

# iHelm Project: Data Analytics for Improving Energy Efficiency in Short Sea Shipping

Mohamed Abuella, Hadi Fanaee, M. Amine Atou, Slawomir Nowaczyk, Simon Johansson, and Ethan Faghani.

**Abstract**—To meet the urgent requirements for the climate change mitigation, several proactive measures of energy efficiency have been implemented in maritime industry. Many of these practices depend highly on the onboard data of vessel's operation and environmental conditions. In this project, a high resolution onboard data from passenger vessels in short-sea shipping (SSS) have been collected and preprocessed. We first investigated the available data to deploy it effectively to model the physics of the vessel, and hence the vessel performance. Since in SSS, the weather measurements and forecasts might have not been in temporal and spatial resolutions that accurately representing the actual environmental conditions. Then, We proposed a data-driven modeling approach for vessel energy efficiency. This approach addresses the challenges of data representation and energy modeling by combining and aggregating data from multiple sources and seamlessly integrates explainable artificial intelligence (XAI) to attain clear insights about the energy efficiency for a vessel in SSS. After that, the developed model of energy efficiency has been utilized in developing a framework for optimizing the vessel voyage to minimize the fuel consumption and meeting the constraint of arrival time. Moreover, we developed a spatial clustering approach for labeling the vessel paths to detect the paths for vessels with operating routes of repeatable and semi-repeatable paths. This report summarizing the major methods and findings of iHelm project.

**Index Terms**—Short-sea shipping, energy efficiency, spatio-temporal aggregation, voyage speed optimization, Time-series analysis, spatial clustering, vessel path identification, maritime transportation,

## I. INTRODUCTION

### A. General Background

Short-Sea Shipping (SSS) is a commercial transportation mode that does not involve intercontinental cross-ocean. The SSS provides a cost-efficient and environment-friendly alternative for transportation by utilizing inland and coastal waterways to transport the commercial freight [1].

On the other hand, the SSS produces some negative effects on the natural habitats and polluting the air along the coasts of populated cities [2]. As a response to this, the International Maritime Organization (IMO) have conducted many studies and recommended standards and imposed policies for the maritime sector to reduce the carbon dioxide ( $\text{CO}_2$ ) to 40%

Manuscript received March 1, 2024; revised March 16, 2024.

This work was supported by the Sweden's innovation agency (Vinnova). (Corresponding author: Mohamed Abuella.)

Mohamed Abuella, Hadi Fanaee, Amine Atoui, and Slawomir Nowaczyk are with Center for Applied Intelligent Systems Research (CAISR), Halmstad University, Kristian IV:s väg 80523, Halmstad, 30118 Sweden (e-mail: {mohamed.abuella; hadi.fanaee; amine.atoui; Slawomir.Nowaczyk}@hh.se). Simon Johansson and Ethan Faghani are with CetaSol AB, Gothenburg, 41251, Sweden (e-mail: {simon.johansson and ethan.faghani}@cetasol.com)

by 2030 and cut 50% of all GHGs by 2050, based on the emissions in 2008 [3].

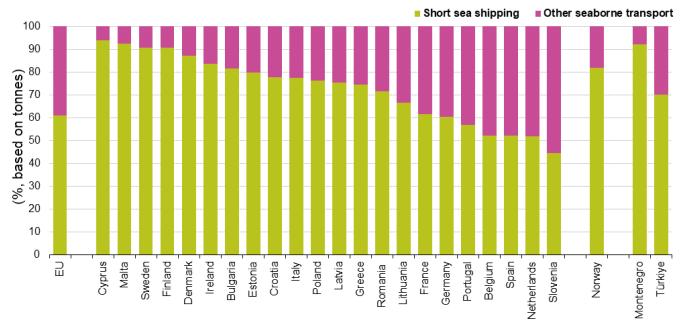


Fig. 1: European short-sea shipping of freight versus total sea transport, in 2021 [4].

Furthermore, COVID-19 pandemic has accelerated the digitization of the entire shipping industry globally, and hence attracted a profound consideration to data collection and preparation stages [5]. The operational and some environmental conditions can be accessed through an Automatic Identification System (AIS) messages, which is a service developed by the International Maritime Organization (IMO) in 2002 to record the sensor measurements and send the vessel position information for the traffic between other ships and neighboring shores [6].

In a broader perspective, for improving the vessel's energy efficiency and harnessing more fuel savings and less GHG emissions, there are mainly two procedures. The first strategy is in the ship design stage, where the ship is built to obtain a body and equipped machinery that work efficiently. The second strategy is during the ship operation over the water or at the ports. This latter procedure can be achieved by adopting energy management plans that optimally enhance the energy efficiency and fuel consumption [7].

Maritime transportation is crucial for global trade, generating extensive vessel trajectory data that reveals intricate spatial and temporal navigation patterns. Understanding these patterns is vital for effective maritime traffic surveillance and management [8].

It is crucial to distinguish between trajectory and path when studying movement data.

The term *path* refers to the specific course taken by the object. While, a *trajectory* refers to a sequence of consecutive geographical points, each representing a specific location at a given timestamp [9]. Thus, the trajectory typically denotes the movement of an object over time. In other words, a

trajectory is a path with schedule and speed information. For instance, if a vessel travels from origin port to destination port, its trajectory is the sequence of geographical points it passes through, while its path is the specific course taken, such as a fairway in a small river or through a specific canal. While the vessel trajectory is its navigation including time and space information, such as time schedule, speed, and location [10], [11].

Trajectories, observed in various scenarios such as pedestrian movements, vehicular routes, and natural events like wildlife migrations or hurricanes, involve time-evolving position data. Trajectory mining aims to uncover significant patterns within datasets, enabling tasks like path classification, anomaly detection for accidents or traffic congestions, surveillance for suspicious activities, and prediction of vessel trajectories in different landscapes [12].

Additionally, the term *route* is synonymous with path or trajectory as long as it shares the same origin and destination. Conversely, any difference in either the origin or destination represents another route. For example, when a vessel travels from an origin port to three destination ports, we say this vessel has three distinct routes to reach its three destinations.

Moreover, a *Voyage* is generally means the period between a departure from a port to the departure from the next port. Voyage is commonly used in reference to sea travel, much like the term 'trip' is used in the context of air travel.

Path clustering - same as trajectory clustering, but we do not consider the time information. Our focus is primarily on spatial information of the moving object. Path clustering, a versatile technique, involves grouping paths into clusters based on their similarity, demonstrating its effectiveness in a myriad of practical applications. In the realm of navigation, path identification empowers systems to generate clear and detailed instructions for users seeking their way. Traffic analysis benefits from path clustering as it facilitates the identification of diverse traffic patterns, such as the smooth flow of traffic on highways and the congestion often encountered on city streets. Path identification proves equally valuable in route planning, enabling the optimization of routes for transportation systems, including public and maritime transportation services [13].

In the scope of the maritime industry, path identification from Automatic Identification System (AIS) data is a challenging task due to the high spatial freedom and, especially in coastal areas, the high frequency of ship's navigation maneuvers. Thus, it is imperative to develop a path identification tool that integrates with route planning systems for improving maritime safety and optimizing vessel routing. As data-driven approaches from AIS data continue to grow and evolve, path clustering will undoubtedly play an increasingly important role in understanding vessel behavior and supporting decision-making in maritime transportation [14].

This project proposes a data-driven frameworks for modeling, optimizing, and path identification approaches to improve energy efficiency in short-sea shipping. In addition, a framework for vessel path identification has been developed for route planning and management of resources in maritime industry.

### B. Main Outcomes

The main outcomes of this project can be summarized as follow:

- Modeling of energy efficiency: Develop a data-driven model for voyage energy efficiency, including:  
A spatiotemporal aggregation of operation and navigation data from onboard and external sources to capture the impact of both spatial and temporal factors on voyage energy efficiency.  
Introduce an efficiency score that considers both total fuel consumption and voyage duration to measure the voyage energy efficiency.
- Data clustering: Clustering the data of voyages and sorting them based on their efficiency scores. This clustering enables the voyage optimization algorithm to learn more insights for better actions, either by selecting the best voyages or by eliminating the worst voyages.
- Time-series analysis models and comparative analysis: Four time-series based models are implemented as algorithms of voyage speed optimization. Then, a rigorous evaluation of their performance is conducted across different data clusters and using metrics that account for voyage efficiency.
- Practical implication: Demonstrate the significant effectiveness and practicality of the proposed approach for fixed-route vessels in short sea shipping, where the options for obtaining efficient voyages are limited. The approach also aligns with the guidance of domain experts, adhering to safety and traffic considerations.
- The proposed clustering approach of vessel paths requires only position information, specifically longitude and latitude.
- The clustering approach has a proven added value for clustering challenging unseen or unknown paths.
- The approach is robust and interpretable by applying a similarity measure that reduces the influence of noise or outliers and offers a clear interpretation of path clustering.
- The approach has a customizable parameter to determine the number of path clusters, thereby enhancing the flexibility and adaptability of the framework and allowing users to tailor it to their specific needs.
- The approach also includes a method to study and analyze the patterns within specific segments of a path.
- It is a data-driven solution that can be used as a valuable asset for informed decision-making in route planning and optimization, traffic management, and resource allocation.

