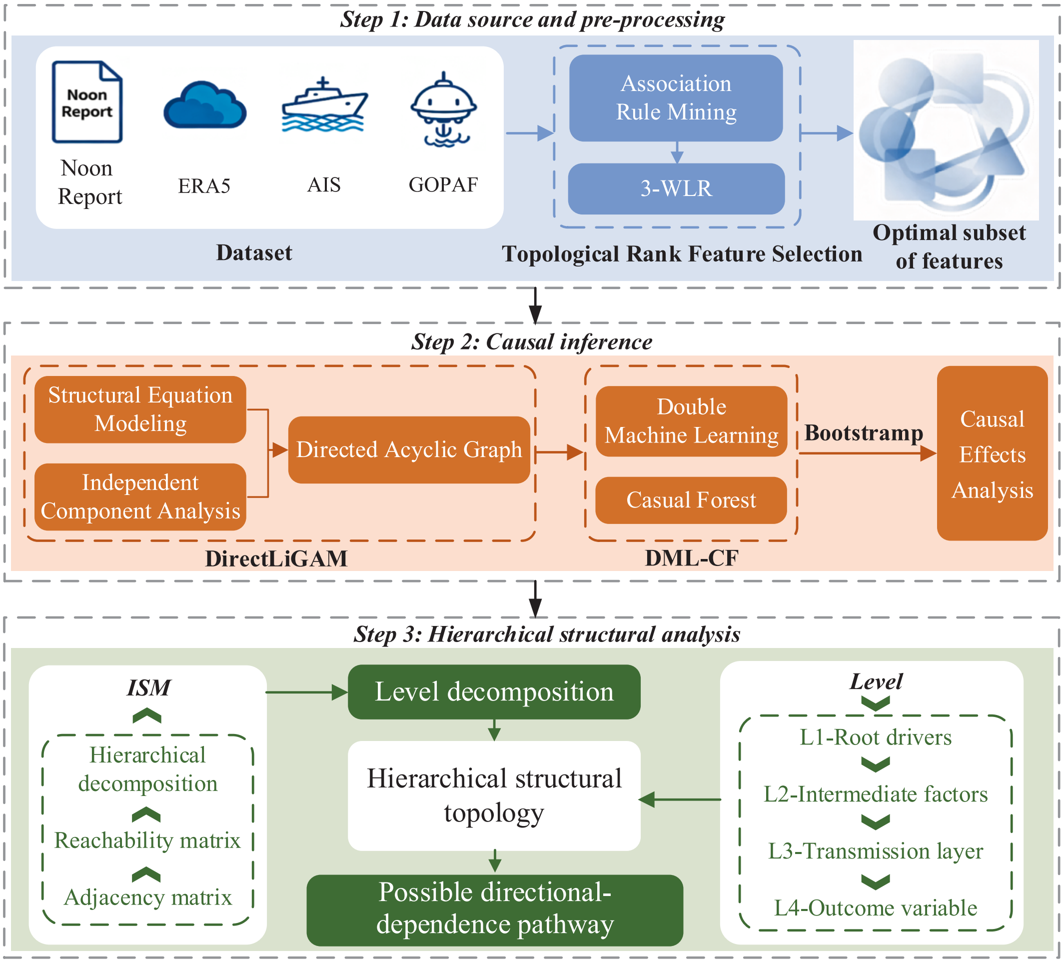

Fig. 1. The proposed flowchart of this study.

this study, TopoRankFS is used as a dimensionality-reduction and noise-filtering procedure for constructing the main analytical variable set, rather than as a method that guarantees causal sufficiency or proves that every possible confounder has been retained.

In the second phase, the retained feature subset is used to construct a mechanism-oriented directional-dependence network through DirectLiNGAM. Operating on a parsimonious subset improves the interpretability and numerical stability of causal discovery, but it does not replace domain knowledge or eliminate the possibility of excluded confounding variables. The DML-CF model is then applied to estimate the ATEs of selected operational variables on SFC, with confounding adjustment conducted using graph-derived observed-covariate adjustment sets constructed from the DirectLiNGAM-estimated DAG. Specifically, the adjustment procedure excludes descendants of each treatment variable, variables located on proper directed causal paths from the treatment to the outcome, and other graph-forbidden nodes, and then derives minimal and expanded adjustment specifications through d-separation checking in the proper back-door graph. Therefore, the subsequent causal estimates should be interpreted as conditional on the retained graph-derived observed adjustment sets rather than as unconditional causal effects under complete confounder preservation.

Finally, ISM is employed to hierarchically decompose the identified causal relationships. This final phase is crucial for transforming the complex, web-like causal network into an ordered, multi-level topology structure. This decomposition is essential for revealing the fundamental transmission pathways and distinguishing foundational root drivers from intermediate variables and final outcomes, thereby clarifying the structural levels within the influencing factor network.

3.2. Data sources and pre-processing

3.2.1. The research data

A robust and comprehensive dataset is fundamental to any data-driven investigation. This study addresses the limitations of relying on a single data source by constructing a fused dataset from multiple heterogeneous sources, thereby establishing a solid foundation for subsequent causal inference.

The initial dataset was systematically compiled from four distinct sources. Ship noon reports provided the core operational data, comprising 479 daily ship-operation records for one ocean-going bulk carrier between 14 January 2021 and 30 July 2024. In this study, the basic observational unit is therefore a daily ship-operation record rather than an independent voyage-level cross-sectional sample. Each record represents the vessel's operational and environmental state over a noon-report reporting interval. These reports contain essential parameters such as SFC, sailing speed, draught, and loading conditions. While the analysis is focused on a single vessel, this approach ensures a high degree of data consistency and mitigates the fleet-level heterogeneity that often introduces confounding variables. It therefore provides a robust and standardized environment to validate the proposed causal inference module, with the vessel serving as a representative case for its class. To overcome the temporal resolution limitations and potential subjectivity of noon reports, high-frequency AIS data for the same period, comprising 162,036 valid track points, were integrated to supplement and validate navigational parameters with high-accuracy geospatial coordinates and timestamps (Luo et al., 2024; Wang et al., 2023a).

Furthermore, to comprehensively capture the vessel's operating environment, high-quality external meteorological and oceanographic data were incorporated. The European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) dataset supplied high-resolution atmospheric and wave variables (Cai et al., 2024; Luo et al., 2023; Yan et al., 2024), while the Global Ocean Physics Analysis and Forecast (GOPAF) dataset provided key oceanographic variables (Ren et al., 2025). To ensure consistency, absolute wind and wave directions from the external datasets were converted to directions relative to the ship's heading, following established methodologies in the field. This multi-source data fusion process yields a rich, high-dimensional dataset that accurately captures the complex interplay between ship operations and environmental conditions. Importantly, the fused dataset provides a consistent, time-aligned operational–environmental state representation, which is essential for subsequent causal discovery and causal effect estimation rather than correlation-only modelling. The specific data pre-processing and multi-source data fusion methods follow established approaches in the literature (Cai et al., 2024; Luo et al., 2023, 2024; Wang et al., 2023a; Yan et al., 2020; Yan et al., 2024). Detailed descriptions of feature variables are provided in Table A of Appendix A.

3.2.2. The feature selection procedures

Following the construction of a SFC dataset by integrating multi-source heterogeneous information, a key challenge arises: the high dimensionality of the feature space. This space contains numerous potential predictor variables, which may include noise, redundant information, and features with low correlation to the target variable, Main engine fuel consumption (Carrasco et al., 2025). Directly incorporating all original features into the modelling process poses two major issues. On the one hand, it can cause a substantial increase in computational complexity and dimensionality problems, which impair the generalisation and stability of the predictive model. On the other hand, irrelevant and redundant variables can significantly hinder the accurate identification of core drivers of SFC, diminishing model interpretability and affecting the validity of subsequent causal inference (Islam et al., 2025). In the SFC context, high dimensionality is not only a predictive challenge but also a causal one, because redundant or weakly relevant variables may inflate confounding adjustment noise and obscure truly operationally actionable drivers (Feng et al., 2024). Therefore, feature selection here is designed to support both robust prediction and reliable causal identification in the subsequent stages (Feng et al., 2024).

To address the challenge of high dimensionality inherent in the fused dataset, a rigorous feature selection strategy is essential. Based on the three-weight LeaderRank (3-WLR) algorithm (Feng et al., 2024), a method whose foundational principles have been validated on maritime datasets, this study proposes a Topological Rank Feature Selection (TopoRankFS) method, specifically tailored for SFC analysis. The adaptation involves two key modifications: (1) optimising the support and confidence thresholds of the association rule mining component to the distributional properties of SFC data; and (2) modifying the weight configuration within the 3-WLR algorithm to more effectively capture SFC-specific influence patterns. Specifically, the original 3-WLR algorithm integrates association-rule strength information through three complementary metrics (support, confidence, and lift) to construct weighted, directed relationships for subsequent network-based ranking. However, SFC determinants often exhibit strong operational–environmental coupling and nonlinear interaction effects, meaning that rules with high support may simply reflect "routine operating regimes" rather than genuinely influential mechanisms. To mitigate this issue, TopoRankFS rebalances the weight configuration to place greater emphasis on confidence and lift, which better characterise (i) directional dependence and (ii) non-trivial synergistic associations beyond marginal prevalence, respectively, while comparatively down-weighting support to avoid dominance by ubiquitous but weakly informative patterns (Agrawal et al., 1993; Feng et al., 2024). The resulting influence network is better optimised to highlight SFC-relevant transmission structures and boundary-sensitive drivers, thereby improving the interpretability and causal suitability of the retained feature subset. Through these optimisations, the TopoRankFS method systematically identifies and ranks a parsimonious set of core variables, providing a robust and focused data foundation for the subsequent causal inference analysis.

It should be emphasised that feature selection before causal modelling is a pragmatic response to the high dimensionality, redundancy and noise introduced by multi-source data fusion, but it is not equivalent to proving causal sufficiency. Predictive relevance, topological relevance and confounding relevance are not identical. A variable with weak marginal predictive contribution may still act as a confounder if it jointly affects a treatment variable and SFC. Therefore, TopoRankFS is used here to construct a parsimonious main-analysis feature subset, rather than to guarantee the preservation of all possible confounders. The retained subset provides an interpretable and computationally stable basis for subsequent DirectLiNGAM and DML-CF modelling, while the resulting causal estimates are interpreted as conditional on the observed covariates retained after feature selection.

3.2.3. Observational unit, temporal alignment and temporal-dependence considerations

Given the sequential nature of ship operation data, the observational unit and temporal alignment procedure are explicitly clarified in this study. The basic observational unit is a daily ship-operation record for the studied bulk carrier, rather than a vessel-level cross-sectional sample. Each observation corresponds to one noon-report reporting interval and represents the vessel's daily operational state, including variables such as daily sailing hours, daily sailing distance, ship course, main engine RPM, and main engine fuel consumption. Therefore, the 479 records used in this study should be interpreted as repeated daily operational records collected from the same vessel over the study period, not as 479 mutually independent vessels or independent voyage-level samples.

The noon report was used as the reference temporal scale for data integration because the outcome variable, main engine fuel consumption, and several key operational variables were recorded at the daily reporting level. High-frequency AIS data were first matched to the same vessel and then aggregated or summarised within the corresponding noon-report interval to derive daily navigational descriptors, including position-, speed-, course-, and draught-related information. ERA5 and GOPAF variables were then spatially and temporally matched to the AIS-derived vessel positions and the corresponding reporting interval, so that meteorological and oceanographic information was represented at the same daily operational scale. Through this procedure, the fused dataset provides a time-aligned daily operational–environmental state representation for subsequent feature selection and causal analysis.

It should be acknowledged that repeated daily observations from the same vessel are not fully independent in a strict statistical sense. Consecutive records may share persistent voyage-level and vessel-specific conditions, such as route planning, loading practice, hull condition, maintenance status, and weather persistence. This study therefore does not interpret the dataset as an independent cross-sectional panel of different ships. Instead, it treats the data as a longitudinal operational record of a single vessel and focuses on dominant daily-scale directional dependencies among time-aligned variables. The use of one vessel helps reduce fleet-level heterogeneity caused by differences in hull form, engine configuration, ship age, and management practice, but temporal dependence within the vessel's operational sequence may still remain.

Accordingly, the causal relationships inferred in this study should be interpreted at the daily operational scale. They do not represent second-level or real-time mechanical feedback between ship systems. Rather, they describe dominant directional dependencies among daily operational, navigational, meteorological, and oceanographic variables after temporal alignment to the noon-report scale.

3.3. Causal inference modelling

Following the preparation of the high-dimensional dataset and the identification of key variables, the analytical focus shifts from correlational analysis to the formal identification and quantification of causal relationships. The module integrates two core components: directed graph-based causal discovery and observed-covariate conditional effect estimation, providing exploratory structural evidence and quantitative support for ship energy-efficiency analysis.

3.3.1. Causal graphs discovery based on DirectLiNGAM

To overcome the limitation of traditional ML models that cannot identify directional causality among variables, this study employs the DirectLiNGAM algorithm to construct an initial causal map. This specific algorithm was chosen over other causal discovery families, such as constraint-based methods or score-based methods, for several compelling reasons. First, this algorithm is based on the assumption of non-Gaussian distributions and the principle of independent component analysis, enabling it to infer causal orderings without relying on prior knowledge (Kotoku et al., 2020; Leyder et al., 2024). Unlike many conventional causal discovery methods that rely on conditional independence tests, DirectLiNGAM does not assume Gaussian-distributed variables and is thus better suited to the asymmetric distribution patterns commonly observed in ship operation data. For example, the variable Daily sailing hours is naturally bounded between 0 and 24 h and commonly exhibits a skewed distribution in operational datasets, with observations tending to cluster near 24 h during sea-passage days, while lower values arise from port stays and manoeuvring periods. This bounded, asymmetric pattern is distinctly non-Gaussian, which further supports the use of DirectLiNGAM for causal ordering and structure discovery in the SFC setting.

Secondly, and most critically for the integrated module, DirectLiNGAM produces a fully specified Directed Acyclic Graph (DAG). This output is not merely a list of relationships; it provides the essential causal architecture that dictates the subsequent analytical steps. The DirectLiNGAM algorithm establishes causal ordering through iterative identification of the most exogenous variable, the "root cause" in the system. Unlike methods relying on simple correlations, it assesses statistical independence between candidate variables and the regression residuals of others (Shimizu et al., 2011). A variable is identified as exogenous if it is independent of the residuals of all other variables, indicating it is not caused by them. Once the most exogenous variable is identified, its influence is mathematically removed from the remaining dataset, and the process is repeated to find the next variable in the causal sequence. This stepwise "peeling" process continues until a complete causal order is determined. With the directed acyclic structure established, the strength of each causal link is quantified through a series of linear regressions, yielding a fully specified causal network that maps the pathways of influence on SFC.

Finally, the resulting DAG is pivotal to the entire analytical framework: it provides the data-driven basis for selecting the observed covariates to be adjusted in the subsequent DML-CF analysis and motivates the final application of ISM for hierarchical decomposition and interpretation. Nevertheless, this study does not treat the estimated DirectLiNGAM graph as definitive causal proof. To address the identification assumptions more explicitly, additional diagnostic checks were conducted for non-Gaussianity, acyclicity, linearity approximation, residual independence, temporal residual dependence and bootstrap edge stability. The detailed DirectLiNGAM diagnostic results are reported in Appendix C. These checks are used to evaluate the compatibility between the DirectLiNGAM assumptions and the present daily ship-operation dataset, rather than to claim complete causal identification under all possible data-generating mechanisms.

3.3.2. Causal Forest-based Double Machine Learning

Building upon the causal network constructed via the DirectLiNGAM algorithm, the analysis transitions from structural identification to the quantitative estimation of causal effects. A crucial preliminary step in this quantification stage is the principled selection of the key operational parameters to be designated as treatment variables. This selection is guided by a dual-criteria framework requiring both high controllability in maritime operations and significant predictive influence on SFC. To rigorously assess the latter, this study employs SHapley Additive exPlanations (SHAP). The choice of SHAP is justified by its strong theoretical underpinnings in cooperative game theory, which provides a consistent and reliable method for quantifying a feature's global importance by considering its marginal contribution across all possible feature combinations. This robust approach ensures that the subsequent causal analysis is focused on variables that are both operationally relevant and empirically significant (Cao et al., 2026a).

To address the limitations of traditional regression methods in controlling confounders and modelling non-linear relationships, this study proposes a DML-CF model to estimate the ATE of key variables. This method integrates the orthogonalisation strategy of DML with the nonparametric estimation framework of CF, enabling robust quantification of causal effects while effectively mitigating potential confounding bias.

Before implementing the DML-CF estimation, the adjustment set for each treatment variable was formally derived from the DirectLiNGAM-estimated DAG. Depending on the graph structure, direct parents may include variables that are not required for adjustment, while valid adjustment may require variables that block non-causal back-door paths without conditioning on mediators, descendants of the treatment or collider-related structures. Therefore, this study adopted a graph-derived adjustment procedure based on the generalised adjustment logic. For each treatment-outcome pair, the proper back-door graph was first constructed by removing the outgoing causal edges from the treatment variable. Candidate covariates were then restricted to observed variables in the retained feature subset, excluding the treatment, the outcome, descendants of the treatment, and variables located on proper directed causal paths from the treatment to Main engine fuel consumption. A minimal adjustment set was obtained through sequential pruning and d-separation checking. In addition, an expanded DAG-derived adjustment set was constructed by retaining theoretically relevant upstream operational and environmental covariates that satisfied the graph-based exclusion rules. The detailed adjustment sets and robustness results are reported in Appendix E.

(1) Double Machine Learning method

DML is an effective method for estimating causal effects in high-dimensional data settings. Its core principle is to eliminate endogeneity bias caused by confounding variables through orthogonalisation (Hu and Xu, 2025). In this study, the treatment variable T is a continuous operational variable rather than a binary treatment indicator. Specifically, T denotes one of the four selected operational treatments, namely Daily sailing hours, Main engine RPM, Daily sailing distance or Ship course. The outcome variable Y denotes Main engine fuel consumption, and X denotes the graph-derived observed covariates used for the corresponding treatment-outcome pair. Therefore, the estimated effect should be interpreted as an original-unit marginal effect of a one-unit increase in T, rather than as a binary treated-control contrast, a discretised high-versus-low comparison, or a standardised effect. The DML estimates the ATE through the following steps:

Firstly, the continuous-treatment partially linear model is described by Eqs. (1) and (2):

$$T = g_0(X) + V \tag{1}$$

$$Y = T\theta_0 + m_0(X) + U \tag{2}$$

where $g_0(X)$ denotes the nuisance function for the continuous treatment model, and $m_0(X)$ denotes the nuisance function for the outcome model. V and U represent residual components with zero mean after conditioning on the retained observed covariates, while θ_0 is the target continuous-treatment effect parameter. In this formulation, θ_0 represents the average partial effect of a one-unit increase in T on Y, conditional on X.

Secondly, DML addresses confounding from observed covariates through the following three key steps:

Step 1: Nuisance function specification and estimation.

In the DML, the population nuisance functions are defined as shown in Eqs. (3) and (4):

$$g_0(X) = \mathbb{E}[T|X] \tag{3}$$

$$m_0(X) = \mathbb{E}[Y|X] \tag{4}$$

Since $g_0(\cdot)$ and $m_0(\cdot)$ are unknown in practice, they are estimated from the observed sample to obtain $\widehat{g}(X)$ and $\widehat{m}_0(X)$, which are subsequently used for residualisation and ATE estimation. Here, $g_0(X)$ models the conditional expectation of the continuous treatment variable given the graph-derived observed covariates, while $m_0(X)$ models the conditional expectation of Main engine fuel consumption given the same covariates. Throughout, the "hat" notation indicates sample-based (data-driven) estimators of the corresponding population quantities and should not be interpreted as implying exact equality. Accordingly, Eqs. (3) and (4) define the population nuisance functions, whereas $\widehat{g}(\cdot)$ and $\widehat{m}_0(\cdot)$ are their data-driven (cross-fitted) estimators used in the subsequent residualisation step.

Step 2: Computation of residuals.

Calculate the orthogonalised residuals of the treatment and outcome variables using Eqs. (5) and (6):

$$\widetilde{T} = T - \widehat{g}(X) \tag{5}$$

$$\widetilde{Y} = Y - \widehat{m}_0(X) \tag{6}$$

This step removes the components of T and Y that can be explained by the retained graph-derived observed covariates. The resulting residualised treatment $\widetilde{T}$ therefore represents the remaining variation in the continuous operational treatment after adjustment for X, while $\widetilde{Y}$ represents the corresponding residual variation in Main engine fuel consumption.

Step 3: Estimation of the ATE.

Treatment effects were estimated using orthogonalised residuals, as illustrated in Eq. (7):

$$\widehat{\theta}_{DML} = \frac{\mathbb{E}\left[\widetilde{T}\widetilde{Y}\right]}{\mathbb{E}\left[\widetilde{T}^2\right]} = \frac{\sum_{i=1}^{n}\widetilde{T}_i\widetilde{Y}_i}{\sum_{i=1}^{n}\widetilde{T}_i^2} \tag{7}$$

where $\widehat{\theta}_{DML}$ denotes the DML estimate of the continuous-treatment effect parameter $\widehat{\theta}_0$ (Chernozhukov et al., 2018). In this study, $\widehat{\theta}_{DML}$ is interpreted as the average change in Main engine fuel consumption associated with a one-unit increase in the corresponding treatment variable, after orthogonalisation with respect to the retained graph-derived observed covariates. Accordingly, the unit interpretation is tonnes per day per additional hour for Daily sailing hours, tonnes per day per additional r/min for Main engine RPM, tonnes per day per additional nautical mile for Daily sailing distance, and tonnes per day per additional degree for Ship course. This residualisation strategy reduces bias associated with the retained observed adjustment covariates and improves the robustness of effect estimation. However, it does not eliminate bias from unobserved confounders or from variables excluded during feature selection.

