Transportation Research Part C 188 (2026) 105659

Contents lists available at ScienceDirect

Transportation Research Part C

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

An integrated method of advanced optimisation and adaptive ensemble learning for ship fuel consumption prediction

Wenjie Cao [a,b], Xinjian Wang [a,b,c,*], Yaqing Shu [b], Huanhuan Li [b,d], Jingen Zhou [e], 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] School of Engineering, University of Southampton, Southampton SO17 1, UK
[e] Department of Operations Management, Supply Chain and Information Systems, KEDGE Business School, Marseille 13288, France

ARTICLE INFO

Keywords:
Maritime transport
Fuel consumption
Machine learning
Data fusion
Feature selection
Adaptive ensemble learning

ABSTRACT

Accurate prediction and interpretable analysis of Ship Fuel Consumption (SFC) are critical for optimising maritime operations and supporting decarbonisation efforts in maritime transport, yet existing approaches face significant challenges including limited model generalisation, redundant features due to multi-source data integration, and a lack of transparency in model outputs. These limitations stem from fragmented modelling pipelines that fail to holistically address the challenges of data heterogeneity, feature relevance, parameter tuning, and model adaptability in complex maritime environments. To address these challenges, this study develops an integrated framework that comprises advanced optimisation techniques with adaptive ensemble learning, structured through a synergised pipeline of four technical components. Firstly, multi-source data fusion employs spatio-temporal alignment to integrate ship noon reports, Automatic Identification System data, ECMWF Reanalysis v5, and Global Ocean Physics Analysis and Forecast data, to construct a comprehensive feature space. Secondly, a SHAP-based Weighted Feature Selection algorithm leverages multi-model SHapley Additive exPlanations (SHAP) value assessment with recursive feature elimination to identify and eliminate redundant features, enhancing model generalisation and prediction efficiency. Thirdly, a Hierarchical Adaptive Parameter Space Exploration algorithm integrates global random search and local grid search for efficient hyperparameter optimisation. Finally, an Adaptive Cluster-based Multi-Ensemble model incorporates data clustering and model fusion to capture operational heterogeneity and adaptively assign optimal weights across different data clusters. Comparative experiments demonstrate that the proposed model significantly outperforms six mainstream machine learning models and three classical ensemble methods across multiple evaluation metrics. Moreover, SHAP-based interpretability analysis quantifies feature contributions and reveals specific effects of value changes,

* Corresponding author at: Navigation College, Dalian Maritime University, Dalian 116026, PR China.
** Corresponding author at: Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK.
E-mail addresses: wangxinjian@dlmu.edu.cn (X. Wang), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.trc.2026.105659
Received 1 August 2025; Received in revised form 23 January 2026; Accepted 20 March 2026
Available online 2 April 2026

enhancing model transparency for decision support. This framework provides a robust technical solution for SFC prediction, offering reliable data-driven tools for energy efficiency management and sustainable maritime operations. The source code is publicly available at: https://github.com/AdvMarTech/ship_fuel_consum_predic.

1. Introduction

The global shipping industry is entering a critical phase of digital and environmental transformation (Cao et al., 2025c; Cao et al., 2026). As pressure mounts to meet international decarbonisation targets while controlling rising operational costs, fuel consumption prediction and optimisation have emerged as both an economic imperative and a regulatory requirement (Li et al., 2025d). Ship Fuel Consumption (SFC) accounts for a major proportion of total voyage costs—often exceeding 50%—and significantly contributes to greenhouse gas and sulfur oxide emissions from the maritime transport sector (Lan et al., 2025; Statista, 2024; Wang et al., 2018; Yan et al., 2024). In response to these challenges, regulatory frameworks such as the Energy Efficiency Existing Ship Index (EEXI), Carbon Intensity Indicator (CII), and Energy Efficiency Operational Indicator (EEOI) have been adopted by the International Maritime Organization (IMO), alongside regional mechanisms such as the European Union Emissions Trading System (EU ETS) and the FuelEU Maritime regulations for sustainable development (International Maritime Organization, 2025; Official Journal of the European Union, 2023). These policy shifts have heightened demand for advanced, data-driven tools that can support accurate, interpretable, and actionable SFC prediction.

The approaches developed to address this predictive task generally fall into three established paradigms: statistical and physics-based models, machine learning (ML) approaches, and hybrid data-driven frameworks (Li et al., 2025d; Wang et al., 2022). Statistical and physics-based models are typically built upon naval architecture principles and hydrodynamic formulations, which quantify SFC through explicit functional relationships involving vessel speed, draught, and hull resistance (Wang et al., 2025a). In contrast, ML approaches focus on extracting patterns and nonlinear dependencies from historical operational data, establishing predictive mappings without requiring a priori physical constraints (Li et al., 2024b; Wang et al., 2024). Hybrid data-driven frameworks represent an intermediary strategy, seeking to integrate deterministic physical laws with stochastic data analytics to enhance representational accuracy across diverse navigational environments (Wang et al., 2025b).

Recent research has increasingly focused on applying ML techniques to SFC prediction, owing to their superior nonlinear modelling capabilities and adaptability to diverse operational scenarios (Jesus et al., 2024; Wang et al., 2023; Yan et al., 2025). These methods benefit from access to rich and heterogeneous data sources—including ship noon reports, Automatic Identification System (AIS) data, and environmental datasets such as ECMWF Reanalysis v5 (ERA5) and Global Ocean Physics Analysis and Forecast (GOPAF)—allowing the construction of high-resolution spatio-temporal feature spaces (Luo et al., 2024; Yan et al., 2024). However, despite promising advancements, several core limitations remain: (1) traditional single-model approaches struggle to generalise across diverse ship types, sea states, and operating conditions; (2) multi-source data integration introduces feature redundancy and misalignment; (3) hyperparameter optimisation remains computationally intensive and suboptimal; and (4) the “black-box” nature of complex models limits transparency and practical adoption.

To address these challenges, this study presents a novel and integrated methodological framework that combines multi-source data fusion, advanced optimisation, and adaptive ensemble learning with a strong emphasis on model interpretability. This holistic design aligns with recent research trends that call for explainable and generalisable ML systems tailored to complex transport environments. This study makes four key contributions (C1-C4) encompassing both algorithmic innovations and novel methodological integrations:

C1: Integrated framework for high-precision SFC prediction.

State of the art: While it is well acknowledged that SFC is influenced by a multitude of interacting variables, existing methods often fall short in capturing this complexity. Single-model approaches lack robustness across operational regimes, and traditional ensemble methods use global static weights, failing to reflect local variations (Wang et al., 2022; Hu et al., 2025).

Contribution: This study proposes a unified data fusion architecture that aligns AIS, noon reports, ERA5, and GOPAF datasets through spatio-temporal referencing and interpolation. To address model generalisation, an Adaptive Cluster-based Multi-Ensemble (ACME) model is introduced, which applies unsupervised clustering to capture operational heterogeneity and dynamically assigns weights to base models across data clusters, significantly enhancing accuracy and robustness.

C2: Optimising high-dimensional feature space with SHAP-Weighted Feature Selection (SWFS).

State of the art: Multi-source integration improves representational richness but also introduces redundant and collinear features, increasing computational burden and overfitting risk. Existing feature selection (FS) approaches are often single-model-based and prone to bias (Li et al., 2025b).

Contribution: A novel adaptation of the SHapley Additive exPlanations (SHAP) framework, termed the SWFS algorithm, is developed, combining SHAP from multiple base models with recursive feature elimination and performance-weighted voting. This multi-model consensus approach ensures robust feature ranking, reduces dimensionality, and enhances model generalisation and interpretability.

C3: Efficient Hyperparameter Tuning via Hierarchical Adaptive Parameter Space Exploration (HAPSE).

State of the art: Hyperparameter optimisation (HPO) is critical for ML model performance but faces trade-offs between accuracy and computational cost. Grid search ensures thoroughness but is computationally prohibitive, while random search lacks directional guidance (Yang and Shami, 2020; Muñoz et al., 2025).

Contribution: A novel HAPSE algorithm is proposed to efficiently explore the parameter space. It integrates global random search for broad exploration and local grid search for fine-tuning, coupled with an adaptive space reduction mechanism to concentrate search efforts. This hybrid strategy significantly improves tuning efficiency without compromising performance.

C4: Deep model interpretation with a SHAP-based dual-layer framework.

State of the art: The opacity of high-performing ML models limits their adoption in safety–critical and regulation-sensitive applications. Most prior studies offer limited post hoc interpretability that is neither scalable nor actionable (Wang et al., 2023).

Contribution: This study introduces a dual-layer interpretability framework leveraging SHAP. At the global level, average absolute SHAP values quantify the importance of each feature across all samples; at the local level, SHAP value distributions offer instance-level insights into prediction mechanisms. This approach transforms the predictive model into a transparent decision-support tool, enhancing stakeholder trust and practical utility.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive literature review on SFC prediction and identifies principal research gaps. Section 3 details the proposed systematic framework, encompassing the multi-source data fusion process, the novel SWFS and HAPSE algorithms, and the architecture of the ACME model. Section 4 presents and analyses the experimental results, including a comparative performance evaluation of the ACME model, an in-depth analysis of its internal mechanisms, and a detailed feature-level interpretability study. Section 5 discusses the theoretical and practical implications of the findings for ship energy efficiency management. Finally, Section 6 summarises the key conclusions and outlines potential directions for future research.

2. Literature review

This study reviews the existing literatures in two aspects closely related to SFC prediction: (1) the research on SFC prediction using traditional approaches, and (2) the research on SFC prediction using advanced technologies. The purpose of this section is to systematically summarise the progress of existing research, identify current gaps, and establish a foundation for positioning the present study.

2.1. SFC prediction using traditional methods

SFC, as a key decision variable in route planning, not only directly influences the operating cost structure of shipping companies but is also closely linked to the increasingly stringent carbon emission regulations imposed by the IMO (Luo et al., 2025b). With the rapid advancement of data science and the growing availability of heterogeneous, multi-source ship data, researchers are increasingly turning to advanced data analytics to enhance the accuracy and reliability of predictive models (Luo et al., 2024; Yan et al., 2024). Existing research models in this domain can be broadly classified into two categories: white-box and black-box models. These two model types differ significantly in terms of construction principles, performance characteristics, and inherent limitations, all of which have been extensively examined in the existing literatures.

White-box models, typically developed based on statistical principles or ship physics mechanisms, offer the core advantage of transparent structures and parameters with clear physical interpretations, thus ensuring strong interpretability. For instance, Cepowski and Drozd (2023) established explicit quantitative relationships between SFC and key operational parameters such as speed and vessel load using formulations grounded in physical laws. Similarly, Adland et al. (2020) quantitatively analysed the effects of environmental factors, such as wind, waves, and currents, on SFC through statistical methods. The strength of such models lies in their ability to clearly explain the influence of input variables on output results, making it easy for shipping managers to understand the underlying mechanisms of SFC and optimise decision-making accordingly. However, white-box models are generally limited in capturing complex, nonlinear interactions among variables, particularly under variable sea states and dynamic operational conditions, and their prediction accuracy is typically lower than that of black-box models (Wang et al., 2022; Yan et al., 2021).

In contrast, black-box models primarily rely on ML algorithms and have gained widespread adoption in recent years due to their superior ability to model complex nonlinear relationships and enhance predictive accuracy. Various ML techniques have been employed by researchers to predict SFC. For example, Yan et al. (2024) applied Artificial Neural Networks (ANNs) to process ship noon report data, achieving accurate predictions of SFC rates and integrating these into a speed optimisation framework. Nguyen et al. (2023) utilised the XGBoost algorithm to capture the intricate nonlinear relationships between navigational states and meteorological variables, significantly improving prediction accuracy and model robustness. Uyanık et al. (2020) further validated the efficacy of complex ML modelling by comparing ensemble learning methods such as Random Forest and Gradient Boosting in SFC prediction tasks. Moreover, multi-source data fusion has emerged as a key approach to enhance model performance. Integrating ship AIS data, ERA5 reanalysis data, and traditional noon reports enables a comprehensive representation of both operational conditions and environmental contexts, thereby significantly improving model generalisation and predictive robustness (Li et al., 2022; Luo et al., 2024; Luo et al., 2025b; Yan et al., 2024). However, the limited interpretability of black-box models remains a major obstacle to their practical deployment. Their internal complex mechanisms make it difficult for domain experts and decision-makers in the maritime industry to intuitively understand the rationale behind the predictions, which undermines model credibility and limits its applicability in operational decision-making (Wang et al., 2023).

Given the respective advantages and limitations of white-box and black-box models, recent studies have begun to explore hybrid modelling strategies and Explainable Artificial Intelligence (XAI) techniques to achieve a more effective balance between predictive accuracy and interpretability. For instance, Zhang et al. (2024) attempted to incorporate domain knowledge into ML models to enhance their transparency. However, research aimed at improving model interpretability remains in its early stages and has yet to

establish a mature, systematic theoretical framework or a widely validated application model.

2.2. SFC prediction using emerging technologies

Although notable progress has been achieved in ML-based SFC prediction models, this field still faces several critical challenges in striving for higher accuracy, greater generalisation capability, and enhanced practical applicability. This underscores the urgent need to introduce and develop new advanced techniques to address these issues.

The inherently dynamic nature of ship operating environments presents a significant challenge to the generalisation capability of SFC prediction models. Hu et al. (2022) demonstrated that SFC patterns of the same ocean-going cargo ship vary substantially across different sea conditions and load states, making it difficult for a single model trained on global data to accurately capture and adapt to such complexity. To enhance predictive accuracy under varying conditions, ensemble learning frameworks have been extensively studied for their theoretical advantages in reducing model variance and improving stability. For instance, Hu et al. (2025) revealed that three core ensemble strategies, Voting, Stacking, and Blending, can mitigate the impact of stochastic errors on SFC predictions through comparative evaluation. However, as Iqbal et al. (2025) critically pointed out, traditional ensemble methods typically employing a global static weighting mechanism, often overlook potential local structures and differing operational modes within the data, and thereby fail to fully exploit the inherent heterogeneity.

To comprehensively characterise the operational status and environmental conditions of ships, researchers increasingly adopt strategies that fuse heterogeneous data from multiple sources (Luo et al., 2023). However, this data fusion inevitably leads to a substantial increase in feature dimensionality, resulting in the construction of high-dimensional feature spaces. Such high dimensionality not only significantly raises the computational complexity of model training but also introduces the risk of a “curse of dimensionality” and severe overfitting, both of which tend to compromise the generalisation performance of predictive models (Feng et al., 2024a). Although a richer feature set intuitively seems to provide more useful information, empirical studies have demonstrated that the relationship between the number of features and model performance is not monotonically positive. Wang et al. (2021) experimentally confirmed that once the number of features exceeds a certain threshold, prediction accuracy may actually decline. To investigate this phenomenon, Afshar and Usefi (2020) identified feature covariance, information redundancy, and noise interference, common in high-dimensional data, as key factors responsible for performance bottlenecks and degradation.

Beyond these established methodologies, the maritime sector's pursuit of innovation is evident in its active exploration of other advanced data science methods to address the challenges of SFC prediction. Transfer learning offers promising solutions to the problem of data scarcity. For example, Luo et al. (2025b) applied a pre-trained model to transfer knowledge from ship types with abundant data to those with limited samples, thereby significantly improving prediction accuracy in small-sample scenarios. Meanwhile, federated learning provides a viable approach to ensuring data privacy. Han et al. (2024) proposed a distributed SFC prediction framework based on federated learning, which enables multiple shipping companies to collaboratively train models without sharing raw data, thus achieving a balance between data utility and privacy protection.

2.3. Research gaps

Despite growing interest in applying ML techniques to SFC prediction, several critical research gaps continue to limit progress in achieving accurate, generalisable, and interpretable models for practical deployment:

(1) Lack of integrated frameworks that capture operational heterogeneity.

Current predictive approaches are often based on single-model architectures or traditional ensemble methods with globally fixed weighting schemes. These models struggle to generalise across diverse operational scenarios such as different ship types, voyage routes, weather conditions, and sea states. Moreover, existing ensemble learning strategies typically fail to exploit local data structures or account for the complex heterogeneity inherent in maritime operations, resulting in suboptimal accuracy and limited robustness.

(2) Redundant and high-dimensional feature spaces from multi-source data fusion.

While the integration of multi-source data, such as AIS, noon reports, and environmental datasets, enhances the descriptive power of input features, it also introduces substantial redundancy, multi-collinearity, and the curse of dimensionality. These issues increase computational complexity, hinder model interpretability, and negatively impact generalisation performance. Most existing feature selection (FS) methods rely on single-model evaluations, which are prone to bias and insufficiently robust for high-dimensional maritime datasets.

(3) Inefficient hyperparameter optimisation for complex ML models.

Hyperparameter tuning is essential for maximising ML performance, particularly for ensemble and deep learning models. However, conventional optimisation methods such as grid search or random search either incur prohibitive computational costs or fail to explore the parameter space effectively. Given the complexity and scale of maritime datasets, there is a pressing need for more efficient and adaptive optimisation strategies that can balance accuracy with computational feasibility.

(4) Limited interpretability of high-performing models.

While advanced ML models can achieve high predictive accuracy, their opaque, “black-box” nature limits practical adoption, especially in safety–critical or regulation-sensitive maritime environments. Existing studies tend to prioritise predictive performance without offering sufficient insights into the internal reasoning or decision pathways of the models. This lack of transparency undermines trust among stakeholders and restricts the models' utility for operational decision-making and policy compliance.

In light of these gaps, this study proposes a unified, multi-component framework that combines advanced data fusion, robust feature selection, efficient hyperparameter tuning, and deep model interpretability. By addressing these challenges in an integrated

manner, the proposed approach offers both methodological innovation and practical value for enhancing ship fuel efficiency and sustainable maritime operations.