## II. RELATED WORK

As digitalization and automation become increasingly prevalent in the maritime sector, the research addressing new challenges has been growing rapidly, particularly with regard to developing frameworks for energy efficiency and Maritime Situational Awareness (MSA) in cross-ocean shipping. On the other hand, the research progress has not kept pace for vessels operating in coastal areas. Thus, this literature review will

focus on our primary area of interest, which is research that is related to short-sea shipping.

### A. Vessel Energy Efficiency Modeling

Recent research studies [15], [16] have explored energy-efficient routing for an electric ferry in Western Norway. They rely on operational data from onboard measurements and environmental conditions from the Norwegian Meteorological Institute, interpolated to the nearest temporal and spatial resolutions of the vessel's onboard data. Similarly, the researchers from Napa Ltd. in Finland conducted several studies on voyage optimization, including two cases [17], [18] where environmental conditions were collected from the weather forecasts. Other studies in the literature have also processed environmental data from different weather providers to match the vessel's operational onboard data, as reviewed in [19]. However, such approaches do not account for weather factors that influence both fuel consumption and the Estimated Time of Arrival (ETA), which is a crucial constraint when optimizing the vessel's voyage, especially in SSS.

The maritime industry increasingly adopts digitization and Machine Learning (ML) techniques; however, their black-box nature remains a significant challenge. While ML can provide valuable insights, the reasoning behind the predictions made by such models is often difficult to comprehend due to their lack of explainability. To address this issue, Shapley additive explanations (SHAP) [20] were developed, providing a way to determine the contribution of each input feature toward the model's output. SHAP is commonly used as a solution to the explainability issue in ML. A recent study [21] analyzed feature importance for the power consumption of a chemical tanker. The results indicate that the ship's speed through the water is the most influential feature, while ship heading and other weather features have relatively minor influences. Kim et al. [22] utilized SHAP in combination with an anomaly detection algorithm to detect and interpret anomalies in onboard data from a cargo vessel. It allowed the identification of the specific sensor variable responsible for an anomaly, and SHAP-based clustering was used to interpret and group common anomaly patterns. A validation study for explainability in the maritime time-series data [23] compared two common model-agnostic XAI approaches, SHAP for a global method and LIME as a local method. A literature review on XAI [24] discusses the importance of XAI as a key component in modern AI techniques. The authors present a taxonomy of existing contributions related to the explainability of different machine learning models. Overall, the use and development of ML techniques in the maritime industry requires a careful balance between performance gain and explainability.

### B. Vessel Voyage Optimization

Various techniques can be applied for the voyage optimization problem, as demonstrated in prior studies such as [25]–[28]. Some of these approaches have been implemented specifically for short-sea shipping, as indicated in the domain of [19] and [29].

Different modeling and optimization algorithms for ship weather routing have been investigated thoroughly in [30]. It was found that the effectiveness of these algorithms strongly depends on specific requirements concerning the objectives, control variables and constraints as well as the implementation. The majority of these voyage optimization approaches are applied to ocean-crossing ships by controlling mainly the vessel speed and its route. Whereas, in the SSS, especially for fixed-route vessels like ours, there are fewer options available for voyage optimization. Therefore, the scope of this project is mainly focused on short-sea shipping and the pertinent research literature.

In the third IMO GHG study 2014 [31], it was assumed that weather effects alone would be responsible for 15% of additional power margin on top of the theoretical propulsion requirements of ocean-going ships, and a 10% additional power requirement for coastal ships.

In a recent adaption of the Ship Traffic Emission Assessment Model (STEAM), propelling power is determined by wave height and directions, accounting for the environmental conditions in a highly detailed manner [6].

Recent research studies [15], [16] have explored energy-efficient routing for an electric ferry in Western Norway. They rely on operational data from onboard measurements and environmental conditions from the Norwegian Meteorological Institute, and proposed a hybrid physics-guided machine learning model for optimizing the ship route. Based on their findings, the hybrid model was showing an energy reduction of 3.7% compared to the actual consumption, simply by applying minor route and speed profile alterations as guided by the provided weather forecasts.

Researchers from Napa Ltd. in Finland conducted several studies on voyage optimization. In two of their studies [17] and [18]. These experts stated that their products of voyage optimization can achieve a fuel cost reduction of more than 10% with 2% to 4% savings from trim optimization and 6% to 8% from speed and route optimization. [32].

In the study conducted by Huotari et al. [33], where they used a combined model with both dynamic programming and convex optimisation to obtain optimal speed profiles. The fuel savings were around 1.1% and for voyages with substantial variance in environmental conditions, the fuel savings reached as high as 3.5%.

In the review paper by Wang et al. [34] numerous studies are explored, including coastal and inland shipping, also revealing a diverse range of fuel savings outcomes.

Meanwhile, regarding weather routing, specifically through speed and route optimization, it has been demonstrated in [35] that the potential savings of carbon dioxide emissions and fuel costs are in range of 5% to 10%.

### C. Vessel Path Identification

Path clustering can be done using a variety of different methods [36]. We will explore the related works of these various methods.

Clustering is gaining popularity for route extraction. Machine learning (ML) has recently been applied extensively for

vessel path identification by learning patterns from historical data. Lee et al., in [12] introduced TRACLUS, a trajectory clustering algorithm employing a partition-and-group framework to discover common sub-trajectories. Demonstrating efficacy through formal trajectory partitioning and density-based clustering and efficiently identifies shared patterns in real trajectory data. The study presented in [37] introduced a framework, Traffic Route Extraction and Anomaly Detection (TREAD), which utilizes unsupervised learning for maritime route extraction. The primary emphasis is on anomaly detection and route prediction, highlighting the crucial role of AIS data in enhancing maritime situational awareness. The work specifically addresses challenges related to intermittency and persistence in AIS data. Another method of route extraction was proposed in [38], transforming ship trajectories into ship trip semantic objects (STSO) and utilizing graph theory for route extraction. The method proves robustness in extracting traffic routes for merchant ships but may have limitations for vessels with frequent navigation behavior changes, such as fishing vessels. The approach in [39], on the other hand, adopts a dynamic time warping (DTW) distance as a similarity measure and considers vessel course changes to analyze its trajectories. Experiments demonstrated its high accuracy in distinguishing and detecting similar vessel trajectories, outperforming existing methods in accuracy and cluster degree evaluation. Moreover, [40] presents a machine learning framework for maritime vessel trajectory analysis, incorporating clustering, classification, and outlier detection. It employs piecewise linear segmentation for compression and alignment kernels to integrate geographical domain knowledge, enhancing task performance. Results show reduced computation time without compromising accuracy.

Capobianco et al. [41] proposed a deep learning approach using recurrent neural networks, employing a Bidirectional Long Short-Term Memory (BiLSTM) layer as an encoder and a Unidirectional Long Short-Term Memory (LSTM) layer as a decoder, for vessel trajectory prediction. Their model outperforms baseline approaches, showcasing the effectiveness of sequence-to-sequence neural networks. In their study, Li et al. [42] present an AIS data-based machine learning method for feature extraction and unsupervised route planning for Maritime Autonomous Surface Ships (MASS). The method uses Automatic and Adaptive Dynamic Time Warping (AADTW), Spectral Clustering with Affinity Feature (SCAF), and a route optimization algorithm based on dynamic programming to extract features, obtain movement patterns, and plan routes. The proposed method outperforms existing methods by considering the impact of hidden factors and providing different routes for different types of MASS. The work in [43] systematically analyzes the performance of twelve ship trajectory prediction methods, including classical machine learning and emerging deep learning techniques. It compares twelve methods across three AIS datasets, representing different maritime traffic scenarios, and evaluates their effectiveness based on six indexes. The study concludes that traditional machine learning-based trajectory prediction methods struggle to meet the rising demands for accuracy and real-time performance, leading to increased interest in and promising results from

deep learning-based approaches. EnvClus\*, introduced in [44], it is an innovative unsupervised data-driven framework for vessel trajectory forecasting, achieving a 33% improvement over state-of-the-art methods. EnvClus\* excels in accurately predicting vessel routes, particularly in long trips, showcasing its effectiveness in mobility analytics and trajectory prediction.