Finally, to prevent estimation bias caused by overfitting, DML employs a K-fold cross-fitting strategy (with K = 5 in this study) (Cao et al., 2026a). The dataset is divided into K non-overlapping subsets; for each subset, the prediction model is trained on the remaining $K - 1$ subsets, residuals are computed, and the local treatment effect is estimated. The K local estimates are then combined to produce the final ATE estimate.

In this study, LASSO regression was employed to implement predictive models $\widehat{g}(X)$ and $\widehat{m}(X)$. Its selection is justified by its efficacy in high-dimensional settings, where its L1 regularisation concurrently performs variable selection and prevents overfitting, thereby producing a parsimonious yet robust predictive model (Chernozhukov et al., 2018). This enables automated variable selection and helps capture potential non-linear relationships between confounding variables and treatment or outcome variables (Fang et al., 2025). The objective function of LASSO regression is shown in Eq. (8):

$$\widehat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{8}$$

where $x_i \in \mathbb{R}^p$ denotes the $p-$dimensional confounder vector for observation i, i.e., the i^{-th} row of the design matrix $\boldsymbol{X}$, x_i is the i^{-th} row vector of $\boldsymbol{X}$, y_i denotes the dependent variable in the corresponding nuisance model, i.e., T for estimating $\widehat{g}(\cdot)$ and Y for estimating $\widehat{m}_0(X)$, n is the number of observations used to fit the nuisance model, p is the number of confounders, $\beta \in \mathbb{R}^p$ is the coefficient vector, β_j is its j^{-th} element, and λ is a regularisation parameter determined via cross-validation, balancing model complexity and goodness of fit.

(2) Causal forest method

CF is a variant of Random Forest specifically developed to estimate heterogeneous treatment effects by recursively partitioning the covariate space into subgroups with different effect patterns (Wang et al., 2025c). In this study, because T is a continuous operational variable, the CF component is used to estimate heterogeneous marginal effects rather than binary treated-control differences. Specifically, CF estimates the Conditional ATE (CATE) through the following steps:

Step 1: Data partitioning and tree generation.

For each decision tree, training samples are drawn from the original dataset with replacement, and a random subset of covariates is selected. The tree construction recursively partitions the covariate space so that observations with similar operational-environmental profiles are grouped into the same terminal leaf. In this study, the splitting criterion is used to identify subgroups in which the marginal effect of the continuous treatment T on the outcome Y may differ. This enables the model to capture heterogeneous fuel-consumption responses under different combinations of operational and environmental conditions.

Step 2: Estimation of the CATE.

For a given covariate vector X, the CATE is defined as the conditional marginal effect of the continuous treatment variable on the outcome. The corresponding equations are presented as Eq. (9):

$$\tau(X) = \frac{\partial E[Y|T = t, X]}{\partial t} \tag{9}$$

where $\tau(X)$ represents the expected change in Main engine fuel consumption associated with a one-unit increase in T for observations with covariate profile X. This definition is different from a binary-treatment CATE, because it does not compare $T = 1$ with $T = 0$. Instead, it quantifies the local marginal response of Y to a small unit change in the continuous operational treatment.

The CATE estimated by the b-th tree is denoted by $\widehat{\tau}_b(X)$, and the forest-level estimate is obtained by averaging across all trees. The corresponding equations are presented as Eq. (10):

$$\widehat{\tau}(X) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\tau}_b(X) \tag{10}$$

where B denotes the number of trees in the forest.

Step 3: Calculation of the ATE.

The overall ATE was obtained by averaging the CATE estimates for all samples, as illustrated in Eq. (11):

$$\widehat{\tau}_{ATE} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\tau}(X_i) \tag{11}$$

where $\widehat{\tau}_{ATE}$ is the sample-average continuous-treatment effect. For consistency with causal-machine-learning terminology, this quantity is still reported as an ATE in the empirical results. However, its interpretation is explicitly continuous and original-unit based.

(3) Procedure for the hybrid DML-CF model

To further enhance the robustness of causal effect estimation and effectively identify potential effect heterogeneity, this study constructs a hybrid analysis module that integrates DML and CF methods. This DML-CF model fully integrates the orthogonalisation processing advantages of DML in bias control with the modelling capabilities of CF in portraying individualised treatment effects, and the integrated modelling is achieved through the following steps:

Step 1: Data partitioning.

In the DML-CF implementation, standard five-fold cross-fitting was employed to reduce overfitting in the estimation of nuisance functions. The folds were used to ensure that the residuals for each held-out subset were obtained from models trained on the remaining data, thereby improving the robustness of causal effect estimation. However, because the available dataset is a longitudinal operational record from a single vessel, this standard cross-fitting procedure should not be interpreted as fully removing temporal dependence among adjacent daily observations.

Step 2: Orthogonal processing.

For each subset Hk, train the LASSO regression model using all other subsets except Hk to predict the continuous treatment variable T and the outcome variable Y, respectively, based on the graph-derived observed covariates X. The residualised continuous treatment $\widetilde{T}_k$ and residualised outcome $\widetilde{Y}_k$ are then computed for subset Hk.

Step 3: CF construction.

Based on the full residuals $\left\{\widetilde{T}_i, \widetilde{Y}_i\right\}_{i=1}^{n}$ and the graph-derived observed covariates X, the CF model is trained to estimate the heterogeneous marginal effect function $\tau(X)$ defined in Eq. (9) for the continuous treatment variable T.

Step 4: ATE estimation and uncertainty assessment.

To assess the uncertainty of the ATE estimates, this study employed temporal block-bootstrap resampling with 300 successful replications to construct empirical 95% confidence intervals and visualise the probability density distributions of the estimates. Compared with ordinary independent and identically distributed (IID) bootstrap resampling, temporal block-bootstrap resampling partially preserves local dependence among adjacent daily observations by resampling consecutive blocks of records (Huang et al., 2024). This procedure therefore provides a more conservative uncertainty diagnostic for the temporally ordered single-vessel operational dataset.

In addition to temporal block-bootstrap uncertainty assessment, this study further conducted leave-one-covariate-out sensitivity analysis and permutation placebo tests for the DML-CF estimates under the expanded DAG-derived adjustment specification. These supplementary diagnostics are reported in Appendix D. They are used to examine whether the estimated ATEs are sensitive to individual adjustment covariates and distinguishable from random placebo effects. Given the repeated daily observations from a single vessel and the possibility of unobserved operational confounding, these diagnostics are interpreted as robustness and sensitivity evidence rather than as proof of complete causal identification.

3.4. Hierarchical structural analysis based on ISM

To further elucidate the structural dependencies and transmission mechanisms among variables in the SFC system, this study employs ISM to hierarchically decompose the initial causal diagrams generated by DirectLiNGAM. ISM is a systematic engineering method that transforms the elements and relationships within a complex system into a clear multi-level structural model, facilitating the transition from relationship identification to structured understanding (Cao et al., 2024; Wang et al., 2025d).

The implementation of ISM begins by converting the directed causal graph into an adjacency matrix, which represents the direct relationships between variables (Cao et al., 2024). Through a process of Boolean matrix multiplication, this initial matrix is evolved into a “reachability matrix” (Cao et al., 2024). This crucial step systematically identifies not only the direct connections but also all indirect, transitive pathways within the network, thereby capturing the complete relational structure of the system. The reachability

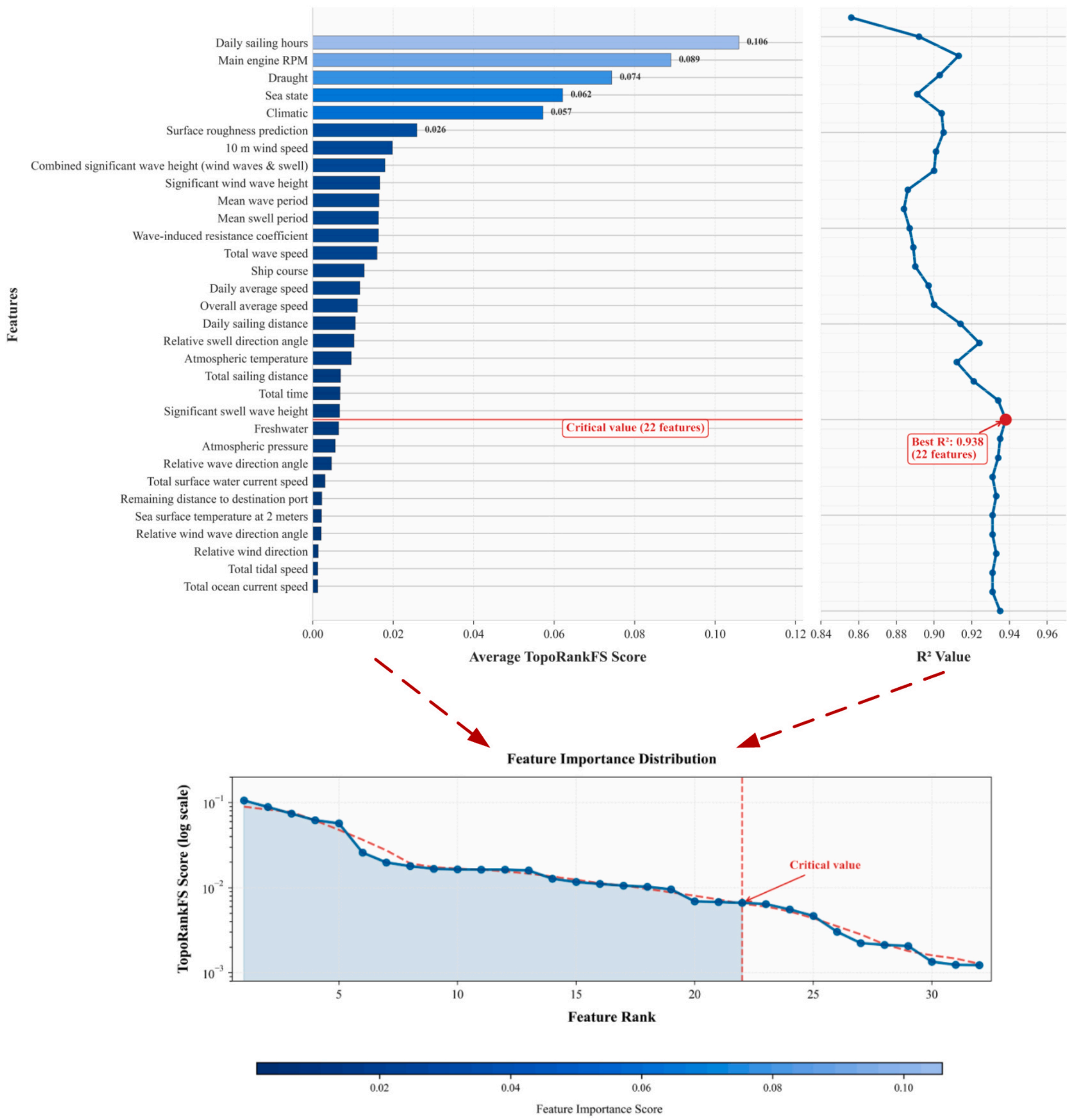

Fig. 2. Final TopoRankFS feature ranking and regression performance. **Note:** The red point on the R^2 value curve (top right) indicates the maximum coefficient of determination achieved. The vertical red dashed line in the feature importance distribution plot (bottom) marks the selected cutoff point of 22 features, corresponding to the identified “elbow” where the marginal importance of additional features diminishes.

matrix then serves as the basis for an iterative partitioning process. By analysing the "reachable set" (all variables influenced by a given variable) and the "antecedent set" (all variables that influence it), the algorithm stratifies all factors into distinct hierarchical levels (Wang et al., 2025d). Specifically, level partitioning is performed iteratively using two sets derived from the reachability matrix: (1) the reachability set, which contains the variable itself and all variables it can influence through direct or indirect paths; and (2) the antecedent set, which contains the variable itself and all variables that can influence it. At each iteration, a variable is assigned to the current level when the set of variables it can reach is fully contained within the variables that can reach it, meaning that it does not act as an upstream driver for any remaining variable outside this shared set. After assigning a level, the corresponding variables are removed and the same procedure is repeated on the reduced system until all variables are allocated to a level. For consistency with the causal interpretation adopted in this study, the resulting levels are reported in Section 4.4 as L1 (root drivers with strongest causal dominance) through L4 (final outcome variable), in line with established ISM-based hierarchy analyses in maritime systems (Cao et al., 2024; Wang et al., 2025d; Wang et al., 2026). This structured decomposition transforms the initially complex network into an ordered, multi-level topology structure, which explicitly reveals the propagation pathways of influence from foundational root drivers to the final system outcome.

4. Results and discussion

This section presents the empirical results obtained from the multi-phase analytical framework detailed in the methodology section. The analysis begins with systematic feature selection to establish a robust and parsimonious dataset, identifying the most influential predictors of SFC from the high-dimensional dataset. Building upon this optimised feature set, the investigation proceeds sequentially through the core stages of causal inference. First, a multi-dimensional causal network constructed using the DirectLiNGAM algorithm, which identifies the directed causal relationships among the key variables influencing SFC. Second, the ATEs estimated via the DML-CF model are detailed, providing a quantitative assessment of the causal impacts of critical operational parameters. Finally, the DirectLiNGAM-derived graph is reorganised using ISM to provide a hierarchical structural representation, which helps interpret the relative upstream, intermediate and downstream positions of the retained variables within the estimated directional-dependence network.

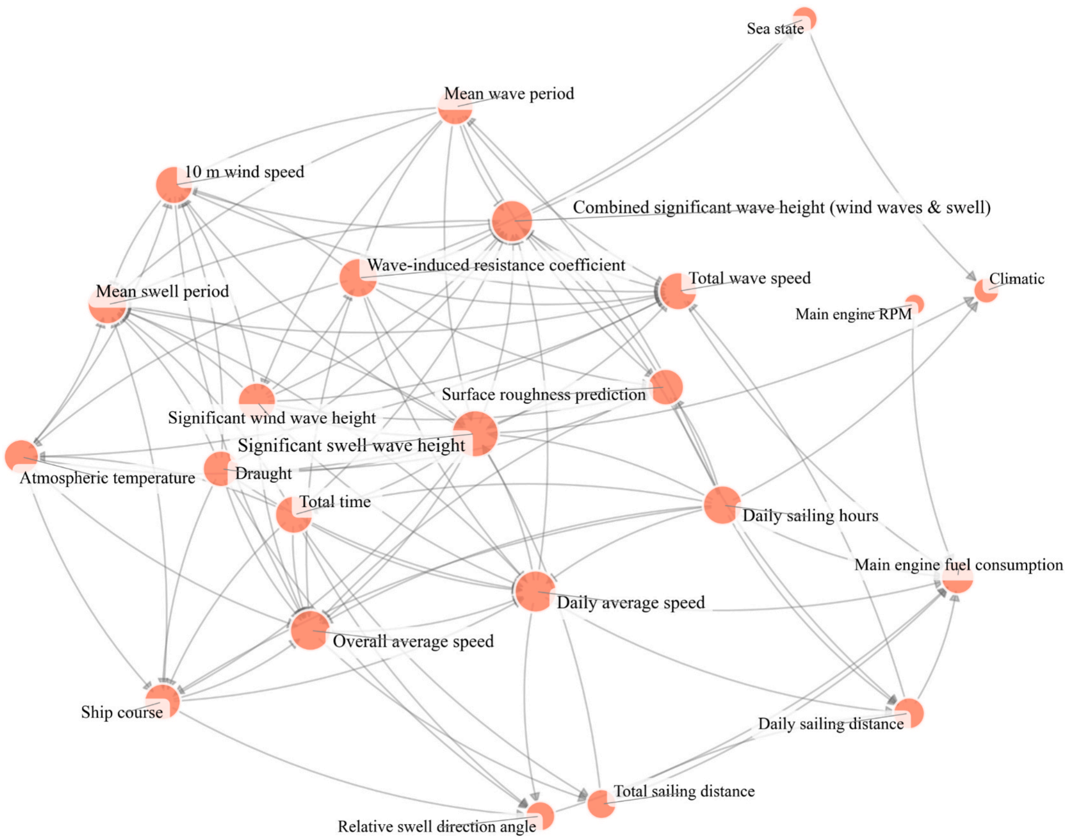

Fig. 3. Causal graph constructed by DirectLiNGAM.

4.1. The feature selection result based on TopoRankFS method

To ensure the accuracy and generalization capability of the prediction model, the initial feature set was systematically screened using the TopoRankFS method. Fig. 2 illustrates the complete feature selection process and results. The average TopoRankFS scores indicate that Daily sailing hours, Main engine revolutions per minute (RPM), and Draught are the most influential variables for predicting SFC. An analysis of the trade-off between model performance and complexity reveals that the coefficient of determination reaches its maximum when the feature count is 22. The logarithmic-scale feature importance distribution curve exhibits a distinct elbow point at the 22nd feature, suggesting that these features capture the core information within the dataset, while the marginal contribution of subsequent features declines sharply. Consequently, the first 22 features were selected as the optimal subset, effectively reducing model complexity without significantly compromising prediction accuracy, and providing a solid data foundation for subsequent analyses.

The selection of 22 variables should be interpreted as the construction of a parsimonious main-analysis subset rather than as evidence that all possible confounders have been retained. This distinction is particularly important in this study because the fused dataset integrates noon-report variables, AIS-derived variables, ERA5 meteorological and wave variables, and GOPAF oceanographic variables. During spatial–temporal interpolation and matching, several environmental descriptors may contain overlapping physical information, for example among wind speed, wave height, wave period, current speed and derived resistance-related variables. Retaining all weakly informative and highly collinear variables would increase the complexity and instability of DirectLiNGAM-based causal discovery, especially under the limited sample size of daily ship-operation records. Accordingly, TopoRankFS was used to remove redundant information and preserve the most informative operational-environmental representation for the main analysis. However, because feature selection based on predictive or topological relevance cannot itself prove causal sufficiency, the subsequent DirectLiNGAM, DML-CF and ISM results are interpreted as observed-covariate conditional findings rather than as unconditional causal effects under complete confounder preservation.

4.2. The causal graph constructed based on DirectLiNGAM

To uncover the underlying causal mechanisms of SFC, the DirectLiNGAM algorithm is applied to construct a multi-dimensional causal network structure, as illustrated in Fig. 3. Based on the assumptions of non-Gaussian distributions and conditional independence, the algorithm identifies causal sequences by maximizing the statistical independence of variables. This enables the inference of complex directed dependencies in ship operation data without requiring prior knowledge (Leyder et al., 2024). The resulting causal network exhibits a highly intricate topology, explicitly positioning Main engine fuel consumption as the core influenced variable, which is influenced through multiple causal pathways. Structurally, the direction of arrows in the graph denotes potential causal flows, while the density of connections between nodes reflects the complexity of interactions among variables. The network also reveals several critical intermediary hubs, such as “Overall average speed” and “Combined significant wave height”. These nodes are pivotal as they are themselves heavily influenced by numerous upstream factors while simultaneously propagating these combined effects downstream, ultimately converging on the main engine fuel consumption.

Fig. 3 visualises the inferred directional-dependence network among the ship’s operational variables, highlighting both the key factors directly linked to main engine fuel consumption and the complex interdependencies among these variables. The arrows in Fig. 3 should be interpreted as data-driven directional dependencies inferred from time-aligned daily operational records, rather than instantaneous physical control relationships. Given the daily temporal resolution of the dataset, the causal interpretation is restricted to dominant daily-scale operational mechanisms. Specifically, Main engine RPM, Daily average speed, and Daily sailing hours form a direct influencing path to SFC. In addition, combined significant wave height indirectly affects Total wave speed by influencing the Wave-induced resistance coefficient and Surface roughness prediction, ultimately impacting fuel efficiency. Environmental parameters such as 10 m wind speed, Atmospheric temperature, Mean wave period, and Mean swell period constitute a network of external conditions that affect the ship's sailing state. Together with ship-specific parameters, such as Draught and Ship course, these factors influence Overall average speed and Daily average speed. Regarding causal pathways, a strong link is observed among Daily sailing hours, Daily sailing distance, and Total sailing distance, which aligns with the physical principles and operational logic of maritime transport. Notably, Fig. 3 shows an inferred directional association between Relative swell direction angle and Total sailing distance, suggesting that route planning may need to account for the influence of sea state on navigational efficiency. This multi-level, multi-path directional structure is consistent with the theory proposed by Guo et al. (2022), which posits that SFC is influenced by the coupling of multiple factors. By identifying these critical daily-scale pathways, this study provides a structured basis for interpreting the operational mechanisms associated with SFC and for informing targeted energy-saving, emission-reduction, and voyage-optimisation strategies.