3. Methodology

In this study, a systematic framework for SFC prediction and analysis is developed, integrating multi-source data fusion, weighted FS, adaptive hyperparameter optimisation, multi-model ensemble learning, and interpretability analysis. As illustrated in Fig. 1, the

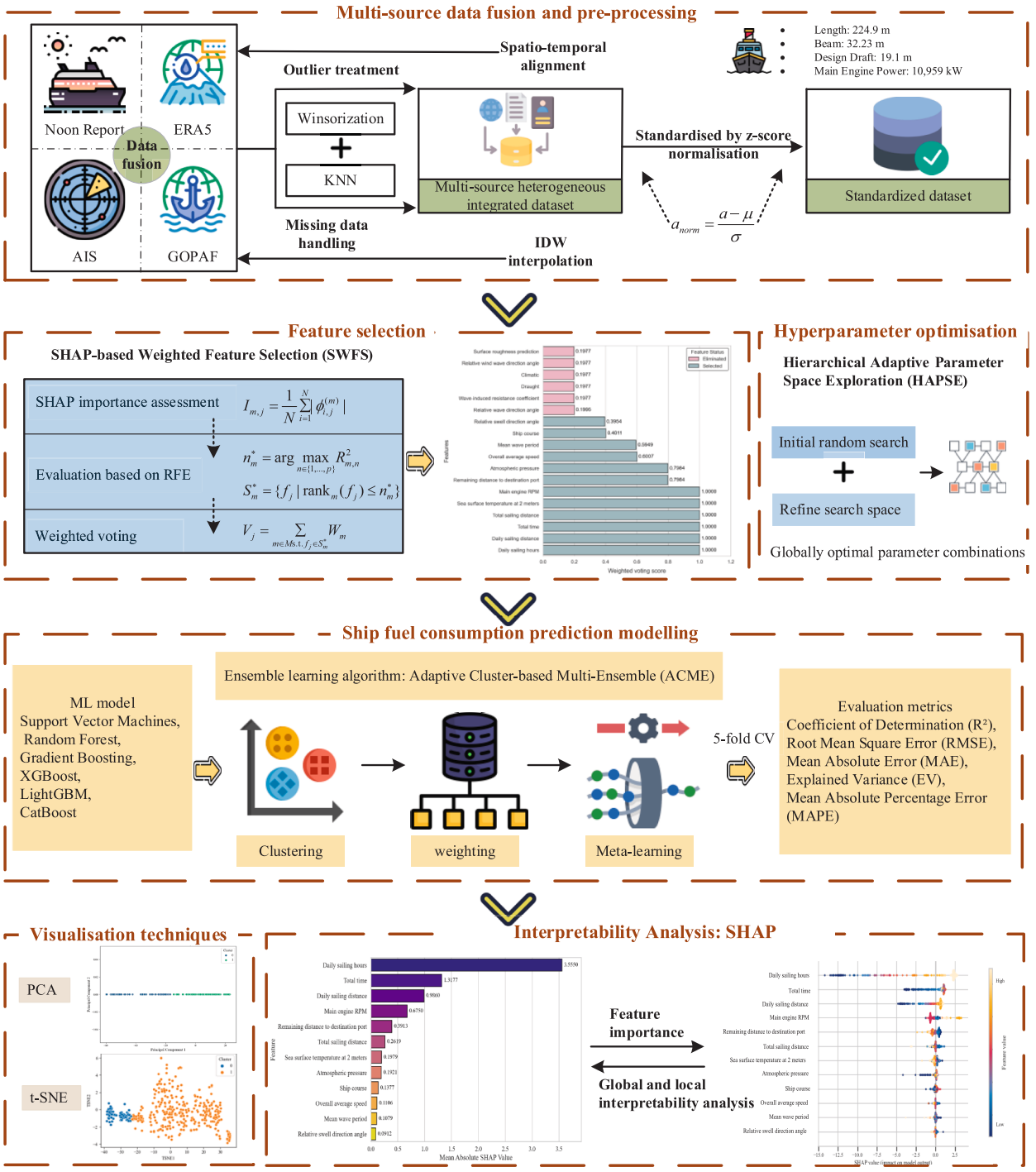


Fig. 1. The fuel consumption prediction framework in this study.

methodological framework encompasses the complete technical pipeline comprising novel algorithmic developments and the sophisticated integration of established methodologies.

3.1. Multi-source data fusion process

The SFC prediction model developed in this study is primarily based on the ship noon reports, supplemented and validated by multiple data sources, including AIS, ERA5, and GOPAF. To resolve the challenges of data heterogeneity and ensure high-fidelity inputs, a systematic four-step data fusion process was implemented: (1) Data acquisition and source identification; (2) Anomaly detection and outlier treatment; (3) KNN-based missing data imputation; and (4) Spatio-temporal synchronisation. The technical specifications and attributes of the data sources involved in this process are summarised in Table 1.

As a standardised document in the shipping industry, the ship's noon report records the vessel's daily operational status. The report includes key information such as the ship's navigational status and environmental meteorological parameters. Notably, the meteorological data are primarily recorded manually by the crew based on onboard instrument readings or visual observations, which introduces subjectivity and limitations, potentially affecting the accuracy and completeness of the data. This study also incorporates AIS data to supplement the ship's noon report. By capturing high-frequency positional information such as latitude and longitude, AIS data provides strong support for the spatial and temporal alignment between the noon reports and external meteorological data, thereby enhancing the accuracy and completeness of data fusion. To address the limitations caused by the low frequency and manual nature of meteorological data in the noon reports, the ERA5 dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) is introduced. ERA5 is widely regarded as the most accurate and comprehensive global climate reanalysis product, offering hourly estimates of atmospheric, ocean wave, and land surface conditions since 1979, which supports the study of environmental parameters at high spatial and temporal resolutions (Yan et al., 2024). Furthermore, to enhance the representation of marine environmental conditions, this study employs the GOPAF dataset, provided by the European Copernicus Marine Service and generated using the Mercator Ocean Analysis and Forecasting System (MOHFAS). The dataset offers 3D global ocean analyses and forecasts since 2020, providing a detailed foundation for analysing ship operational characteristics in dynamic marine environments (Cai et al., 2024; Luo et al., 2024; Yan et al., 2024).

In the data pre-processing stage, a systematic strategy for outlier detection and missing value imputation was implemented to ensure data integrity and reliability. For outlier identification, the Isolation Forest algorithm was employed. This method constructs an ensemble of decision trees based on randomly selected subspaces and evaluates the degree of abnormality by quantifying the path length required to isolate individual samples, effectively identifying observations that deviate significantly from the data distribution (Lesouple et al., 2021). For the flagged anomalies, the Winsorisation technique was applied to constrain extreme values by replacing those beyond pre-set quantile thresholds with the corresponding boundary values. This approach preserves the overall distribution characteristics while reducing the influence of outliers on subsequent analyses. Regarding missing data, the K-Nearest Neighbours (KNN) method was adopted for imputation. By leveraging the topological structure of the dataset and measuring sample similarity, missing values were estimated through weighted averages of neighbouring observations in the feature space. Compared to simple mean or median imputation, KNN better captures underlying data correlations and substantially improves interpolation accuracy (Memon et al., 2023). These outlier detection and missing value imputation techniques have been widely validated in ML applications, demonstrating their effectiveness in maintaining data quality and enhancing model robustness (Lesouple et al., 2021; Memon et al., 2023).

Following data pre-processing, a multidimensional data fusion framework is constructed based on spatio-temporal alignment of multiple heterogeneous data sources, including the ship's noon report, AIS, ERA5, and GOPAF datasets. Firstly, a standardised georeferencing system is employed to normalise each data source by establishing a unified temporal reference and spatial coordinate system. Secondly, Inverse Distance Weighting (IDW) interpolation is applied to harmonise the spatial resolutions across datasets, while temporal resampling techniques are used to standardise the time intervals. Finally, the high-precision positional information from AIS trajectory data serves as the primary reference for building a spatio-temporal index structure. This structure enables accurate alignment and association of navigational state parameters from the noon reports with meteorological variables from ERA5 and marine environmental data from GOPAF, thereby generating a comprehensive, multidimensional integrated feature vector. The proposed data fusion approach not only preserves the critical information and distinctive strengths of each heterogeneous data source but also effectively mitigates the inherent limitations of any single source through complementary integration. This process ensures the provision of high-quality, multidimensional data support for the subsequent development of the SFC prediction model. For further details on the fusion algorithm, readers are referred to the study by Luo et al. (2023). The multi-source data fusion feature sources and their corresponding feature names are provided in Table A of Appendix A.

The comprehensive multidimensional feature vector generated through this fusion process forms the foundation for the subsequent FS method, where the enriched feature space necessitates systematic dimensionality reduction to optimise model performance whilst maintaining predictive relevance.

3.2. Feature selection method

In the context of SFC prediction, although multi-source data fusion enriches the feature space, it often introduces challenges such as multi-collinearity among features, information redundancy, and varying predictive contributions across features. To address these challenges, this study proposes a FS method termed SHAP-based Weighted Feature Selection (SWFS). Building upon the multidimensional integrated dataset established in Section 3.1. This approach aims to optimise the input dimensionality, enhance prediction

Table 1
Technical details of the multi-source data fusion process.

Data Source	Key Attributes	Sampling Frequency	Preprocessing Techniques	Fusion & Synchronisation Method
Ship Noon Report	Main engine fuel consumption, main engine RPM, etc.	Daily (24-hour interval)	Isolation Forest, Winsorisation, KNN imputation	Ground-truth framework for SFC targets and navigational states
AIS Data	Latitude, longitude, etc.	High-frequency (Seconds to minutes)	Trajectory cleaning, KNN imputation	Primary spatio-temporal baseline for voyage alignment
ERA5 (ECMWF)	10 m wind speed, significant wave height, etc.	Hourly (0.25° spatial resolution)	Z-score normalisation, KNN imputation	Inverse Distance Weighting (IDW) based on AIS coordinates
GOPAF	Total ocean current speed, total tidal speed, etc.	Daily / Hourly (0.08° spatial resolution)	Z-score normalisation, KNN imputation	IDW interpolation based on AIS coordinates, followed by temporal resampling and synchronisation with noon-report timestamps.

accuracy, and, to some extent, provide a theoretical foundation for improving model interpretability.

The core of the SWFS method is to integrate SHAP evaluation results from multiple ML models, combined with recursive feature evaluation and a weighted voting mechanism, to comprehensively and objectively quantify each feature's importance, thereby effectively mitigating selection bias that may arise from relying on a single model. Specifically, different models excel at capturing diverse patterns and feature interactions in the data, leading to variability in their assessments of feature importance (Xie et al., 2025). By aggregating SHAP evaluations from multiple models, such as Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost, across various feature subsets and weighting them based on the prediction performance of each subset, a more stable and reliable feature importance ranking can be achieved. It is important to note that the ML models employed during the FS phase are carefully chosen solely to accurately estimate feature contributions through efficient integration with TreeExplainer for SHAP (Wang et al., 2023). These models serve exclusively as evaluation tools within the SWFS algorithm and are not directly linked to the SFC prediction models presented in Section 3.3. The pseudocode of the SWFS algorithm is presented in Table B of Appendix B, and its implementation procedure primarily comprises the following three steps:

Step 1: Individual model training and initial SHAP importance assessment.

A set of ML models with diverse architectures and learning mechanisms are selected and trained independently on the full set of training features to predict the target variables. For each trained model, SHAP values are calculated on the training data. The initial feature importance for each model is determined by calculating the average absolute SHAP value of each feature. The feature importance $I_{m,j}$ is calculated as shown in Eq. (1).

$$I_{m,j} = \frac{1}{Ntr} \sum_{i=1}^{N} \phi_{i,j}^{(m)} \tag{1}$$

where Ntr represents the number of training samples, and $\phi_{i,j}^{(m)}$ denotes the SHAP value of model m for the j^{-th} feature of the i^{-th} sample. Based on this metric, all features are ranked in a descending order to produce the initial feature importance list $Rank_m$ for each model m. The detailed principles and calculation methods of the SHAP value, including the definition of the base value, are elaborated in Section 3.5.

Step 2: Performance evaluation based on recursive evaluation

For each model m, recursive feature evaluation is conducted. Based on its ranked feature importance list $Rank_m$, subsets $F_{m,n}$ are sequentially constructed by selecting the top n features, where n ranges from 1 to P' (the total number of input features). Each feature subset $F_{m,n}$ is used to train the corresponding model m (reinitialized to prevent information leakage) on the respective training data subset. The model's predictive performance is then evaluated using 5-fold cross-validation to identify the number of features that maximise the coefficient of determination R^2, denoted as n_m^*. The R^2 metric will be described in more detail in Section 3.3. The procedure is illustrated in Eqs. (2) and (3).

$$n_m^* = \arg \max_{n \in \{1,\ldots,P'\}} R_{m,n}^2 \tag{2}$$

$$S_m^* = \left\{ f_j \middle| \mathrm{rank}_m\left(f_j\right) \le n_m^* \right\} \tag{3}$$

where S_m^* denotes the optimal subset of features, and f_j represents the j^{-th} feature in the model. When $R_{m,max}^2 = R_{m,n_m^*}^2$, it is considered the best performance achieved by model m in the FS task and is used as the weight in the subsequent weighted voting procedure.

To address potential scale-induced dominance and improve fairness, sum normalisation was applied to the model performance scores. Specifically, the raw performance score was first defined as $W_m^{\mathrm{raw}} = R_{m,\max}^2$, and then the normalised voting weight was calculated as Eq. (4):

$$\widetilde{W}m = \frac{W_m^{\mathrm{raw}}}{\sum_{k=1}^{M} W_k^{\mathrm{raw}}} = \frac{R_{m,\max}^2}{\sum_{k=1}^{M} R_{k,\max}^2} \tag{4}$$

where M denotes the number of base models and $\sum_{m=1}^{M} \widetilde{W}_m = 1$.

Step 3: Feature-weighted voting aggregation

The optimal subsets of features selected by all models are aggregated to construct a global feature voting counter. For each feature that appears in any of these subsets, its final weighted voting score is calculated by summing the performance weights of all models that selected it. This process is formally expressed in Eq. (5).

$$V_j = \sum_{m \in \mathcal{M}, f_j \in S_m^*} \widetilde{W}_m \tag{5}$$

where V_j represents the final voting score, $\mathcal{M} = \{1,2,\ldots,M\}$, denotes the index set of base models, while M denotes the number of base models. $\widetilde{W}_m$ denotes the sum-normalised performance weight of model m. The resulting score V_j reflects the aggregated importance of feature f_j from the perspective of multiple high-performing models. Based on the calculated weighted voting scores V_j, all participating features are ranked in descending order to produce the final global feature importance ranking list $Rank_{final}$.

In this study, features with a voting score $V_j > \tau$ are retained as the final set of optimal features for subsequent modelling and analysis. where $\tau = \max_m \widetilde{W}_m$ denotes the maximum single-model contribution under the normalised weighting scheme. The rationale is that, because each model can contribute at most $\widetilde{W}_m$ to any feature, a score exceeding τ implies that the feature must be supported by at least two base models. This criterion significantly increases confidence in the importance and consistency of the selected features across models, suggesting that these features possess strong predictive value from diverse algorithmic perspectives.

The optimal feature subset identified through the SWFS methodology provides the refined input space for the subsequent SFC prediction modelling phase, ensuring that only the most predictive and stable features are utilised in model training and hyper-parameter optimisation.

3.3. SFC prediction modelling

3.3.1. Selected baseline ML models for SFC prediction

In the field of SFC prediction, data-driven ML methods have gained significant attention due to their strong nonlinear modelling capabilities and effective extraction of latent relationships within high-dimensional complex data (Cai et al., 2024; Fan et al., 2022;

Table 2
Comparative analysis of baseline ML models for SFC prediction.

Model	Advantages	Limitations	Reference
Support Vector Regression	• Effective for high-dimensional feature spaces • Robust against overfitting • Capable of capturing complex non-linear relationships • Memory efficient for datasets with moderate sample size • Well-suited for datasets with clear marginal separation	• Computational complexity scales poorly with sample size • Sensitive to feature scaling • Hyperparameter tuning is complex and time-consuming • Performance deteriorates with large datasets • Lacks interpretability	(Wang et al., 2026; Uemoto and Naito, 2022)
Random Forest	• Robust to outliers and noise • Handles high-dimensional data effectively • No need for feature scaling • Less prone to overfitting compared to decision trees • Built-in feature importance assessment • Inherent parallelisation capability	• Reduced interpretability compared to single decision trees • Suboptimal performance in extreme operational conditions • Underestimate high SFC events • Computationally demanding for very large datasets • Less effective for extrapolation beyond training data range	(Borup et al., 2023; Li et al., 2025c; Liu et al., 2026)
Gradient Boosting	• High prediction accuracy • Effective handling of heterogeneous features • Automatic feature interaction detection • Robust to irrelevant features • Flexible loss function selection	• Sequential training process limits parallelisation • Prone to overfitting without proper regularisation • Long training times for complex datasets • Sensitive to noisy data and outliers • Requires careful hyperparameter tuning	(Guillen et al., 2023; Li et al., 2025c; Sheikhmohammadi et al., 2025)
XGBoost	• Superior predictive performance • Built-in regularisation mechanisms • Efficient handling of sparse data • Parallel processing capabilities • Built-in cross-validation functionality • Effectively handles missing values	• Complex hyperparameter tuning process • High memory consumption for large datasets • Less intuitive model interpretation • Overfit with improper regularisation parameters • Computationally intensive hyperparameter optimisation	(Han et al., 2024; Liu et al., 2026; Luo et al., 2023)
LightGBM	• Exceptionally fast training speed • Lower memory utilisation • High efficiency for large-scale datasets • Leaf-wise tree growth strategy optimises accuracy • Supports categorical features natively • GPU acceleration capabilities	• Less stable on small datasets • Sensitive to outliers • Overfit on small-to-medium datasets without constraints • Less effective with high sparsity in features • Requires careful parameter tuning to avoid overfitting	(Alsulamy, 2025; Feng et al., 2024a; Ni et al., 2024)
CatBoost	• Superior handling of categorical features • Robust predictive performance • Less prone to overfitting • Minimal hyperparameter tuning required • Built-in handling of missing values • Effective with heterogeneous feature types	• Slower training speeds compared to LightGBM • High memory consumption during training • Computational resource intensity with numerous features • Limited parallelisation capabilities • Reduced efficiency for extremely large datasets	(Alsulamy, 2025; Zhang et al., 2020)

Yan et al., 2024). Given that SFC is influenced by heterogeneous multi-source features under real operating conditions, traditional linear models often struggle to capture the underlying mechanisms accurately. Therefore, constructing nonlinear supervised learning models with robust generalisation ability is particularly important. In the development of the novel ensemble framework proposed herein, the selection of appropriate base learners is a critical first step. This study selects six widely used supervised learning models, Support Vector Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost, as base predictors to comprehensively evaluate their predictive performance in SFC modelling, thereby establishing a solid foundation for subsequent model fusion and optimisation. Each model offers distinct advantages in terms of modelling approach, feature handling, and computational efficiency. A comparative overview of their applicability and limitations is provided in Table 2. Since advanced research has demonstrated the potential of combining models into bespoke hybrid structures (Zhu et al., 2025), the approach of this study is to begin with these fundamental, well-understood components. This strategy ensures a controlled and interpretable experimental setup, allowing for a rigorous assessment of how the ACME framework dynamically leverages the unique predictive capabilities of each base learner and thereby enabling the specific contribution of the ACME architecture itself to be isolated.

3.3.2. Dataset partitioning and cross-validation strategy

In constructing a SFC prediction model, a scientifically sound data partitioning strategy combined with a systematic cross-validation approach is crucial to ensure the reliability and validity of model evaluation. This study adopts data processing methods consistent with best practices in the field to guarantee objective and accurate assessment of model performance.

In this study, the original dataset is divided into training and testing sets using an 80%/20% split, following widely accepted ML practices. This partitioning ratio has been extensively validated in the field of SFC prediction (Lan et al., 2025; Luo et al., 2023). The training set is utilised for model parameter learning and optimisation, while the independent test set is reserved for evaluating the model generalisation ability. This allocation ensures sufficient training data while maintaining an adequate test sample size for statistically robust validation. To guarantee experimental repeatability and result reproducibility, the dataset split is performed with a fixed random seed, adhering both to ML standards and the fundamental principles of rigorous scientific research (Gundersen et al., 2022).

To optimise model training efficiency and enhance prediction accuracy, this study applies standardised preprocessing of input features. The standardisation is performed using the Z-score method (Ahmed Ouameur et al., 2020), as defined in Eq. (6).

$$a'_{norm} = \frac{a' - \mu}{\sigma} \tag{6}$$

where a'_{norm} represents the normalised feature value, a' is the original feature value, and μ and σ denote the mean and standard deviation of the feature, respectively. Through the standardisation process described in Eq. (6), each input feature is transformed into a standard normal distribution with a mean of 0 and a standard deviation of 1.