A maritime traffic route extraction approach based on multi-dimensional density-based spatial clustering of applications with noise (MD-DBSCAN) was developed in [45]. The approach incorporates trajectory compression, similarity measures, and extraction of ship trajectory clusters. The approach demonstrates effectiveness in noise reduction and route extraction. The authors in [46] proposed a trajectory clustering method based on Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and Hausdorff distance to generate a similarity matrix. The method adapts to shape characteristics and exhibits good clustering scalability and improved clustering results compared to DBSCAN, k-means, and spectral clustering algorithms.

Eljabu et al. proposed spatial clustering methods (SPT-CLUST and SPTCLUST-II) in [47] and [48] respectively, for maritime traffic routes extraction from AIS data. The approach consists of data preprocessing, pathfinding, and route extraction without using traditional clustering algorithms. It achieved high F1-scores, 97% and 99%, for tankers and cargo maritime traffic routes.

The study in [49] enhanced the DBSCAN method by integrating the Mahalanobis distance metric for vessel behavior modeling. The proposed methodology includes clustering historical AIS data and detecting anomalies. The study showcases applicability to diverse water regions, contributing to situational awareness, collision prevention, and route planning.

Farahnakian et al. [50] conducted a comprehensive examination of clustering-based techniques, including k-means, DBSCAN, Affinity Propagation (AP), and the Gaussian Mixtures Model (GMM), for detecting abnormal vessel behaviors from AIS data. Results indicate that k-means is particularly effective in detecting dark ships and spiral vessel movements, which is crucial for enhancing maritime safety. Furthermore, the study [51] proposed two methods for trajectory outlier detection, with the first utilizing DBSCAN clustering and Hausdorff distance, and the second employing Support Vector Machine (SVM) classifier and the Generalized Sequence Pattern algorithm. Both models outperform the baselines, with the SVM approach demonstrating superior performance in the identification of traffic patterns and outliers. The study [10] developed MUITAS, a novel trajectory similarity measure addressing limitations in existing approaches for multiple-aspect trajectories enriched with heterogeneous semantic dimensions. Through evaluation on real datasets, MUITAS demonstrates robustness and outperforms current methods in precision at recall and clustering techniques for diverse mobility data. Moreover, the authors in [52] presented a methodology for extracting navigation network information from vessel trajectories, utilizing AIS data. The proposed model identifies key areas, speed, and course patterns, forming a network abstraction for optimizing ship routing and scheduling in the maritime industry, demonstrated through analysis in the

eastern Mediterranean sea. This model is also useful in an outlier behavior detection.

The research paper [53] offers a detailed survey of visual analytics for vessel trajectory data. The authors discuss a variety of methods, including map-based visualization, timeline-based visualization, and interactive visualization.

The survey [11] delves into the growing focus on semantically rich trajectories in movement data analysis, covering concepts, management issues, and techniques for constructing, enriching, and mining trajectories, with attention to emerging privacy challenges.

The paper [54] comprehensively reviews various approaches for vessel trajectory predicting, including clustering algorithms and machine learning algorithms. It also discusses the challenges and future research directions, such as the uncertainty in the data, the dynamic environment, and the computational complexity.

Among the identified challenges, which are subjects of ongoing research and require additional attention, three are worthy of specific mention: navigating dynamic maritime environments poses a substantial challenge in accurately identifying vessel paths (I); ensuring stability, explainability, and managing the computational cost of the model add further complexity (II); finally, addressing the need for flexibility, scalability, and practical applicability is crucial for a comprehensive solution in the field of vessel path identification (III). Motivated by these challenges, we aim to develop a framework that focuses on vessel path identification and potentially tackling such challenges faced in maritime transportation.

### III. DATA COLLECTION AND PREPARATION

#### A. Data Collection

The ship's onboard data have been received from our industry partner CetaSol AB in Gothenburg [55]. The data has been gathered over a period of 15 months, between January 2020 and March 2021. It has a 3Hz frequency and records about the ship's position, course direction, and speed. It is also including some of operational and meteorological data, such as fuel rate, engine speed, torque, acceleration, wind speed and direction.

Some information about the ship and its voyage can be found on Marine Traffic website [56].

Other weather variables such as wave height and sea current speed and direction have been collected from external sources, Copernicus Marine Service [57] and Stormglass [58] APIs.

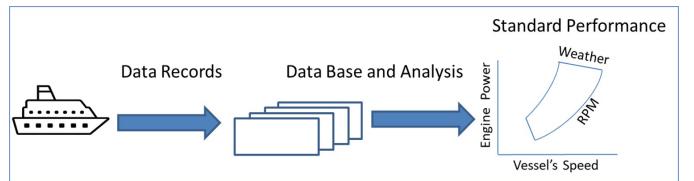
TABLE I: The navigational variables and their data sources.

Variable	Source	Variable	Source
Latitude	Onboard	WindSpeed_cps	Copernicus
Longitude	Onboard	WindDirection_cps	Copernicus
SpeedOverGround	Onboard	WaveHeight	Copernicus
HeadingMagnetic	Onboard	WaveDirection	Copernicus
Pitch	Onboard	WindSpeed_sg	Stormglass
Roll	Onboard	WindDirection_sg	Stormglass
WindSpeed_onb	Onboard	CurrentSpeed	Stormglass
WindDirection_onb	Onboard	CurrentDirection	Stormglass

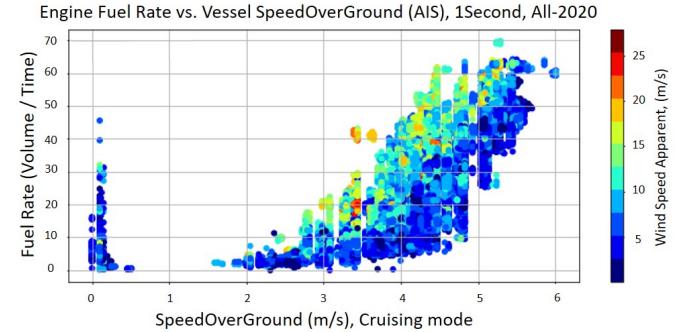
#### B. Data Preparation and Validation

The external weather data are past forecasts (hindcasts), which have reanalysed to become hourly in temporal resolution and with 0.25 to 0.5 degree as a spatial resolution. Trilinear interpolation in time and space dimensions has been applied on external weather data to be more suitable for time and position frames of the given vessel routing. Therefore, the weather and onboard data are used in this analysis with a temporal resolution of 1-minute in average.

The data validation is conducted through the cruising-speeds mode is to reduce the other vessel effects on the fuel consumption, and thus, producing graphs that can be then compared with the general ship's standard performance. The operational and weather data validation is carried out visually, as shown in Figure 2.



(a) General data analytics for ship's standard performance [59].



(b) Vessel Buro's standard performance

Fig. 2: Vessel's data analytics and standard performance graph for the case study

### IV. METHODOLOGY

Energy efficiency modeling and data clustering is a vital component of our framework. The primary objective is to identify and sort the voyages based on their efficiency scores. Then, train the models with the sorted data clusters iteratively, to distill insights from the voyages with different behaviours. Thus, the trained models will gain valuable insights into the performance and operational patterns of the vessel.

#### A. Energy Efficiency Modeling

This part presents a mathematical and visual overview of the fundamental theoretical background that forms the basis for modeling vessel energy efficiency—an indispensable element within our comprehensive framework.

To estimate the vessel's energy efficiency in this framework of voyage optimization, we employed a previously developed

model equipped with artificial intelligence (XAI) and machine learning techniques. More details about this energy efficiency modeling approach can be found in [60].

The efficiency score ( $\text{Eff}_{\text{score}}$ ) is calculated from the normalized total fuel and time for every voyage, as following:

$$\text{Eff}_{\text{Score}} = 1 - \frac{2 \times [\text{Fuel}_{Tl_{Nm}} \times \text{Time}_{Tl_{Nm}}]}{[\text{Fuel}_{Tl_{Nm}} + \text{Time}_{Tl_{Nm}}]} \quad (1)$$

The efficiency score considers the proportional reduction in both fuel consumption and time, assessing the vessel's efficient use of resources during the voyage.