4.3. The causal inference analysis based on DML-CF

Building on the constructed causal network, the DML-CF model is employed to quantitatively estimate the ATE of key variables. The selection of treatment variables follows two primary criteria: (1) high controllability in practical shipping operations, and (2) significant predictive influence, as quantified by the SHAP feature importance analysis (As shown in Appendix B). Here, “controllability” refers to the extent to which a variable can be intentionally and routinely adjusted by shipboard personnel or shore-based managers during voyage execution, e.g., via propulsion setting and route/heading decisions, without requiring structural retrofits or redesign. This definition is consistent with the operational energy-efficiency measures highlighted in the SEEMP framework (Li et al., 2025). Moreover, to remain fully consistent with Phase I, candidate treatment variables were restricted to the 22 features retained after TopoRankFS (i.e., these above the critical threshold in the feature-selection results). The final interventions were chosen from this subset. Consequently, Daily sailing hours, Main engine RPM, Daily sailing distance, and Ship course are selected as the core treatment variables for causal effect estimation. Because these four treatment variables are continuous operational variables, the reported ATEs are interpreted as original-unit marginal effects. Specifically, they represent the expected change in Main engine fuel consumption associated with a one-unit increase in the corresponding treatment variable, conditional on the retained graph-derived observed adjustment covariates. They do not represent binary treated-control comparisons, discretised high-versus-low contrasts, or standardised effects. By contrast, exogenous meteorological and oceanographic variables within the retained feature set (e.g., wind and wave descriptors, atmospheric temperature) are not directly controllable by operators and are therefore treated as contextual or confounding covariates rather than intervention targets. These variables are chosen not only for their statistically significant explanatory power regarding SFC, but also for their potential to be adjusted through operational interventions or optimisation strategies to effectively reduce fuel use. For each treatment variable, the DirectLiNGAM-estimated DAG in Fig. 3 was used to construct the corresponding proper back-door graph. Variables located on proper directed causal paths from the treatment to Main engine fuel consumption, descendants of the treatment, the treatment itself and the outcome variable were excluded from the adjustment candidates. The remaining observed variables were then screened using d-separation checking to derive the minimal DAG-derived adjustment set. To provide a more conservative reporting and robustness specification, an expanded DAG-derived adjustment set was also constructed by retaining theoretically relevant upstream operational and environmental covariates that satisfied the graph-based exclusion rules. The resulting adjustment sets are reported in Table E1 of Appendix E. Accordingly, the DML-CF estimates reported in this section should be interpreted as conditional effect estimates under the retained graph-derived observed adjustment sets, rather than as definitive causal effects under complete confounder control.

The DML-CF model was re-estimated under the expanded DAG-derived adjustment specification. This specification was selected as the main reporting specification because it retains theoretically relevant upstream operational-environmental information while avoiding adjustment for treatment descendants, pathway mediators and graph-forbidden variables. In this setting, the DML-CF model

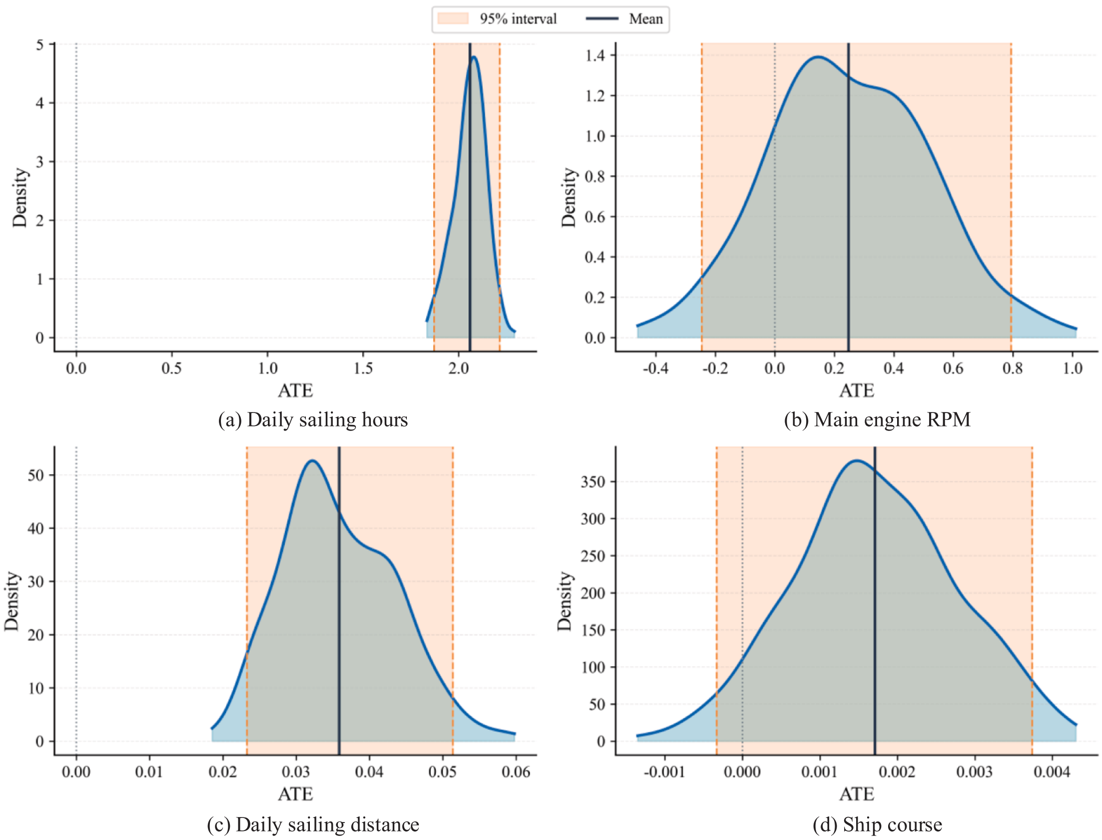

Fig. 4. Temporal block-bootstrap distributions of continuous-treatment DML-CF ATE estimates under the expanded DAG-derived adjustment specification. **Note:** The shaded areas represent the empirical 95% confidence intervals, and the vertical solid lines indicate the bootstrap mean estimates. The horizontal axis denotes the original-unit marginal effect of each continuous treatment variable on main engine fuel consumption. The corresponding units are tonnes per day per additional hour for Daily sailing hours, tonnes per day per additional r/min for Main engine RPM, tonnes per day per additional nautical mile for Daily sailing distance, and tonnes per day per additional degree for Ship course. Temporal block-bootstrap resampling was conducted with 300 successful replications.

still follows a two-stage orthogonalisation logic. In the first stage, LASSO regression is used to estimate the nuisance relationships between the treatment variable, Main engine fuel consumption and the graph-derived adjustment covariates. The resulting residualised components reduce the influence of the retained observed confounders. In the second stage, the non-parametric CF is applied to model the conditional distribution of individual treatment effects, from which the ATE is obtained by averaging the individual-level effect estimates.

Considering that the dataset consists of repeated daily observations from a single vessel, temporal block-bootstrap resampling was further used to assess the uncertainty of the estimated ATEs. Unlike ordinary IID bootstrap resampling, temporal block-bootstrap resampling preserves short consecutive blocks of observations and therefore partially retains local temporal dependence among adjacent daily ship-operation records. Fig. 4 presents the resulting temporal block-bootstrap distributions of the ATE estimates under the expanded DAG-derived adjustment specification. The shaded intervals indicate the empirical 95% confidence intervals, while the vertical solid lines denote the bootstrap mean estimates. Consistent with the use of bootstrap-based uncertainty diagnostics in causal machine-learning analysis, the shape of each density distribution provides an intuitive indication of estimation stability: a narrow and concentrated density suggests lower resampling variability, whereas a broader distribution indicates greater uncertainty under temporal resampling (Huang et al., 2024). These results should therefore be interpreted as conditional effect estimates under the retained graph-derived observed-covariate adjustment sets, rather than as definitive intervention effects under complete confounder control.

The specific results indicate that Daily sailing hours remains the most influential operational-exposure variable. As shown in Fig. 4 (a), its expanded DAG-adjusted ATE is 2.0577, with a temporal block-bootstrap 95% confidence interval of [1.8723, 2.2152]. The distribution is relatively concentrated and entirely located above zero, suggesting that the positive association between sailing duration and Main engine fuel consumption is stable after graph-derived adjustment and temporal resampling. From an operational perspective, this result is consistent with voyage-based fuel consumption studies showing that sailing time is closely related to accumulated fuel use during ship operation (Le et al., 2020). As a time-exposure variable, Daily sailing hours reflects the duration over which the propulsion system remains in active operation, and its cumulative effect on energy use is also consistent with real-time ship energy-consumption and emission monitoring evidence (Yeh et al., 2022). However, the estimated value should be understood as a conditional effect under the retained observed covariates, rather than as a universal mechanical fuel-saving coefficient.

For Main engine RPM, Fig. 4(b) shows a positive point estimate but substantially greater uncertainty. The expanded DAG-adjusted ATE is 0.2681, with a temporal block-bootstrap 95% confidence interval of [-0.2461, 0.7936]. Compared with Daily sailing hours, the distribution is wider and crosses zero, indicating that the propulsion-speed-related effect is more sensitive to temporal resampling and adjustment specification. This uncertainty is plausible in the SFC context because Main engine RPM is not determined independently; it is jointly shaped by required sailing speed, sea state, voyage planning, engine load, safety constraints and schedule pressure. From an engineering perspective, the positive point estimate remains consistent with the general expectation that increased propulsion demand tends to increase fuel use (Işıklı et al., 2020). It is also compatible with thermodynamic analyses of marine diesel engines, where engine load and operating state affect fuel efficiency and energy conversion performance (Yao et al., 2019; Zhu et al., 2020).

Daily sailing distance also shows a stable positive effect, although its unit effect magnitude is much smaller than that of Daily sailing hours. As shown in Fig. 4(c), the expanded DAG-adjusted ATE is 0.0344, with a temporal block-bootstrap 95% confidence interval of [0.0233, 0.0514]. The confidence interval remains above zero, indicating that the distance-related effect is comparatively robust under this adjustment strategy. This finding is reasonable because Daily sailing distance represents the spatial dimension of daily operational exposure. After adjusting for relevant upstream operational-environmental covariates, each additional nautical mile is associated with a modest increase in Main engine fuel consumption. The smaller unit effect does not imply practical irrelevance, because daily sailing distance can accumulate substantially over long voyages. Rather, it indicates that the marginal contribution of distance is more moderate once sailing time, speed-related variables and environmental conditions are considered. This interpretation is consistent with the physical relationship between ship speed, power and fuel consumption, while also recognising that regression-based interpretations of distance, speed and fuel use require caution when these variables are mechanically coupled (Psaraftis and Lagouvardou, 2023).

For Ship course, Fig. 4(d) reports a very small effect magnitude. The expanded DAG-adjusted ATE is 0.0017, with a temporal block-bootstrap 95% confidence interval of [-0.0003, 0.0037]. Although the distribution is relatively concentrated, the confidence interval is close to zero and slightly crosses zero, suggesting limited statistical stability and weak practical magnitude. This result is consistent with the role of Ship course as a heading-related operational variable. Changes in course may influence fuel consumption through the relative angle between the vessel and wind, waves or currents, but this influence is indirect and strongly dependent on the surrounding environmental field and route-planning constraints. From a hydrodynamic perspective, variations in heading may alter wind resistance and the interaction between hull motion and environmental loads (Grlj et al., 2023), while ship–environment and ship–ship interaction studies further show that directional and hydrodynamic effects can be highly context dependent (Zhou et al., 2023). Therefore, Ship course should be treated as a contextual and heading-related operational factor rather than as a dominant independent driver of SFC.

By systematically comparing the four treatment variables, the DML-CF results show a clearer and more cautious effect pattern than the original analysis. Daily sailing hours exhibits the strongest and most stable positive conditional effect, while Daily sailing distance shows a smaller but robust positive effect. In contrast, Main engine RPM retains a positive point estimate but shows greater uncertainty under temporal block-bootstrap resampling, and Ship course exhibits only a weak and near-zero conditional effect. These findings suggest that time- and distance-related operational exposure variables provide more stable evidence for energy-management intervention than propulsion-speed and heading variables under the expanded DAG-derived adjustment specification. This pattern is also consistent with the operational reality of ship navigation, where sailing duration and travelled distance directly determine exposure to fuel-consuming operation, whereas RPM and course are more strongly constrained by voyage planning, weather, safety requirements and navigational practice.

To further examine the robustness of these DML-CF estimates, additional sensitivity analyses were conducted and reported in Appendix D. The temporal block-bootstrap results assess whether the ATE estimates remain stable when local temporal dependence among adjacent daily observations is partially preserved. The leave-one-covariate-out analysis evaluates the influence of individual adjustment variables on the estimated effects. The permutation placebo tests further examine whether the observed ATEs are distinguishable from effects generated by randomly permuted treatment variables. These supplementary results provide additional support for the main effect patterns, while also indicating that the estimated ATEs should be interpreted as conditional on the retained observed covariates rather than as definitive intervention effects under complete confounder control.

4.4. The hierarchical analysis of causal relationships based on ISM

Building on the DirectLiNGAM-derived directional-dependence network in Section 4.2 and the DML-CF conditional effect estimates in Section 4.3, this section uses ISM to provide a hierarchical structural representation of the estimated network. This methodological step is crucial, as it transforms the network of causal relationships into an ordered, multi-level topology structure, thereby revealing the propagation of influence from foundational drivers to the final outcome. As illustrated in Fig. 5, the ISM method categorizes the influencing factors of SFC into four hierarchical levels (L1–L4). The assignment of variables to L1–L4 follows the iterative ISM partitioning rule described in Section 3.4, based on the reachability and antecedent sets calculated from the reachability matrix. In brief, variables are placed closest to the outcome when their downstream influence does not extend beyond the variables that also influence them within the remaining system, and the procedure is repeated to progressively identify higher-level upstream drivers (Cao et al., 2024). Accordingly, L4 contains the final outcome variable, whereas L1 contains the most upstream root drivers identified in the final iteration of the partitioning process (Wang et al., 2026). Nodes at each level are distinguished by different colours to construct a

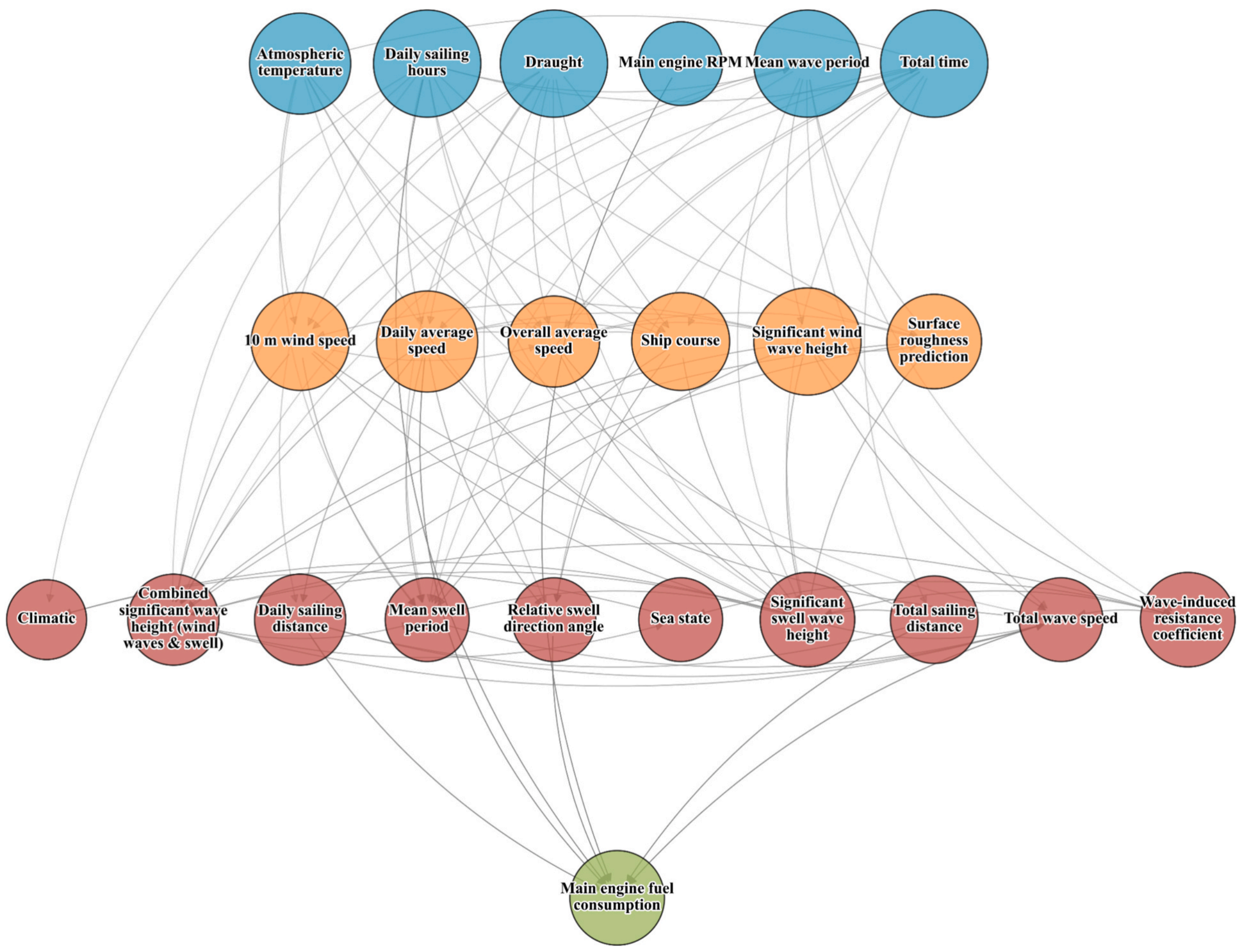

Fig. 5. ISM-based hierarchical structural representation of the DirectLiNGAM-derived network. **Note:** Different node colours represent distinct hierarchical levels identified by ISM. Blue nodes indicate upstream variables, orange and red nodes represent intermediate and proximal variables, respectively, and the green node indicates the final response variable, Main engine fuel consumption.

comprehensive hierarchical topology.

As shown in Fig. 5, the ISM analysis categorizes the factors influencing SFC into four distinct levels. The blue nodes (L1 level) represent root drivers, including Daily sailing hours, Mean wave period, and Main engine RPM. These occupy the highest position in the graph, exerting initial influence on downstream variables without being constrained by other factors. The orange nodes (L2 level) correspond to intermediate transduction factors, such as 10 m wind speed, overall average speed, and Ship course, which are influenced by upstream drivers and transmit effects downstream. The red nodes (L3 level) indicate proximal factors near the response variable, including Daily sailing distance, Total sailing distance, and Significant swell wave height. Finally, the green node (L4 level) represents the final response variable, Main engine fuel consumption. Drivers at a higher level (L1) generally exhibit strong dominance, whereas lower-level factors show high dependence (Cao et al., 2024; Wang et al., 2025d). This pattern is well validated in the SFC system. Daily sailing hours, as a core driver in the L1 level, corresponds to the strongest expanded DAG-adjusted conditional effect estimate in Section 4.3 (ATE = 2.0577), confirming the critical role of temporal exposure in SFC variation under the retained observed adjustment set. Notably, the ISM hierarchical structure provides systematic cross-validation and theoretical complementarity to the causal inference results in Section 4.3. From a multi-method perspective, Daily sailing hours demonstrate remarkable consistency: it ranks highest in SHAP analyses, indicating the greatest marginal contribution to predictive accuracy; it exhibits the strongest statistical significance in ATE estimation; and it occupies the foundational driver position in the ISM hierarchy. This convergence across methodological dimensions collectively corroborates its consistently important role under the retained observed-covariate specification from the perspectives of predictive power, causal strength, and structural position. Similarly, Main engine RPM, another key L1 variable, retains a positive point estimate under the expanded DAG-derived adjustment specification (ATE = 0.2681), although the temporal block-bootstrap interval indicates greater uncertainty. Its identification as a base-level driver in the ISM structure therefore supports its structural relevance, while its quantitative effect estimate should be interpreted cautiously. This multi-method cross-validation enhances the robustness of the findings and provides a multi-level integrated perspective for understanding the driving mechanisms of ship energy efficiency.

Among several candidate representative structural paths, this study selects the sequence “Daily sailing hours (L1) − Overall average speed (L2) − Daily sailing distance (L3) − Main engine fuel consumption (L4)” as a representative case. The selection of this pathway is predicated on its clear operational logic and its capacity to comprehensively illustrate the transmission of effects from a principal root driver (L1) through key intermediary variables to the final outcome (L4). This path is used to examine the cascading effect and its comprehensive impact on SFC. It effectively illustrates the transfer mechanism linking the time dimension, operational status, and sailing distance, serving as a critical bridge between the ship’s fundamental operational indicators and SFC response.

In this pathway, Daily sailing hours function as the top-level driver, reflecting the cumulative base time of a ship in daily operation. This indicator not only directly records the ship’s operating hours but also partially reflects fluctuations in the ship’s power system demand. Within this transmission process, Daily sailing hours act as a fundamental variable whose effect is amplified and accumulated through subsequent mediating variables. Overall average speed serves as a mediator, influenced by Daily sailing hours while directly reflecting the ship’s response to actual operating and subtle hydrodynamic conditions (Lang and Mao, 2020). Daily sailing distance operates as a key transfer node, translating operational status from the upper level into specific spatial change metrics. It represents an important measure of ship operational efficiency and indirectly indicates the resistance and loading conditions encountered during a voyage segment. As sailing distance increases, the complexity of managing energy consumption challenges also accumulates. Ultimately, these hierarchical transmission effects converge on Main engine fuel consumption. Through this cascade: from time, to speed, to distance, Main engine fuel consumption is affected both directly by Daily sailing hours and indirectly through the mediating variables of Overall average speed and Daily sailing distance. This path confirms the dual mechanism of sailing hours’ influence on SFC: a direct effect and an indirect effect via operating conditions and sailing distance. From a practical standpoint, identifying this structural pathway provides a theoretical foundation for hierarchical interventions in fuel efficiency management. Firstly, sailing time, as the initiating driver, can be optimised through rational ship scheduling and port stay management. Secondly, voyage speed, as a critical intermediary, may be dynamically adjusted via intelligent voyage speed planning systems. Finally, optimizing sailing distance requires integrating route planning with meteorological navigation techniques to achieve systematic SFC reduction.