Normalisation is particularly important in SFC prediction due to the significant differences in magnitude and numerical ranges among input features such as speed, wind speed, air pressure, and wave height. To systematically assess model stability and minimize the impact of random variability, this study employed a rigorous 5-fold cross-validation strategy on the training dataset (Feng et al., 2024a). This approach significantly enhances the reliability of model evaluation, reduces potential bias from single random splits, and provides a robust empirical basis for model selection and optimisation (Feng et al., 2024a).

3.3.3. Hierarchical adaptive parameter space exploration (HAPSE)

In the construction of SFC prediction models, hyperparameter optimisation (HPO) plays a crucial role in enhancing model performance. Among traditional HPO methods, such as grid search (GS) and random search (RS), each has distinct advantages; however, a trade-off exists between efficiency and accuracy when applied to high-dimensional parameter spaces. GS offers comprehensive exploration but suffers from exponentially increasing computational complexity as the number of parameters grows (Yang and Shami, 2020). Conversely, RS is more efficient but lacks systematic coverage of the parameter space (Muñoz et al., 2025). To address these limitations, this study proposes the Hierarchical Adaptive Parameter Space Exploration (HAPSE) method, which integrates the strengths of GS and RS by employing a hierarchical search structure combined with an adaptive parameter space reduction strategy, thereby efficiently identifying globally optimal parameter combinations.

The overall design of HAPSE employs a two-stage search strategy. Initially, global random sampling is used to rapidly identify potential high-performance regions. Subsequently, a local grid-based fine-grained search is conducted within these regions, enabling a progressive exploration of the parameter space from coarse to fine resolution. Concurrently, the method incorporates an adaptive parameter space reduction mechanism that dynamically adjusts the local search boundaries based on preliminary results. By setting a space compression factor, HAPSE effectively concentrates the search on subspaces likely to contain the optimal solution, thereby improving computational efficiency while minimising the risk of missing the global optimum. Furthermore, HAPSE features a multi-scale parameter mapping system that accurately enables a transform between the standardised search space and the model's actual parameter space. This ensures optimal tuning of hyperparameters of varying scales within a unified framework, enhancing the method's adaptability and robustness for complex prediction tasks. To further increase computational efficiency, the method supports parallelisation of parameter evaluations, fully leveraging multi-core processor resources, making it particularly well-suited for computationally demanding applications, e.g. SFC prediction. The pseudocode of the HAPSE algorithm is presented in Table C.1 of Appendix C, where a detailed performance evaluation and discussion are also provided. Furthermore, the optimised hyperparameter configurations resulting from applying HAPSE to the six ML models used in this study are detailed in Table C.2 of Appendix C. The

HAPSE implementation procedure primarily comprises the following two steps:

Step 1: Initial random search stage.

Firstly, an initial hyperparameter search space $\mathscr{S}$ is defined. Subsequently, n_{random} sets of hyperparameter combinations $\left\{p_1, p_2, \ldots, p_{n_{\text{random}}}\right\}$ are randomly sampled from this space. For each parameter set p_i, the model performance is evaluated on the training and validation sets using cross-validation to obtain the corresponding evaluation error $\mathscr{E}(p_i),$. The process of determining the parameter combinations with optimal performance at the current stage is formalized in Eqs. (7), (8) and (9).

$$\mathscr{S} = \{[l_1, u_1], [l_2, u_2], \ldots, [l_d, u_d]\} \tag{7}$$

$$\mathscr{E}(p_i), = \text{Eval}(\text{model}, p_i, X_{\text{train}}, y_{\text{train}}, X_{\text{test}}, y_{\text{test}}) \tag{8}$$

$$p^* = \underset{p_i}{\text{argmin}}\, \mathscr{E}(p_i), \tag{9}$$

where n_{random} represents the number of random samples, d denotes the hyperparameter dimensionality, l_j and u_j are the lower and upper bounds of the j^{th} hyperparameter, respectively, and p^* indicates the optimal solution obtained from the initial random search.

Step 2: Refine search space stage.

Based on the optimal solution p^* obtained in the initial stage, a further local search is conducted. Centred on p^*, the search space is reduced by a scaling factor $\alpha \in (0, 1)$ to define the local parameter space $\mathscr{S}'$. Each parameter dimension within this space is divided into n_{grid} equidistant points, forming a gridded hyperparameter set $G \subset \mathscr{S}'$. For each hyperparameter combination $p_k \in G$, the model performance is evaluated again, and the combination achieving the best overall performance in the global context is selected. This process is illustrated in Eqs. (10), (11) and (12).

$$\left[l'_q, u'_q\right] = \left[\max\left(l_q, p_q^* - \alpha(u_q - l_q)\right), \min\left(u_q, p_q^* + \alpha(u_q - l_q)\right)\right] \tag{10}$$

$$\mathscr{E}(p_k) = \text{Eval}(\text{model}, p_k, X_{\text{train}}, y_{\text{train}}, X_{\text{test}}, y_{\text{test}}) \tag{11}$$

$$p^{**} = \underset{p_k}{\text{argmin}}\, \mathscr{E}(p_k) \tag{12}$$

where α represents the search space contraction factor, n_{grid} denotes the number of divisions per dimension in the refined search, and $\left[l'_q, u'_q\right]$ defines the new search range for the q^{th} parameter. The evaluation error after model training and validation for each hyperparameter combination p_k is indicated by $\mathscr{E}(p_k)$, while p^{**} corresponds to the finalized globally optimal hyperparameter combination.

Concretely, HAPSE parallelises the evaluation of candidate hyperparameter configurations: each call to $\text{Eval}\left(\text{model}, p_i, X_{\text{train}}, y_{\text{train}}, X_{\text{test}}, y_{\text{test}}\right)$ in the random-search stage (Eq. (7)) and each evaluation of p_k in the refined-grid stage is mutually independent, and can therefore be dispatched to separate CPU workers and executed concurrently. After all workers return their evaluation errors $\{\mathscr{E}(p)\}$, HAPSE performs a lightweight reduction step (Eqs. (8) and (11)) to select the best configuration via $\underset{p_k}{\text{argmin}}$. As shown in Table C.2 (Appendix C), the wall-clock time is reduced because a large set of independent configuration evaluations, each involving model training and validation, can be executed concurrently across multiple CPU cores, rather than sequentially. In practice, the runtime scales approximately with the number of available cores until the serial fraction and parallel overhead (e.g., task scheduling and inter-process communication) become non-negligible, in line with Amdahl's law (Schryen, 2024). In implementation, we adopt configuration-level (outer-loop) parallelism as the default, while controlling model-internal multithreading to avoid nested parallelism/CPU oversubscription, consistent with standard practice in widely used ML toolchains (Colagrande and Benini, 2025; Liu et al., 2025).

Importantly, in this study α is treated as a fixed, user-defined contraction factor and is set to $\alpha = 0.20$ for all models and all hyperparameter dimensions. The factor itself is neither learned nor adjusted based on model type, and it is not explicitly tuned as a function of the hyperparameter dimensionality d. Instead, Stage-2 refinement contracts each hyperparameter interval independently using Eq. (10), where the absolute contraction width is $\alpha(u_q - l_q)$. Therefore, any apparent model-dependent behaviour arises indirectly from model-specific original bounds $\left[l'_q, u'_q\right]$, while the dimensionality affects only the resulting refined hypervolume rather than the value of α itself. A dedicated sensitivity analysis is provided in Appendix C (Fig. C2), showing that $\alpha = 0.20$ achieves the minimum MSE or falls within the within-2% low-MSE zone across all six base learners, supporting the robustness of this setting.

The hyperparameter configurations optimised through HAPSE ensure that each base model achieves its maximum predictive potential, thereby establishing the foundation for effective ensemble learning where well-tuned individual models contribute complementary strengths to the overall prediction framework.

3.3.4. *Ensemble learning algorithm for prediction enhancement*

SFC prediction is a complex nonlinear regression task, where a single ML model often struggles to capture all relevant features and patterns in the data. Model fusion, a fundamental approach in ensemble learning, enhances prediction stability and accuracy by

combining the strengths of multiple base models. The theoretical foundation of model fusion primarily stems from statistical learning theory and information fusion theory (Wang et al., 2025c; Zadeh et al., 2020). According to the bias-variance decomposition, integrating several high-quality, weakly correlated models can substantially reduce prediction variance while maintaining low bias, thereby improving overall predictive performance. This principle is mathematically expressed in Eq. (13):

$$E\left[(y-\widehat{f}(x))^2\right] = (Bias[\widehat{f}(x)])^2 + Var[\widehat{f}(x)] + \sigma^2 \tag{13}$$

where $\widehat{f}(x)$ represents the predicted value from the model, y denotes the actual observation, and σ^2 is the non-commensurable error.

In SFC prediction, influencing factors are diverse and complex, encompassing ship technical parameters, navigational environmental conditions, operational factors, and their intricate interactions (Wang et al., 2023). Different ML models exhibit distinct strengths in handling these complexities: Support Vector Regression excels in modelling high-dimensional nonlinear relationships; Random Forest is robust to outliers and accommodates various feature types; while gradient boosting algorithms (e.g., XGBoost, LightGBM, and CatBoost) effectively capture complex decision boundaries (Chen et al., 2019; Yu et al., 2020). Consequently, model fusion can leverage these complementary advantages to enhance prediction robustness and accuracy. Traditional fusion methods primarily include Voting, Stacking, and Blending. However, these approaches often assume uniform data distribution or apply globally consistent weights during fusion, overlooking structural heterogeneity within the data and regional differences in model performance (Cao et al., 2025a,b; Wang et al., 2025d; Xie et al., 2025; Iqbal et al., 2025). For SFC, data samples may cluster into distinct groups representing different operating states or environmental conditions, where a global fusion strategy may fail to optimally integrate predictions from individual models.

To address this challenge, this study proposes an Adaptive Cluster-based Multi-Ensemble (ACME) model. The core innovation of ACME lies in integrating data clustering with model fusion to achieve region-specific accurate fusion by adaptively assigning different model weights to distinct data clusters. The method first extracts distributional features for prediction, then clusters samples into groups based on these features, optimises fusion weights for each group individually, and finally refines the fusion outcomes through *meta*-learning. The pseudocode of the ACME algorithm is provided in Table D of Appendix D. The mathematical formulation of ACME is as follows:

Given N training samples $(X_i, y_i)_{i=1}^{N}$ and M base models $f_{jj=1}^{M}$, each model f_j produces a predicted value $\widehat{y}_{ij}$ for sample X_i. The objective of ACME is to identify the optimal fusion function F that minimizes the prediction error relative to the true value. This process is described by Eq. (14).

$$\widehat{y}_i = F(\widehat{y}_{i1}, \widehat{y}_{i2}, \ldots, \widehat{y}_{iM}) \tag{14}$$

where X_i denotes the feature vector of the i^{-th} training sample, y_i represents the true target value of the i^{-th} training sample, and $\widehat{y}_i$ is the final fused prediction for the i^{-th} sample produced by the fusion function F.

ACME categorizes the aforementioned model into three hierarchical levels:

1) Clustering levels: Features are extracted from the model predictions, and the samples are clustered into K groups based on these features. For each sample i, the cluster assignment is denoted by $c_i \in 1,2,\ldots,K$. A feature vector z_i is defined for each sample, comprising statistical features z_i^{stat} and correlation features z_i^{corr}. The mathematical formulations are presented in Eqs. (15) and (16).

$$z_i^{stat} = [\mu_i, \sigma_i, R_i, med_i, q1_i, q3_i] \tag{15}$$

$$z_i^{corr} = \left[\rho_{c_i,j,r}\right]_{1\le j<r\le M} \tag{16}$$

where μ_i denotes the mean of the model predictions, σ_i is the standard deviation, R_i is the extreme deviation, and med_i is the median. Parameters $q1_i$ and $q3_i$ represent the 25th and 75th percentiles, respectively. $\rho_{c_i,j,r}$ denotes the predicted correlation coefficient between model j and r within the cluster c_i which sample i belongs.

Based on the feature vector z_i, the samples are clustered into K groups using the K-means algorithm. To make the selection of K transparent and reproducible, the optimal number of clusters K^* is evaluated by maximising the average silhouette coefficient $S(K)$. Specifically, candidate values $K \in \{2, 3, \ldots, K_{\max}\}$ are evaluated, where $K_{\max} = \min(10, N-1)$ and N is the number of training samples used for clustering. For each candidate K, K-means is repeated under multiple random initialisations, and the run that achieves the largest $S(K)$ is retained. The final choice is $K^* = \arg\max_{K\in\{2,\ldots,K_{\max}\}} S(K)$, and when several K values yield indistinguishable $S(K)$, the smaller K is selected to favour parsimony and avoid over-fragmentation.

To ensure each cluster contains a sufficient number of samples for reliable weight optimisation, clusters with fewer samples than a predefined threshold θ_{min} are merged into the nearest larger cluster. In this study, $\theta_{\min} = 5$, consistent with Algorithm D.3 in Appendix D (Here, $\theta_{\min} = 5$ is adopted as a pragmatic minimum cluster-size constraint to avoid spurious micro-clusters and to stabilise the cluster-specific statistics used for fusion-weight optimisation, consistent with common practice in clustering methods that explicitly enforce a minimum cluster size (Xin et al., 2025)). The corresponding process is formulated in Eqs. (17, 18, 19, 20 and 21).

$$c_i = \text{argmin}_{k\in 1,2,\ldots,K} |z_i - \mu_k|^2 \tag{17}$$

$$S(K) = \frac{1}{N}\sum_{i=1}^{N}\frac{b(i)-a(i)}{\max a(i), b(i)} \tag{18}$$

$$d(c', c) = \|\boldsymbol{\mu}c - \boldsymbol{\mu}c'\| \tag{19}$$

$$a(i) = \frac{1}{|C_{c_i}| - 1}\sum_{j \in C_{c_i}} d(\boldsymbol{z}_i, \boldsymbol{z}_j) \tag{20}$$

$$b(i) = \min_{k \neq c_i}\frac{1}{|C_k|}\sum_{j \in C_k} d(\boldsymbol{z}_i, \boldsymbol{z}_j) \tag{21}$$

where μ_k denotes the centroid of the k^{-th} cluster, $a(i)$ represents the average intra-cluster distance of sample i from other samples within the same cluster, and $b(i)$ indicates the average distance of sample i to the nearest distinct cluster. c denotes a cluster index, and $\boldsymbol{\mu}c$ denotes the centroid of cluster c. The process selects the number of clusters corresponding to the maximum silhouette coefficient value as the optimal choice. Additionally, the value K that yields the largest silhouette coefficient $S(K)$ is selected as the optimal number of clusters. C_k denotes the set of samples in cluster k, $|C_k|$ is its size, and $d(\cdot,\cdot)$ is the Euclidean distance in the clustering feature space.

For the small-cluster merging step, $d(c', c)$ in Eq. (19) measures the Euclidean distance between centroids of a small cluster c' and a candidate larger cluster c; samples in c' are reassigned to the nearest c (with size $\geq \theta_{min}$) and the centroids are updated iteratively until all clusters satisfy the size constraint.

2) Integration levels: The objective of this phase is to determine the optimal set of fusion weights for each data cluster k identified in the previous steps. These weights are used to integrate predictions or information from M distinct sources. By optimising the fusion weights, the aim is to achieve a final prediction that is more accurate than any individual source. The procedure for computing the optimal fusion weight vector $\boldsymbol{wk} \in \{w_{k1}, w_{k2}, \ldots, w_{kM}\}$ is presented in Eqs. (22) and (23).

$$\boldsymbol{w}_k = \underset{\boldsymbol{w}_k}{\operatorname{argmin}}\sum_{i:c_i=k}\left(y_i - \sum_{j=1}^{M} w_{kj}\widehat{y}_{i,j}\right)^2 \tag{22}$$

$$\text{s.t.}\sum_{j=1}^{M} w_{kj} = 1, \quad w_{kj} \geq 0, \quad j = 1, \ldots, M \tag{23}$$

where $\boldsymbol{w}_k$ is the cluster-specific fusion weight vector; $\widehat{y}_{i,j}$ is the prediction of base model j for sample i; y_i is the ground truth; and w_j is the weight assigned to the j^{-th} model. Eq. (23) defines the constraints that must be satisfied, ensuring that the calculated weight vector w_j constitutes a valid and efficient proportional allocation.

3) Meta-learning levels: In this stage, rather than directly utilising the original base model predictions, the model employs the base fusion results derived in the previous step, along with selected statistical characteristics of these base predictions, to construct a new and more informative *meta*-feature vector for each data point i. This *meta*-feature vector is then used as the input to a new ML model, which generates the final predicted output. The mathematical expressions for this process are provided in Eqs. (24), (25) and (26).

$$\boldsymbol{mi} = \left[y_i^{base}, \sigma_i, \min_j \widehat{y}ij, \max_j \widehat{y}ij\right] \tag{24}$$

$$y_i^{base} = \sum_{j=1}^{M} w_{c_i j}\widehat{y}_{ij} \tag{25}$$

$$y_i^{final} = g(\boldsymbol{m}_i) \tag{26}$$

where $\boldsymbol{mi}$ denotes the *meta*-feature vector of the i^{-th} data point, which serves as the input to the *meta*-model g. The *meta*-feature vector includes y_i^{base}, the base fusion prediction of the i^{-th} data point, which is the weighted average prediction calculated based on the optimal weight w_{c_i} of the cluster c_i to which it belongs. This represents the best fused estimate. σ_i is the standard deviation of the base predictions, while $\max_j \widehat{y}ij$ and $\min_j \widehat{y}ij$ represent the maximum and minimum predicted values, respectively. Finally, y_i^{final} is the final predicted value of the i^{-th} data point.

3.3.5. Performance evaluation metrics

In constructing the evaluation system for SFC prediction models, the selection of scientifically sound and appropriate performance metrics is crucial for a comprehensive and objective assessment of both prediction accuracy and generalisation capability (Shu et al., 2024). Considering the specific characteristics of SFC prediction and the implementation requirements of the proposed HAPSE-based HPO method and ACME model, this study develops a multi-dimensional cross-validation evaluation framework. The framework systematically evaluates the model across multiple dimensions, including accuracy, bias, and stability. Specifically, it incorporates five core metrics: Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute

Percentage Error (MAPE), and Explained Variance (EV) (Kolassa, 2020; Yan et al., 2024).

The R^2 quantifies the proportion of variance in the dependent variable explained by the model; of which the values closer to 1 indicate stronger predictive performance (Feng et al., 2024b). RMSE, an absolute error metric, is sensitive to outliers and effectively evaluates model accuracy under extreme conditions (Duan et al., 2025). MAE offers a more robust error measure that is less influenced by outliers and visually represents the average prediction bias (Li et al., 2024a). EV assesses the variance component of prediction errors, providing insight into the stability of model predictions (Luo et al., 2025a). MAPE, a relative error metric, mitigates the impact of absolute value differences and is particularly suitable for evaluating prediction consistency across varying vessel operating conditions (Li et al., 2025a).