Figure 3 facilitates to visualize the process of aggregation for the vessel's voyages. First by illustrating the voyages in terms of space (i.e., latitude and longitude) as shown in Figure 3a, and second by representing the aggregated voyages as points in new dimensions of Efficiency Score versus total fuel and total time. The aggregated data and its new dimensions are projected as in Figure 3b.

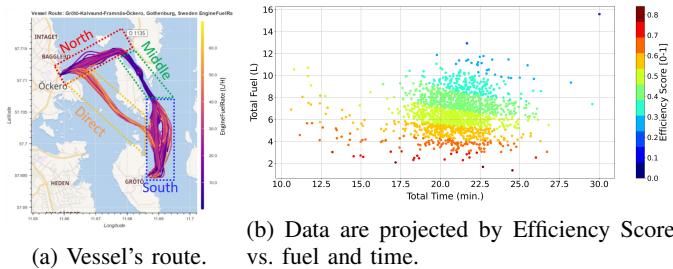


Fig. 3: The vessel voyages and the aggregated data projected by the Efficiency Score.

The representation of voyages in terms of fuel and time is done by adopting the concept of the Efficiency Score (Eff-Score). The efficiency scores for all vessel's voyages are presented in Figure 3b. It is evident that the voyage with lower fuel and shorter time have higher efficiency scores, and vice versa.

There are four voyage data clusters, namely  $\text{Top75Pr}$ ,  $\text{Top50Pr}$ ,  $\text{Top25Pr}$ , and  $\text{Top10Pr}$ , as shown in Figure 4. These clusters are categorized on their respective Eff-Score percentiles, enabling a structured analysis of voyage efficiency across various percentile groups.

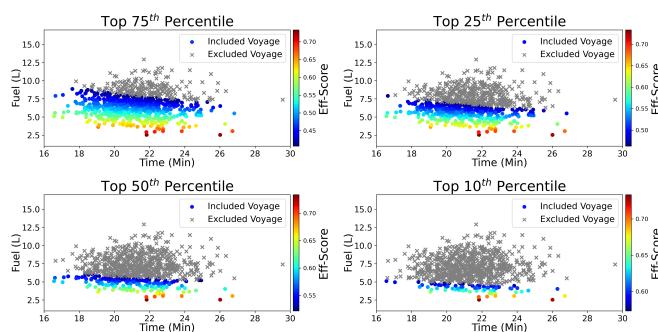


Fig. 4: Four clusters of voyages based on their efficiency.

We introduced a data-driven approach for modeling of vessel energy efficiency, by integrating data from various

---

**Algorithm 1:** Modeling of Energy Efficiency and Clustering of Voyages Data Based on Their Energy Efficiency

---

**Data:** Voyages data of the vessel

**Result:** Clusters of Voyages

Load the operational and navigational data, including speed, course, fuel, position, distance, and weather;

Tag the datapoints to its corresponding voyage,  $V_{id}$ ;

**foreach** voyage  $V_i$  in voyages data **do**

    Calculate total fuel consumption and time for  $V_i$ ;  
    Normalize total fuel and time for  $V_i$  based on their maximum values of all voyages;

    Calculate the  $\text{Eff-Score}$  as described in Eq. (1), and assign it to all datapoints of this voyage  $V_i$ ;

Initialize four empty lists for each cluster:  $\text{Top75Pr}$ ,  $\text{Top50Pr}$ ,  $\text{Top25Pr}$ ,  $\text{Top10Pr}$  (Percentiles of Eff-Scores);

**foreach** data point in all data **do**

    Extract the Eff-Score of the data point;

**if**  $\text{Eff-Score}$  is in the top 75% **then**

        Append the data point to  $\text{Top75Pr}$ ;

**else if**  $\text{Eff-Score}$  is in the top 50% **then**

        Append the data point to  $\text{Top50Pr}$ ;

**else if**  $\text{Eff-Score}$  is in the top 25% **then**

        Append the data point to  $\text{Top25Pr}$ ;

**else if**  $\text{Eff-Score}$  is in the top 10% **then**

        Append the data point to  $\text{Top10Pr}$ ;

sources and employing explainable artificial intelligence (XAI) and artificial neural network (ANN) to gain clear insights into a vessel's energy efficiency in SSS. For more details, refer to our publication [60].

The workflow for modeling of vessel energy efficiency is shown in Figure 5.

The source codes for this approach of vessel energy efficiency modeling by using Python 3.9.7 to produce the results in this case study are available at: <https://github.com/MohamedAbuella/ST4EESSS>.

### B. Framework of Voyage Optimization

One of the main objective of this project is to improve the vessel voyage by optimizing its speed to enhance the vessel's energy efficiency. In other words, reducing the vessel's fuel consumption within constrained arrival time. More details about the voyage optimization approach can be found in [61]. The framework of the developed approach for improving the vessel voyages is depicted in Figure 6.

For the purpose of vessel voyage optimization, we need a model to optimize the vessel's speed profile for improving the vessel's energy efficiency. This ideal model should mainly be able to:

- Model the temporal dependencies in vessel speed profiles.
- Incorporate external variables, such as weather conditions, into the modeling process.

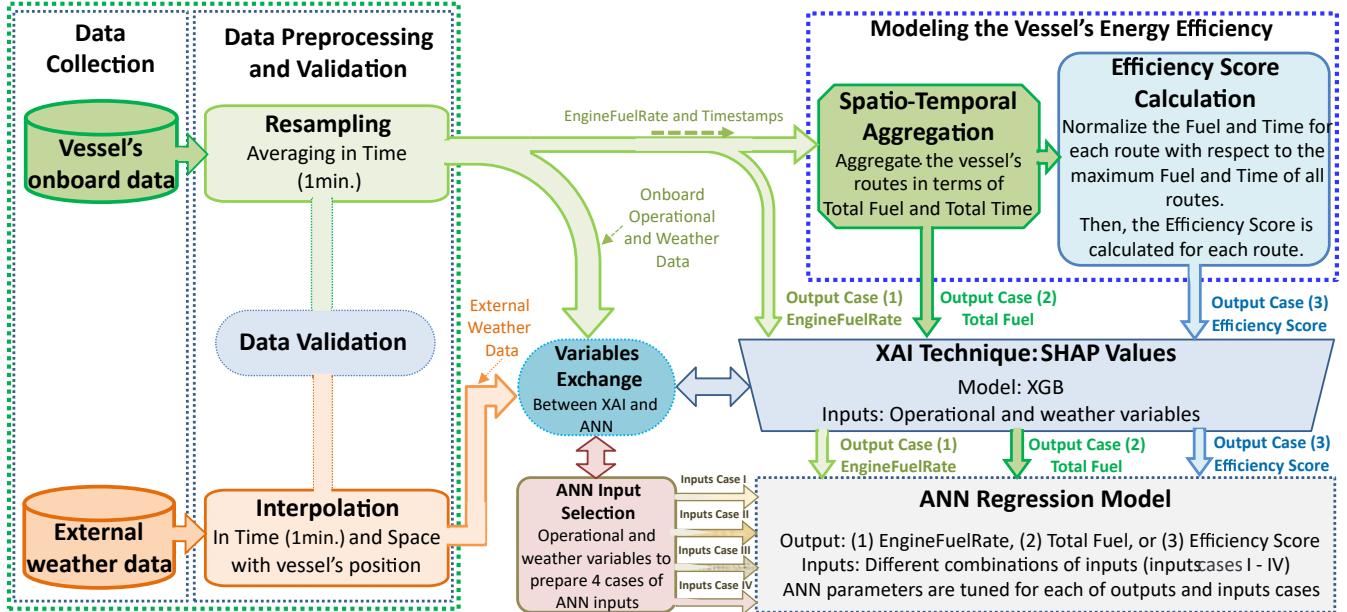


Fig. 5: Workflow of modeling and analysis of energy efficiency in short-sea shipping

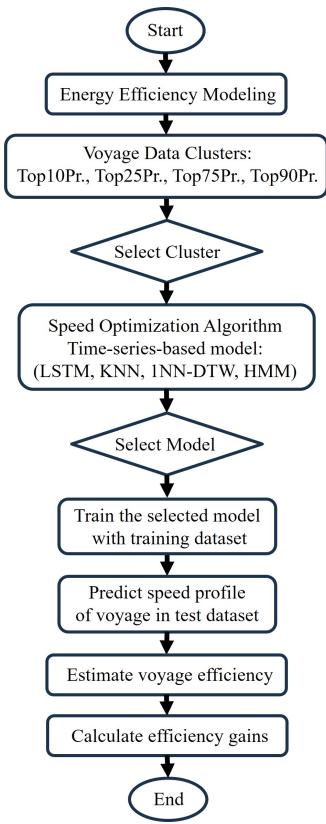


Fig. 6: Framework of vessel voyage optimization.