In summary, this study presents a comprehensive causal deconstruction of the mechanisms influencing SFC by integrating DirectLiNGAM and DML-CF methods. Additionally, it employs the ISM to analyse the multi-level transmission mechanisms of influencing factors. DirectLiNGAM provides the estimated directional-dependence structure, DML-CF estimates observed-covariate conditional effect patterns, and ISM organises the inferred structure into hierarchical levels, revealing cascading effects from foundational drivers through mediators to the main engine’s fuel consumption. This framework confirms the dual role of Daily sailing hours, which exerts both a direct effect and an indirect influence via mediating variables. It also emphasises the critical regulatory roles of operating conditions and sailing distance in the accumulation of energy consumption. Consequently, this study provides a robust quantitative foundation and theoretical support for optimising ship energy efficiency and fuel management strategies. Its methodological rigor and innovation offer new practical avenues for future research in related fields.

5. Implications

This study constructs a multi-phase analytical framework to comprehensively investigate the mechanisms underlying SFC by integrating advanced ML methods with causal inference techniques. The framework combines rigorous data processing, statistical modelling, and causal analysis, offering a scientifically robust and methodologically sound foundation for SFC modelling. It also provides theoretical support and practical guidance for improving energy efficiency and promoting sustainable development in the shipping industry. The research findings hold significant methodological and practical value, which are reflected in the following aspects:

Firstly, by constructing a fused dataset from multiple heterogeneous sources and applying feature selection techniques, this study addresses the limitations of relying on a single data source and establishes a solid foundation for subsequent causal inference. Building on this foundation, this study evaluates the predictive performance of nine ML models and identifies the most effective model for accurately predicting SFC. This systematic process involving data fusion, feature selection, and model optimisation not only improves predictive accuracy but also enhances the model's capacity to generalize across complex and variable maritime conditions. Moreover, by providing a framework that moves beyond correlational accuracy to deliver causal insights, this study supports shipowners and operators in formulating data-driven sailing strategies and energy management plans, contributing to the reduction of both operational costs and carbon emission intensity.

Secondly, the directional relationships identified through DirectLiNGAM and the conditional effect estimates obtained from DML-CF provide graph-informed intervention targets and a conditional quantitative basis for optimising ship energy efficiency. By constructing a multi-dimensional directional-dependence network, this study identifies key pathways associated with SFC and illustrates the complex interactions among environmental parameters, ship characteristics and operational variables. In particular, the conditional effects of Daily sailing hours, Main engine RPM, Daily sailing distance and Ship course are quantified under the expanded DAG-derived observed-covariate adjustment specification, rather than under an assumption of complete confounder control. These results indicate that, in real-world operating conditions, rational planning of sailing time and route execution appears to provide more stable energy-management leverage than treating propulsion setting or heading as isolated control variables. For instance, the conditional ATE estimate suggests that a one-hour reduction in Daily sailing hours is associated with an average reduction of approximately 2.06 tonnes per day in Main engine fuel consumption under the retained graph-derived observed-covariate adjustment set. This estimate should not be interpreted as a universal mechanical fuel-saving rate, a binary sailing/non-sailing intervention effect, or a standardised effect, but as an original-unit conditional marginal estimate within the present single-vessel daily operational dataset. This intervention strategy, grounded in causal inference, not only overcomes the limitations of traditional correlation-based approaches but also offers a scientifically valid pathway for achieving more refined and effective energy efficiency management in the maritime sector.

Finally, this study proposes a structured intervention strategy for the systematic optimisation of SFC, grounded in the hierarchical analysis structure developed using the ISM method. The influencing factors are categorized into four distinct levels (L1 to L4), clearly revealing the full chain from foundational driving variables to mediating factors that ultimately influence SFC. In particular, the analysis identifies a key transmission pathway: Daily sailing hours (L1) influences Overall average speed (L2), which in turn affects Daily sailing distance (L3), ultimately impacting Main engine fuel consumption (L4). This cascade clarifies the transmission mechanism linking time, speed, and distance. The findings support a stepwise intervention framework for practical implementation, including 1) optimised sailing time through better port scheduling; 2) application of intelligent speed management systems to control voyage speed; and 3) reduction of sailing distance through advanced route planning and meteorological navigation. This structured approach enables precise and coordinated energy efficiency management, while also enhancing the economic and environmental performance of ship operations. Ultimately, it contributes to advancing the maritime industry toward low-carbon and sustainable development.

In summary, this study establishes a comprehensive theoretical framework that extends from correlation analysis to causal mechanism exploration by integrating ML within a robust causal inference. This offers a multi-dimensional and multi-level scientific foundation for the prediction and optimisation of SFC. This moves beyond the limitations of purely predictive "black-box" models by elucidating the underlying causal structure of the system, providing an insightful pathway to support the green and low-carbon transition of the shipping industry. These findings hold both theoretical value and practical significance.

6. Conclusions

This study develops a holistic analytical framework that integrates data fusion, feature selection, causal inference, and hierarchical analysis to investigate the key factors influencing SFC and their complex interrelationships. By combining ML and causal inference methods, the framework facilitates a systematic transition from statistical correlation to causal understanding, offering a robust scientific basis and decision-making support for improving energy efficiency and advancing the green, low-emission transformation of the shipping industry. The new findings of this study are summarised as follows:

First, this study finds that a core set of variables are the primary drivers of SFC, with temporal and operational parameters demonstrating the most significant predictive influence. The application of the TopoRankFS feature selection method identified a parsimonious set of variables from the high-dimensional fused dataset, thereby establishing an interpretable main-analysis feature subset for the subsequent causal analysis. However, this subset should not be interpreted as guaranteeing complete confounder preservation. Second, this study reveals the distinct causal hierarchy and quantitative impact of key operational parameters on SFC. The DirectLiNGAM algorithm identified a daily-scale directional-dependence network, upon which the DML-CF model estimated conditional ATEs under the retained observed-covariate set. The findings suggest that temporal exposure, represented by Daily sailing hours, shows the strongest expanded DAG-adjusted conditional effect estimate, while Daily sailing distance remains positive and stable. Main engine RPM retains a positive point estimate but exhibits greater uncertainty under temporal block-bootstrap resampling, and Ship course shows only a small conditional effect. This causal effect quantification provides a reliable, data-driven foundation for evaluating the energy-saving potential of different operational strategies while controlling for confounding variables. Finally, this study uncovers the structured, multi-level transmission mechanism through which foundational drivers ultimately affect SFC. By applying the ISM method to conduct a hierarchical decomposition of the causal network, this study reveals a clear topology of upstream directional-dependence factors, intermediate mediators, and final responses. This structural analysis clarifies the recursive relationships and transmission pathways within the system, highlighting the most effective prioritization and intervention points for multi-objective, multi-stage system optimisation in ship energy management.

The framework proposed in this study demonstrates substantial analytical value in the domain of ship energy efficiency and exhibits strong generalisability and scalability in its core concepts and methodological integration. These features make it applicable as a methodological reference for analysing multi-factor coupling mechanisms in other complex systems, such as transportation logistics and industrial energy consumption. Despite the notable advancements achieved, several limitations remain that warrant further investigation. On the one hand, the limited geographical scope and temporal coverage of the current dataset may constrain the generalisability of the model's conclusions. In particular, the empirical analysis is based on repeated daily operational observations from a single bulk carrier. Although this single-vessel setting helps reduce fleet-level heterogeneity caused by differences in hull form, engine configuration, ship age, and management practice, temporal dependence among adjacent operational records may not be fully accounted for by the present modelling framework. Therefore, the inferred causal relationships should be interpreted as dominant daily-scale directional dependencies rather than high-frequency dynamic feedback mechanisms. On the other hand, although DML can reduce bias from observed confounders, two sources of residual uncertainty remain. First, feature selection based on predictive or topological relevance may remove variables with limited predictive contribution but potential confounding roles. Therefore, the retained 22-variable subset should not be interpreted as fully preserving all possible causal confounders, and the causal discovery and ATE estimation results should be understood as conditional on the retained observed covariates. Second, unobserved or coarsely documented operational states, such as hull fouling/surface roughness and trim/ballast condition, may simultaneously affect propulsion settings and SFC, thereby introducing residual confounding. Future research could therefore focus on: (1) expanding data dimensions by incorporating a broader range of ship types, operational scenarios, multi-vessel longitudinal datasets, and extended time series; (2) exploring more robust nonlinear causal discovery techniques and dynamic causal modelling approaches, including lagged variable structures and temporal block-based validation strategies, to better capture the complexities and sequential dependence of ship operational dynamics; and (3) integrating additional maintenance/condition information, such as hull-cleaning or dry-docking records and trim/ballast monitoring data, or suitable proxies to further reduce the risk of unobserved and feature-selection-induced confounding.

CRediT authorship contribution statement

Wenjie Cao: Writing – original draft, Methodology, Investigation, Conceptualization. **Xinjian Wang:** Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Wei Zhang:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. **Huanhuan Li:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Siming Fang:** Writing – review & editing, Validation, Software, Investigation. **Zaili Yang:** Writing – review & editing, Visualization, Supervision, Funding acquisition, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors sincerely thank Dr. Evelin Krmac (Faculty of Maritime Studies and Transport, University of Ljubljana) for her valuable intellectual contributions during manuscript preparation and revision. Her insightful revisions substantially improve this work, including refining the causal interpretation of results, clarifying research findings, and strengthening the methodological justification of the DirectLiNGAM and DML-CF approaches.

The authors gratefully acknowledge support from Natural Science Foundation of Liaoning Province, China [Grant No. 2026-MS-093], the European Union H2020 European Research Council Consolidator Grant program [TRUST Grant No. 864724].

Appendix A:. Feature lists

Table A
Features in each data source.

Data source	Features	Units
Ship noon report	Daily sailing distance	nmi
	Total sailing distance	nmi
	Remaining distance to destination port	nmi
	Daily sailing hours	hours
	Total time	hours
	Freshwater	t

(continued on next page)

Table A (*continued*)

Data source	Features	Units
	Climatic	N.A.
	Atmospheric temperature	°C
	Overall average speed	knots
	Daily average speed	knots
	Ship course	degrees
	Relative wind direction	degrees
	Atmospheric pressure	hPa
	Sea state	N.A.
	Main engine RPM	r/min
	Main engine fuel consumption	t/day
AIS	Longitude	degrees
	Latitude	degrees
	Draught	m
ERA5	10 m wind speed	m/s
	Sea surface temperature at 2 m	K
	Surface roughness prediction	m
	Combined significant wave height (wind waves & swell)	m
	Wave-induced resistance coefficient	N.A.
	Mean swell period	s
	Mean wave period	s
	Significant swell wave height	m
	Significant wind wave height	m
	Relative wave direction angle	degrees
	Relative swell direction angle	degrees
	Relative wind wave direction angle	degrees
GOPAF	Total ocean current speed	m/s
	Total tidal speed	m/s
	Total wave speed	m/s
	Total surface water current speed	m/s

Appendix B:. The selection processes of treatment variables

The primary objective of this appendix is to detail the rigorous, data-driven methodology employed for the principled selection of treatment variables, which serve as the focus of the causal inference analysis in the main body of this study. This selection process is predicated on identifying the operational factors with the most significant predictive influence on SFC. To achieve this robustly, a two-stage procedure was implemented. First, a comprehensive comparative analysis of various ML algorithms was conducted to establish the optimal, highest-fidelity prediction model for SFC. Second, leveraging the optimal model, a feature importance analysis using the SHapley Additive exPlanations (SHAP) framework was performed to quantitatively rank the contribution of each variable. This systematic approach, underpinned by the robust theoretical guarantees of SHAP, ensures that the variables selected for causal analysis are grounded in empirical evidence of their predictive power.

In the first stage, this study constructs a comprehensive ML model framework for predicting SFC. A rigorous data partitioning strategy is employed, dividing the dataset into training and test sets at a ratio of 80% to 20%, respectively, to ensure objective and valid model evaluation (Feng et al., 2024). Prior to training, all feature variables are standardized using Z-score normalization to mitigate the effects of scale differences on model performance while enhancing algorithm convergence speed and numerical stability (Nkikabahizi et al., 2022). To comprehensively evaluate the performance of different algorithms for SFC prediction, nine well-established ML models were selected for comparative analysis: Linear Regression, Ridge, Lasso, ElasticNet, Support Vector Machine, Random Forest, Gradient Boosting, Extra Trees, and XGBoost (Uyanık et al., 2020; Yan et al., 2020). These models span a broad spectrum of algorithmic complexity, ranging from simple linear to complex nonlinear approaches, enabling effective capture of the intricate relationships between SFC and its influencing factors. To ensure optimal performance of each model, five-fold cross validation combined with Bayesian optimisation was employed to systematically tune the hyperparameters (Garrido-Merchán et al., 2023). By constructing a probabilistic model of the parameter space, Bayesian optimisation efficiently identifies optimal hyperparameter combinations with fewer iterations, demonstrating greater computational efficiency than traditional grid and random search methods (Joy et al., 2020).

To systematically assess the performance of each candidate model, this study employs Taylor diagrams for multi-dimensional visualisation and comparison (Fig. B1). Taylor diagrams simultaneously present three key metrics: correlation coefficient, centred root mean square error, and standard deviation in a single chart, providing a visual measure of each model's similarity to the observed data. This rigorous assessment is essential for identifying the model that most accurately captures the complex, non-linear relationships within the dataset, a critical prerequisite for conducting a reliable feature importance analysis.

The results in Fig. B1 clearly indicate that the Random Forest model (highlighted with red diamonds) achieves optimal performance across all key evaluation metrics. Its data points lie closest to the reference point representing a perfect model, demonstrating that it simultaneously satisfies three optimisation criteria: (1) the highest correlation between predicted and observed values; (2) the closest standard deviation to the observed data; and (3) the smallest centred root mean square error. In contrast, other models exhibit either lower correlation or greater discrepancies in variability compared to the observed data. This combination of strengths enables the

Random Forest model to significantly outperform all other candidates and justifies its selection as the high-fidelity base model for the subsequent SHAP analysis.

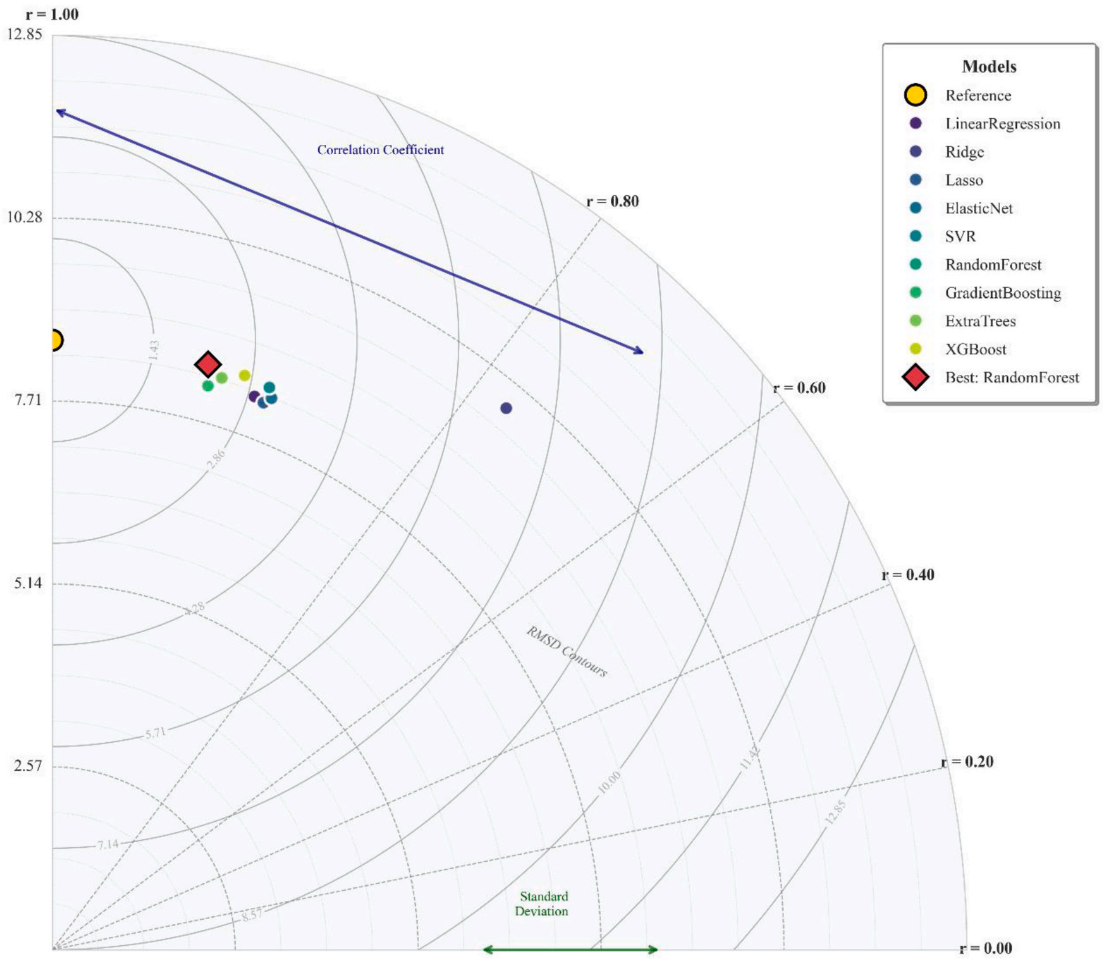

Fig. B1. Taylor diagram: Model performance comparison.

In the second stage, this study employs the SHAP method to rigorously quantify the predictive importance of each feature. SHAP is an interpretable artificial intelligence technique grounded in game theory that assigns contribution values to each predictor by calculating the marginal effect of each feature across all possible feature combinations (Abdulrashid et al., 2024). This method possesses a solid theoretical foundation and consistency guarantees, effectively quantifying both the magnitude and direction of each feature's influence on the model output (Xie et al., 2024).

To implement this data-driven selection principle, the Random Forest algorithm, identified as the optimal prediction model, was first trained on the dataset. The SHAP method was then applied to this model to derive a global feature importance ranking. As shown in Fig. B2, the SHAP results clearly quantify the contribution of each variable to the prediction of SFC, indicating their relative influence magnitudes. The variables exhibiting the highest SHAP values were consequently identified as strong candidate variables for the causal analysis, thereby providing the empirical foundation for the second selection criterion: significant predictive influence. The final selection of treatment variables, as detailed in Section 4, was then made by evaluating these candidates against the primary criterion of high controllability in practical shipping operations, ensuring the causal analysis is focused on factors that are both impactful and operationally relevant.

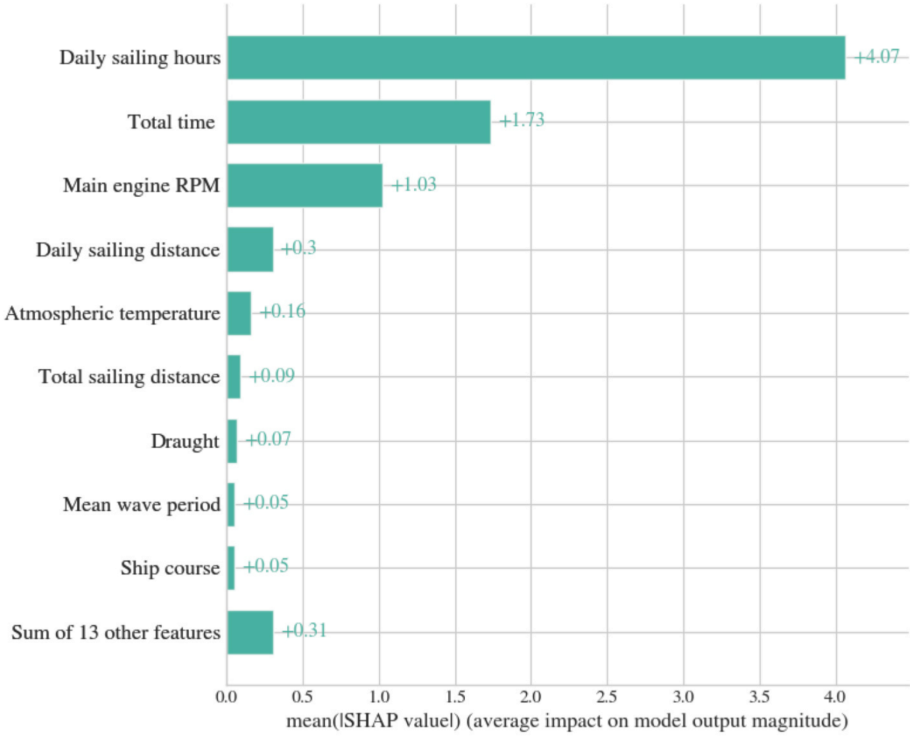

Fig. B2. SHAP global feature importance analysis for SFC model.

Appendix C:. Assumption diagnostics and stability checks for DirectLiNGAM

The purpose of the Appendix C is to provide a transparent assessment of the key assumptions underlying the DirectLiNGAM-based causal discovery stage. DirectLiNGAM relies on a linear non-Gaussian acyclic structural equation framework with mutually independent error terms (Shimizu et al., 2006; Shimizu et al., 2011). In the context of SFC analysis, these assumptions are strong because operational variables, weather conditions, voyage planning and vessel condition may be jointly determined. Therefore, this appendix does not aim to prove that all causal-identification assumptions are fully satisfied. Instead, it reports a set of assumption-diagnostic and stability checks to clarify the extent to which the analysis set comprising the 22 retained explanatory variables and the outcome variable, Main engine fuel consumption, i.e., 23 variables in total, is compatible with the use of DirectLiNGAM as a linear directional-dependence discovery tool.

Specifically, five groups of diagnostics were conducted. First, non-Gaussianity was examined using Shapiro-Wilk tests, Jarque-Bera tests, skewness and excess kurtosis (Chowdhury et al., 2022; Monter-Pozos and González-Estrada, 2024). Second, the acyclicity of the estimated directed graph was verified using directed-cycle detection and topological sorting. Third, the linearity approximation was assessed by comparing cross-validated structural-equation errors between linear regression and Random Forest regression for each node with retained parent variables. Fourth, the independent-error assumption was examined through pairwise Spearman dependence screening among the structural residuals. Finally, because the data consist of repeated daily observations from a single vessel, temporal residual dependence was assessed using Ljung-Box and Durbin-Watson diagnostics. In addition, bootstrap edge-frequency analysis was performed to evaluate the stability of the inferred edge directions under both independent and identically distributed (IID) resampling and temporal block resampling.