3.4. Visualisation techniques for model diagnostics

To interrogate the internal behaviour of the ACME model, specifically the efficacy of its data clustering stage, a set of post hoc diagnostic tools is required. This analysis is distinct from the predictive pipeline and serves to provide visual evidence for the model's mechanistic assumptions. For this purpose, a complementary pairing of Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE) is selected as the most suitable technical approach (Zheng et al., 2025). Importantly, these dimensionality-reduction techniques are used solely for post hoc visual diagnosis, and do not contribute to the clustering procedure itself, nor to the determination of the number of clusters. The cluster partition is obtained in the original fused feature space, and the optimal cluster number K^* is selected quantitatively by maximising the average silhouette coefficient $S(K)$ (Xin et al., 2025; Yue et al., 2026), with multiple random initialisations as described in Section 3.3.4. Accordingly, the "consistency" between clustering and visualisation is assessed by projecting the same cluster labels onto the PCA and t-SNE embeddings. The embeddings are calculated from the same fused feature vectors. Separation patterns are then checked in PCA (global linear structure) and in t-SNE (local neighbourhood structure). This comparison provides a conservative sanity check that the observed grouping is not an artefact of a single projection (Pieraccioli et al., 2026; Zheng et al., 2025).

In this study, PCA is chosen for its strength in identifying broad, linear patterns and capturing the principal axes of variance within the fused feature space. This makes it the optimal technique for testing the hypothesis that the data is structured around a dominant, linearly-influenced factor (Zheng et al., 2025). However, PCA is less effective at preserving the complex, non-linear local structures that often define the boundaries between clusters. To address this limitation, t-SNE is employed. As a non-linear visualisation technique, t-SNE excels at mapping data of high dimensionality to a low dimensional space while preserving local similarities (Zheng et al., 2025). It is therefore the superior method for examining the separability of clusters and the nature of samples at their transitional boundaries. The combined application of PCA and t-SNE thus provides a comprehensive and multifaceted interrogation of the clustering results.

3.5. Unveiling model behaviour: An interpretability study

With the increasing complexity of SFC prediction models, model interpretability has become a crucial factor in evaluating their reliability and practical applicability (Wang et al., 2023). Although traditional "black-box" prediction approaches can achieve high accuracy, they often fail to elucidate the underlying decision-making process and the contribution of individual features. This limitation significantly constrains the practical value and credibility of such models in shipping decision support systems (Wang et al., 2023). The sophisticated ensemble predictions generated by the ACME framework, whilst achieving enhanced accuracy, require systematic interpretability analysis to elucidate the underlying decision-making mechanisms and quantify the marginal contribution of each feature within the optimised feature space to the final SFC predictions. To address this issue, this study employs the SHAP method to develop a systematic model interpretation framework (Ma et al., 2023). This framework aims to uncover the internal mechanisms of diverse prediction models and quantify the marginal contribution of each feature to SFC. The primary advantage of SHAP lies in its rigorous mathematical foundation in game theory, enabling the decomposition of complex model predictions into the additive contributions of individual features.

Specifically, for the trained ACME predictor and an input sample, SHAP meets the additive decomposition, as shown in Eq. (27):

$$\widehat{Q}_i = \phi_0 + \sum_{j=1}^{J} \phi_{ij}, \tag{27}$$

where $\widehat{Q}_i$ denotes the predicted SFC for sample i, ϕ_{ij} denotes the SHAP contribution of feature j for sample i, ϕ_0 is the base value defined as the expected model output over a background dataset, and J denotes the number of input features (Cao et al., 2026). In this study, the background dataset is taken as the preprocessed training set, such that ϕ_0 corresponds to the empirical mean prediction under the training distribution. Consequently, positive (negative) ϕ_{ij} values indicate that feature j increases (decreases) the predicted SFC relative to ϕ_0, providing a consistent baseline for interpreting feature contributions across all SHAP-based analyses.

4. Experimental results and discussion

This section presents a systematic evaluation and in-depth analysis of the performance of the proposed ACME model for SFC prediction. To provide a clear empirical foundation for this analysis, the data sources, and experimental configuration are first detailed.

The standalone performance of the ACME model is then presented, followed by a rigorous comparative evaluation against mainstream methods to benchmark its effectiveness. Finally, the model's internal mechanisms and feature contributions are analysed to provide a deep, interpretable understanding of its behaviour.

4.1. Description of research data

To empirically validate the performance of the proposed integrated framework, this study utilises a comprehensive, real-world dataset from a single vessel. The subject of this study is an ocean-going bulk carrier, with principal dimensions of 224.9 m in length, 32.23 m in breadth, a design draft of 19.1 m, and a main engine power of 10,959 kW. The operational profile of this vessel provides a representative basis for evaluating fuel consumption prediction models in a practical maritime context.

The primary dataset, provided by an international shipping company (anonymized due to a confidentiality agreement), comprises 479 days of voyage records extracted from the ship's noon reports. This dataset covers the operational period from 14 January 2021 to 30 July 2024. As detailed in the methodology (Section 3.1), these noon reports are augmented and spatiotemporally aligned with high-frequency data from the AIS and meteorological and oceanographic data from ERA5 and GOPAF sources. This fusion process yields a high-fidelity, multi-source feature space used for all subsequent model training, validation, and testing, ensuring that the experimental evaluation is grounded in a rich and realistic operational environment.

4.2. Model performance evaluation

The evaluation of the ACME model begins with an assessment of its direct application to the test dataset, thereby establishing its fundamental predictive accuracy and stability. Fig. 2 illustrates the prediction consistency, plotting the model's predicted SFC values against the actual observed values.

Regarding the prediction consistency, Fig. 2 shows that most data points are densely distributed near the y = x diagonal, with the density hotspot closely aligning with the ideal line. This observation directly confirms the ACME model's high accuracy in fitting SFC under regular sailing conditions. Fig. 2 also shows a small number of data points with local deviations from the diagonal, primarily concentrated in extreme operating conditions or regions with sparse samples. This phenomenon is closely related to the local instability of SFC caused by complex environmental interactions (such as high speeds, full loads, and severe weather) and may also reflect measurement noise and systematic errors in data collection (Fan et al., 2022; Yuan et al., 2021). Nevertheless, the limited number and overall impact of these deviations do not significantly affect the model's overall predictive performance. Furthermore, Fig. 2 demonstrates that the ACME model consistently captures the dominant data regions while maintaining reasonable accuracy in edge

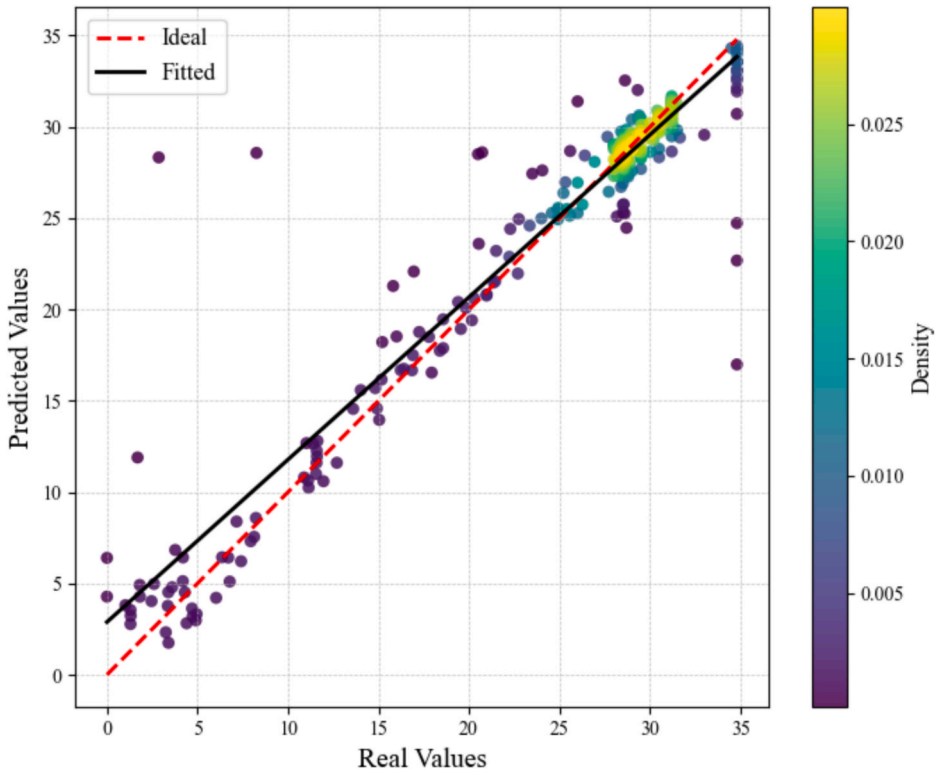


Fig. 2. Consistency distribution of ACME model predicted values vs. actual values.

regions, indicating strong global and local adaptability. This capability is crucial for SFC prediction, as ships exhibit distinct SFC patterns during different sailing phases, requiring models to adapt to diverse operating conditions (Zheng et al., 2019).

4.2.1. Performance comparison among models

In this study, six mainstream models for SFC prediction serve as benchmarks, alongside three commonly used fusion methods selected for comparison with the proposed innovative fusion method, ACME. All these models are evaluated using 5-fold cross-validation to ensure the stability of the results. To ensure a rigorously fair comparison, the *meta*-learners employed in the Stacking and Blending baselines were optimised using the same HAPSE procedure as the individual base models. This ensures that all benchmarked ensemble methods achieved their maximum predictive potential, allowing the performance gap to be attributed solely to architectural differences. Table 3 summarises the combined performance of the models.

In terms of overall performance, the ACME model demonstrates significant advantages across all evaluation metrics. Specifically, its R2 reaches 0.9088, substantially surpassing all comparative models, indicating that it explains more than 90% of the total variance in the observed data. This not only reflects the model's ability to capture the primary influencing factors of SFC but also demonstrates excellent data fitting. Meanwhile, the RMSE and MAE values of the ACME model are 1.7817 and 0.8322, respectively, further validating its superior performance in controlling absolute and relative errors. Additionally, with a MAPE of only 2.73%, the model exhibits robustness and accuracy in handling high-dimensional, nonlinear SFC prediction problems. Notably, the EV metric, which measures how well the model explains overall data variance, attains a high value of 0.9096, confirming that the ACME model effectively captures and explains key data features through clustering partitioning combined with adaptive weight adjustment in a high-dimensional nonlinear context. As noted by Kong and Ge (2022), fully leveraging local feature information is critical for improving prediction accuracy in multivariate regression problems in complex industrial processes. The ACME model achieves a technological breakthrough guided by this principle.

In this study, the traditional fusion methods (Voting, Stacking, and Blending) are less effective than individual base models, despite some improvement in overall performance. This primarily reflects the inherent limitations of these methods in handling the complex, multifactorial, coupled, and localised nature of SFC prediction—a challenging task. The Voting method employs a simple weighted averaging strategy, where weights are assigned based on global statistical information. This approach struggles to finely tune local subspaces of the data, failing to adequately capture the localised strengths of each base model under specific operating conditions (Iqbal et al., 2025). This limitation is particularly critical because SFC is influenced not only by explicit factors such as speed and load but also by implicit interaction effects under complex environmental conditions. Conversely, the Stacking approach can potentially enhance prediction accuracy by employing a secondary learner to recombine base-model outputs. In this study, the Stacking baseline used an XGBoost regressor as the *meta*-learner, and its hyperparameter space was exhaustively explored via the HAPSE algorithm to prevent suboptimal performance. However, its effectiveness heavily depends on the design of the secondary learner and the dimensionality-reduction strategy applied to high-dimensional features (Xie et al., 2025). The results indicate that even with an optimised *meta*-learner, the global Stacking approach remains inferior to ACME, as it fails to explicitly account for the localised operational heterogeneity inherent in maritime data. Given the significant data diversity and strong nonlinear relationships, secondary learners are prone to overfitting, which reduces generalisation ability and weakens fusion performance. Furthermore, although the Blending method attempts to overcome the limitations of global weight allocation by using a validation set to calculate model weights, improper partitioning of training and validation sets, especially with limited samples or uneven distributions in ship sailing data, can lead to information leakage and underfitting. This impedes the full synergistic effect of the base models (Wang et al., 2025c).

4.2.2. Analysis of the ACME mechanism

To further demonstrate the advantages of the ACME model in SFC prediction, this section presents an in-depth discussion of its fusion mechanism through multi-perspective visualisation. By integrating the analysis of clustering results and fusion weights, this study focuses on several key dimensions of the model's mechanism: firstly, the rationale for subspace modelling derived from the non-uniform data distribution; secondly, the adaptive fusion strategy that assigns differential importance to base learners according to local data characteristics; and finally, the discriminative power of the learned feature representation as visualised through dimensionality reduction.

Table 3
Performance comparison of different models.

Model	R^2	RMSE	MAE	EV	MAPE (%)
Support Vector Regression	0.8512	1.8775	1.0069	0.8545	3.97
Random Forest	0.8446	1.9045	0.8411	0.8493	3.27
Gradient Boosting	0.8043	2.1612	1.0909	0.8073	4.26
XGBoost	0.8441	1.9237	0.9082	0.8476	3.56
LightGBM	0.8487	1.8712	0.8708	0.8520	3.36
CatBoost	0.8668	1.8328	0.8776	0.8692	3.45
Voting	0.8596	1.8205	0.8287	0.8626	3.24
Stacking	0.8519	1.8719	0.9334	0.8554	3.66
Blending	0.8567	1.8421	0.8513	0.8586	3.32
ACME	**0.9088**	**1.7817**	**0.8322**	**0.9096**	**2.73**

Note: The results presented are obtained from the test set and reflect the average performance across 5-fold cross-validation.

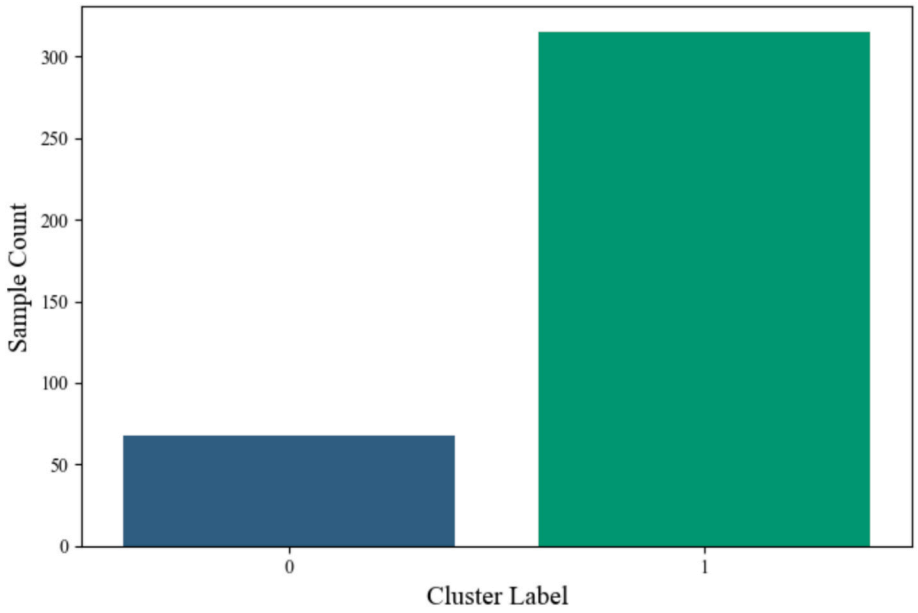


Fig. 3. Distribution of samples in clusters.

Clustering methods partition the data into distinct subspaces, enabling the capture of localised variations in SFC features. As illustrated in Fig. 3, cluster 1 (labelled 1) contains significantly more samples than cluster 0 (labelled 0), with approximately 316 and 67 samples, respectively. This uneven distribution reflects inherent characteristics of SFC data: Under routine sailing conditions, SFC patterns are relatively consistent (cluster 1), whereas under specific or abnormal environmental conditions (cluster 0), they exhibit unique feature structures. This non-uniform data distribution offers critical insights for model design, highlighting the challenge of using a single global model to adequately address diverse subspace patterns. Consequently, a clustering-based partitioned modelling strategy is both theoretically justified and practically necessary. Furthermore, such distributions support subsequent local model fusion and indicate that clusters with small sample sizes require careful handling during training to avoid overfitting or biased weight allocation.

In the ACME model architecture, the weight assigned to each base model's contribution across different clusters is a crucial factor in explaining overall performance. As shown in Fig. 4, this study analyses the average fusion weights of six base models within two clusters. In Cluster 0, Gradient Boosting and XGBoost receive relatively high weights (approximately 0.172), while LightGBM has a lower weight (0.156). In Cluster 1, CatBoost and Gradient Boosting are more prominent, with weights of approximately 0.180 and 0.172, respectively, whereas Support Vector Regression holds the lowest weight (0.152). This differential weight allocation exemplifies the core advantage of the ACME model—its ability to adaptively adjust each base model's contribution according to local data characteristics, thereby enhancing overall prediction accuracy. Notably, CatBoost achieves higher weights in both clusters, reflecting its inherent strength in handling categorical features and data with unknown distributions. Conversely, although Support Vector

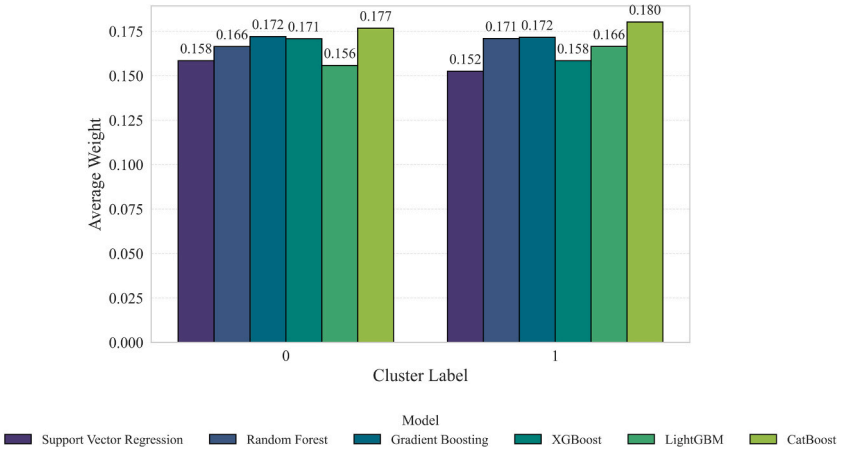


Fig. 4. Average fusion weights by cluster.

Regression receives a lower weight in Cluster 1, it performs moderately well in Cluster 0, demonstrating its specific advantage in managing structured data. This clustering-based differentiated weighting mechanism enables the ACME model to fully leverage each base model's strengths within specific data subspaces, leading to substantial improvements in overall prediction performance.

To gain a deeper understanding of the ACME model's optimisation mechanism, this study applies PCA to the fused features and visualises the results using cluster labels as colour indicators, as shown in Fig. 5. Fig. 5 shows that although the two clusters are partially separated in the principal component space, some overlap remains at the boundary. Specifically, in the positive semi-axis of the first principal component (between 0 and 20), Cluster 1 (green points) is predominant, whereas Cluster 0 (blue points) is more concentrated in the negative semi-axis (between −60 and −20). This distribution reveals a clear clustering tendency in the SFC data after dimensionality reduction. It is noteworthy that the variation in the second principal component is extremely limited, indicating that the data exhibit minimal variability in this direction and that the majority of the information is captured by the first principal component. This suggests the presence of a dominant factor in the SFC data, aligned with the direction of the first principal component, which largely determines the clustering structure. From a practical perspective, this dominant factor is likely related to key operational parameters of the ship such as speed or load that significantly influence SFC (Cepowski and Drozd, 2023). Furthermore, Fig. 5 reveals that samples located near the cluster boundaries are generally more complex to predict. These samples often correspond to transitional states between different operating conditions or to special cases influenced by multi-factor interactions. The ACME model's ability to effectively handle these boundary samples contributes significantly to its superior overall prediction performance compared to traditional ensemble methods. By applying an optimally configured set of model weights to each cluster, the ACME model accurately captures the characteristics of boundary region samples, thereby reducing prediction error.