- Adapt to changing conditions to provide real-time optimized speed profiles.

In order to meet these requirements, we adopt several time-series analysis models, specifically they are: Long Short-Term Memory (LSTM), k-Nearest Neighbors (kNN), 1-Nearest

Neighbor with Dynamic Time Warping (1NN-DTW), and Hidden Markov Model (HMM). These models are described with some details as follow.

1) *Long Short-Term Memory (LSTM)*: It is a type of recurrent neural network (RNN) architecture designed for sequential data analysis. It is particularly useful for modeling and predicting time-series data, making it a valuable tool for improving energy efficiency in vessel operations [62].

LSTM networks are well-suited for tasks involving time-series sequences of varying lengths, capturing long-term dependencies, and handling irregular temporal patterns. The LSTM architecture deals specifically to address the problem of vanishing gradients that often occurs in other RNN structures [63].

The LSTM's capability to learn the hidden insights from selected data, particularly in terms of vessel's efficiency scores, comes with a trade-off. The data clustering may result in a smaller dataset, which potentially limiting the LSTM's deep learning capabilities. Additionally, data clustering increases the susceptibility of LSTM to overfitting.

LSTM can be employed to model and predict the vessel's speed to improve its fuel efficiency, taking into account various factors such as weather conditions and operational parameters

2) *k-Nearest Neighbors (kNN)*: The kNN algorithm is a non-parametric model that makes no assumptions and operates on the principle of determining an unknown observation's class by measuring distances to nearby observations, attributing the observation to the majority class of its nearest neighbors. The parameter  $k$  represents the number of neighboring observations considered when classifying a given observation, and its value is determined through a search for the optimal choice that maximizes accuracy in the training set.

The paper by Cover and Hart [64] is a foundational work in pattern recognition and machine learning, introduced the KNN algorithm for pattern classification by utilizing the nearest neighbors to classify data points based on majority vote.

KNN can be employed to analyze and predict energy consumption patterns based on similar historical data. By considering the nearest neighbors of a given operational scenario, KNN provides a straightforward approach to making energy-efficient decisions. The choice of the parameter ' $k$ ', representing the number of neighbors, plays a crucial role in the accuracy of KNN-based models.

**3) 1-Nearest Neighbor with Dynamic Time Warping (1NN-DTW)** : This model is a distinctive variant of kNN with ( $k=1$ ) and with a distance-wise measure that utilizes Dynamic Time Warping (DTW) algorithm for a comparison of two time-series sequences that may have different lengths, time shifts, and speed variations [65]–[69]. The 1NN-DTW algorithm can be utilized to measure the similarity between the given speed profile and other measured vessel speed profiles with higher efficiency scores. 1NN-DTW is capable to capture the temporal dependency of speed profiles, even in short-sea shipping where the options of control the ship to improve its energy efficiency are limited.

Thus, the 1NN-DTW identifies the efficient observed speed profile that most similar to the given speed profile to be chosen as suggested an optimized speed profile for the current journey of the vessel that can lead to improved energy efficiency.

**4) Hidden Markov Model (HMM):** This model is based on probabilistic modeling and uses a variety of techniques from the statistical modeling, and they are widely used in time-series analysis and pattern recognition.

The work by Rabiner in 1989 [70] is an essential reference in the application of HMM. This tutorial delved into the formulation of a statistical method for representing speech, showcasing a successful HMM system implementation with a focus on discrete or continuous density parameter distributions.

In the context of improving energy efficiency in constrained environments, HMMs can be applied to model the underlying patterns and transitions in ship operational data, allowing for more informed decisions on optimizing energy consumption. The HMM estimates three main weather states and dynamically adjusts the SpeedOverGround (SOG) predictions. During calm weather, it selects the maximum observed speed in the historical data at that condition. In moderate conditions, it uses the average speed, while in rough weather, it relies on the minimum speed profiles. For instance, Eq. 2 indicates that the predicted speed ( $SOG_{Pred_i}$ ) by HMM is the minimum value of measured speed profiles in rough weather state, and likewise for other weather states.

$$SOG_{\text{Pred},i} = \min(SOG_{\text{Meas.}} \mid \text{Rough weather}) \quad (2)$$

This adaptability to changing weather conditions enhances voyage optimization, making it a crucial component of the overall framework.

### C. Framework of Path Identification

The theoretical background and description of the underlying methodology of our proposed framework are covered within this section. Additional details about the framework of path identification can be found in [71].

The framework of vessel path identification is depicted in Figure 7.

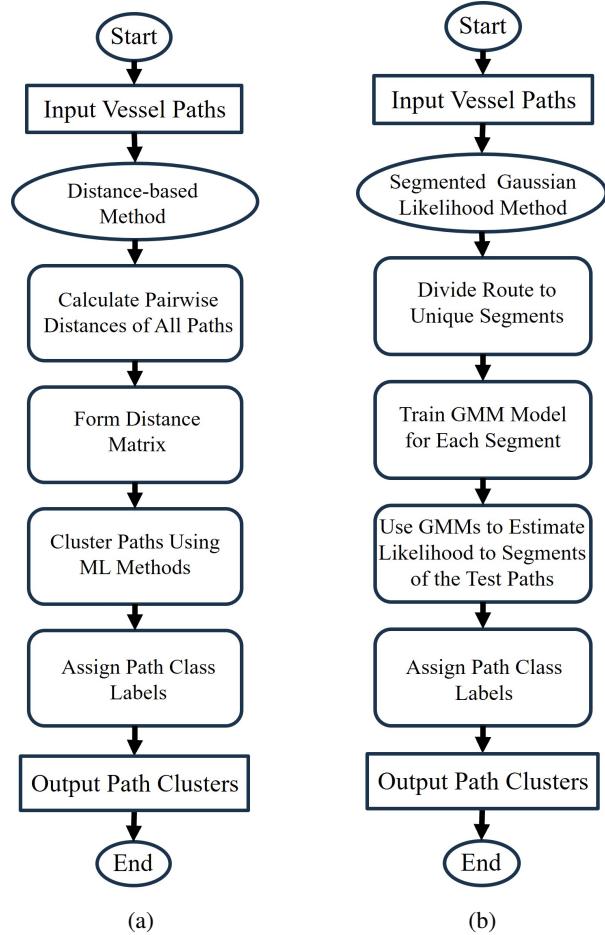


Fig. 7: Framework of vessel path identification. (a) Flowchart of distance-based method. (b) Flowchart of segmented Gaussian likelihood method.

### 1) Problem Formulation:

The equations (3-5) serve as a mathematical representation to describe the clustering of vessel paths.

It is worth mentioning that the clustering process is conducted sequentially, point by point, while the labeling of path clusters is performed for the entire voyage. Therefore, each voyage has a single label of path cluster.

$$Voyages \in Path\ Clusters \quad (3)$$

$$Voyages = \{ts_1[p_1, \dots, p_n], \dots, ts_i[p_1, \dots, p_n]\} \quad (4)$$

$$Path\ Cluster\ Set = \{cluster_1, \dots, cluster_k\} \quad (5)$$

where:

*Voyages*: a collection of time series, each representing a voyage of the vessel taken following a given path, with a predicted cluster.

$ts_j$ : a time series corresponding to voyage  $j$ , i.e., a sequence of  $n$  data points, where each data point  $p$  represents the vessel position and is defined by a pair of coordinates, namely latitude and longitude.

$n$ : the number of time steps (duration) of each voyage, which can differ from one voyage to another.

$j$ : a total number of voyages.

*Path Cluster Set*: a set of  $k$  clusters into which the path of voyages are being clustered.

*Remark*: In the definition of  $k$  clusters, it is important to clarify that this study is for labeling or identifying predefined fairways for vessels navigating in confined waters rather than open sea, resulting in a predetermined number of path clusters,  $k$ .

### 2) Distance-Based Method:

The similarity between two paths is measured by the average nearest neighbor distance (ANND), as shown in Eq. (6).

$$ANND(i, j) = \frac{1}{n_i} \sum_{k=1}^{n_i} Distance(P_i^k, NN(P_j^k)) \quad (6)$$

where:

$ANND(i, j)$ : is the average nearest neighbor distance between path  $i$  and path  $j$ , present in the distance matrix at row  $i$  and column  $j$ . It is a symmetric, meaning that  $ANND(i, j)$  is the same as  $ANND(j, i)$

$Distance(P_i^k, NN(P_j^k))$  : The distance between the  $k^{th}$  point in path  $i$ , denoted as  $P_i^k$ , and its corresponding nearest neighbor point in path  $j$ , indicated as  $NN(P_j^k)$ .  $n_i$  is the total number of points in path  $i$ .