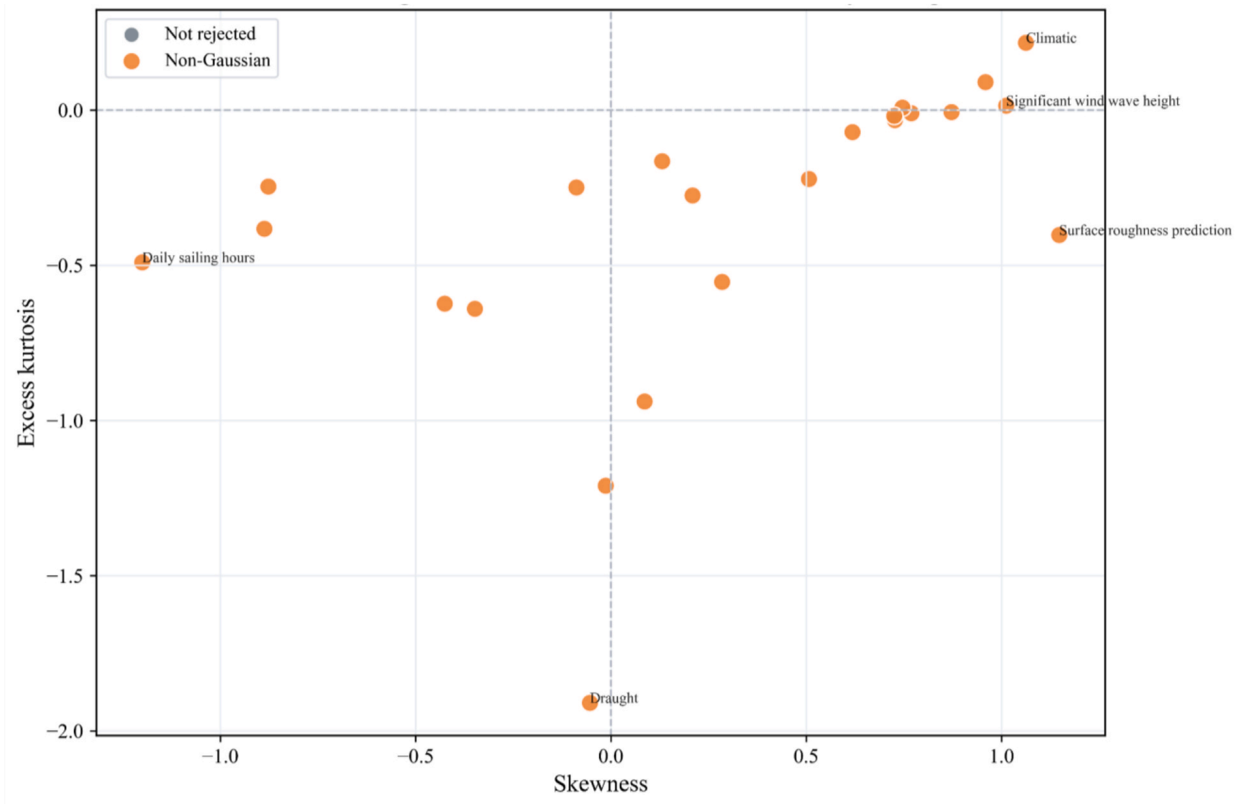

Fig. C1. Distributional diagnostics for the non-Gaussianity assumption of DirectLiNGAM. **Note:** All plotted variables are marked as "Non-Gaussian", indicating that Gaussianity is rejected by at least one of the formal normality diagnostics reported in Table C1. The "Not rejected" category is retained in the legend only to clarify the classification criterion, although no variable falls into this category in the present dataset.

Fig. C1 visualises the skewness and excess kurtosis of the analysis variables, while Table C1 reports the corresponding Shapiro-Wilk and Jarque-Bera test results. The purpose of combining graphical and formal diagnostics is to avoid relying on a single normality test. As shown in Table C1, all 23 variables, including the 22 retained explanatory variables and the outcome variable, are identified as non-Gaussian by at least one formal test. Several operational variables, such as Daily sailing hours, Main engine fuel consumption and Daily sailing distance, also exhibit clear skewness, which is consistent with the bounded and regime-dependent nature of ship-operation records. These results provide empirical support for the distributional condition required by DirectLiNGAM. However, this finding should not be over-interpreted, because non-Gaussianity alone is not sufficient for causal identification. It only indicates that the retained variables contain distributional asymmetry and/or tail behaviour compatible with a LiNGAM-type modelling framework (Harada and Fujisawa, 2021).

Table C1
Non-Gaussianity diagnostics for the retained variables.

Variables	N	Skewness	Excess kurtosis	Shapiro-Wilk p	Jarque-Bera p	Non-Gaussian
Daily sailing hours	479	−1.200	−0.490	8.19e-33	1.03e-26	Yes
Surface roughness prediction	479	1.147	−0.402	3.39e-28	2.98e-24	Yes
Climatic	479	1.062	0.217	2.00e-23	1.72e-20	Yes
Significant wind wave height	479	1.011	0.014	3.76e-19	1.86e-18	Yes
Mean wave period	479	0.959	0.090	7.97e-17	1.06e-16	Yes
Draught	479	−0.054	−1.909	2.57e-27	1.43e-16	Yes
Main engine fuel consumption	479	−0.887	−0.382	5.19e-23	5.22e-15	Yes
Daily sailing distance	479	−0.877	−0.246	7.47e-19	2.59e-14	Yes
Combined significant wave height (wind waves & swell)	479	0.871	−0.007	2.08e-15	6.84e-14	Yes
Wave-induced resistance coefficient	479	0.769	−0.010	8.35e-13	5.64e-11	Yes
Total sailing distance	479	0.746	0.008	1.77e-12	2.29e-10	Yes
Total time	479	0.727	−0.032	3.08e-12	6.93e-10	Yes
Total wave speed	479	0.725	−0.018	2.92e-11	7.76e-10	Yes
Significant swell wave height	479	0.618	−0.071	2.47e-10	2.24e-07	Yes

(continued on next page)

Table C1 (*continued*)

Variables	N	Skewness	Excess kurtosis	Shapiro-Wilk p	Jarque-Bera p	Non-Gaussian
Relative swell direction angle	479	−0.013	−1.210	2.15e-11	4.53e-07	Yes
Main engine RPM	479	−0.426	−0.623	2.80e-23	1.50e-05	Yes
10 m wind speed	479	0.507	−0.222	2.37e-07	2.15e-05	Yes
Ship course	479	0.086	−0.939	2.53e-12	1.13e-04	Yes
Atmospheric temperature	479	−0.349	−0.640	1.50e-09	1.32e-04	Yes
Sea state	479	0.284	−0.553	2.80e-21	1.86e-03	Yes
Mean swell period	479	0.209	−0.274	2.54e-03	8.30e-02	Yes
Overall average speed	479	0.131	−0.164	7.42e-07	3.85e-01	Yes
Daily average speed	479	−0.089	−0.249	4.23e-03	3.93e-01	Yes

After confirming the distributional compatibility of the retained variables, the next diagnostic step concerns whether the estimated graph satisfies the directed acyclic structure assumed by DirectLiNGAM. Table C2 summarises the acyclicity check. The estimated graph contains 23 nodes and 89 retained directed edges, and no directed cycle is detected. This indicates that the estimated structure satisfies the formal graph-theoretic requirement of a directed acyclic graph. Therefore, from a graph-theoretic perspective, the DirectLiNGAM output is suitable for subsequent hierarchical decomposition using ISM.

Table C2
Acyclicity check for the DirectLiNGAM graph.

Diagnostic	Result
Is directed acyclic graph	True
Number of nodes	23
Number of retained edges	89
Directed cycle detected	No

Although acyclicity is satisfied, the DirectLiNGAM framework also assumes linear structural equations (Harada and Fujisawa, 2021). Therefore, Fig. C2 and Table C3 further examine whether the estimated parent–child relationships can be reasonably approximated by linear structural equations. Fig. C2 provides the overall comparison between linear regression and Random Forest regression, while Table C3 lists representative nodes with different degrees of non-linearity. If Random Forest substantially reduces RMSE relative to the linear model, this indicates that the corresponding node may contain non-linear structural behaviour. The 5% and 15% thresholds in Fig. C2 are used as heuristic diagnostic cut-offs for screening potential non-linearity, rather than as formal statistical significance thresholds.

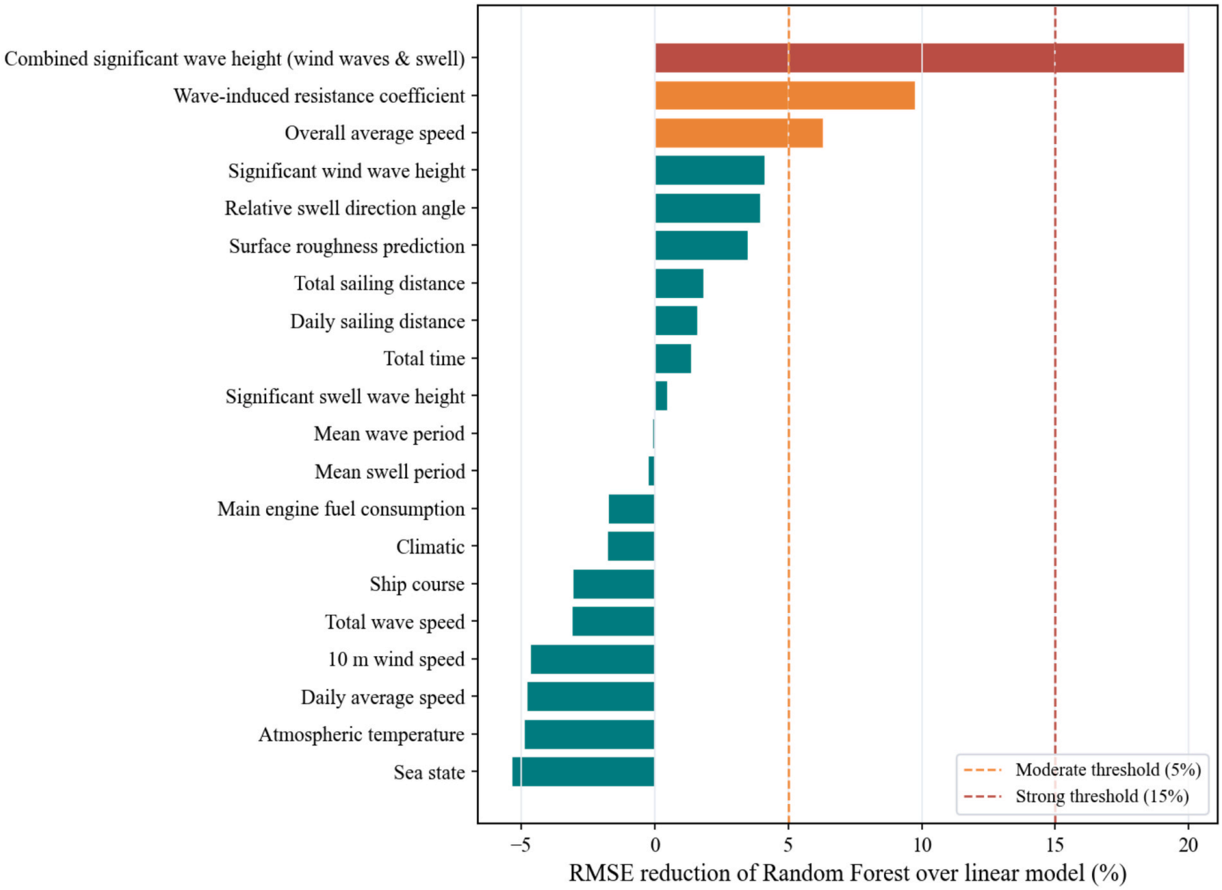

Fig. C2. Linearity-approximation diagnostics for DirectLiNGAM structural equations.

As shown in Table C3, most structural equations can be regarded as acceptable linear approximations, with only modest RMSE improvement when Random Forest is used. However, the structural equation for Combined significant wave height shows a 19.831% RMSE reduction, indicating relatively strong non-linearity. This result does not invalidate the use of DirectLiNGAM, but it requires caution in interpretation. Specifically, the DirectLiNGAM graph is best understood as a linear directional-dependence approximation among the retained observed variables, rather than as a complete non-linear physical model of SFC.

Table C3
Linearity-approximation diagnostics for selected structural equations.

Node	No. of parent variables	Cross-validated linear RMSE	Cross-validated RF RMSE	RF improvement over linear (%)	Diagnostic interpretation
Combined significant wave height (wind waves & swell)	4	0.087	0.070	19.831	Strong non-linearity; interpret DirectLiNGAM as exploratory
Wave-induced resistance coefficient	2	0.000	0.000	9.756	Moderate non-linearity
Overall average speed	9	0.904	0.847	6.298	Moderate non-linearity
Significant wind wave height	1	0.180	0.173	4.099	Linear approximation acceptable
Relative swell direction angle	5	83.064	79.791	3.940	Linear approximation acceptable
Surface roughness prediction	5	0.000	0.000	3.481	Linear approximation acceptable
Total sailing distance	3	101.697	99.840	1.826	Linear approximation acceptable
Daily sailing distance	2	17.460	17.183	1.591	Linear approximation acceptable
Total time	1	25.543	25.193	1.369	Linear approximation acceptable
Significant swell wave height	13	0.230	0.229	0.456	Linear approximation acceptable
Mean wave period	1	0.876	0.877	−0.094	Linear approximation acceptable
Mean swell period	10	0.493	0.494	−0.269	Linear approximation acceptable

Note: The diagnostic categories are based on heuristic cut-offs: <5%, acceptable linear approximation; 5–15%, moderate non-linearity; and > 15%, strong non-linearity. These categories are used only to assess the plausibility of the linear approximation and should not be interpreted as formal statistical significance tests.

Based on the acyclicity and linearity-approximation checks above, Fig. C3 provides a diagnostic visualisation of the retained DirectLiNGAM directional-dependence network. Fig. C3 is not intended to provide a detailed interpretation of every individual edge, but to show the overall structure used for assumption checking and stability assessment. The graph retains a connected multi-factor structure, including operational variables, environmental variables, intermediate variables and the outcome variable. This indicates that the estimated graph is not a set of isolated pairwise relationships, but a system-level directional-dependence structure. Detailed interpretation of the main network structure is provided in Section 4.2, while Appendix C focuses on diagnostic validity and stability. However, a visually interpretable network does not necessarily imply that every edge is stable. Therefore, Fig. C4 and Table C4 further examine the robustness of edge directions under resampling. Fig. C4 visualises the bootstrap frequency of inferred edges, while Table C4 reports the most stable edges under both IID bootstrap and temporal block bootstrap resampling. In this context, temporal block bootstrap provides a more conservative diagnostic because it partially preserves local temporal dependence among adjacent daily observations (Chen et al., 2025).

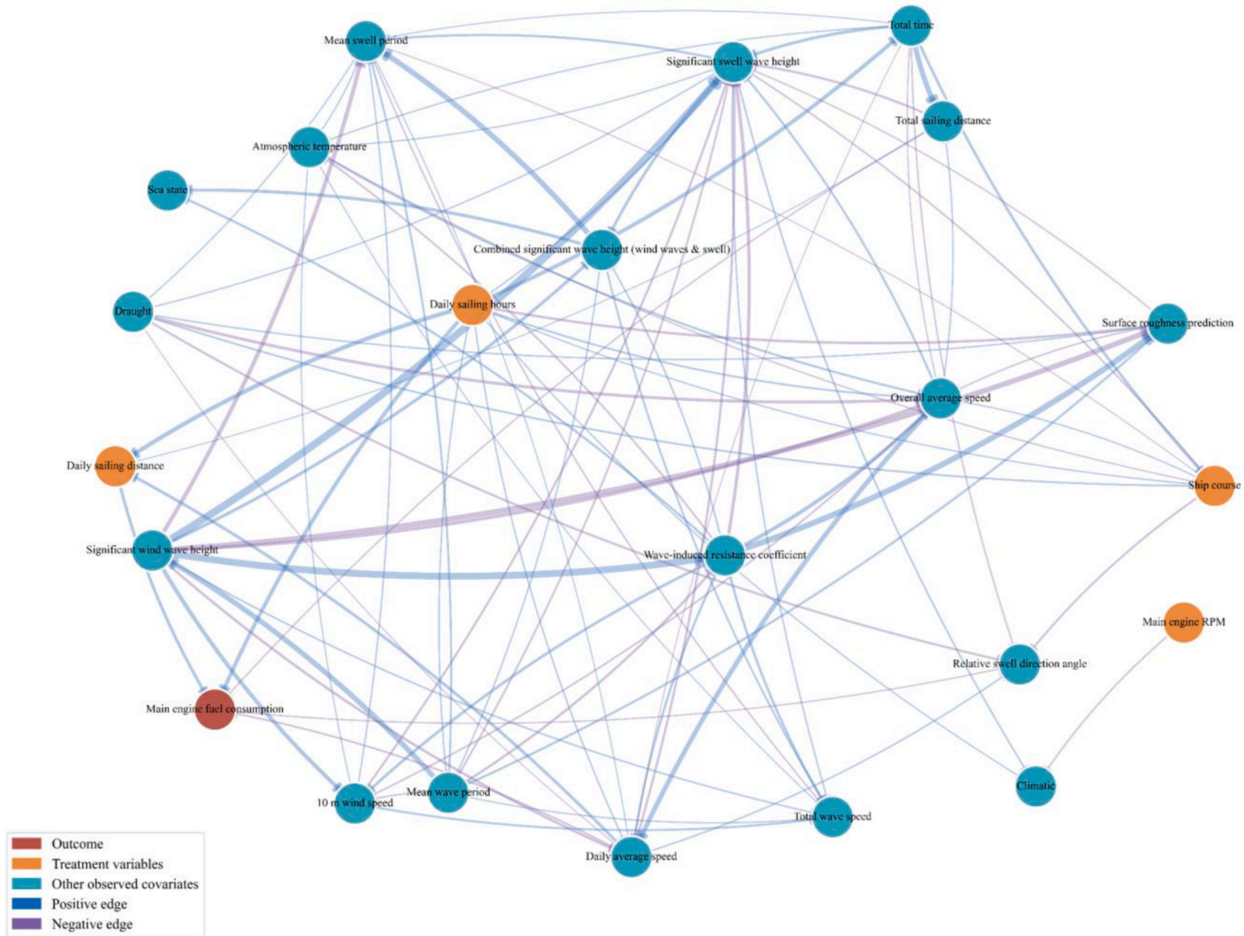

Fig. C3. Diagnostic visualisation of the DirectLiNGAM directional-dependence network.

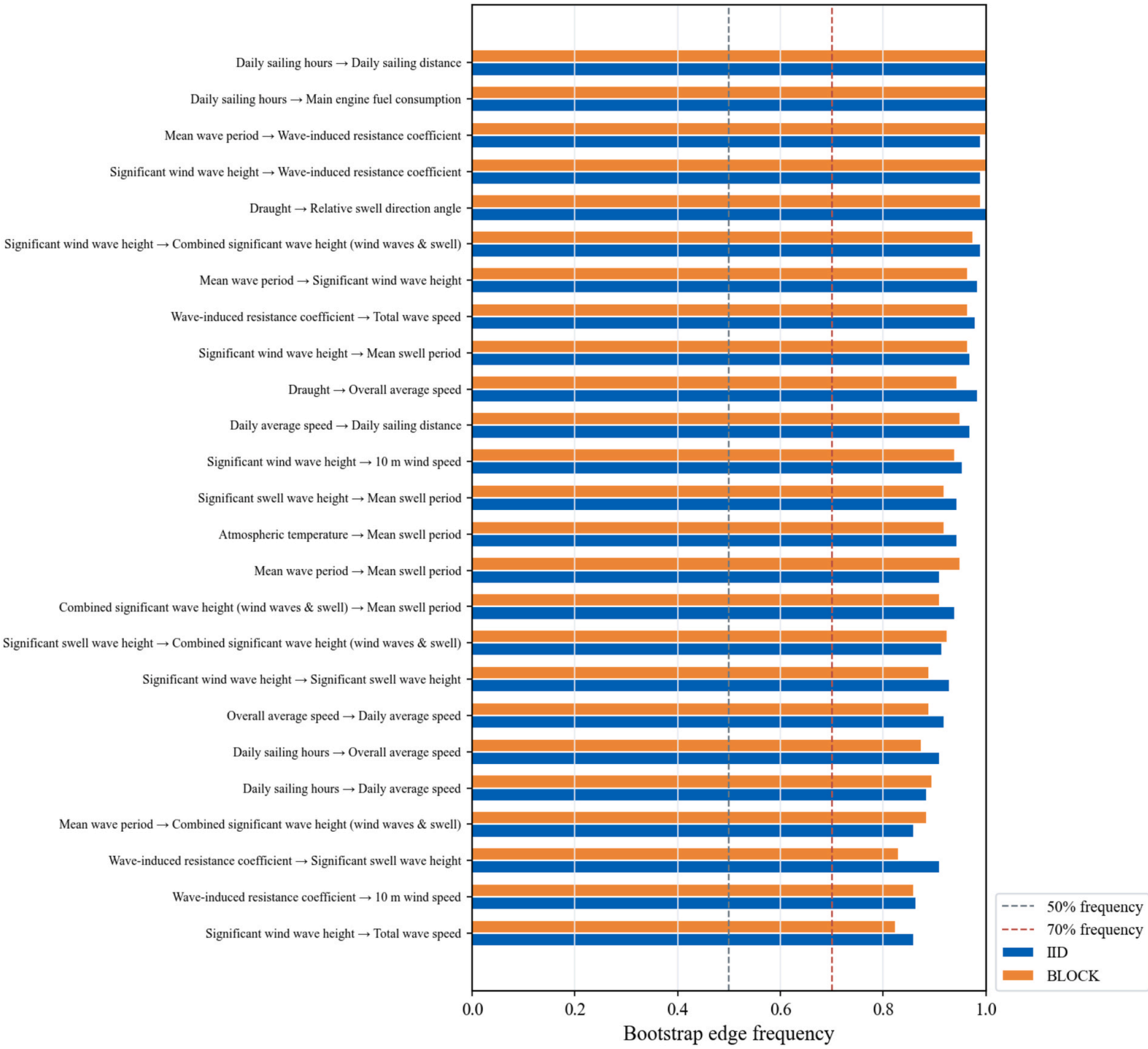

Fig. C4. Bootstrap stability of DirectLiNGAM edge directions.