To further investigate and validate the distribution characteristics of SFC data in the fused feature space, this study employs both a histogram of the PC1 distribution and a t-SNE-based nonlinear dimensionality reduction visualisation (Zheng et al., 2025). Fig. 6(a) illustrates a clear divergence between the two clusters along the PC1 axis: Cluster 1 and Cluster 0 exhibit distinct peaks in the positive and negative semi-axes, respectively. This observation strongly aligns with the clustering trend identified in Fig. 5 and reinforces the dominant role of PC1 in distinguishing different ship operating states. In contrast to the limited variation captured by PC2 in PCA, Fig. 6(b) uses t-SNE to map the data structure onto a two-dimensional plane, revealing a more intricate clustering pattern. While the two clusters generally form compact and distinct regions, there remains noticeable overlap at the boundaries. This overlap corresponds to the transitional samples observed in the PCA results and reflects the underlying complexity of the data due to multi-factor coupling. In real-world ship operations, such complexity often arises from the interplay between primary variables (e.g., engine power and speed) and secondary factors (e.g., meteorological conditions) (Wang et al., 2018). While PC1 captures the dominant influences directly related to SFC, t-SNE highlights the subtle effects of secondary variables. Together, these findings demonstrate that the ACME model not only effectively captures the global impact of dominant variables but also models the nuanced behavior of boundary samples, thereby significantly enhancing prediction accuracy.

In summary, this study demonstrates the superior performance of the ACME model in predicting SFC through a comprehensive

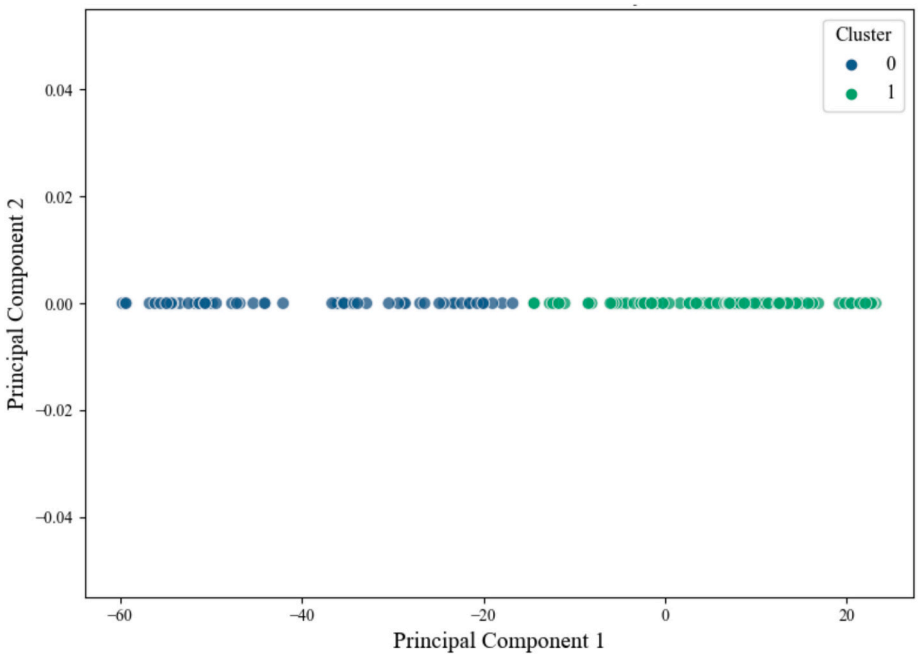
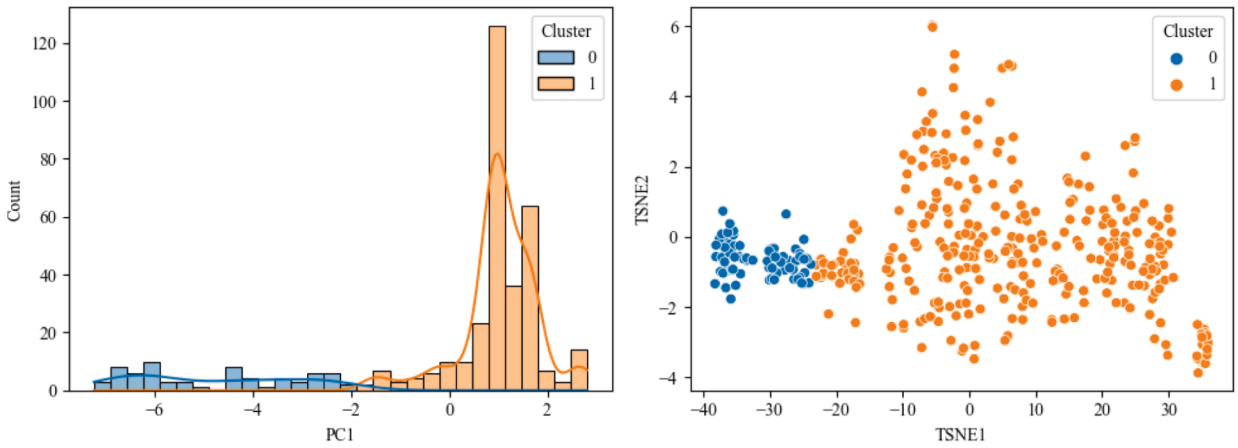


Fig. 5. The PCA of fusion features coloured by cluster. Note: The percentages of variance explained by each principal component are reported for completeness (PC1 ≈ 98.9%; PC2 ≈ 1.1%), indicating that the fusion-feature space is effectively dominated by a single principal direction, with negligible variability captured along PC2.



(a) Distribution of PC1 by cluster (b) t-SNE Visualization

Fig. 6. Complementary dimensionality reduction analysis of SFC data.

comparison of base models and fusion methods across multiple evaluation metrics. By systematically examining the clustering distribution of samples, analysing fusion weights, and applying principal component analysis to the fused feature space, the ACME model is shown to effectively address the limitations of traditional ensemble methods in handling data heterogeneity and local anomalies. By fully leveraging localised data characteristics and implementing differentiated modelling strategies for distinct subspaces, the model significantly enhances prediction accuracy. The successful application of the ACME model not only offers a novel approach for accurate SFC prediction but also provides a robust methodological framework for solving similar high-dimensional, nonlinear regression problems.

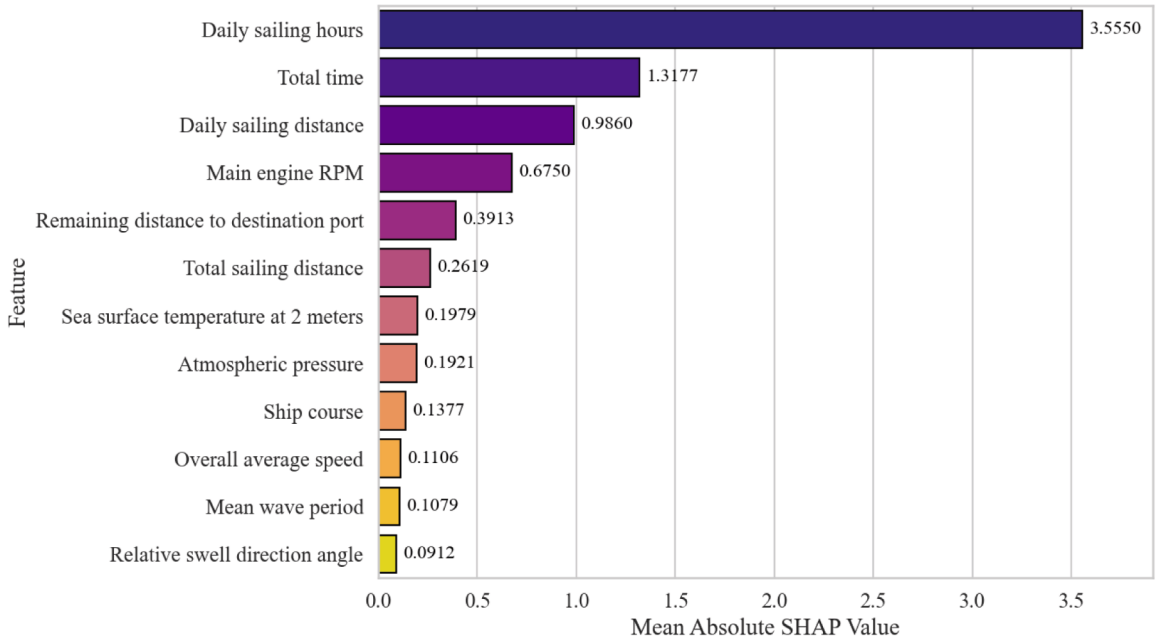
4.3. Feature importance and explainability analysis

In Section 4.2, this study investigates the strong performance of the ACME model in predicting SFC, focusing primarily on overall predictive accuracy and model mechanisms. To further enhance understanding of the model's decision-making across different feature dimensions, this section examines the model's interpretability using the SHAP method. The SHAP feature importance analysis results, shown in Fig. 7, intuitively illustrate both the magnitude and direction (positive or negative) of each feature's influence on SFC prediction. This analysis transforms the ACME model from a "black box" into a more transparent and interpretable system, revealing not only the contribution of each feature but also how it impacts the prediction outcomes.

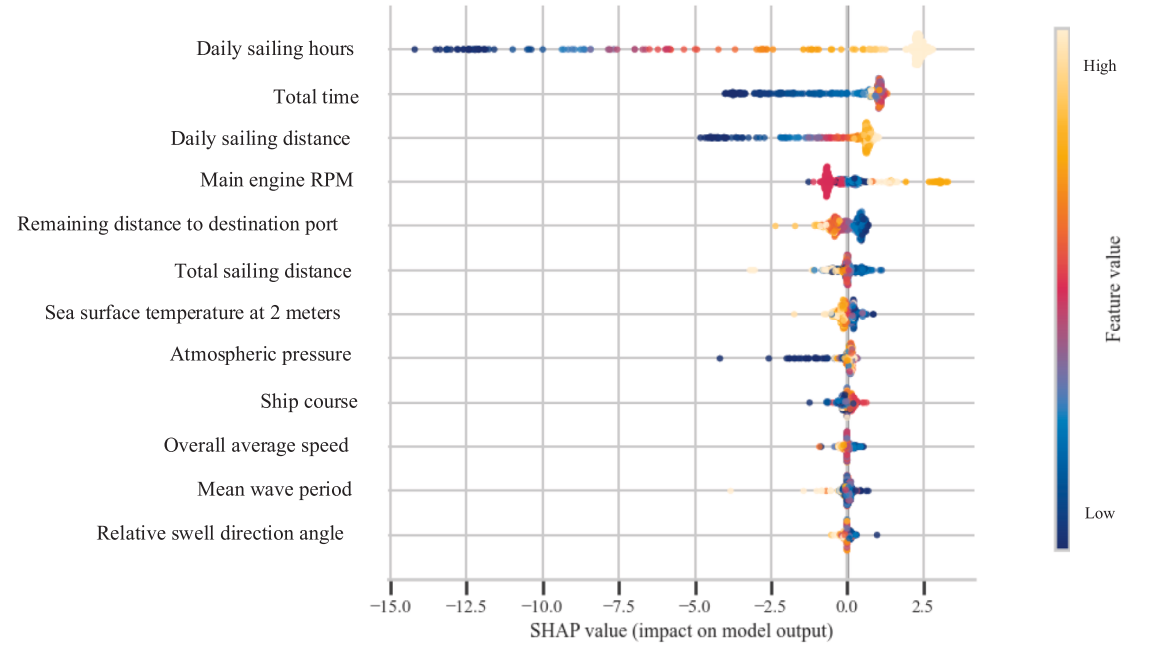
4.3.1. Overall feature importance analysis

Fig. 7(a) presents the results of the global feature importance analysis using the SHAP method, clearly revealing significant differences in the contribution of each feature to SFC prediction. Among the various factors influencing SFC, Daily sailing hours, Total time, and Daily sailing distance rank highest, aligning closely with empirical observations in the maritime industry. Sailing hours and sailing distance are the most direct and influential determinants of SFC—longer durations or distances generally lead to higher fuel usage (Cepowski and Drozd, 2023). Notably, although voyage speed is widely recognised in ocean engineering as a primary factor affecting SFC (Psaraftis and Lagouvardou, 2023), this study finds that time-related features exhibit greater predictive power. This discrepancy may be attributed to the integrative nature of sailing time, which captures the cumulative effects of speed variations, course changes, and intermittent sailing states such as berthing. Consequently, it serves as an indirect indicator of a ship's overall energy consumption under complex operational conditions (Bialystocki and Konovessis, 2016). Additionally, Main engine RPM ranks high in feature importance, underscoring the direct and critical role of the propulsion system's operational state in determining fuel demand. Higher RPM generally indicates greater power output and SFC, consistent with the thermodynamic principles and functional characteristics of marine engines (Psaraftis and Lagouvardou, 2023).

At a medium importance level, Remaining distance to destination port and Total sailing distance have SHAP values of 0.3913 and 0.2649, respectively, indicating a significant impact on the prediction results. Although these route-planning features are somewhat less influential than time and power parameters, they indirectly regulate SFC by affecting the captain's route choices and adjustments. As noted by Mollaoglu et al. (2023), route characteristics can reduce SFC indirectly by optimising the sailing path to minimize additional drag; the findings of this study quantitatively support this mechanism. Environmental factors, such as Sea surface temperature at 2 m and Atmospheric pressure, also demonstrate moderate importance in the SHAP analysis. While less influential than the aforementioned variables, their indirect effects on SFC are non-negligible. These environmental conditions can substantially alter a ship's resistance and navigational efficiency, thereby exerting cascading effects on SFC (Cepowski and Drozd, 2023). This aligns with Bialystocki and Konovessis (2016), who suggest that environmental factors typically play a moderating rather than determining role under routine sailing conditions. Furthermore, SHAP values for Ship course, Overall average speed, Mean wave period, and Relative



(a) SHAP global feature importance



(b) SHAP summary plot

Fig. 7. SHAP Feature importance analysis for SFC prediction. Note: SHAP values are reported as deviations from the model base value calculated over the training (background) distribution.

swell direction angle are all below 0.14, with Relative swell direction angle at only 0.0912, indicating a limited contribution to the prediction model. This finding contrasts with Bilgili (2023) assertion of the critical influence of swell direction on additional drag. The discrepancy may be explained by the fact that the navigational conditions in this study primarily involve relatively calm sea states, thereby diminishing the relative impact of this feature.

The overall influence of each feature on the prediction results is visualised by computing the average absolute SHAP value across

the entire sample. This analysis is cross-validated with the SWFS FS results to ensure that the selected feature subset possesses both statistical significance and robust physical interpretability.

4.3.2. SHAP distribution and local feature contribution

Fig. 7(b) presents the distribution of SHAP values for each feature using a swarm scatter plot, where the colour gradient (blue to yellow) visually represents the magnitude of each sample's value for the corresponding feature. In Fig. 7(b), SHAP values are expressed as deviations from the base value, namely the expected SFC prediction of the trained ACME model over the training distribution (Section 3.5). In contrast to the global average importance shown in Fig. 7(a), Fig. 7(b) reveals more detailed local contribution patterns and feature interactions, offering a multidimensional perspective for a deeper understanding of the SFC mechanism.

Fig. 7(b) shows that the SHAP value of Daily sailing hours, the most influential variable, exhibits a clear positive correlation with SFC. High values (yellow points) consistently drive an increase in SFC, whereas low values (blue points) significantly reduce it. This observation aligns with the principle of energy conservation and fundamental ship dynamics (Ma et al., 2023), and provides a quantitative basis for optimising sailing hours as a core ship fuel management strategy. Similarly, the SHAP distribution of Total time demonstrates a comparable but more dispersed pattern. This scattering or "stratification" reveals variations in ship fuel efficiency across different sailing phases and suggests potential interactions among features, consistent with Cepowski and Drozd (2023) findings on phase-dependent SFC characteristics. The SHAP distribution for Daily sailing distance exhibits a pronounced bimodal pattern, indicating a possible threshold effect or nonlinear response across sailing stages. This complexity may reflect differences in fuel efficiency between offshore and oceanic operations. Meanwhile, Main engine RPM shows a consistently increasing positive contribution with rising values, corroborating the analysis in Section 4.3.1 and the relationship between engine load and SFC described by Ma et al. (2023). Additionally, extreme or anomalous points in Fig. 7(b) tend to cluster where high main engine speeds coincide with extended daily sailing hours, resulting in significantly elevated local SHAP values. Although these samples represent a small portion of the dataset, they challenge model robustness. As Musulin et al. (2024) note, addressing these extremes is critical to ensuring model interpretability and reliability. Meanwhile, the SHAP distributions of Remaining distance to destination port and Total sailing distance display bidirectional effects, reflecting the complex role of route planning in regulating speed and engine load management strategies, as explained by Pelić et al. (2023). This compounding effect further complicates the SFC prediction mechanism.

It is noteworthy that, among environmental characteristics, the SHAP distributions of Sea surface temperature at 2 m and Atmospheric pressure exhibit concentrated but oppositely oriented effects. This suggests a complementary influence on propulsive efficiency, primarily through the local modulation of water and air resistance. However, their full-scale effects are often masked by dominant navigational parameters, consistent with the findings of Bialystocki and Konovessis (2016). Feature interactions are reflected in the SHAP scatterplots by the vertical dispersion of points, which is particularly pronounced in key variables such as Sailing Distance and Sailing Time. This dispersion indicates potential nonlinear variations in SFC under conditions of low sailing duration and high speed, offering data-driven insights for optimal operation scheduling and voyage planning. Although the SHAP contributions of Ship course, Overall average speed, Mean wave period, and Relative swell direction angle are relatively small, the positive SHAP trend of Overall average speed with increasing values remains consistent with the physical principle that resistance increases with the square of speed (Gao and Hu, 2021). This underscores that features with low global importance may still play significant roles under specific operating conditions.

In summary, the SHAP-based interpretability analysis of the ACME model not only quantifies the influence of each feature on SFC, but also elucidates the intricate mechanisms of feature interactions and nonlinear impact patterns. These insights offer a robust foundation for advancing the understanding of ship energy efficiency, optimising ship fuel management strategies, and designing more precise monitoring systems. When combined with the strong predictive performance demonstrated by the ACME model in Section 4.2, this study achieves an effective integration of high-accuracy prediction and deep interpretability.

5. Implications

5.1. Theoretical implications

This study presents a systematic theoretical advancement in SFC prediction by developing a high-precision, reliable, and scientifically interpretable predictive framework. Through the integration of multi-source heterogeneous data, advanced FS, HPO, an ensemble learning framework, and model interpretability analysis, it offers a novel analytical perspective for ship energy efficiency management.

Firstly, this study develops and validates a multi-source heterogeneous data fusion framework based on a unified spatio-temporal reference. The framework integrates ship noon reports, AIS data, ERA5 data, and GOPAF data, achieving high-precision alignment and correlation of heterogeneous sources through the application of IDW spatial interpolation and temporal resampling. A rigorous data quality control strategy enhances the completeness, accuracy, and spatio-temporal resolution of the modelling dataset. This approach offers a novel solution to address the challenges of data heterogeneity, sparsity, and quality that are commonly encountered in modelling complex systems in the maritime domain.