The measure  $Distance$  is an Euclidean distance. However, for longer curved routes, Haversine or Great-circle distance would be more suitable.

The ANND, as expressed in Eq. (6), is computed by averaging the distances between each point in one path and its nearest neighbor in the other path.

Then, the similarity value (i.e., ANND) of this pair of paths is stored as an element in the distance matrix.

A lower ANND indicates that the paths within a cluster are more similar. The distance matrix will have dimensions  $(m \times m)$ , where  $m$  is the number of paths.

For instance, the computed distance matrix for a set of 12 paths is illustrated in Figure 8.

After the construction of the distance matrix, the machine learning (ML) technique is applied to cluster the paths based on their corresponding values in this distance matrix. The ML techniques that we used are k-means, Gaussian Mixture Model (GMM), and hierarchical clustering.

### 3) Segmented Gaussian Likelihood Method:

In addition to identifying the vessel's path, to better understand how the vessel changes its paths, we employ Gaussian distributions on several distinct segments of the vessel route. This technique can be summarized as follows:

- Utilize a training dataset comprising vessel position information that should adequately represent all potential paths of the vessel route.
- Divide the route into different distinct segments.
- Train a single GMM model for each segment to find the Gaussian distributions of all segments of the route.

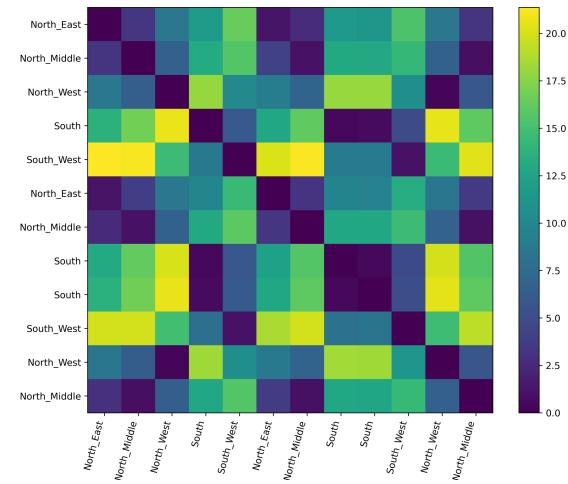


Fig. 8: Heatmap of the distance matrix for 12 paths of Cinderella II vessel.

- Estimate likelihoods of the segments by using the trained GMM models with their corresponding segments of each vessel voyage in the test dataset.
- Label the path clusters based on the estimated likelihood at the unique segments of the route.

## V. IMPLEMENTATION OF FRAMEWORKS

In this section, we present specific case studies demonstrating how the frameworks of the project, applied with real-world data, provide practical insights and solutions.

### A. Case Study of Voyage Efficiency Modeling and Optimization

The implementation of our approach of a time-series analysis-based voyage optimization framework for a fixed-route vessel of our case study is depicted in Figure 6, and the step-by-step process is described by algorithms 1 and 2. In this case study, we utilized a dataset collected from a passenger ship named Buro, which operates in Sweden. The dataset from the vessel Cinderella II was also available; however, it was not utilized for vessel voyage optimization. This decision was made due to the dataset's duration being limited to five months, encompassing only the summer and autumn seasons. As a result, it does not cover the full spectrum of seasonal variability in weather impacts.

For more detailed information about setting up the models and their specifications including various parameters, you may refer to the source codes, which are developed in Python 3.9.7 to produce the results of this study. These source codes are available at: <https://github.com/MohamedAbuella/ST4EESSS> and <https://github.com/MohamedAbuella/TSA4EESSS>.

### B. Case Study of Path Identification

In this section, we describe the case study, including the data collection, preprocessing, and analysis.

**Algorithm 2:** Speed Optimization Models for Improving Voyage Efficiency

---

**Data:** Refer to **Algorithm 1** for data processing and clustering.

Add  $SOG_{Meas.}$  and Weathers to Inputs.

**foreach**  $C_k$  in  $CT$  (sorted by Eff-Score) **do**

**foreach** model in [*LSTM*, *kNN*, *INN-DTW*, *HMM*] **do**

Train **model** with voyages in  $C_k$ ;

**foreach** Voyage  $v_i$  in test dataset  $\notin C_k$  **do**

**if** **model** is *LSTM* **then**

Predict  $SOG_{Pred,i}$  using trained *LSTM*;

**else if** **model** is *kNN* **then**

Predict  $SOG_{Pred,i}$  using trained *kNN*;

**else if** **model** is *INN-DTW* **then**

Predict  $SOG_{Pred,i}$  based on the most similar voyage in  $C_k$ ;

**else if** **model** is *HMM* **then**

Predict  $SOG_{Pred,i}$  based on weather states (Calm, Moderate, Rough);

Estimate efficiency for  $SOG_{Meas,i}$  and  $SOG_{Pred,i}$ ;

Evaluate efficiency gains for test voyages;

---

In this study, we utilized datasets collected from two passenger ships, named Cinderella II and Buro, operating in Sweden. The vessels are shown in Figure 9, additional information about vessels Cinderella II and Buro can be found in [72].

Cinderella II operates in Stockholm archipelago, east of Sweden. The data of Cinderella II spans over five months (July to November 2022). It comprises information on 124 voyages of this vessel, connecting the two main ports of Vaxholm in the east and Sodra in the west.

While Buro works in Gothenburg, west of Sweden. Its dataset has been gathered over a period of 15 months (between January 2020 and March 2021). The data of Buro has 1755 voyages, between two main ports, Grotto in the south and Ockero in the northwest.



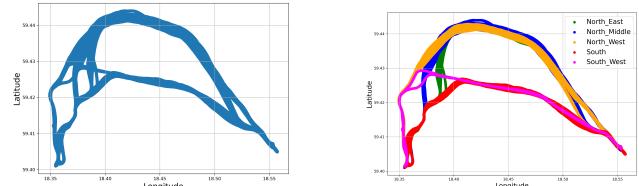
(a) Cinderella II.



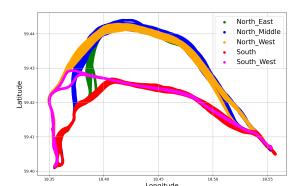
(b) Buro.

Fig. 9: Images of two vessels that are used in the case study [72].

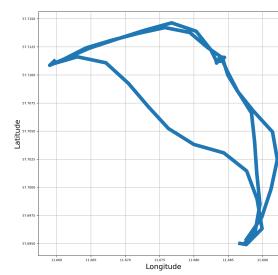
In our approach, we emphasize the significance of data representation. As a result, we group the path points based on their timestamps with a resolution of one second and store these grouped path points with distinctive Voyage IDs. The routes of Cinderella II and Buro, along their chosen paths,



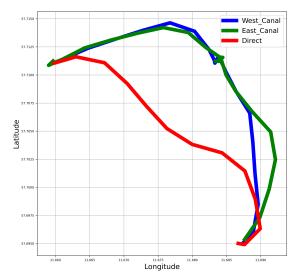
(a) Cinderella II route.



(b) Five clusters of path.



(c) Buro route.



(d) Three clusters of path.

Fig. 10: Routes of the two vessels beside their all possible path clusters, which will be identified by applying proposed framework.

are depicted in Figure 10. Path cluster distribution for both vessels are illustrated by the histograms in Figure 11.

Afterward, these paths are ready to be processed by the path clustering approach to determine the overall path cluster.

For further details regarding the identification framework, the reader may refer to the source codes employed to produce the case study results. These codes are accessible at: [https://github.com/MohamedAbuella/Path\\_Clustering](https://github.com/MohamedAbuella/Path_Clustering).

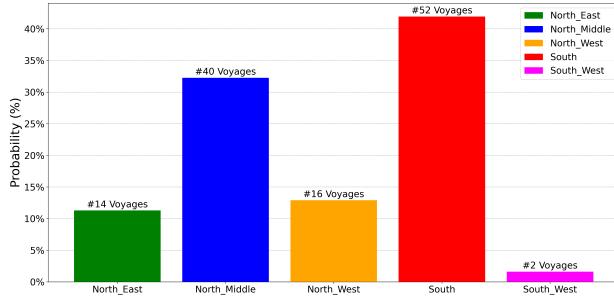
## VI. RESULTS AND DISCUSSION

The vessel has a fixed-route which starts from the southern port to the northern port or vice versa. This route can be divided into four segments, specifically North, Middle, South, and Direct, as depicted in Figure 3a. Cruising speeds are more common in North, South, and Direct segments of the vessel's route. Meanwhile, in the Middle segment, the vessel typically operates at maneuvering speeds, due to the presence of two ports located on west and east sides of the canal.