The results in Fig. C4 and Table C4 show that several physically and operationally interpretable edges are highly stable. For example, Daily sailing hours → Daily sailing distance and Daily sailing hours → Main engine fuel consumption appear with a frequency of 1.000 under both resampling schemes. Environmental and wave-related links, such as Significant wind wave height → Wave-induced resistance coefficient and Mean wave period → Wave-induced resistance coefficient, also exhibit high stability. These stable links support the credibility of the main directional patterns used in the subsequent analysis. At the same time, edges with lower bootstrap frequencies are treated as exploratory and are not over-interpreted as definitive causal pathways.

Table C4
Most stable DirectLiNGAM edges under IID and temporal block bootstrap resampling.

Cause	Effect	Block frequency	IID frequency	Mean frequency
Daily sailing hours	Daily sailing distance	1.000	1.000	1.000
Daily sailing hours	Main engine fuel consumption	1.000	1.000	1.000
Significant wind wave height	Wave-induced resistance coefficient	1.000	0.990	0.995
Mean wave period	Wave-induced resistance coefficient	1.000	0.990	0.995
Draught	Relative swell direction angle	0.990	1.000	0.995

(continued on next page)

Table C4 (*continued*)

Cause	Effect	Block frequency	IID frequency	Mean frequency
Significant wind wave height	Combined significant wave height (wind waves & swell)	0.975	0.990	0.982
Mean wave period	Significant wind wave height	0.965	0.985	0.975
Wave-induced resistance coefficient	Total wave speed	0.965	0.980	0.972
Significant wind wave height	Mean swell period	0.965	0.970	0.968
Draught	Overall average speed	0.945	0.985	0.965
Daily average speed	Daily sailing distance	0.950	0.970	0.960
Significant wind wave height	10 m wind speed	0.940	0.955	0.948
Atmospheric temperature	Mean swell period	0.920	0.945	0.932
Significant swell wave height	Mean swell period	0.920	0.945	0.932
Mean wave period	Mean swell period	0.950	0.910	0.930
Combined significant wave height (wind waves & swell)	Mean swell period	0.910	0.940	0.925
Significant swell wave height	Combined significant wave height (wind waves & swell)	0.925	0.915	0.920
Significant wind wave height	Significant swell wave height	0.890	0.930	0.910
Overall average speed	Daily average speed	0.890	0.920	0.905
Daily sailing hours	Overall average speed	0.875	0.910	0.893

The previous diagnostics evaluate the distributional assumption, graph acyclicity, linear approximation and edge stability. The final group of diagnostics examines the residual-related assumptions, which are especially important for DirectLiNGAM (Shahbazinia et al., 2023). Fig. C5 reports the pairwise Spearman correlation matrix of the structural residuals. The heatmap indicates that most residual correlations are weak in magnitude, but the residual-independence assumption is not fully supported. In the pairwise screening, 53 residual pairs are significant at $p < 0.05$ before multiple-testing adjustment. This means that statistical dependence can be detected for some residual pairs even though the absolute residual correlations are generally small. This result does not invalidate the analysis, but it shows that the estimated graph should not be interpreted as a fully identified causal structure with mutually independent errors. Rather, it should be understood as a data-driven approximation of dominant directional dependencies among retained observed variables.

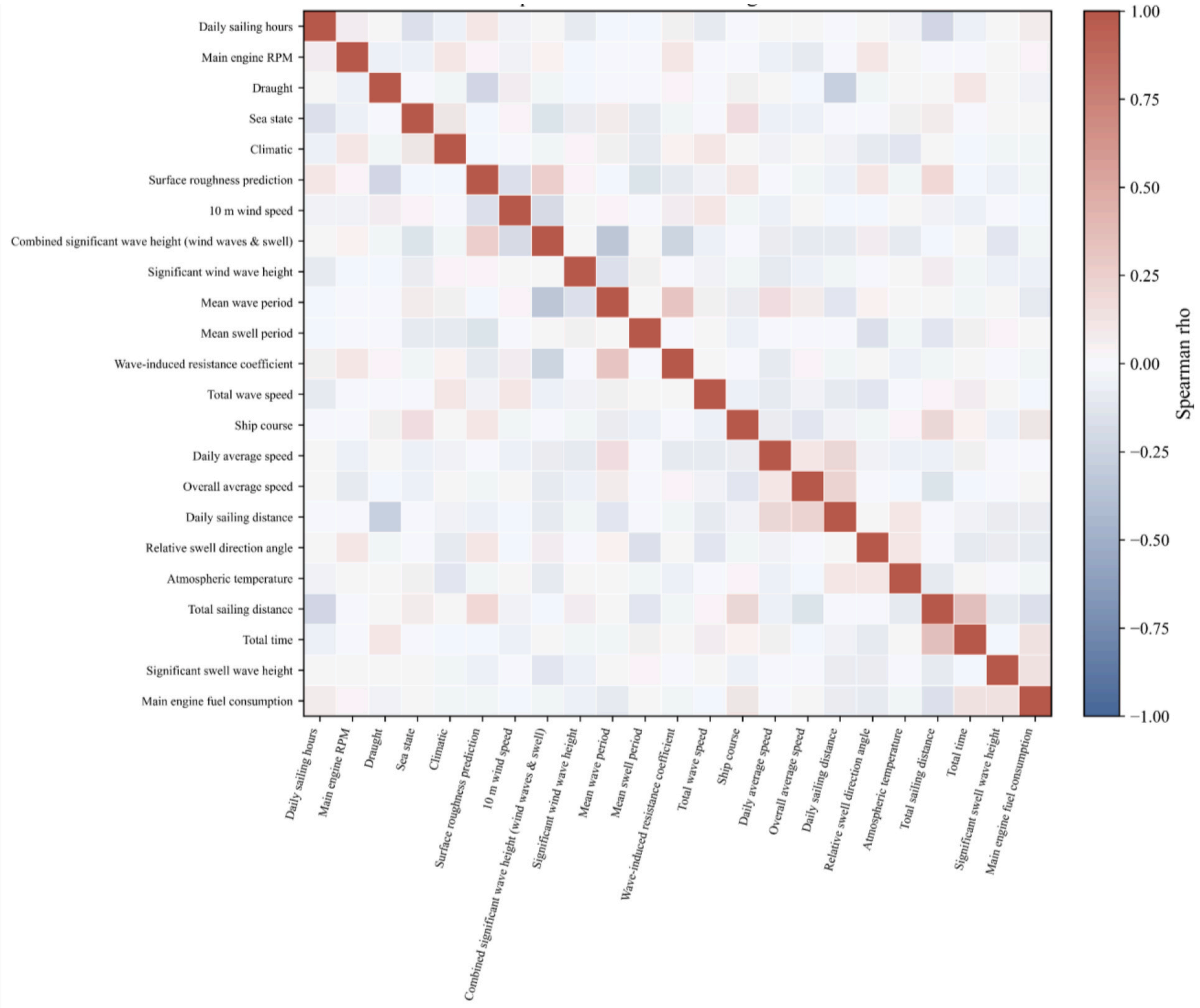

Fig. C5. Pairwise dependence among DirectLiNGAM structural residuals.

Because the dataset consists of repeated daily observations from a single vessel, residual independence also needs to be examined from a temporal perspective. Fig. C6 presents the Ljung–Box temporal-dependence diagnostics for structural residuals at lags 1, 3, 7 and 14, while Table C5 summarises the residual variables showing the strongest temporal dependence. In Fig. C6, the plotted value is − log10 (p-value), which is a transformed visualisation of the Ljung–Box test p-value rather than a separate significance statistic. Larger values indicate smaller Ljung–Box p-values and therefore stronger evidence of residual temporal dependence. For plotting purposes only, p-values smaller than 1×10^{-12} were truncated to 1×10^{-12} before transformation, so the maximum displayed value is 12. This upper limit is a numerical visualisation cap and should not be interpreted as a statistical threshold. Temporal dependence was identified using the Ljung–Box p-value criterion of $p < 0.05$. The results show that temporal dependence remains in a considerable proportion of residual series. For example, Atmospheric temperature, Main engine RPM, Main engine fuel consumption, Draught and Sea state show temporal dependence across all tested lags. This pattern is consistent with the operational reality of ship data, where adjacent daily records may share persistent voyage-level, weather-related and vessel-condition characteristics.

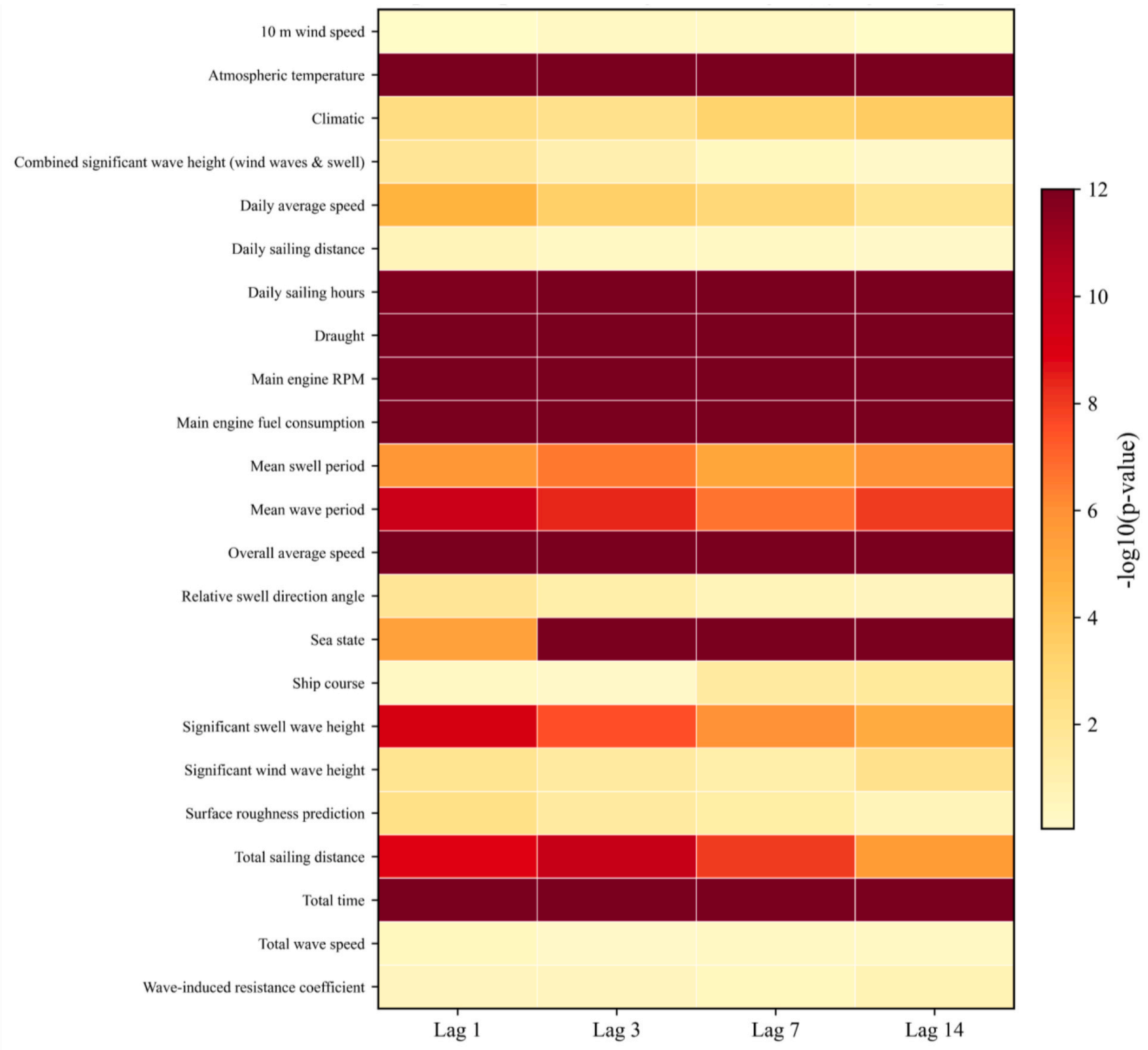

Fig. C6. Temporal-dependence diagnostics for DirectLiNGAM structural residuals.

Table C5
Summary of temporal-dependence diagnostics for selected residual variables.

Residual variable	Tests with temporal dependence	Durbin-Watson	Minimum Ljung-Box p
Atmospheric temperature	4	0.323	0.00e + 00
Main engine RPM	4	0.211	0.00e + 00
Main engine fuel consumption	4	1.086	5.78e-218
Draught	4	0.968	5.50e-77
Sea state	4	1.579	2.97e-43
Overall average speed	4	0.965	1.03e-38
Total time	4	1.298	1.06e-32
Daily sailing hours	4	2.643	1.44e-22
Total sailing distance	4	1.450	1.47e-10
Mean wave period	4	1.425	2.79e-10
Significant swell wave height	4	1.438	8.12e-10
Mean swell period	4	1.566	2.46e-07
Daily average speed	4	1.612	2.20e-05
Climatic	4	1.724	2.49e-04
Significant wind wave height	3	1.764	6.38e-03

Note: The Ljung–Box p-value is the p-value obtained from the Ljung–Box test for residual autocorrelation. It is the same type of p-value as that visualised in Fig. C6, where the values are transformed as − log10 (p-value) for graphical clarity.

Overall, Appendix C shows that the retained variables are strongly non-Gaussian and that the estimated graph satisfies the acyclicity requirement. Several stable edges also show meaningful physical and operational consistency. However, the linearity, residual-independence and temporal-dependence diagnostics indicate that the DirectLiNGAM output should be interpreted as an exploratory linear directional-dependence structure conditional on the retained observed variables, rather than as conclusive evidence of full causal identification.

Appendix D:. Robustness and sensitivity diagnostics for DML-CF under graph-derived adjustment

The Appendix D reports robustness and sensitivity diagnostics for the DML-CF effect-estimation stage under the expanded DAG-derived adjustment specification. The expanded adjustment sets used here were formally derived from the DirectLiNGAM-estimated graph, as reported in Appendix E, rather than selected using the simplified direct-parent or parent-plus-domain rule. Therefore, the diagnostics in the Appendix D evaluate the stability of the revised DML-CF estimates after graph-based covariate adjustment. Three complementary checks are conducted: temporal block-bootstrap resampling, leave-one-covariate-out sensitivity analysis and permutation placebo testing. These diagnostics are particularly important because the present dataset consists of repeated daily observations from a single vessel, and because unobserved operational factors such as voyage planning, loading condition, hull condition, maintenance state and hull fouling may still affect both operational decisions and fuel consumption. Accordingly, the results are used to support a cautious interpretation of the DML-CF estimates as conditional effect estimates under the retained observed covariates, rather than as definitive causal effects under complete confounder control (Li et al., 2024b; Nachtigall et al., 2025).

The first robustness check concerns temporal dependence. Although the main DML-CF estimates are obtained after graph-derived covariate adjustment, the observations are repeated daily operational records from the same vessel and may therefore contain local temporal dependence. To address this issue, temporal block-bootstrap resampling was used to assess the uncertainty of the ATE estimates. Compared with ordinary IID bootstrap resampling, temporal block-bootstrap resampling partially preserves local dependence by resampling consecutive blocks of observations, thereby providing a more conservative diagnostic for the repeated daily ship-operation dataset (Ma and Zhang, 2024). The corresponding temporal block-bootstrap distributions are reported in Fig. 4 in the main text, while Table D1 summarises the bootstrap statistics under the expanded DAG-derived adjustment specification.

Table D1
Temporal block-bootstrap robustness of DML-CF estimates under the expanded DAG-derived adjustment specification.

Treatment	Adjustment specification	Block size	Successful bootstrap runs	Bootstrap ATE mean	Bootstrap median	95% CI lower	95% CI upper	Std
Daily sailing hours	Expanded DAG-derived adjustment set	14	300	2.0596	2.0679	1.8723	2.2152	0.0835
Main engine RPM	Expanded DAG-derived adjustment set	14	300	0.2472	0.2332	−0.2461	0.7936	0.2666
Daily sailing distance	Expanded DAG-derived adjustment set	14	300	0.0359	0.0346	0.0233	0.0514	0.0076
Ship course	Expanded DAG-derived adjustment set	14	300	0.0017	0.0016	−0.0003	0.0037	0.0010

Note: The estimates were obtained under the expanded DAG-derived adjustment specification. The confidence intervals were calculated from temporal block-bootstrap resampling with 300 successful runs. These intervals should be interpreted as uncertainty diagnostics for observed-covariate conditional effects, rather than as proof of complete causal identification.

As shown in Table D1 and Fig. 4, Daily sailing hours and Daily sailing distance remain positive with confidence intervals above zero under temporal block-bootstrap resampling. This indicates that the effects of time- and distance-related operational exposure variables are comparatively stable when local temporal dependence is partially retained. Daily sailing hours shows the largest and most stable positive conditional effect, with a temporal block-bootstrap mean of 2.0596 and a 95% confidence interval of [1.8723, 2.2152]. Daily sailing distance also remains positive and stable, with a temporal block-bootstrap mean of 0.0359 and a 95% confidence interval of [0.0233, 0.0514]. By contrast, Main engine RPM retains a positive mean estimate, but its confidence interval crosses zero, indicating greater uncertainty under the more conservative temporal resampling scheme. Ship course has a very small effect magnitude, and its confidence interval is close to zero. These findings support the main conclusion that sailing-time- and distance-related operational exposure variables provide more stable evidence for fuel-consumption management than propulsion-speed and heading-related variables under the expanded DAG-derived adjustment specification.

The second diagnostic evaluates whether the estimated treatment effects are sensitive to individual adjustment covariates. This check is necessary because the DML-CF estimates are conditional on the retained observed adjustment set, and ship-operational variables such as sailing time, sailing distance, speed and propulsion setting are mechanically and operationally coupled (Li et al., 2024b). For each treatment variable, one adjustment covariate was removed at a time from the expanded DAG-derived adjustment set, and the DML-CF ATE was re-estimated. Fig. D1 visualises the relative change in ATE after removing individual covariates, while Table D2 reports the most influential removed covariates for each treatment variable.

The leave-one-covariate-out results reveal heterogeneous sensitivity across the four treatment variables. The estimate for Daily sailing hours is relatively stable, because removing Draught changes the ATE by only 4.01%. For Main engine RPM, the largest relative changes are caused by removing Daily sailing distance, Daily sailing hours and Ship course, indicating that propulsion-related effects are linked to broader operational-exposure and heading conditions. For Daily sailing distance, removing Daily sailing hours produces a substantial relative increase in the estimated ATE, while removing Daily average speed produces a large negative change. This finding is expected because sailing time, speed and distance jointly describe daily voyage execution and operational exposure. For Ship course, the largest relative changes are observed when Total time, Daily sailing hours or Draught are removed. However, these percentage changes should be interpreted cautiously because the baseline ATE of Ship course is very small. Overall, the sensitivity analysis reinforces that the DML-CF estimates are conditional on the specified graph-derived observed-covariate set and should not be interpreted as universal intervention effects independent of adjustment specification.