Secondly, to resolve the trade-off between high-dimensional FS and model interpretability, this study proposes and implements the SWFS method. Unlike traditional approaches that rely on a single model to assess feature importance, the core theoretical contribution of SWFS lies in its integrative design: It combines evaluation results from multiple ML models with different learning mechanisms and employs recursive evaluation along with a weighted voting strategy based on cross-validation performance to finalize FS. SHAP, grounded in cooperative game theory, provides a fair and additive quantitative measure of feature contributions, while the recursive

evaluation and weighted voting processes ensure that the selected features consistently exhibit strong predictive value across diverse model perspectives. This methodological innovation effectively reduces selection bias associated with single-model approaches, improves the robustness and generalizability of the selected feature subset, and offers a novel solution for FS in high-dimensional and multi-source data environments.

Thirdly, this study designs and evaluates a novel HAPSE algorithm to address the trade-off between efficiency and accuracy in HPO for complex ML models. The HAPSE algorithm theoretically integrates the broad search capability of global RS with the deep local exploration strength of GS, and efficiently focuses on high-potential hyperparameter regions through an adaptive dynamic reduction mechanism in the parameter space. This approach overcomes the limitations of traditional GS (high computational cost) and RS (susceptibility to local optima), offering a structured and theoretically robust solution for efficient and systematic optimisation in high-dimensional, non-convex hyperparameter spaces. Experimental results demonstrate that HAPSE significantly reduces computation time while achieving prediction accuracy comparable to or better than GS, providing theoretical support for the efficient deployment of complex models.

In addition, the core theoretical innovation of this study lies in the development of the ACME model. ACME explicitly acknowledges the inherent heterogeneity in SFC data—namely, the existence of multiple data patterns corresponding to different navigational states or environmental conditions—and extends traditional ensemble learning theory accordingly. Unlike conventional ensemble methods that adopt static or globally uniform model weights, ACME introduces a novel integration framework that deeply couples data clustering with model fusion strategies. Specifically, it dynamically and adaptively assigns the optimal combination of base model weights to each identified data cluster and incorporates *meta*-learning in the second-layer fusion. This enables the model to intelligently leverage the relative strengths of each base learner within specific data regions or operating conditions. At the level of ensemble learning theory, ACME demonstrates that explicitly recognizing and utilising structural differences within data—followed by localised, adaptive model integration—can significantly enhance predictive performance and robustness. The success of ACME offers a more refined theoretical paradigm for addressing time series prediction problems characterised by complex patterns and intrinsic heterogeneity.

Finally, to address the interpretability challenges associated with “black box” models, this study systematically applies the SHAP method grounded in game theory and establishes a global–local two-tier interpretability framework to analyse the optimal ACME model in depth. By combining global feature importance rankings with local swarm scatter distributions, this study effectively transforms the complex ensemble model into a transparent and comprehensible “white box”. This not only enhances the credibility and trustworthiness of the model outputs but also provides data-driven insights into the physical and operational mechanisms underlying SFC. Furthermore, the interpretability analysis complements and validates existing domain knowledge, while illustrating the theoretical value of uncovering latent predictive patterns directly from data.

In summary, this study systematically constructs and validates a high-precision and interpretable theoretical system for SFC prediction from four key dimensions: Multi-source data fusion, FS, model optimisation and integrated learning, and model interpretability. This theoretical framework not only significantly enhances the accuracy and robustness of SFC prediction but also introduces new theoretical perspectives and analytical tools for ship energy efficiency research. The findings contribute meaningfully to both the theoretical advancement and methodological innovation in the broader domain of maritime informatics and complex system modelling.

5.2. Practical implications

The results of this study demonstrate strong potential for practical application across several key dimensions, including ship operation and management optimisation, energy efficiency decision support, cost control strategy development, and sustainable maritime practices. These contributions offer concrete technical support and a data-driven foundation for decision-making, enabling the shipping industry to better address the dual challenges of economic pressure and environmental sustainability.

(1) Ship operations management optimisation: The ACME model developed in this study offers a powerful decision support for key stakeholders in the shipping industry, including ship operators, management companies, charterers, and fuel suppliers, due to its high prediction accuracy. It delivers reliable SFC prediction for specific voyages and varying operating conditions. Its practical applications include: 1) optimising voyage planning and execution by providing an accurate quantitative basis for selecting optimal economic speeds, designing routes, and estimating voyage durations, thereby balancing fuel costs with operational efficiency; 2) establishing dynamic benchmarks for real-time vessel performance monitoring and diagnosis. This enables timely identification of performance degradation or abnormal operating behaviour through comparison with actual SFC data, supports discretionary maintenance and crew performance evaluations, and offers objective references for setting and verifying SFC guarantees in chartering contracts, ultimately helping to reduce commercial disputes. To illustrate the operational utility of the ACME model, consider a hypothetical scenario where a shore-based manager monitors a bulk carrier encountering deteriorating sea states. Under standard practice, the crew may try to adhere to the planned Estimated Time of Arrival (ETA) by holding a steady engine setting (i.e., near-constant RPM) and, when sea states deteriorate, incrementally increasing RPM to compensate for the resulting speed loss; this typically raises the required propulsive power non-linearly because added resistance in waves and wind must be overcome, leading to a marked increase in SFC. However, the ACME framework provides a more nuanced intervention by processing real-time AIS and meteorological feeds to classify the current voyage leg into a specific operational cluster. In this state, the model adaptively re-weights its base learners to prioritise those with superior performance in high-resistance regimes. If the model predicts a non-linear spike in fuel demand that threatens the voyage’s carbon intensity targets, the manager can issue a data-informed directive to reduce speed by a specific increment or optimise the vessel's trim. This proactive adjustment, grounded in cluster-specific predictive insights, enables a

dynamic trade-off between fuel economy and schedule integrity that surpasses the "one-size-fits-all" approach of conventional maritime operations.

(2) Decision support for energy efficiency improvement: This study goes beyond merely providing predictive results by offering profound data-driven insights and practical methodologies for shipping companies to enhance energy efficiency management. It achieves this through a systematic approach that integrates heterogeneous data from multiple sources with advanced model interpretation techniques. On the one hand, this study validates the effectiveness of multi-source data fusion in overcoming the limitations of single data sources, thereby establishing a more comprehensive and reliable benchmark for ship energy efficiency, which underpins the calculation of indicators, such as EEOI and CII. On the other hand, SHAP-based feature importance analysis accurately identifies and quantifies the primary drivers of SFC, highlighting the dominant roles of operational factors—such as sailing time and main engine parameters—and the moderating effects of environmental variables. This enables management to prioritize optimisation efforts effectively. Furthermore, SHAP local swarm scatter analysis uncovers complex relationships and potential threshold effects among key features influencing SFC. These detailed insights assist crew members and shore-based managers in understanding the mechanisms behind energy efficiency variations under different operating conditions, facilitating the development of more refined ship maneuvering strategies and energy management practices.

(3) Accelerating the digital transformation of the industry: The methodological innovations in ML presented in this study, particularly the proposed HAPSE algorithm and ACME model, hold substantial practical significance for enhancing the efficiency of model development and deployment, thereby accelerating the digital transformation of the shipping industry. The HAPSE algorithm significantly reduces the computational time and resources required for HPO of complex models while maintaining or even improving model performance. This lowers the technical barriers to deploying high-performance predictive models, making it especially suitable for scenarios demanding rapid iteration or operating under resource constraints. Concurrently, SHAP interpretability analysis applied to the high-performance ACME model transforms the originally opaque "black box" into a transparent "white box" with clearer decision-making logic. This advancement greatly increases the credibility of model predictions and fosters greater acceptance and trust among end-users, such as shipmasters, engineers, and shore-based supervisors. Such trust is essential for promoting a culture of data-driven decision-making within the traditionally conservative shipping industry.

(4) Supporting regulatory compliance and sustainable development strategies: The results of this study offer substantial practical support for addressing the increasingly stringent regulatory framework on greenhouse gas (GHG) emissions reduction in the global shipping industry. These regulations include, but are not limited to, the GHG reduction strategies developed by the IMO, the Carbon Intensity Indicator (CII) rating system, the existing Energy Efficiency Existing Ship Index (EEXI) requirements (International Maritime Organization, 2025), as well as the European Union's Carbon Emissions Trading System (EU ETS) and the Marine Fuel Sustainability Regulation (FuelEU Maritime) (Kotzampasakis, 2025; Official Journal of the European Union, 2023). Specifically, high-precision SFC prediction forms the essential foundation for the accurate calculation and management of ship CII, enabling shipping companies to proactively adjust operational plans or retrofit technologies to ensure compliance with annual requirements and avoid potential commercial penalties (Yan et al., 2025). Furthermore, the benchmarking models developed herein provide a reliable quantitative basis for assessing the fuel savings and emission reduction potential of alternative fuels, energy-saving technologies, or hull optimisation measures, thus supporting informed investment decisions. Additionally, the models proposed in this study supply fundamental data and analytical tools to support key activities such as energy demand forecasting, route energy efficiency simulation, and emission reduction evaluation within the framework of green shipping corridors (Jesus et al., 2024).

6. Conclusion

This study describes the development of a holistic framework for SFC prediction that integrates high-precision forecasting with deep interpretability, offering a novel theoretical perspective and a robust data-driven tool to enhance ship energy efficiency management. The framework is built upon a synergised pipeline of four technical components: A spatio-temporal data fusion technique, a SWFS algorithm, a HAPSE method, and an ACME model.

The results reveal the superior performance of the ACME model, which significantly outperforms a suite of mainstream ML models and classical ensemble methods. Crucially, the integration of a dual-layer SHAP-based analysis transforms the predictive model from an opaque "black-box" to a transparent and interpretable system. This not only quantifies the global and local contributions of each input feature but also provides actionable, data-driven insights into the complex dynamics of SFC. From a practical standpoint, this integrated framework offers the maritime industry a reliable tool for optimising operations, managing energy efficiency, and supporting regulatory compliance, thereby advancing both economic and environmental sustainability.

Despite the significant findings of this study, several limitations remain. The primary constraint lies in the fact that the empirical validation was conducted using a high-fidelity dataset from a single ocean-going bulk carrier. While the 479-day operational period provides substantial temporal depth to capture seasonal and operational variations, the lack of vessel diversity means that the specific cluster boundaries and optimised model weights may not immediately transfer to ship categories with fundamentally different

hydrodynamic and operational profiles, such as tankers or ultra-large container ships. Nevertheless, it is important to emphasise that the architectural design of the integrated framework, comprising SWFS for feature refinement, HAPSE for efficient tuning, and ACME for adaptive ensemble learning, is inherently vessel-agnostic. Future research should aim to expand the dataset to include a heterogeneous fleet, which will be essential to substantiate the universal robustness of the proposed optimisation pipeline and to refine the adaptive mechanisms across diverse maritime contexts. On the other hand, although SHAP offers a robust interpretability tool, it remains challenging to explain higher-order feature interactions in complex nonlinear models and to differentiate correlation from causation. Future studies could explore integrating ML interpretability methods with causal inference techniques to identify causal relationships between features and SFC, rather than merely predictive associations. Specifically, a promising direction involves the integration of the proposed framework with causal discovery algorithms and Structural Causal Models (SCM). Such methodologies would enable the formalisation of a causal directed acyclic graph (DAG) to represent the maritime operational domain, allowing researchers to disentangle the direct and indirect causal effects of controllable variables (e.g., vessel speed and trim) and exogenous environmental stressors (e.g., significant wave height and swell direction) on fuel demand. By transitioning from correlation-based SHAP explanations to causal-based counterfactual reasoning, the framework could provide more robust 'what-if' decision support. This would allow shore-based managers to quantify the true impact of operational interventions with greater epistemic certainty, ensuring that decarbonisation strategies are grounded in physical and operational causality rather than purely statistical associations.

CRediT authorship contribution statement

Wenjie Cao: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Xinjian Wang:** Writing – original draft, Visualization, Supervision, Resources, Funding acquisition, Data curation, Conceptualization. **Yaqing Shu:** Writing – review & editing, Validation, Supervision, Software, Resources, Investigation, Formal analysis. **Huanhuan Li:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Jingen Zhou:** Writing – review & editing, Validation, Resources, Investigation, Formal analysis. **Zaili Yang:** Writing – original draft, Validation, Supervision, Investigation, Funding acquisition, Formal analysis, 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 gratefully acknowledge support from the MarRI-UK Core Project [Grant Agreement No. MarRI-UK_CORE_2024_02]. This work was also supported by the EU H2020 ERC 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
	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

(*continued on next page*)

Table A (*continued*)

Data source	Features	Units
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. Shap-based weighted feature selection (SWFS) algorithm: Pseudocode and feature selection results

Table B
Pseudocode for SHAP-based Weighted Feature Selection (SWFS) Algorithm.

Algorithm 1: SHAP-based Weighted Feature Selection (SWFS)

Input: *X_train*: Training feature matrix (*N* samples, *p* features); *y_train*: Training target vector (*N* samples); *F*: Set of all features [*f_1, …, f_p*]; *M*: Set of base machine learning models [*m_1, …, m_k*] compatible with SHAP; *k_cv*: Number of folds for cross-validation (e.g., 5)
Output: *Rank_final*: Final ranked list of features based on weighted votes; *S_final*: Final selected feature set for subsequent modelling (consensus-based)
// Initial SHAP Importance Assessment for each model
For each model m in *M*:
Train m on (*X_train, y_train*)
Calculate SHAP values **for** *m* on *X_train*
Compute mean absolute SHAP value *I_[m,j]* for each feature *f_j*
Rank features based on *I_[m,j]* in descending order −> *Rank_m*
Store *Rank_m* in SHAP_Importance[*m.name*]
End for
// Recursive Feature Evaluation via Cross-Validation
For each model *m* in *M*:
Initialize *current_R2_profile* = []
Ranked_features_m = SHAP_Importance[*m.name*]
For *n* = 1 to *p*:
// Select top *n* features based on *Rank_m*
Feature_subset_{mn} = *Ranked_features_m[:n]*
X_train_subset = *X_train*[*Feature_subset_{mn}*]
// Perform k-fold cross-validation using model *m*'s class
Initialize model_sample = new sample of m's class (with fixed random_state)
avg_*R2*_score = cross_val_score(model_sample, *X_train_subset, y_train*, cv = *k_cv*, scoring='*r2*').mean()
Append (*n*, avg_*R2*_score) to current_*R2*_profile
End for
Store current_*R2*_profile in *R2*_Profiles[*m.name*]
End for
// Determine Best Subset and Raw Performance Weight for each model
For each model *m* in *M*:
Find (*n_m**, *R2_m_max*) = pair in *R2*_Profiles[*m.name*] with the highest *R2_m_max*
Set $W_m^{raw} = R2_{m,\max}$

(*continued on next page*)

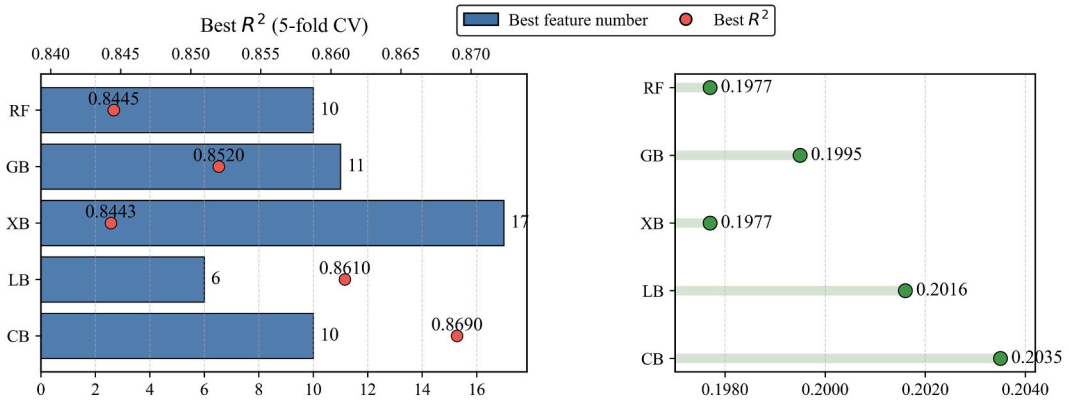
Table B (*continued*)

Algorithm 1: SHAP-based Weighted Feature Selection (SWFS)

28: Set Best_Subsets[*m.name*] = SHAP_Importance[*m.name*][:*n_m**]
29: **End for**
30: // Sum-normalise model weights to ensure fairness and reproducibility
31: **Compute** $W^{sum} = \sum_{m \in M} W_m^{raw}$
32: **For each** model *m* in *M*:
33: **Set** $\widetilde{W}_m = W_m^{raw} / W^{sum}$
34: **End for**
35: // Weighted Voting Aggregation (using normalised weights)
36: Initialise $Weighted\_Votes\left[f_j\right] = 0$ for all $f_j \in F$
37: For each model *m* in *M*:
38: $Weight_m = \widetilde{W}_m$
39: Optimal_features_*m* = Best_Subsets[*m.name*]
40: **For each** feature *f_j* in Optimal_features_*m*:
41: Weighted_Votes[*f_j*] = Weighted_Votes.get(*f_j*, 0) + Weight_*m*
42: **End for**
43: **End for**
44: // Generate Final Feature Ranking
45: Sort features in Weighted_Votes based on their scores in descending order −> *Rank_final*
46: // Consensus-based feature retention (logic preserved under normalisation)
47: **Set** $\tau = \max_{m \in M} \widetilde{W}_m$
48: **Set** $S_{\text{final}} = \left[f_j \in F \middle| Weighted\_Votes\left[f_j\right] > \tau\right]$
49: **Return** *Rank_final*

The initial stage of the experimental investigation involved applying the proposed SWFS algorithm to identify the most significant features for SFC prediction. This procedure was executed to construct an optimised, high-utility feature space prior to model training.

Fig. B1 summarises model-wise (a) optimal feature subset sizes identified via recursive evaluation under 5-fold cross-validation and the corresponding best R^2, and (b) the resulting sum-normalised weights used in the SWFS weighted voting stage. This visualisation improves transparency of the weighting scheme and enables readers to verify that no single base learner can dominate the voting solely due to the magnitude of its raw performance score.



(a) Optimal feature subset size and best R^2

(b) Sum-normalised model weights for weighted voting

Fig. B1. Model-wise optimal feature subset sizes and performance-based normalised weights used in SWFS weighted voting.