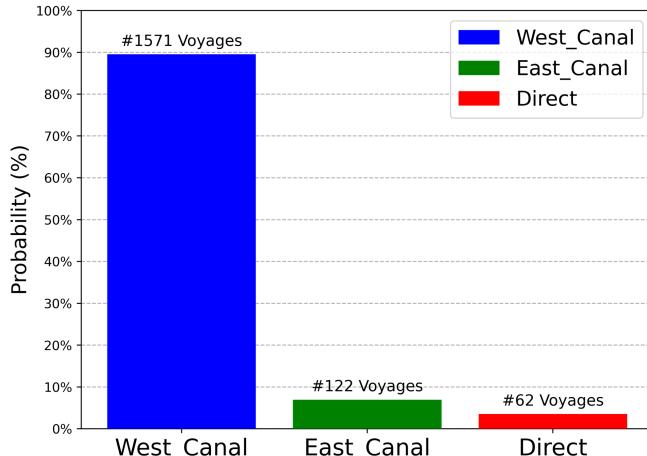
### A. Statistical Analysis

We have first conducted a statistical analysis on the dataset. Figure 12 illustrates some important statistic for all aggregated voyage, with regard to the accumulated fuel, time, and distance at different route segments. These statistics represent the routes of vessel Buro, as depicted in Figure 10.

As it can be seen from Table II, the difference of fuel consumption of the and cruising speeds is 5.47%, so that and also based on the recommendations from domain experts in compliance with maritime regulations including safety and traffic considerations, it might be more practical to primarily focus on optimizing cruising speed.



(a) Distribution of Cinderella II voyages across the five path clusters.



(b) Distribution of Buro voyages across the three path clusters.

Fig. 11: Distribution of voyages if both vessels across the the path clusters.

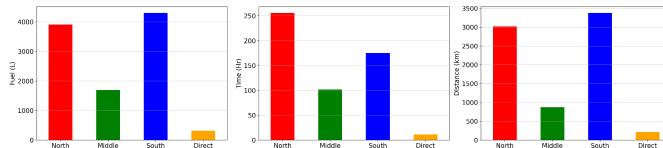


Fig. 12: Barplots for statistics of Fuel, Time, and Distance in different route segments, for vessel Buro.

### B. Modeling of Energy Efficiency

The first step in optimizing the model is to identify the best set of input parameters. We consider four cases of ANN inputs, where each case consists of different combinations of operational and weather variables. Further details about these ANN input cases are provided in Tables III and I.

The Beeswarm plot in Figure 13b indicates that the vessel's location has the most significant impact on the Efficiency Score. Therefore, a spatial analysis was conducted to identify the impact of various combinations of operational and weather variables on the Efficiency Score concerning the vessel's

TABLE II: Statistics of the dataset for different speed modes

Variable	Speeds		Difference (%)
	All	Cruising	
Fuel, total (Liter)	1329.2	1256.62	5.47%
Time, total (Hour)	48.04	22.32	53.57%
Distance, total (km)	608.72	349.2	42.68%
Speed, average (m/s)	2.67	1.67	37.5%

TABLE III: Description of the four input cases of ANN.

Inputs Case	Operational Variables	Weather Variables	
		onboard data	external sources
I	Vessel's location, speed, and direction	wind	—
II	—	—	wind, wave, and current
III	—	wind	wave and current
IV	wind	wind	wind, wave, and current

location.

The results are shown as heatmaps in Figure 14, revealing that the direct route from south to north or vice versa, located on the open sea, is particularly susceptible to the impact of weather conditions. Thus, for this direct section of vessel routes, the estimation of Efficiency Score, as shown in Figure 14b, has the highest accuracy with different input combinations.

Meanwhile, in the north section, where strong either head or tail wind is more frequent (in this area, west winds dominate) with respect to the vessel route, the Efficiency Score estimation has the second highest accuracy, as in Figure 14b.

In the other case, when it comes to estimating EngineFuelRate, as shown in Figure 14a, the results are not accurate. For instance, the direct sections of the route are not achieving the highest accuracy, even though they are supposed to experience more weather conditions than other sections due to these sections being the most similar to an open sea.

### C. Improving Voyage Efficiency

After that we have implemented the framework for improving the vessel's energy efficiency, as presented in Section IV-B. Then, we evaluate the results, and for a sake of a fair comparison, we are injected both the actual measured and optimized speed profiles into the same estimation model of energy efficiency to predict fuel and time before and after the improving framework of energy efficiency is being implemented. Once the fuel and time estimated, we compute the efficiency to determine how much energy has been saved.

One of our main metrics to evaluate the model performance of voyage efficiency optimization is the gain of efficiency scores, as represented in (7).

$$Eff.Gain = \frac{Eff.Score_{Pred.} - Eff.Score_{Meas.}}{Eff.Score_{Meas.}} \times 100 \quad (7)$$

Where  $Eff.Score_{Meas.}$  and  $Eff.Score_{Pred.}$  represent the voyage efficiency obtained with measured and predicted speed profiles, respectively.

As shown in Table IV, the HMM model achieves the highest average efficiency gain of 6.00%, followed by the 1NN-DTW model (2.82%), the kNN model (1.30%), and the LSTM model (1.28%). In terms of the number improved voyages out of 162 voyages in test dataset, the HMM model also improves the

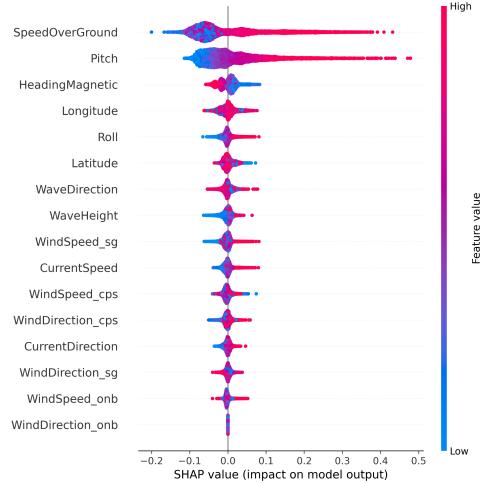
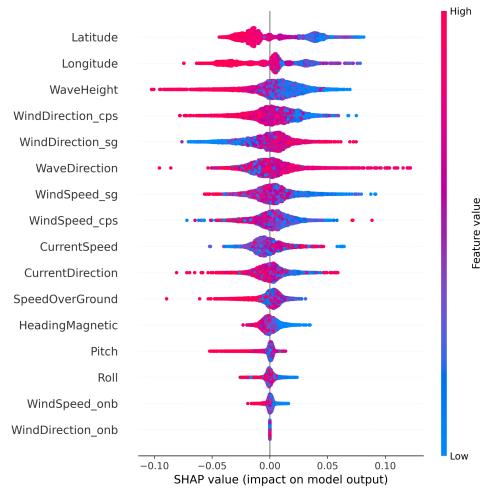
(a) Model's output is EngineFuelRate, ( $R^2=0.7615$ )(b) Model's output is Efficiency Score, ( $R^2=0.8324$ )

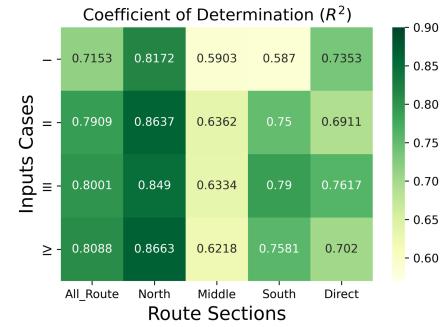
Fig. 13: Beeswarm plots of SHAP values for XGBoost regression model with different outputs

TABLE IV: Average efficiency gains (Eff. Gains %, see Eq. 7) and counts of improved voyages (IV Count #) out of 162 voyages in the test dataset.