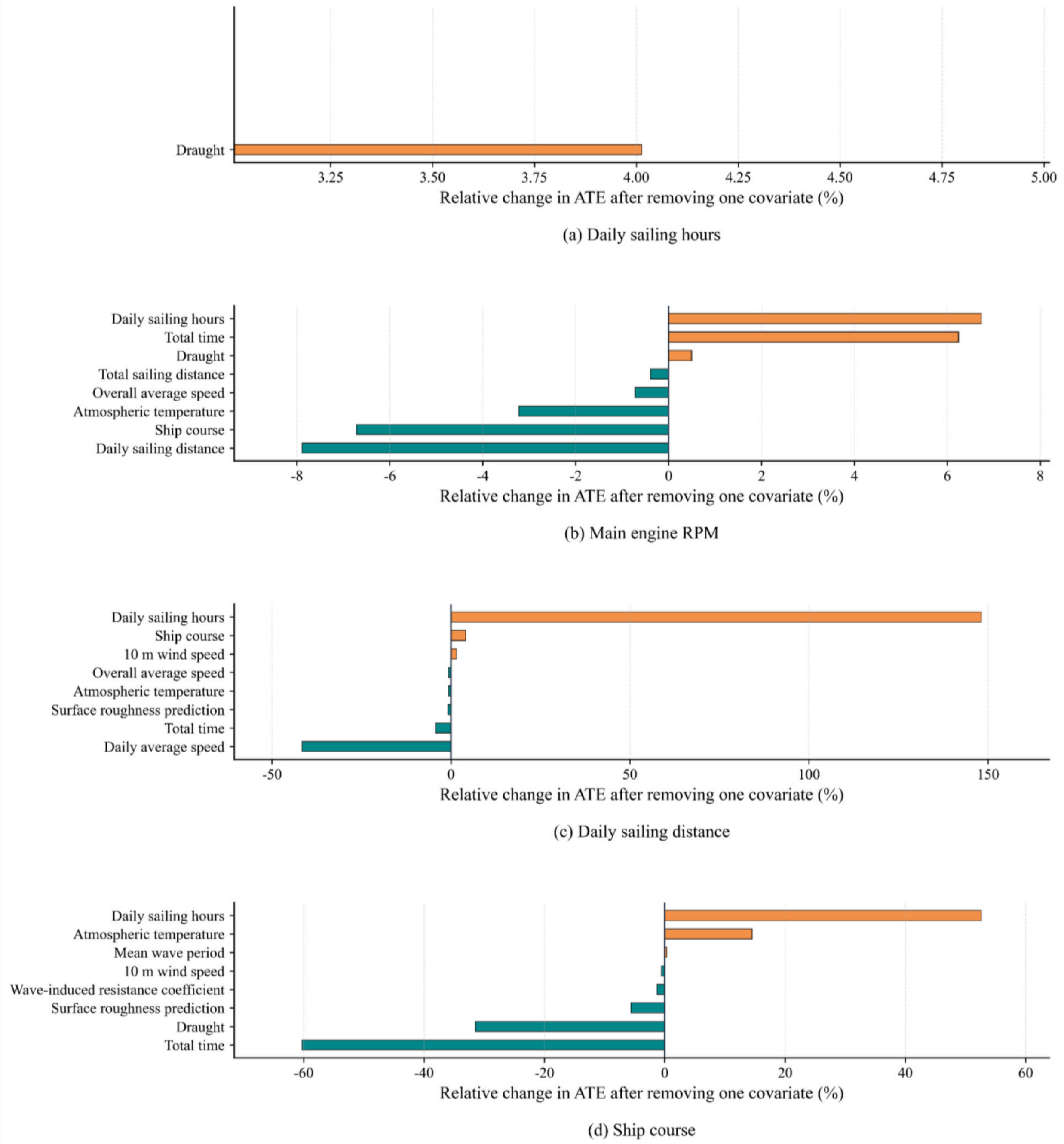

Fig. D1. Leave-one-covariate-out sensitivity of DML-CF ATE estimates under the expanded DAG-derived adjustment specification. **Note:** Each bar reports the relative change in ATE after removing one adjustment covariate. Larger absolute values indicate stronger dependence of the estimated treatment effect on the corresponding covariate.

Table D2
Main leave-one-covariate-out sensitivity results under the expanded DAG-derived adjustment specification.

Treatment	Removed covariate	Base ATE	ATE after removal	Relative change (%)	Interpretation
Daily sailing hours	Draught	2.0509	2.1332	4.01	Low sensitivity to the single adjustment covariate
Main engine RPM	Daily sailing distance	0.2654	0.2444	−7.89	Moderate sensitivity to distance-related operational exposure
Main engine RPM	Daily sailing hours	0.2654	0.2833	6.73	Moderate sensitivity to sailing-time exposure
Main engine RPM	Ship course	0.2654	0.2476	−6.72	Moderate sensitivity to heading-related operation
Daily sailing distance	Daily sailing hours	0.0355	0.0880	148.17	High sensitivity, reflecting strong coupling with sailing-time exposure
Daily sailing distance	Daily average speed	0.0355	0.0207	−41.60	High sensitivity to speed-related adjustment
Daily sailing distance	Total time	0.0355	0.0339	−4.26	Limited sensitivity to cumulative time information
Ship course	Total time	0.0017	0.0007	−60.28	High relative sensitivity caused by small baseline effect magnitude
Ship course	Daily sailing hours	0.0017	0.0025	52.64	High relative sensitivity caused by small baseline effect magnitude
Ship course	Draught	0.0017	0.0011	−31.48	High relative sensitivity caused by small baseline effect magnitude

Note: Relative change was calculated as 100 × (ATE after removal − base ATE) / base ATE. The sensitivity levels were classified using the absolute relative change: <5% as low or limited sensitivity, 5–15% as moderate sensitivity, and ≥ 15% as high relative sensitivity. These thresholds are heuristic diagnostic cut-offs rather than formal statistical significance criteria.

The third diagnostic evaluates whether the observed DML-CF estimates are distinguishable from effects generated by randomised treatment assignments. In each placebo test, the treatment variable was randomly permuted while the outcome and graph-derived adjustment covariates were kept unchanged. The DML-CF model was then re-estimated using the permuted treatment variable. This process was repeated 200 times for each treatment. If the observed ATE is clearly separated from the placebo distribution, this suggests that the estimated effect is unlikely to be produced by random treatment assignment alone. However, the placebo test does not rule out unobserved confounding, omitted operational states or all forms of temporal dependence. It should therefore be interpreted as a robustness diagnostic rather than as proof of complete causal identification.

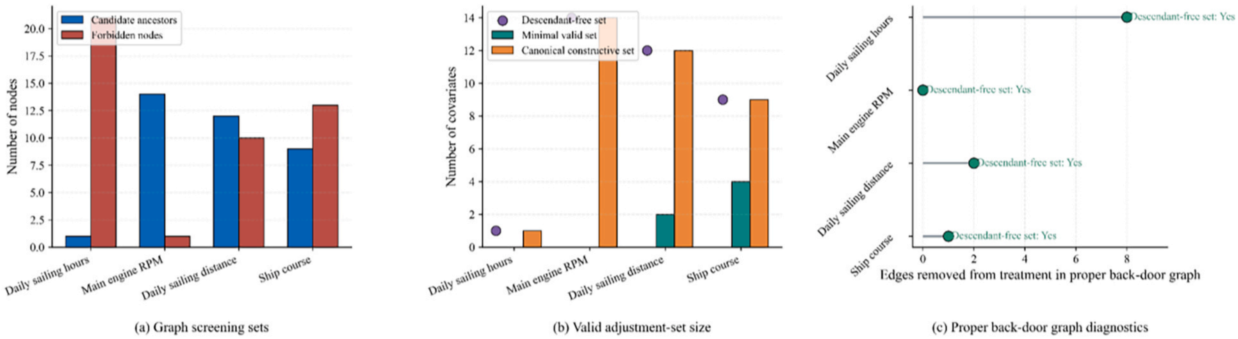

Fig. D2. Permutation placebo distributions of DML-CF ATE estimates under the expanded DAG-derived adjustment specification. **Note:** The vertical solid lines indicate the observed ATEs, while the shaded distributions represent placebo estimates obtained after randomly permuting each treatment variable.

As shown in Fig. D2 and Table D3, the placebo distributions are centred close to zero for all four treatment variables. The observed ATEs for Daily sailing hours, Main engine RPM and Daily sailing distance are clearly separated from their corresponding placebo intervals, with empirical two-sided p-values of 0.0050. This provides additional evidence that the estimated effects of these variables reflect structured relationships in the observed ship-operation data rather than random treatment-assignment artefacts. However, Ship course is not clearly distinguishable from its placebo distribution, with an empirical two-sided p-value of 0.1443. This result is consistent with the temporal block-bootstrap results, which show that the estimated effect of Ship course is small and close to zero. Therefore, Ship course should be interpreted as a contextual heading-related variable rather than as a dominant independent driver of fuel consumption.

Table D3
Permutation placebo diagnostics under the expanded DAG-derived adjustment specification.

Treatment	Observed ATE	Placebo mean	Placebo 2.5%	Placebo 97.5%	Empirical two-sided p	Successful permutations	Interpretation
Daily sailing hours	2.0467	0.0096	−0.1662	0.2381	0.0050	200	Observed estimate is clearly separated from the placebo distribution
Main engine RPM	0.2719	−0.0003	−0.1835	0.1707	0.0050	200	Observed estimate exceeds the placebo range, but temporal-bootstrap uncertainty remains
Daily sailing distance	0.0358	−0.0001	−0.0034	0.0027	0.0050	200	Observed estimate is clearly separated from randomised-treatment effects
Ship course	0.0017	−0.0001	−0.0023	0.0020	0.1443	200	Observed estimate is not clearly separated from the placebo distribution

Note: For each treatment variable, the treatment values were randomly permuted 200 times while keeping the outcome and graph-derived adjustment covariates unchanged. The empirical two-sided p-value measures the proportion of placebo estimates whose absolute magnitude is at least as large as the observed estimate.

Overall, the robustness diagnostics provide a more cautious and internally consistent interpretation of the DML-CF estimates under the expanded DAG-derived adjustment specification. The temporal block-bootstrap results indicate that Daily sailing hours and Daily sailing distance retain positive and stable conditional effects, whereas Main engine RPM and Ship course show greater uncertainty under temporal resampling. The leave-one-covariate-out analysis further shows that the Daily sailing hours estimate is relatively insensitive to its single adjustment covariate, while the Daily sailing distance estimate is sensitive to the removal of Daily sailing hours and Daily average speed, reflecting the strong operational coupling between sailing time, speed and distance. The placebo diagnostics show that the observed estimates for Daily sailing hours, Main engine RPM and Daily sailing distance are clearly separated from randomised-treatment benchmarks, whereas Ship course is not clearly distinguishable from its placebo distribution. Taken together, these results support the main conclusion that time- and distance-related operational exposure variables provide more stable evidence for fuel-consumption management than propulsion-speed and heading-related variables. At the same time, the results reinforce the need to interpret the DML-CF estimates as observed-covariate conditional effect estimates, rather than as definitive causal effects under complete confounder control.

Appendix E:. Graph-derived adjustment-set construction for DML-CF effect estimation

The Appendix E reports the graph-derived adjustment-set construction used for the DML-CF effect-estimation stage. The purpose is to address the concern that selecting the direct parent nodes of each treatment variable is not, in general, sufficient to satisfy the back-door criterion. A valid adjustment set must be derived from the graph structure by blocking non-causal back-door paths while avoiding adjustment for treatment descendants, pathway mediators and collider-related structures. Therefore, for each treatment variable, this study constructed the proper back-door graph from the DirectLiNGAM-estimated DAG, excluded forbidden nodes according to the generalised adjustment logic, and derived both a minimal DAG-derived adjustment set and an expanded robustness-oriented adjustment set.

Table E1 reports the resulting adjustment sets. Fig. E1 summarises the corresponding graph-screening diagnostics, including candidate ancestors, forbidden nodes, adjustment-set size and proper back-door graph construction. The expanded DAG-derived adjustment sets reported in Table E1 were subsequently used as the unified adjustment specification for the DML-CF robustness and sensitivity diagnostics reported in Appendix D. This separation avoids conflating adjustment-set construction with effect-estimation robustness: Appendix E explains how the adjustment sets were derived, while Appendix D evaluates the stability of DML-CF estimates under these graph-derived adjustment sets.

Table E1
DAG-derived adjustment sets for DML-CF effect estimation.

Treatment	Outcome	Minimal DAG-derived adjustment set	Expanded DAG-derived adjustment set used for robustness
Daily sailing hours	Main engine fuel consumption	None required after graphical pruning in the estimated DAG	Draught
Main engine RPM	Main engine fuel consumption	None required after graphical pruning in the estimated DAG	10 m wind speed; Atmospheric temperature; Daily average speed; Daily sailing distance; Daily sailing hours; Draught; Mean wave period; Overall average speed; Ship course; Significant wind wave height; Surface roughness prediction; Total sailing distance; Total time; Wave-induced resistance coefficient
Daily sailing distance	Main engine fuel consumption	Daily average speed; Daily sailing hours	10 m wind speed; Atmospheric temperature; Daily average speed; Daily sailing hours; Draught; Mean wave period; Overall average speed; Ship course; Significant wind wave height; Surface roughness prediction; Total time; Wave-induced resistance coefficient
Ship course	Main engine fuel consumption	Atmospheric temperature; Daily sailing hours; Draught; Total time	10 m wind speed; Atmospheric temperature; Daily sailing hours; Draught; Mean wave period; Significant wind wave height; Surface roughness prediction; Total time; Wave-induced resistance coefficient

Note: The minimal adjustment set was obtained through graphical pruning and d-separation checking in the proper back-door graph derived from the DirectLiNGAM-estimated DAG. “None required” indicates that no additional observed covariate was required after graphical pruning in the estimated DAG. The expanded DAG-derived adjustment set was used as a conservative robustness specification by retaining theoretically relevant upstream operational and environmental covariates that were not forbidden by the graph-based adjustment procedure.

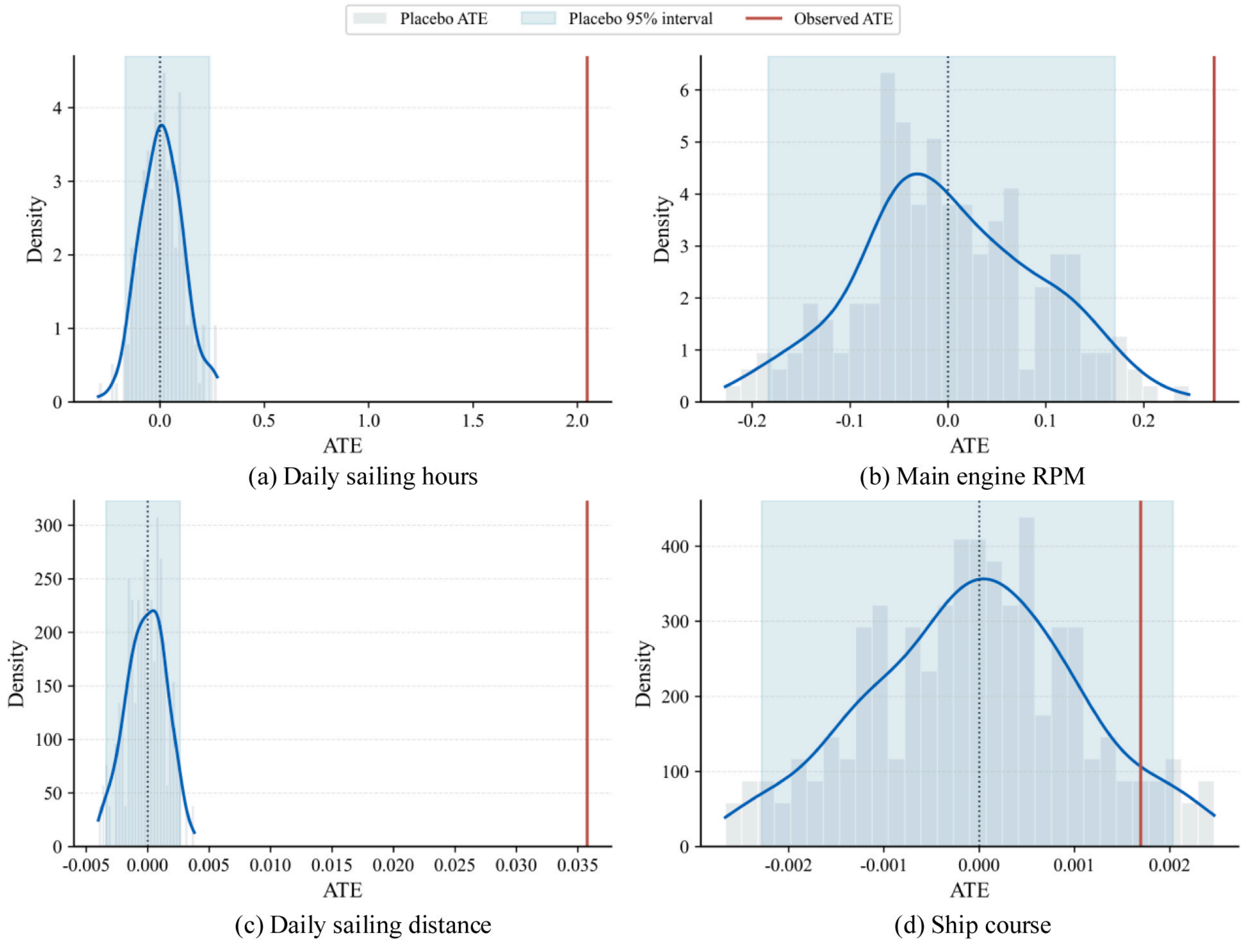

Fig. E1. Graphical diagnostics for DAG-derived adjustment-set construction.

The Fig. E1(a) reports the number of candidate ancestors and forbidden nodes considered for each treatment variable. Fig. E1(b) compares the sizes of the descendant-free, minimal and canonical constructive adjustment sets. Fig. E1(c) shows the number of treatment-outgoing edges removed when constructing the proper back-door graph and confirms whether descendant-free adjustment sets were retained.

Overall, the graph-derived adjustment results indicate that the valid adjustment set is not equivalent to the direct parent nodes of each treatment variable. For Daily sailing hours and Main engine RPM, no additional observed covariate was required in the minimal DAG-derived adjustment set after graphical pruning, whereas the expanded robustness specification retained upstream operational-environmental covariates. For Daily sailing distance and Ship course, the minimal sets include operational and environmental covariates that block non-causal paths in the estimated DAG. These results directly address the limitation of the simplified direct-parent rule and provide the graph-derived adjustment basis for the DML-CF robustness diagnostics in Appendix D.

Data availability

Data will be made available on request. The source code is publicly available at: https://github.com/AdvMarTech/ship_fuel_consum_causalinfer.

References

Abdulrashid, I., Zanjirani Farahani, R., Mammadov, S., Khalafalla, M., Chiang, W.-C., 2024. Explainable artificial intelligence in transport Logistics: risk analysis for road accidents. Transportation Research Part E: Logistics and Transportation Review 186, 103563. https://doi.org/10.1016/j.tre.2024.103563.

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. https://doi.org/10.1016/j.tre.2020.101972.

Agrawal, R., Imieliński, T., Swami, A., 1993. In: Mining Association Rules between Sets of Items in Large Databases. Association for Computing Machinery, Washington, D.C., USA, pp. 207–216.

Arif, S., MacNeil, M.A., 2023. Applying the structural causal model framework for observational causal inference in ecology. Ecol. Monogr. 93 (1), e1554.

Balcombe, P., Brierley, J., Lewis, C., Skatvedt, L., Speirs, J., Hawkes, A., Staffell, I., 2019. How to decarbonise international shipping: Options for fuels, technologies and policies. Energ. Conver. Manage. 182, 72–88. https://doi.org/10.1016/j.enconman.2018.12.080.

Bell, M.G.H., Liu, X., Rioult, J., Angeloudis, P., 2013. A cost-based maritime container assignment model. Transp. Res. B Methodol. 58, 58–70. https://doi.org/10.1016/j.trb.2013.09.006.

Bialystocki, N., Konovessis, D., 2016. On the estimation of ship's fuel consumption and speed curve: a statistical approach. J. Ocean. Eng. Sci. 1 (2), 157–166. https://doi.org/10.1016/j.joes.2016.02.001.

Borén, C., Castells-Sanabra, M., Grifoll, M., 2022. Ship emissions reduction using weather ship routing optimisation. Proceedings of the Institution of Mechanical Engineers, Part m: Journal of Engineering for the Maritime Environment 236 (4), 856–867. https://doi.org/10.1177/14750902221082901.

Borkowski, T., Kowalak, P., Wiśnicki, B., Przywarty, M., 2019. Slow steaming of container carriers. Maritime University of Szczecin Press, A study.

Cai, Z., Li, L., Yu, L., Li, C., Sun, M., 2024. Diversity, quality, and quantity of real ship data on the black-box and gray-box prediction models of ship fuel consumption. Ocean Eng. 291, 116434. https://doi.org/10.1016/j.oceaneng.2023.116434.

Cao, W., Wang, X., Feng, Y., Zhou, J., Yang, Z., 2026a. Improving maritime accident severity prediction accuracy: a holistic machine learning framework with data balancing and explainability techniques. Reliab. Eng. Syst. Saf. 266, 111648. https://doi.org/10.1016/j.ress.2025.111648.

Cao, W., Wang, X., Li, J., Zhang, Z., Cao, Y., Feng, Y., 2024. A novel integrated method for heterogeneity analysis of marine accidents involving different ship types. Ocean Eng. 312, 119295. https://doi.org/10.1016/j.oceaneng.2024.119295.

Cao, W., Wang, X., Shu, Y., Li, H., Zhou, J., Yang, Z., 2026b. An integrated method of advanced optimisation and adaptive ensemble learning for ship fuel consumption prediction. Transp. Res. Part C Emerging Technol. 188, 105659. https://doi.org/10.1016/j.trc.2026.105659.

Cao, Y., Iulia, M., Majumdar, A., Feng, Y., Xin, X., Wang, X., Wang, H., Yang, Z., 2025a. Investigation of the risk influential factors of maritime accidents: a novel topology and robustness analytical framework. Reliab. Eng. Syst. Saf. 254, 110636. https://doi.org/10.1016/j.ress.2024.110636.

Cao, Y., Xin, X., Jarumaneeroj, P., Li, H., Feng, Y., Wang, J., Wang, X., Pyne, R., Yang, Z., 2025b. Data-driven resilience analysis of the global container shipping network against two cascading failures. Transportation Research Part E: Logistics and Transportation Review 193, 103857. https://doi.org/10.1016/j.tre.2024.103857.

Cao, Y., Xin, X., Wang, X., Wang, J., Yang, Z., 2025c. Multi-objective resilience-oriented optimisation for the global container shipping network against cascading failures. Transp. Res. A Policy Pract. 200, 104659. https://doi.org/10.1016/j.tra.2025.104659.

Carrasco, M., Ivorra, B., López, J., Ramos, A.M., 2025. Embedded feature selection for robust probability learning machines. Pattern Recogn. 159, 111157. https://doi.org/10.1016/j.patcog.2024.111157.