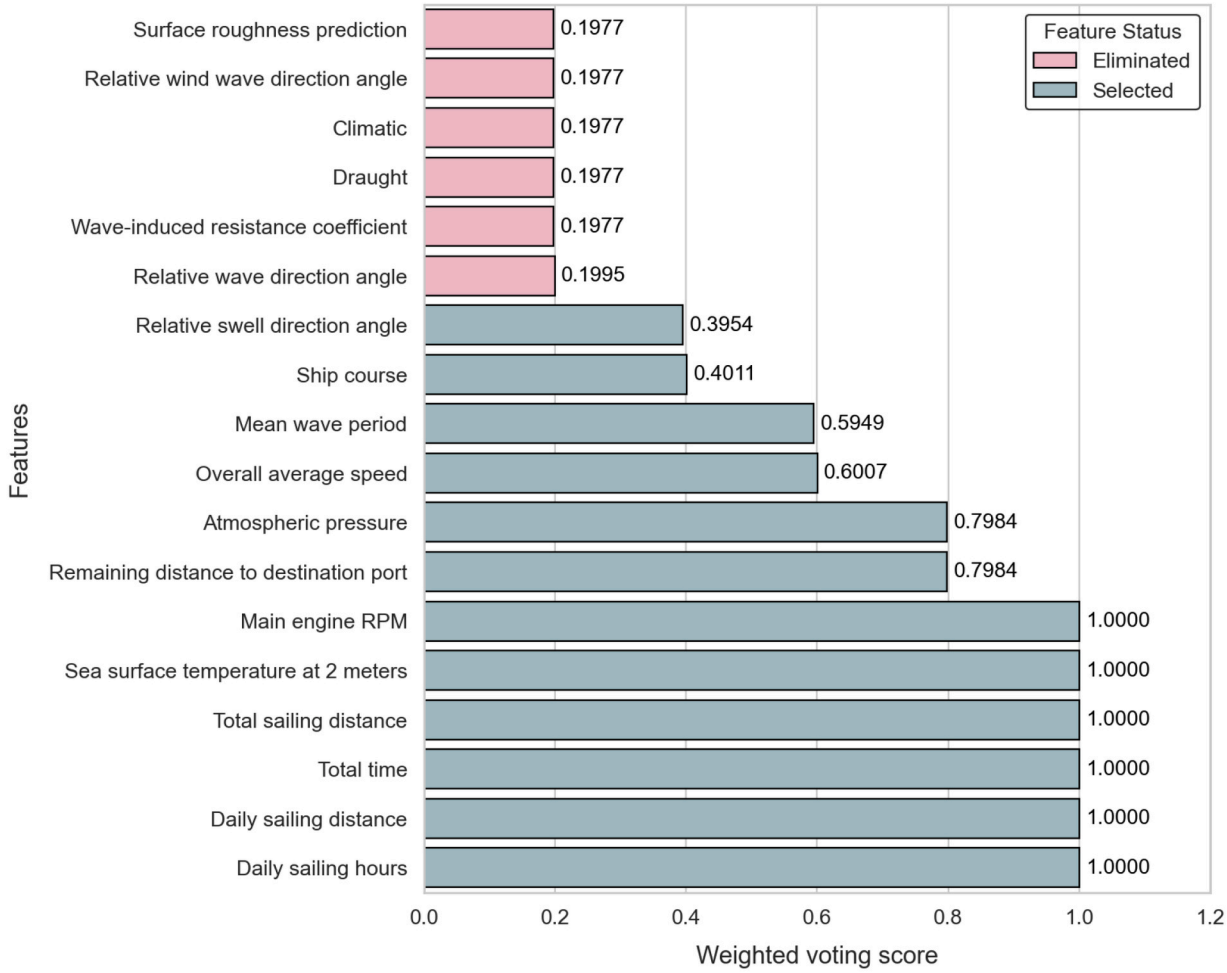


Fig. B2. Feature ranking and selection based on SWFS weighted voting scores. **Note:** Weighted voting scores are rounded to four decimal places. Features not shown in the figure were not selected as part of the optimal feature subset by any of the base models during the recursive evaluation step, and thus received a weighted voting score of zero.

The outcome of the SWFS algorithm is presented in Fig. B2, which displays the final weighted voting score for each candidate feature and indicates its selection status. A subset of 12 features was retained based on the threshold, forming the basis for all subsequent predictive modelling.

Appendix C. HAPSE algorithm: pseudocode, evaluation, and optimised hyperparameters

This study systematically evaluates the performance advantages of the HAPSE method compared to traditional approaches for HPO in ML. Under stringent computational resource constraints, comprehensive comparative experiments are conducted on three representative ML models—Random Forest, Gradient Boosting, and CatBoost—using four optimisation methods: GS, RS, Bayesian optimisation (BO), and HAPSE. The statistical significance and reliability of the results are ensured through 5-fold cross-validation.

The experimental results visualise the performance distributions of the four HPO methods across Mean Squared Error (MSE) and computational cost, as shown in Fig. C1. For the Random Forest model (Fig. C1(a)), HAPSE achieves prediction accuracy comparable to GS while significantly outperforming RS and BO. In terms of computational efficiency, HAPSE markedly surpasses GS, demonstrating a

substantial improvement. Similarly, experiments on the Gradient Boosting model (Fig. C1(b)) and the CatBoost model (Fig. C1(c)) indicate that HAPSE maintains prediction accuracy comparable to GS while substantially reducing computation time and error relative to RS and BO.

The comprehensive analysis demonstrates that the HAPSE method achieves an optimal balance between accuracy and efficiency in the HPO task for SFC prediction. On average, HAPSE reduces computation time by approximately 50% compared to traditional GS while maintaining comparable or superior accuracy, and improves prediction performance by about 5% relative to RS and BO. The logarithmic scales in Fig. C1 clearly illustrate that HAPSE decreases optimisation time from thousands to hundreds of seconds, representing an order-of-magnitude improvement of significant practical value. The experiments further show that HAPSE consistently maintains performance advantages across different model architectures, underscoring its robustness and versatility. Notably, its outstanding performance with the advanced integrated CatBoost model highlights HAPSE's adaptability in handling complex nonlinear relationships. This capability is crucial for complex prediction tasks such as SFC influenced by multiple factors and provides reliable technical support for enhancing both prediction accuracy and model deployment efficiency.

Conceptually, the superior efficiency of HAPSE stems from its structured approach to the exploration–exploitation trade-off. While traditional GS is predominantly exploitative but computationally prohibitive (Yang and Shami, 2020), and RS is more exploratory but offers limited directional convergence (Muñoz et al., 2025), HAPSE reconciles the competing objectives of broad exploration and targeted exploitation through a hierarchical framework. In the first stage, global random sampling functions as a robust exploratory mechanism to cover the non-convex parameter manifold. In the second stage, the algorithm transitions to intensive exploitation via an adaptive space reduction mechanism. This re-centres and contracts the search boundaries based on initial optima, a logic that parallels the "expected improvement" strategy in BO (Li et al., 2025a). In contrast to BO, HAPSE distinguishes itself by avoiding the heavy *meta*-computational overhead associated with stochastic surrogate models. Unlike BO, which requires fitting and updating complex Gaussian Processes (GP) and Tree-structured Parzen Estimators (TPE) (Bischl et al., 2023; Li et al., 2025a), HAPSE employs a deterministic refinement process. This makes it significantly faster and more transparent, particularly for high-dimensional SFC prediction tasks where rapid model re-calibration is essential. By systematically narrowing the search manifold without complex probabilistic modelling, HAPSE achieves a convergence rate comparable to state-of-the-art HPO methods while maintaining superior computational economy.

In summary, the HAPSE method effectively resolves the "accuracy–cost" trade-off in the HPO process through its hierarchical adaptive search mechanism, offering a practical and efficient solution for parameter optimisation in complex deep models and large-scale data scenarios.

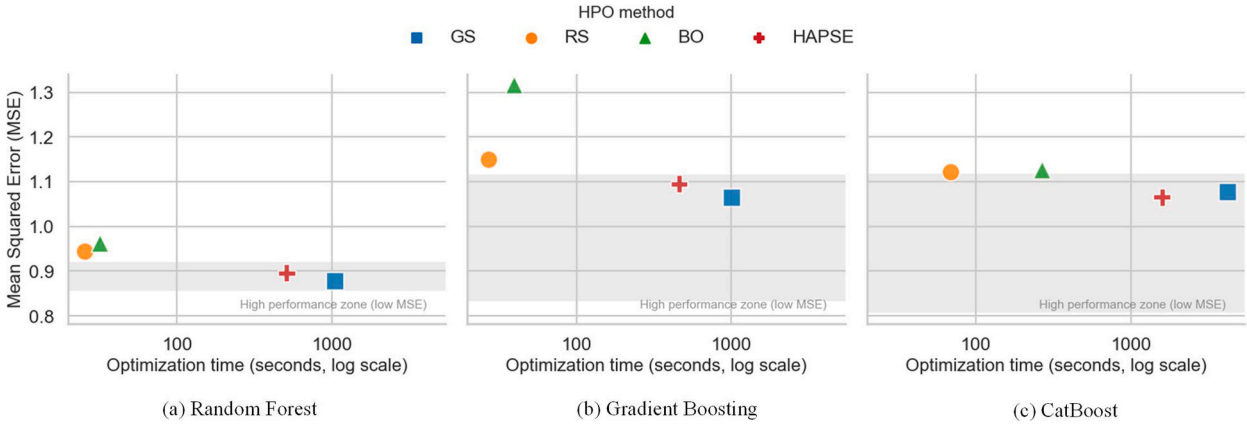


Fig. C1. HPO performance vs. computational cost comparison.

To examine the sensitivity of the Stage-2 local contraction factor α, as shown in Eq. (10), a one-factor sensitivity analysis was conducted by varying $\alpha \in \{0.10, 0.15, 0.20, 0.25, 0.30, 0.40\}$ while keeping all other modelling, data-splitting, and optimisation settings unchanged. For each representative base learner, e.g. Random Forest, Gradient Boosting, Support Vector Regression, XGBoost, LightGBM, and CatBoost, the evaluation was performed under a fixed random seed to ensure deterministic comparability. Fig. C2 reports the resulting MSE curves; the grey band denotes the model-specific within-2% region of the minimum MSE (low-MSE zone). Overall, the results indicate that performance is stable for moderate α values, and$\alpha = 0.20$ consistently attains the minimum or lies inside the within-2% low-MSE zone across all six models, supporting$\alpha = 0.20$ as a robust default.

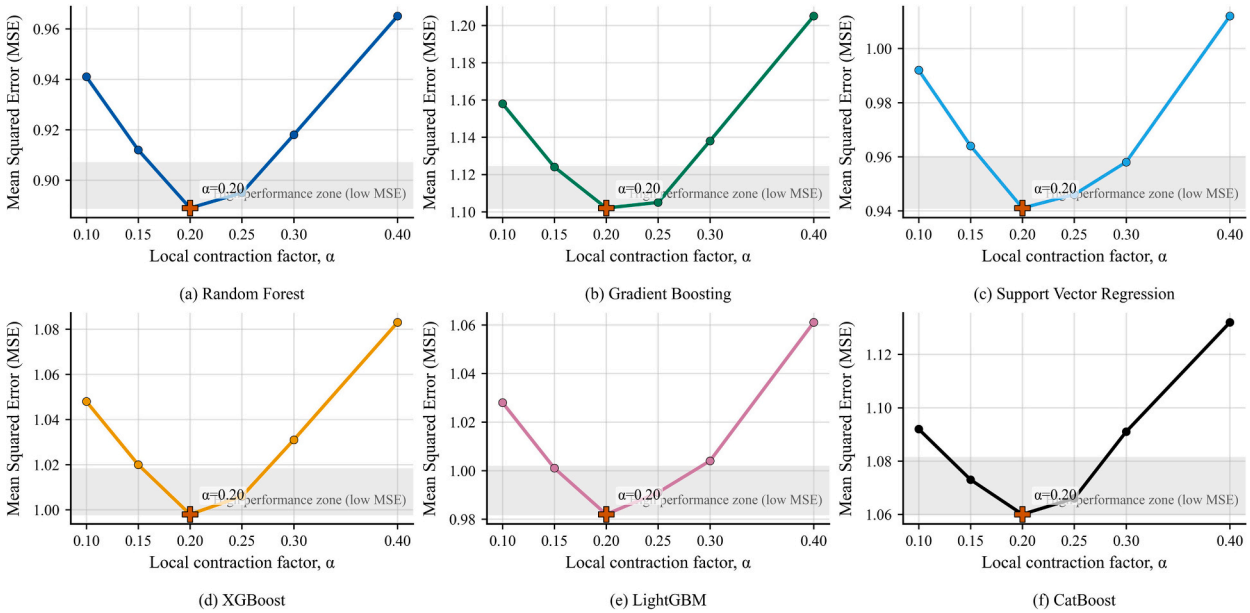


Fig. C2. Sensitivity analysis of the local search space contraction factor α. Note: Deterministic evaluation under a fixed random seed. The grey band denotes the model-specific within-2% region of the minimum MSE (low-MSE zone).

Table C1
Pseudocode for Hierarchical Adaptive Parameter Space Exploration (HAPSE) Algorithm.

Algorithm 2: Hierarchical Adaptive Parameter Space Exploration (HAPSE)

Input: *Parameter_bounds*: Dictionary of parameter ranges for each model; *X_train, y_train*: Training data; *X_test, y_test*: Testing data; *n_iter_random*: Number of random search iterations; *grid_steps*: Number of steps for grid search refinement; *α*: Parameter space reduction factor; *n_workers*: Number of CPU workers for parallel evaluation; *random_state*: Random seed for reproducibility
Output: *best_params*: Optimised hyperparameters; *best_score*: Corresponding performance score
Function HAPSE (*model_name, X_train, y_train, X_test, y_test, n_iter_random, grid_steps, α, n_workers, random_state*):
// Stage 1: Random search for promising region
best_score ← ∞
best_params ← null
bounds ← *Parameter_bounds[model_name]*
Candidate_Params_random ← []
For *i* = 1 to *n_iter_random* **do**
params ← []
For each (key, (low, high)) in bounds **do**
If isInteger(low) and isInteger(high) **then**
params[key] ← *randomInteger*(low, high, *random_state*)
Else
params[key] ← *randomUniform*(low, high, *random_state*)
End if
End for
If *model_name* = "Support Vector Regression" **then**
params["*kernel*"] ← "*rbf*"
End if
Append params to *Candidate_Params_random*
End for
// Parallel evaluation of random candidates
Scores_random ← *parallelEvaluate*(*model_name, Candidate_Params_random,*
X_train, y_train, X_test, y_test,
n_workers)
idx_best ← *argmin(Scores_random)*
best_score ← *Scores_random[idx_best]*
best_params ← *Candidate_Params_random[idx_best]*
// Stage 2: Refine search space around best parameters
refined_grid ← []
For each (key, (low, high)) in bounds **do**

(*continued on next page*)

Table C1 (*continued*)

Algorithm 2: Hierarchical Adaptive Parameter Space Exploration (HAPSE)

30: *best_val* ← *best_params*[key]
31: *range_width* ← high − low
32: *refined_low* ← max(low, *best_val* − α * *range_width*)
33: *refined_high* ← min(high, *best_val* + α * *range_width*)
34: **If** *isInteger*(low) and *isInteger*(high) **then**
35: *grid_values* ← *uniqueLinSpace*(*refined_low, refined_high, grid_steps, asInteger* = true)
36: **Else**
37: *grid_values* ← *linSpace*(*refined_low, refined_high, grid_steps*)
38: **End if**
39: *refined_grid*[key] ← *grid_values*
40: **End for**
41: // Perform grid search in refined space (parallel)
42: *Candidate_Params_grid* ← *cartesianProduct*(*refined_grid*)
43: **For each** params in *Candidate_Params_grid* **do**
44: **If** *model_name* = "Support Vector Regression" **then**
45: params["*kernel*"] ← "*rbf*"
46: **End if**
47: **End for**
48: *Scores_grid* ← *parallelEvaluate*(*model_name, Candidate_Params_grid*,
49: *X_train, y_train, X_test, y_test,*
50: *n_workers*)
51: *idx_final* ← *argmin(Scores_grid)*
52: *final_best_score* ← *Scores_grid[idx_final]*
53: *final_best_params* ← *Candidate_Params_grid[idx_final]*
54: **Return** *final_best_params, final_best_score*

Table C2
Hyperparameters for each prediction model.

Model	Parameter	Searching space	Selected setting
Support Vector Regression	*C*	[0.1, 100]	21.715
	epsilon	[0.001, 1]	0.199
	gamma	[0.0001, 1]	0.061
	kernel	Fixed ('rbf')	rbf
Random Forest	*n_estimators*	Integer [50, 300]	300
	max_depth	Integer [3, 20]	12
	min_samples_split	Integer [2, 10]	2
	max_features	[0.1, 1.0]	0.440
Gradient Boosting	*n_estimators*	Integer [50, 300]	92
	learning_rate	[0.01, 0.3]	0.262
	max_depth	Integer [3, 10]	5
	subsample	[0.5, 1.0]	0.682
XGBoost	*n_estimators*	Integer [50, 300]	293
	learning_rate	[0.01, 0.3]	0.097
	max_depth	Integer [3, 10]	3
	reg_alpha	[0.0, 1.0]	0.712
	reg_lambda	[0.0, 1.0]	0.907
LightGBM	*n_estimators*	Integer [50, 300]	135
	learning_rate	[0.01, 0.3]	0.035
	num_leaves	Integer [20, 150]	98
	reg_alpha	[0.0, 1.0]	0.275
	reg_lambda	[0.0, 1.0]	0.787
CatBoost	*iterations*	Integer [100, 1000]	813
	learning_rate	[0.01, 0.3]	0.046
	depth	Integer [3, 10]	6
	l2_leaf_reg	[1.0, 11.0]	1.000

Appendix D. Pseudocode for adaptive Cluster-based Multi-Ensemble (ACME) algorithm

Table D
Pseudocode for Adaptive Cluster-based Multi-Ensemble (ACME) Algorithm.

Algorithm 3: Adaptive Cluster-based Multi-Ensemble (ACME)

Input: Training features *X_train* and targets *y_train*; Test features *X_test*; Base models $M = [M\_1, M\_2, \ldots, M\_m]$; Minimum cluster size $\theta = 5$
Output: Final predictions *y_pred* on test set
// Phase 1: Base models training and prediction
Scale *X_train* and *X_test* using StandardScaler
Initialize empty matrices *P_train* and *P_test* of shape [*n_samples, m*]
For each model *M_i* in *M*:
Fit *M_i* on *X_train* and *y_train*
P_train[:, i] = *M_i.predict(X_train)*
P_test[:, i] = *M_i.predict(X_test)*
// Phase 2: Feature extraction for clustering
F_train = ExtractFeatures(*P_train*)
F_test = ExtractFeatures(*P_test*)
// Phase 3: Adaptive clustering
best_k = null, *best_score* = -1
For *k* in range(*2*, min(*11, n_samples*)):
For *rs* in range(*5*): // Try different random states
km = KMeans(*n_clusters* = *k*, *random_state* = *rs*)
labels = *km*.fit_predict(*F_train*)
score = silhouette_score(*F_train*, labels)
If score > *best_score*:
best_score = score
best_k = *k*
best_rs = *rs*
km_final = KMeans(*n_clusters* = *best_k, random_state* = *best_rs*)
train_labels = *km_final*.fit_predict(*F_train*)
test_labels = *km_final*.predict(*F_test)*
// Phase 4: Small cluster merging
train_labels = MergeSmallClusters(*train_labels*, *F_train, θ*)
unique_clusters = unique(*train_labels*)
// Phase 5: Group-wise weight calculation
For each cluster *c* in *unique_clusters*:
cluster_indices = indices where *train_labels* == *c*
If length(*cluster_indices*) < *θ*:
weights[*c*] = [*1/m, 1/m, …, 1/m*] // Equal weights for small clusters
Else:
P_c = *P_train*[*cluster_indices*,:]
y_c = *y_train*[*cluster_indices*]
weights[*c*] = CalculateCorrelationWeights(*P_c, y_c*)
// Phase 6: First-level fusion
y_base_train = zeros(shape = length(*y_train*))
y_base_test = zeros(shape = length(*X_test*))
For each sample *i*:
c = *train_labels*[*i*] for training samples or *test_labels*[*i*] for test samples
y_base_train[i] = dot(*P_train[i,:]*, weights[*c*])
y_base_test[i] = dot(*P_test[i,:],* weights[*c*])
// Phase 7: Meta-features construction
meta_features_train = [y_base_train, var(P_train, axis = 1), min(P_train, axis = 1), max(P_train, axis = 1)]
meta_features_test = [y_base_test, var(P_test, axis = 1), min(P_test, axis = 1), max(P_test, axis = 1)]
// Phase 8: Second-level fusion with *meta*-learner
meta_model = XGBoostRegressor(n_estimators = 150, max_depth = 4, learning_rate = 0.05)
meta_model.fit(meta_features_train, *y_train*)
y_pred = meta_model.predict(meta_features_test)
Return *y_pred*

Data availability

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

References

Adland, R., Cariou, P., Wolff, F.-C., 2020. Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet. Transportation Research Part E: Logistics and Transportation Review 140, 101972. https://doi.org/10.1016/j.tre.2020.101972.

Afshar, M., Usefi, H., 2020. High-dimensional feature selection for genomic datasets. Knowl.-Based Syst. 206, 106370. https://doi.org/10.1016/j.knosys.2020.106370.

Ahmed Ouameur, M., Caza-Szoka, M., Massicotte, D., 2020. Machine learning enabled tools and methods for indoor localization using low power wireless network. Internet Things 12, 100300. https://doi.org/10.1016/j.iot.2020.100300.

Alsulamy, S., 2025. Predicting construction delay risks in Saudi Arabian projects: a comparative analysis of CatBoost, XGBoost, and LGBM. Expert Syst. Appl. 268, 126268. https://doi.org/10.1016/j.eswa.2024.126268.

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.

Bilgili, L., 2023. Determination of the weights of external conditions for ship resistance. Ocean Eng. 276, 114141. https://doi.org/10.1016/j.oceaneng.2023.114141.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., Lindauer, M., 2023. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. WIREs Data Min. Knowl. Discovery 13 (2), e1484. https://doi.org/10.1002/widm.1484.

Borup, D., Christensen, B.J., Mühlbach, N.S., Nielsen, M.S., 2023. Targeting predictors in random forest regression. Int. J. Forecast. 39 (2), 841–868. https://doi.org/10.1016/j.ijforecast.2022.02.010.

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, 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. Transportation Research Part A: Policy and Practice 200, 104659. https://doi.org/10.1016/j.tra.2025.104659.

Cao, W., Wang, X., Feng, Y., Zhou, J., Yang, Z., 2026. 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.

Cepowski, T., Drozd, A., 2023. Measurement-based relationships between container ship operating parameters and fuel consumption. Appl. Energy 347, 121315. https://doi.org/10.1016/j.apenergy.2023.121315.

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environment international 130, 104934. https://doi.org/10.1016/j.envint.2019.104934.

Colagrande, L., Benini, L., 2025. Taming Offload Overheads in a Massively Parallel Open-Source RISC-V MPSoC: Analysis and Optimization. IEEE Trans. Parallel Distrib. Syst. 36 (6), 1193–1205. https://doi.org/10.1109/TPDS.2025.3555718.

Duan, K., Dong, S., Fan, Z., Zhang, S., Shu, Y., Liu, M., 2025. Multimode trajectory tracking control of Unmanned Surface Vehicles based on LSTM assisted Model Predictive Control. Ocean Eng. 328, 121015. https://doi.org/10.1016/j.oceaneng.2025.121015.

Fan, A., Yang, J., Yang, L., Wu, D., Vladimir, N., 2022. A review of ship fuel consumption models. Ocean Eng. 264, 112405. https://doi.org/10.1016/j.oceaneng.2022.112405.

Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Xia, G., Liu, Z., 2024a. 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.

Feng, Y., Wang, X., Luan, J., Wang, H., Li, H., Li, H., Liu, Z., Yang, Z., 2024b. A novel method for ship carbon emissions prediction under the influence of emergency events. Transp. Res. Part C Emerging Technol. 165, 104749. https://doi.org/10.1016/j.trc.2024.104749.

Gao, C.-F., Hu, Z.-H., 2021. Speed Optimization for Container Ship Fleet Deployment considering fuel Consumption. Sustainability 13 (9), 5242. https://doi.org/10.3390/su13095242.

Guillen, M.D., Aparicio, J., Esteve, M., 2023. Gradient tree boosting and the estimation of production frontiers. Expert Syst. Appl. 214, 119134. https://doi.org/10.1016/j.eswa.2022.119134.

Gundersen, O.E., Shamsaliei, S., Isdahl, R.J., 2022. Do machine learning platforms provide out-of-the-box reproducibility? Futur. Gener. Comput. Syst. 126, 34–47. https://doi.org/10.1016/j.future.2021.06.014.

Han, P., Liu, Z., Sun, Z., Yan, C., 2024. A novel prediction model for ship fuel consumption considering shipping data privacy: an XGBoost-IGWO-LSTM-based personalized federated learning approach. Ocean Eng. 302, 117668. https://doi.org/10.1016/j.oceaneng.2024.117668.

Hu, Z., Zhou, T., Zhen, R., Jin, Y., Li, X., Osman, M.T., 2022. A two-step strategy for fuel consumption prediction and optimization of ocean-going ships. Ocean Eng. 249, 110904. https://doi.org/10.1016/j.oceaneng.2022.110904.

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.

International Maritime Organization, 2025. Index of MEPC Resolutions and Guidelines related to MARPOL Annex VI. https://www.imo.org/en/OurWork/Environment/Pages/Index-of-MEPC-Resolutions-and-Guidelines-related-to-MARPOL-Annex-VI.aspx.

Iqbal, S., Qureshi, A.N., Alhussein, M., Aurangzeb, K., Mahmood, A., Razalli Bin Azzuhri, S., 2025. Dynamic selectout and voting-based federated learning for enhanced medical image analysis. Machine Learning: Science and Technology 6 (1), 015010. https://doi.org/10.1088/2632-2153/ada0a6.

Jesus, B., Ferreira, I.A., Carreira, A., Ove Erikstad, S., Godina, R., 2024. Economic framework for green shipping corridors: evaluating cost-effective transition from fossil fuels towards hydrogen. International Journal of Hydrogen Energy 83, 1429–1447. https://doi.org/10.1016/j.ijhydene.2024.08.147.

Kolassa, S., 2020. Why the “best” point forecast depends on the error or accuracy measure. International Journal of Forecasting 36 (1), 208–211. https://doi.org/10.1016/j.ijforecast.2019.02.017.

Kong, X., Ge, Z., 2022. Deep Learning of Latent Variable Models for Industrial Process monitoring. IEEE Transactions on Industrial Informatics 18 (10), 6778–6788. https://doi.org/10.1109/TII.2021.3134251.

Kotzampasakis, M., 2025. Maritime emissions trading in the EU: Systematic literature review and policy assessment. Transport Policy 165, 28–41. https://doi.org/10.1016/j.tranpol.2025.02.014.

Lan, T., Huang, L., Ma, R., Wang, K., Ruan, Z., Wu, J., Li, X., Chen, L., 2025. A robust method of dual adaptive prediction for ship fuel consumption based on polymorphic particle swarm algorithm driven. Applied Energy 379, 124911. https://doi.org/10.1016/j.apenergy.2024.124911.

Lesouple, J., Baudoin, C., Spigai, M., Tourneret, J.-Y., 2021. Generalized isolation forest for anomaly detection. Pattern Recognition Letters 149, 109–119. https://doi.org/10.1016/j.patrec.2021.05.022.

Li, X., Du, Y., Chen, Y., Nguyen, S., Zhang, W., Schönborn, A., Sun, Z., 2022. Data fusion and machine learning for ship fuel efficiency modeling: Part I – Voyage report data and meteorological data. Communications in Transportation Research 2, 100074. https://doi.org/10.1016/j.commtr.2022.100074.

Li, H., Xing, W., Jiao, H., Yuen, K.F., Gao, R., Li, Y., Matthews, C., Yang, Z., 2024a. Bi-directional information fusion-driven deep network for ship trajectory prediction in intelligent transportation systems. Transportation Research Part E: Logistics and Transportation Review 192, 103770. https://doi.org/10.1016/j.tre.2024.103770.

Li, Z., Wang, K., Hua, Y., Liu, X., Ma, R., Wang, Z., Huang, L., 2024b. 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, H., Zhang, Y., Li, Y., Lam, J.S.L., Matthews, C., Yang, Z., 2025a. Deep multi-view information-powered vessel traffic flow prediction for intelligent transportation management. Transportation Research Part E: Logistics and Transportation Review 197, 104072. https://doi.org/10.1016/j.tre.2025.104072.
Li, Q., Kamaruddin, N., Zhang, J., Peng, C., Sui Ki Khoo, A., 2025b. A novel method of bayesian genetic optimization on automated hyperparameter tuning. Scientific Reports 15 (1), 43181. https://doi.org/10.1038/s41598-025-29383-7.
Li, Y., Liu, X., Zhang, X., Gu, X., Yu, L., Cai, H., Peng, X., 2025c. Using solar-induced chlorophyll fluorescence to predict winter wheat actual evapotranspiration through machine learning and deep learning methods. Agricultural Water Management 309, 109322. https://doi.org/10.1016/j.agwat.2025.109322.
Li, Z., Wang, K., Liang, H., Wang, Y., Ma, R., Cao, J., Huang, L., 2025d. 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, Z., Zhang, B., Zhang, M., Wang, H., Lang, X., Mao, W., 2025. A Proactive Decision support Method for Ship Collision Awareness Incorporating Ship Maneuvering. IEEE Transactions on Intelligent Transportation Systems 26 (10), 16559–16573. https://doi.org/10.1109/TITS.2025.3577731.
Liu, X., Wang, X., Gao, Y., Zhang, Y., Chen, X., Dai, Z., Yu, G., Wang, F., 2026. Prediction and optimization of coal ash flow temperature using machine learning approaches. Fuel 404, 136115. https://doi.org/10.1016/j.fuel.2025.136115.
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.
Luo, X., Li, Y., Liu, Y., Mu, J., Quan, J., 2025a. Research on the prediction of mechanical properties of magnesium-silicon-based cement and the mechanism of element interaction based on machine learning. Construction and Building Materials 463, 140062. https://doi.org/10.1016/j.conbuildmat.2025.140062.
Luo, X., Zhang, M., Han, Y., Yan, R., Wang, S., 2025b. Ship fuel consumption prediction based on transfer learning: Models and applications. Engineering Applications of Artificial Intelligence 141, 109769. https://doi.org/10.1016/j.engappai.2024.109769.
Ma, Y., Zhao, Y., Yu, J., Zhou, J., Kuang, H., 2023. An Interpretable Gray Box Model for Ship fuel Consumption Prediction based on the SHAP Framework. Journal of Marine Science and Engineering 11 (5), 1059. https://doi.org/10.3390/jmse11051059.
Memon, S.M.Z., Wamala, R., Kabano, I.H., 2023. A comparison of imputation methods for categorical data. Informatics in Medicine Unlocked 42, 101382. https://doi.org/10.1016/j.imu.2023.101382.
Mollaoglu, M., Altay, B.C., Balin, A., 2023. Bibliometric Review of Route Optimization in Maritime Transportation: Environmental Sustainability and Operational Efficiency. Transportation Research Record 2677 (6), 879–890. https://doi.org/10.1177/03611981221150922.
Muñoz, V., Ballester, C., Copaci, D., Moreno, L., Blanco, D., 2025. Accelerating hyperparameter optimization with a secretary. Neurocomputing 625, 129455. https://doi.org/10.1016/j.neucom.2025.129455.
Musulin, M., Mihanović, L., Balić, K., Musulin, H.N., 2024. The Impact of Container Ship trim on fuel Consumption and Navigation Safety. Journal of Marine Science and Engineering 12 (9), 1658. https://doi.org/10.3390/jmse12091658.
Nguyen, S., Fu, X., Ogawa, D., Zheng, Q., 2023. An application-oriented testing regime and multi-ship predictive modeling for vessel fuel consumption prediction. Transportation Research Part E: Logistics and Transportation Review 177, 103261. https://doi.org/10.1016/j.tre.2023.103261.
Ni, C., Huang, H., Cui, P., Ke, Q., Tan, S., Ooi, K.T., Liu, Z., 2024. Light Gradient Boosting Machine (LightGBM) to forecasting data and assisting the defrosting strategy design of refrigerators. International Journal of Refrigeration 160, 182–196. https://doi.org/10.1016/j.ijrefrig.2024.01.025.
Official Journal of the European Union, 2023. Regulation (EU) 2023/1805 of the European Parliament and of the Council of 13 September 2023 on the use of renewable and low-carbon fuels in maritime transport, and amending Directive 2009/16/EC. https://eur-lex.europa.eu/eli/reg/2023/1805.
Pelić, V., Bukovac, O., Radonja, R., Degiuli, N., 2023. The Impact of Slow Steaming on fuel Consumption and CO2 Emissions of a Container Ship. Journal of Marine Science and Engineering 11 (3), 675. https://doi.org/10.3390/jmse11030675.
Pieraccioli, M., Ciucci, A., Corti, C., Mastrantonio, R., Scarpone, E.K., Cesari, E., Piermattei, A., Minucci, A., Urbani, A., Camarda, F., Fagotti, A., Tamagnone, L., Scambia, G., Nero, C., Sette, C., 2026. Single-cell transcriptome analysis of patient-derived organoids captures inter- and intratumor heterogeneity and uncovers targetable pathways in high grade serous ovarian cancer. Drug Resistance Updates 85, 101354. https://doi.org/10.1016/j.drup.2026.101354.
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.
Schryen, G., 2024. Speedup and efficiency of computational parallelization: a unifying approach and asymptotic analysis. Journal of Parallel and Distributed Computing 187, 104835. https://doi.org/10.1016/j.jpdc.2023.104835.
Sheikhmohammadi, A., Khakzad, P., Rasolevandi, T., Azarpira, H., 2025. Leveraging artificial intelligence models (GBR, SVR, and GA) for efficient chromium reduction via UV/trichlorophenol/sulfite reaction. Results in Engineering 26, 104599. https://doi.org/10.1016/j.rineng.2025.104599.
Shu, Y., Han, B., Song, L., Yan, T., Gan, L., Zhu, Y., Zheng, C., 2024. Analyzing the spatio-temporal correlation between tide and shipping behavior at estuarine port for energy-saving purposes. Applied Energy 367, 123382. https://doi.org/10.1016/j.apenergy.2024.123382.
Statista, 2024. Carbon dioxide emissions from international shipping worldwide from 1970 to 2023. https://www.statista.com/statistics/1291468/international-shipping-emissions-worldwide//.
Uemoto, T., Naito, K., 2022. Support vector regression with penalized likelihood. Computational Statistics & Data Analysis 174, 107522. https://doi.org/10.1016/j.csda.2022.107522.
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, K., Yan, X., Yuan, Y., Jiang, X., Lin, X., Negenborn, R.R., 2018. Dynamic optimization of ship energy efficiency considering time-varying environmental factors. Transportation Research Part D: Transport and Environment 62, 685–698. https://doi.org/10.1016/j.trd.2018.04.005.
Wang, L., Jiang, S., Jiang, S., 2021. A feature selection method via analysis of relevance, redundancy, and interaction. Expert Systems with Applications 183, 115365. https://doi.org/10.1016/j.eswa.2021.115365.
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, H., Yan, R., Wang, S., Zhen, L., 2023. Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. Transportation Research Part C: Emerging Technologies 157, 104361. https://doi.org/10.1016/j.trc.2023.104361.
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., 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., 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, X., Cao, W., Li, T., Feng, Y., Uğurlu, Ö., Wang, J., 2025c. 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, Z., Chen, X., Wu, Y., Jiang, L., Lin, S., Qiu, G., 2025d. A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud. Scientific Reports 15 (1), 218. https://doi.org/10.1038/s41598-024-82062-x.
Wang, X., Yuan, Y., Fang, S., Zhang, Z., Wang, J., 2026. A novel causal inference method of exit choice behaviour analysis for passenger ships during emergency evacuation. Reliab. Eng. Syst. Saf. 274, 112489. https://doi.org/10.1016/j.ress.2026.112489.
Xie, Y., Yu, W., Lan, H., Gong, J., Wen, S., Zhang, H., Wu, G., Gao, W., Song, S., Wang, W., 2025. Thyroid disease diagnosis based on feature interpolation Interaction and dynamic assignment Stacking model. Biomedical Signal Processing and Control 101, 107207. https://doi.org/10.1016/j.bspc.2024.107207.
Xin, X., Cao, Y., Jarumaneeroj, P., Yang, Z., 2025. Vulnerability assessment of International Container Shipping Networks under national-level restriction policies. Transport Policy 167, 191–209. https://doi.org/10.1016/j.tranpol.2025.03.020.

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.

Yan, D., Chen, C., Gan, W., Sasa, K., He, G., Yu, H., 2025. Carbon intensity indicator (CII) compliance: applications of ship speed optimization on each level using measurement data. Marine Pollution Bulletin 212, 117593. https://doi.org/10.1016/j.marpolbul.2025.117593.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Yuan, Z., Liu, J., Zhang, Q., Liu, Y., Yuan, Y., Li, Z., 2021. Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors. Ocean Engineering 221, 108530. https://doi.org/10.1016/j.oceaneng.2020.108530.

Yue, J., Li, X., Sheng, G., Yue, Y., Yue, M., 2026. Prediction of Moisture Content in Gray Bricks of Historical buildings based on Infrared Images and Deep Learning. Journal of Building Engineering, 115199. https://doi.org/10.1016/j.jobe.2026.115199.

Yu, X., Wang, Y., Wu, L., Chen, G., Wang, L., Qin, H., 2020. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. Journal of Hydrology 582, 124293. https://doi.org/10.1016/j.jhydrol.2019.124293.

Zadeh, A., Liang, P.P., Morency, L.-P., 2020. Foundations of Multimodal Co-learning. Information Fusion 64, 188–193. https://doi.org/10.1016/j.inffus.2020.06.001.

Zhang, Y., Zhao, Z., Zheng, J., 2020. CatBoost: a new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. Journal of Hydrology 588, 125087. https://doi.org/10.1016/j.jhydrol.2020.125087.

Zhang, M., Tsoulakos, N., Kujala, P., Hirdaris, S., 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. Engineering Applications of Artificial Intelligence 130, 107425. https://doi.org/10.1016/j.engappai.2023.107425.

Zheng, J., Zhang, H., Yin, L., Liang, Y., Wang, B., Li, Z., Song, X., Zhang, Y., 2019. A voyage with minimal fuel consumption for cruise ships. Journal of Cleaner Production 215, 144–153. https://doi.org/10.1016/j.jclepro.2019.01.032.

Zheng, Z., Liu, K., Zhou, Y., Debliquy, M., Bittencourt, C., Zhang, C., 2025. Response to letter to the editor from Y. Takefuji on “beyond principal component analysis: Enhancing feature reduction in electronic noses through robust statistical methods”. Trends in Food Science & Technology 157, 104918. https://doi.org/10.1016/j.tifs.2025.104918.

Zhu, J., Yin, Y., Ma, T., Wang, D., 2025. A novel maintenance decision model for asphalt pavement considering crack causes based on random forest and XGBoost. Construction and Building Materials 477, 140610. https://doi.org/10.1016/j.conbuildmat.2025.140610.