Chauhan, R.S., Riis, C., Adhikari, S., Derrible, S., Zheleva, E., Choudhury, C.F., Pereira, F.C., 2024. Determining causality in travel mode choice. Travel Behav. Soc. 36, 100789. https://doi.org/10.1016/j.tbs.2024.100789.

Chen, N., Wubs, E.R.J., Bucur, D., 2025. *Dragon*: Data-driven causal discovery for soils in the presence of latent and discrete variables. Eco. Inform. 92, 103527. https://doi.org/10.1016/j.ecoinf.2025.103527.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econ. J. 21 (1), C1–C68. https://doi.org/10.1111/ectj.12097.

Chowdhury, J., Dutta, S., Arellano-Valle, R.B., Genton, M.G., 2022. Sub-dimensional Mardia measures of multivariate skewness and kurtosis. J. Multivar. Anal. 192, 105089. https://doi.org/10.1016/j.jmva.2022.105089.

Coraddu, A., Oneto, L., Baldi, F., Anguita, D., 2017. Vessels fuel consumption forecast and trim optimisation: a data analytics perspective. Ocean Eng. 130, 351–370. https://doi.org/10.1016/j.oceaneng.2016.11.058.

Degiuli, N., Martić, I., Farkas, A., Gospić, I., 2021. The impact of slow steaming on reducing CO2 emissions in the Mediterranean Sea. Energy Rep. 7, 8131–8141. https://doi.org/10.1016/j.egyr.2021.02.046.

Dibaeinia, P., Ojha, A., Sinha, S., 2025. Interpretable AI for inference of causal molecular relationships from omics data. Sci. Adv. 11 (7), eadk0837. https://doi.org/10.1126/sciadv.adk0837.

Fang, Y., Liu, Y., Yang, Y., Lucey, B., Abedin, M.Z., 2025. How do chinese urban investment bonds affect its economic resilience? evidence from double machine learning. Res. Int. Bus. Financ. 74, 102728. https://doi.org/10.1016/j.ribaf.2024.102728.

Farag, Y.B.A., Ölçer, A.I., 2020. The development of a ship performance model in varying operating conditions based on ANN and regression techniques. Ocean Eng. 198, 106972. https://doi.org/10.1016/j.oceaneng.2020.106972.

Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Xia, G., Liu, Z., 2024. Prediction of the severity of marine accidents using improved machine learning. Transportation Research Part E: Logistics and Transportation Review 188, 103647. https://doi.org/10.1016/j.tre.2024.103647.

Garrido-Merchán, E.C., Fernández-Sánchez, D., Hernández-Lobato, D., 2023. Parallel predictive entropy search for multi-objective Bayesian optimization with constraints applied to the tuning of machine learning algorithms. Expert Syst. Appl. 215, 119328. https://doi.org/10.1016/j.eswa.2022.119328.

Grlj, C.G., Degiuli, N., Tuković, Ž., Farkas, A., Martić, I., 2023. The effect of loading conditions and ship speed on the wind and air resistance of a containership. Ocean Eng. 273, 113991. https://doi.org/10.1016/j.oceaneng.2023.113991.

Guo, B., Liang, Q., Tvete, H.A., Brinks, H., Vanem, E., 2022. Combined machine learning and physics-based models for estimating fuel consumption of cargo ships. Ocean Eng. 255, 111435. https://doi.org/10.1016/j.oceaneng.2022.111435.
Halff, A., Younes, L., Boersma, T., 2019. The likely implications of the new IMO standards on the shipping industry. Energy Policy 126, 277–286. https://doi.org/10.1016/j.enpol.2018.11.033.
Hamed, A., 2022. Multi-objective optimization method of trimaran hull form for resistance reduction and propeller intake flow improvement. Ocean Eng. 244, 110352. https://doi.org/10.1016/j.oceaneng.2021.110352.
Harada, K., Fujisawa, H., 2021. Sparse estimation of Linear Non-Gaussian Acyclic Model for Causal Discovery. Neurocomputing 459, 223–233. https://doi.org/10.1016/j.neucom.2021.06.083.
He, S., Luo, Q., Du, R., Zhao, L., He, G., Fu, H., Li, H., 2023. STGC-GNNs: a GNN-based traffic prediction framework with a spatial–temporal Granger causality graph. Physica A 623, 128913. https://doi.org/10.1016/j.physa.2023.128913.
Hu, Z., Fan, A., Mao, W., Shu, Y., Wang, Y., Xia, M., Yi, Q., Li, B., 2025. Ship energy consumption prediction: Multi-model fusion methods and multi-dimensional performance evaluation. Ocean Eng. 322, 120538. https://doi.org/10.1016/j.oceaneng.2025.120538.
Hu, Z., Xu, Y., 2025. Can new energy demonstration cities break through the multiple carbon lock-in? evidence based on double machine learning. Energy Policy 199, 114522. https://doi.org/10.1016/j.enpol.2025.114522.
Huang, G., He, Z., Zhao, P., Zhang, C., 2025a. A review of climate-related disasters impact on global shipping carbon emissions. Transp. Res. Part D: Transp. Environ. 139, 104553. https://doi.org/10.1016/j.trd.2024.104553.
Huang, H., Li, B., Wang, Y., Zhang, Z., He, H., 2024. Analysis of factors influencing energy consumption of electric vehicles: Statistical, predictive, and causal perspectives. Appl. Energy 375, 124110. https://doi.org/10.1016/j.apenergy.2024.124110.
Huang, Z., Dong, H., Liu, Z., Albitar, K., 2025b. Unleashing the empowered effect of data resource on inclusive green growth: based on double machine learning. Economic Analysis and Policy 85, 1270–1290. https://doi.org/10.1016/j.eap.2025.01.018.
Inal, O.B., Charpentier, J.-F., Deniz, C., 2022. Hybrid power and propulsion systems for ships: current status and future challenges. Renew. Sustain. Energy Rev. 156, 111965. https://doi.org/10.1016/j.rser.2021.111965.
International Maritime Organization, 2020. Fourth Greenhouse Gas Study 2020. https://www.imo.org/en/ourwork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx.
International Maritime Organization 2023-IMO-Strategy-on-Reduction-of-GHG-Emissions-from-Ships 2023 https://www.imo.org/en/OurWork/Environment/Pages/2023-IMO-Strategy-on-Reduction-of-GHG-Emissions-from-Ships.aspx.
Işıklı, E., Aydın, N., Bilgili, L., Toprak, A., 2020. Estimating fuel consumption in maritime transport. Journal of Cleaner Production 275, 124142. https://doi.org/10.1016/j.jclepro.2020.124142.
Islam, M.S., Dash, P., Nur, A.M., Ahmad, H., Panda, R.M., Wolfe, J.S., Turnage, G., Hathcock, L., Chesser, G.D., Moorhead, R.J., 2025. Estimation of chlorophyll-a in uncrewed aircraft systems imagery using autonomous surface vessel data with machine learning algorithms and feature selection techniques. Ecological Informatics 85, 102954. https://doi.org/10.1016/j.ecoinf.2024.102954.
Joy, T.T., Rana, S., Gupta, S., Venkatesh, S., 2020. Fast hyperparameter tuning using Bayesian optimization with directional derivatives. Knowledge-Based Systems 205, 106247. https://doi.org/10.1016/j.knosys.2020.106247.
Kim, M., Hizir, O., Turan, O., Day, S., Incecik, A., 2017. Estimation of added resistance and ship speed loss in a seaway. Ocean Engineering 141, 465–476. https://doi.org/10.1016/j.oceaneng.2017.06.051.
Kontovas, C.A., 2014. The Green Ship Routing and Scheduling Problem (GSRSP): a conceptual approach. Transportation Research Part E: Transport and Environment 31, 61–69. https://doi.org/10.1016/j.trd.2014.05.014.
Kotoku, J.i., Oyama, A., Kitazumi, K., Toki, H., Haga, A., Yamamoto, R., Shinzawa, M., Yamakawa, M., Fukui, S., Yamamoto, K., 2020. Causal relations of health indices inferred statistically using the DirectLiNGAM algorithm from big data of Osaka prefecture health checkups. Plos One 15 (12), e0243229. https://doi.org/10.1371/journal.pone.0243229.
Lang, X., Mao, W., 2020. A semi-empirical model for ship speed loss prediction at head sea and its validation by full-scale measurements. Ocean Engineering 209, 107494. https://doi.org/10.1016/j.oceaneng.2020.107494.
Le, L.T., Gunwoo, L., Hwayoung, K., Woo, S.-H., 2020. Voyage-based statistical fuel consumption models of ocean-going container ships in Korea. Maritime Policy & Management 47 (3), 304–331. https://doi.org/10.1080/03088839.2019.1684591.
Leyder, S., Jakob, R., Verdonck, T., 2024. TSLiNGAM: DirectLiNGAM under Heavy Tails. Journal of Computational and Graphical Statistics 1–11. https://doi.org/10.1080/10618600.2024.2394462.
Li, S., Pu, Z., Cui, Z., Lee, S., Guo, X., Ngoduy, D., 2024b. Inferring heterogeneous treatment effects of crashes on highway traffic: a doubly robust causal machine learning approach. Transportation Research Part C: Emerging Technologies 160, 104537. https://doi.org/10.1016/j.trc.2024.104537.
Li, Z., Wang, K., Hua, Y., Liu, X., Ma, R., Wang, Z., Huang, L., 2024c. GA-LSTM and NSGA-III based collaborative optimization of ship energy efficiency for low-carbon shipping. Ocean Engineering 312, 119190. https://doi.org/10.1016/j.oceaneng.2024.119190.
Li, Z., Wang, K., Liang, H., Wang, Y., Ma, R., Cao, J., Huang, L., 2025. Marine alternative fuels for shipping decarbonization: Technologies, applications and challenges. Energy Conversion and Management 329, 119641. https://doi.org/10.1016/j.enconman.2025.119641.
Liu, S., Xu, J., Wang, R., Fu, X., Liu, X., Zhao, Y., Zhang, X., 2025a. Investigating the causal effects of anthropogenic factors on urban streams and lakes water quality by integrating causal inference with interpretable machine learning. Journal of Cleaner Production 488, 144626. https://doi.org/10.1016/j.jclepro.2024.144626.
Liu, Z., Zhang, B., Zhang, M., Wang, H., Lang, X., Mao, W., 2025b. A Proactive Decision support Method for Ship Collision Awareness Incorporating Ship Maneuvering. IEEE Transactions on Intelligent Transportation Systems 1–15. https://doi.org/10.1109/TITS.2025.3577731.
Luo, X., Yan, R., Wang, S., 2023. Comparison of deterministic and ensemble weather forecasts on ship sailing speed optimization. Transportation Research Part D: Transport and Environment 121, 103801. https://doi.org/10.1016/j.trd.2023.103801.
Luo, X., Yan, R., Wang, S., 2024. Ship sailing speed optimization considering dynamic meteorological conditions. Transportation Research Part C: Emerging Technologies 167, 104827. https://doi.org/10.1016/j.trc.2024.104827.
Ma, J., Cao, L., Feng, Y., Karatuğ, Ç., Buber, M., Wang, X., 2026. Intelligent analysis of ship collision accidents via Low-Rank Adaptation-based fine-tuning of medium-scale Large Language Models. Reliability Engineering & System Safety 275, 112774. https://doi.org/10.1016/j.ress.2026.112774.
Ma, R., Zhang, S., 2024. Multiplier subsample bootstrap for statistics of time series. Journal of Statistical Planning and Inference 233, 106183. https://doi.org/10.1016/j.jspi.2024.106183.
Marques, C.H., Belchior, C.R.P., Caprace, J.D., 2018. Optimising the engine-propeller matching for a liquefied natural gas carrier under rough weather. Applied Energy 232, 187–196. https://doi.org/10.1016/j.apenergy.2018.09.155.
Meng, Q., Du, Y., Wang, Y., 2016. Shipping log data based container ship fuel efficiency modeling. Transportation Research Part B: Methodological 83, 207–229. https://doi.org/10.1016/j.trb.2015.11.007.
Moneta, A., Entner, D., Hoyer, P.O., Coad, A., 2013. Causal Inference by Independent Component Analysis: Theory and applications. Oxford Bulletin of Economics and Statistics 75 (5), 705–730. https://doi.org/10.1111/j.1468-0084.2012.00710.x.
Monter-Pozos, A., González-Estrada, E., 2024. On testing the skew normal distribution by using Shapiro–Wilk test. Journal of Computational and Applied Mathematics 440, 115649. https://doi.org/10.1016/j.cam.2023.115649.
Nachtigall, F., Wagner, F., Berrill, P., Creutzig, F., 2025. Built environment and travel: Tackling non-linear residential self-selection with double machine learning. Transportation Research Part D: Transport and Environment 140, 104593. https://doi.org/10.1016/j.trd.2025.104593.
Nguyen, P.-N., Woo, S.-H., Kim, H., 2022. Ship emissions in hotelling phase and loading/unloading in Southeast Asia ports. Transportation Research Part D: Transport and Environment 105, 103223. https://doi.org/10.1016/j.trd.2022.103223.
Nkikabahizi, C., Cheruiyot, W., Kibe, A., 2022. Chaining Zscore and feature scaling methods to improve neural networks for classification. Applied Soft Computing 123, 108908. https://doi.org/10.1016/j.asoc.2022.108908.

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. https://doi.org/10.1016/j.trc.2014.03.001.

Psaraftis, H.N., Lagouvardou, S., 2023. Ship speed vs power or fuel consumption: are laws of physics still valid? Regression analysis pitfalls and misguided policy implications. Cleaner Logistics and Supply Chain 7, 100111. https://doi.org/10.1016/j.clscn.2023.100111.

Ren, P., Jia, Q., Xu, Q., Li, Y., Bi, F., Xu, J., Gao, S., 2025. Oil spill Drift Prediction Enhanced by Correcting Numerically Forecasted Sea Surface Dynamic Fields with Adversarial Temporal Convolutional Networks. IEEE Transactions on Geoscience and Remote Sensing 63, 1–18. https://doi.org/10.1109/TGRS.2025.3528631.

Saber, S.M., Thowai, K.Z., Rahman, M.A., Hassan, M.M., Bari, A.B.M.M., Raihan, A., 2025. High-accuracy prediction of vessels' estimated time of arrival in seaports: a hybrid machine learning approach. Maritime Transport Research 8, 100133. https://doi.org/10.1016/j.martra.2025.100133.

Saydam, A.Z., Küçüksu, G.N., İnsel, M., Gökçay, S., 2022. Uncertainty quantification of self-propulsion analyses with RANS-CFD and comparison with full-scale ship trials. Brodogradnja: an International Journal of Naval Architecture and Ocean Engineering for Research and Development 73 (4), 107–129. https://doi.org/10.21278/brod73406.

Shahbazinia, A., Salehkaleybar, S., Hashemi, M., 2023. ParaLiNGAM: Parallel causal structure learning for linear non-Gaussian acyclic models. Journal of Parallel and Distributed Computing 176, 114–127. https://doi.org/10.1016/j.jpdc.2023.01.007.

Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A., Jordan, M., 2006. A linear non-Gaussian acyclic model for causal discovery. Journal of Machine Learning Research 7 (10). https://doi.org/10.5555/1248547.1248619.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P.O., Bollen, K., Hoyer, P., 2011. DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model. Journal of Machine Learning Research-JMLR 12 (Apr), 1225–1248. https://doi.org/10.17615/wr4c-0434.

Singh, N.K., Borrok, D.M., 2019. A Granger causality analysis of groundwater patterns over a half-century. Scientific Reports 9 (1), 12828. https://doi.org/10.1038/s41598-019-49278-8.

Sun, Q., Zheng, T., Zheng, X., Cao, M., Zhang, B., Jiang, S., 2023. Causal interpretation for groundwater exploitation strategy in a coastal aquifer. Science of the Total Environment 867, 161443. https://doi.org/10.1016/j.scitotenv.2023.161443.

Tang, X., Guo, R., Mo, Z., Fu, W., Qian, X., 2025. Causality-driven candidate identification for reliable DNA methylation biomarker discovery. Nature Communications 16 (1), 680. https://doi.org/10.1038/s41467-025-56054-y.

Theocharis, D., Rodrigues, V.S., Pettit, S., Haider, J., 2019. Feasibility of the Northern Sea Route: the role of distance, fuel prices, ice breaking fees and ship size for the product tanker market. Transportation Research Part E: Logistics and Transportation Review 129, 111–135. https://doi.org/10.1016/j.tre.2019.07.003.

Tillig, F., Mao, W., Ringsberg, J., 2015. Systems modelling for energy-efficient shipping. Chalmers University of Technology.

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. https://doi.org/10.1016/j.trd.2020.102389.

Wang, H., Yan, R., Wang, S., Zhen, L., 2023a. Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. Transportation Research Part C: Emerging Technologies 157, 104361. https://doi.org/10.1016/j.trc.2023.104361.

Wang, K., Li, Z., Zhang, R., Ma, R., Huang, L., Wang, Z., Jiang, X., 2025a. Computational fluid dynamics-based ship energy-saving technologies: a comprehensive review. Renewable and Sustainable Energy Reviews 207, 114896. https://doi.org/10.1016/j.rser.2024.114896.

Wang, K., Liu, X., Guo, X., Wang, J., Wang, Z., Huang, L., 2024. A novel high-precision and self-adaptive prediction method for ship energy consumption based on the multi-model fusion approach. Energy 310, 133265. https://doi.org/10.1016/j.energy.2024.133265.

Wang, K., Wang, J., Huang, L., Yuan, Y., Wu, G., Xing, H., Wang, Z., Wang, Z., Jiang, X., 2022. A comprehensive review on the prediction of ship energy consumption and pollution gas emissions. Ocean Engineering 266, 112826. https://doi.org/10.1016/j.oceaneng.2022.112826.

Wang, K., Wang, Y., Liang, H., Jing, Z., Cong, L., Ma, R., Huang, L., 2025b. Ship energy efficiency optimization considering the influences of multiple complex navigational environments: a review. Marine Pollution Bulletin 216, 117976. https://doi.org/10.1016/j.marpolbul.2025.117976.

Wang, L., Chen, B., Ouyang, J., Mu, Y., Zhen, L., Yang, L., Xu, W., Tang, L., 2025c. Causal-inference machine learning reveals the drivers of China's 2022 ozone rebound. Environmental Science and Ecotechnology 24, 100524. https://doi.org/10.1016/j.ese.2025.100524.

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. https://doi.org/10.1016/j.tre.2011.12.003.

Wang, X., Cao, W., Li, T., Feng, Y., Uğurlu, Ö., Wang, J., 2025d. An integrated multidimensional model for heterogeneity analysis of maritime accidents during different watchkeeping periods. Ocean & Coastal Management 264, 107625. https://doi.org/10.1016/j.ocecoaman.2025.107625.

Wang, X., Xia, G., Zhao, J., Wang, J., Yang, Z., Loughney, S., Fang, S., Zhang, S., Xing, Y., Liu, Z., 2023b. A novel method for the risk assessment of human evacuation from cruise ships in maritime transportation. Reliability Engineering & System Safety 230, 108887. https://doi.org/10.1016/j.ress.2022.108887.

Wang, X., Zhao, G., Cao, L., Yu, J., Li, Y., Wang, H., Cao, W., 2026. Structural causality hierarchy analysis: a novel data-driven integrated method for assessing occupational accidents in maritime transportation. Safety Science 196, 107086. https://doi.org/10.1016/j.ssci.2025.107086.

Xie, C., Hu, J., Vasdravellis, G., Wang, X., Cheng, S., 2024. Explainable AI model for predicting equivalent viscous damping in dual frame–wall resilient system. Journal of Building Engineering 96, 110564. https://doi.org/10.1016/j.jobe.2024.110564.

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. https://doi.org/10.1016/j.tre.2020.101930.

Yan, R., Wang, S., Psaraftis, H.N., 2021. Data analytics for fuel consumption management in maritime transportation: Status and perspectives. Transportation Research Part E: Logistics and Transportation Review 155, 102489. https://doi.org/10.1016/j.tre.2021.102489.

Yan, R., Yang, D., Wang, T., Mo, H., Wang, S., 2024. Improving ship energy efficiency: Models, methods, and applications. Applied Energy 368, 123132. https://doi.org/10.1016/j.apenergy.2024.123132.

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. https://doi.org/10.1007/s10479-019-03183-5.

Yao, Z.-M., Qian, Z.-Q., Li, R., Hu, E., 2019. Energy efficiency analysis of marine high-powered medium-speed diesel engine base on energy balance and exergy. Energy 176, 991–1006. https://doi.org/10.1016/j.energy.2019.04.027.

Yeh, C.-K., Lin, C., Shen, H.-C., Cheruiyot, N.K., Nguyen, D.-H., Chang, C.-C., 2022. Real-time energy consumption and air pollution emission during the transpacific crossing of a container ship. Scientific Reports 12 (1), 15272. https://doi.org/10.1038/s41598-022-19605-7.

Zhang, Q., Ma, Z., Cui, Z., 2025a. A causality-based explainable AI method for bus delay propagation analysis. Communications in Transportation Research 5, 100178. https://doi.org/10.1016/j.commtr.2025.100178.

Zhang, Y., Li, H., Ren, G., 2022. Quantifying the social impacts of the London Night Tube with a double/debiased machine learning based difference-in-differences approach. Transportation Research Part A: Policy and Practice 163, 288–303. https://doi.org/10.1016/j.tra.2022.07.015.

Zhou, J., Ren, J., Bai, W., 2023. Survey on hydrodynamic analysis of ship–ship interaction during the past decade. Ocean Engineering 278, 114361. https://doi.org/10.1016/j.oceaneng.2023.114361.

Zhu, S., Gu, Y., Yuan, H., Ma, Z., Deng, K., 2020. Thermodynamic analysis of the turbocharged marine two-stroke engine cycle with different scavenging air control technologies. Energy 191, 116533. https://doi.org/10.1016/j.energy.2019.116533.