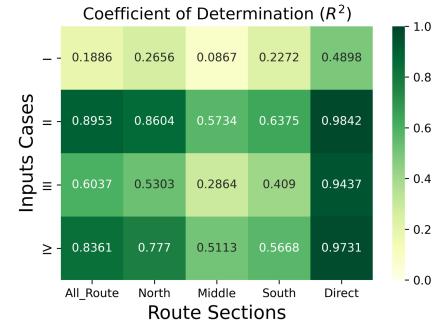
Cluster	Efficiency Score	LSTM	kNN	1NN-DTW	HMM
Top10Pr	Eff. Gains (%)	2.61	2.13	3.20	6.05
	IV Count (#)	134	114	127	139
Top25Pr	Eff. Gains (%)	2.38	1.58	3.23	1.30
	IV Count (#)	129	107	128	107
Top50Pr	Eff. Gains (%)	0.97	0.98	2.58	7.34
	IV Count (#)	100	106	117	140
Top75Pr	Eff. Gains (%)	-0.84	0.50	2.28	9.31
	IV Count (#)	60	93	119	141
Average	Eff. Gains (%)	1.28	1.30	2.82	6.00
	IV Count (#)	105.75	105.00	122.75	131.75

energy efficiency of the most average number of improved voyages (131.75 out of 162 voyages).

The HMM model achieves its best performance when trained on the Top75Pr cluster, which includes voyages with lower Eff-Scores and frequently encountered adverse weather conditions. Such performance underscores the HMM model's



(a) EngineFuelRate



(b) Efficiency Score

Fig. 14: Results ( $R^2$ ) for ANN regression with EngineFuelRate and Efficiency Score as outputs across different input cases in relation to varying vessel's route sections.

TABLE V: Average and deviation of gains (%) of Eff-Scores in three weather states, when models are trained with the four data clusters

Model	LSTM		kNN		1NN-DTW		HMM	
	Weather	Avg	Std	Avg	Std	Avg	Std	Avg
Calm	0.17	4.72	1.17	3.29	2.21	4.30	3.96	6.19
Moderate	1.53	4.12	1.26	4.19	3.4	4.47	5.33	6.56
Rough	1.94	3.48	1.42	3.65	2.84	4.90	8.17	9.08
Average	1.21	4.11	1.28	3.71	2.82	4.56	5.82	7.28

capability to learn the hidden patterns between the vessel speed and weather states, ultimately facilitates for developing more efficient speed profiles.

We also present the results by plots illustrating the predicted speed profiles for a test voyage, which are generated by four time-series based models that incorporate weather data as inputs. These models were trained using data from the Top10Pr cluster.

In summary, the HMM-based model is the most effective model for improving energy efficiency for a vessel voyage in short sea. The HMM model is able to learn the complex relationships between the input features (e.g., speed and weather) and the output feature (Eff-Score), even in different weather conditions.

#### D. Path Identification

In this section, we present the results of our spatial clustering approach for vessel path identification and discuss the implications of these findings. The approach involved

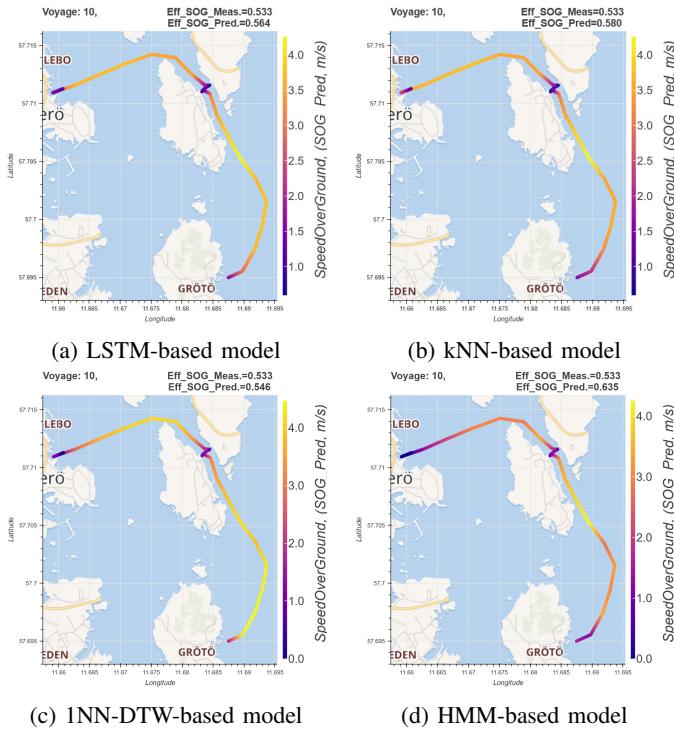


Fig. 15: Predicted speed profile for a test voyage. From four time-series based models incorporate weather data as inputs and are trained by Top10Pr cluster.

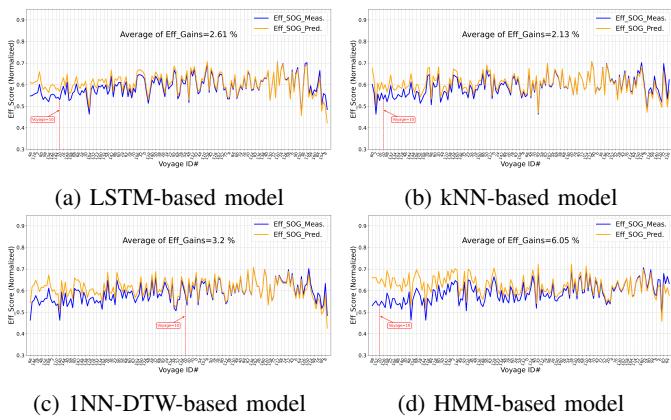


Fig. 16: Sorted Eff-Scores based on their gain, for 162 test voyages. From four time-series based models incorporate weather data as inputs and are trained by Top10Pr cluster,  $\text{Eff\_Gain}$  as in Eq. 7.

the utilization of position information and various clustering techniques, specifically k-means, hierarchical clustering, and Gaussian distributions clustering to the paths of two vessels, Buro and Cinderella II. These vessels operate in distinct locations, leading to diverse datasets.

### 1) Evaluation Metrics for Path Clustering:

The results of vessel path identification are evaluated through visual inspection and tabulation using metrics such as confusion matrix, precision, recall, and F1-score [73].

The hits and misses of path clustering are presented by the confusion matrix. For our results of path clustering, the confu-

sion matrix is a one-vs-one type matrix. Then, the confusion matrix is converted into a one-vs-all type matrix (binary-class confusion matrix) as shown in Eq. (8), for calculating class-wise metrics like precision, recall, and F1-score.

	Pred. Pos.	Pred. Neg.
Act. Pos.	TP	FN
Act. Neg.	FP	TN

(8)

where True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) are determined by comparing the predicted (Pred.) and actual (Act.) path clusters.

The confusion matrix transformation involves considering one class as positive at a time, while combining all other classes as negative. This process is repeated iteratively for each class, resulting in multiple binary-class confusion matrices.

The following performance metrics were used:

- Precision: the ratio of true positives to the total number of predicted positives.
- Recall: the ratio of true positives to the total number of actual positives in the data.
- F1-score: the harmonic mean of precision and recall.

The equations for precision, recall, and F1-score are shown in Eqs. (9), (10), and (11).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{F1-score} = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (11)$$

### 2) Results of Path Clustering for Cinderella II Vessel:

Table VI shows the results of applying k-means or Gaussian Mixture Model (GMM) models to identify the vessel paths from the distance matrix in the distance-based method of the path identification approach. Notably, the paths with clusters of North-West, South, and South-West achieved an F1-score of 1.0, indicating that the approach correctly identified all the paths of these clusters.

In contrast, the North-East and North-Middle paths exhibited lower F1 scores compared to other clusters. The path cluster of North-Middle is the most challenging path to identify since six such paths have been clustered as North-East, as can be seen by comparing Figures 17 and 18, which are the visualization for all the paths, color-marked based on their ground truth clusters. Figure 19 illustrates the probability distribution of mis-clustered paths with respect to latitude and longitude coordinates. It is obvious that these paths have nearly identical coordinates, which makes them challenging paths to cluster with k-means or GMM. This suggests that there is still room for improvement by using other ML clustering methods.

Table VII presents the results of employing hierarchical clustering to the distance matrix in the distance-based method for clustering the vessel paths. In hierarchical clustering, there is a parameter called "Dendrogram Cut-off threshold," and its value should be selected depending on the number of path clusters. Hence, as illustrated in Figure 20, this parameter is

denoted by the Y-axis as a clustering height, and its value is set to 100 for clustering the vessel path into five clusters.

Remarkably, all path clusters achieved an F1-score of 1, indicating that hierarchical clustering successfully identified all paths with high accuracy from the distance matrix using the distance-based method. This suggests that the choice of ML clustering technique with the distance matrix can influence the accuracy of path identification. Table VII displays the outcomes of clustering, now by applying the segmented likelihood Gaussian method. This method achieved perfect precision, recall, and F1-score for all path clusters.

Figures 21, 22, and 23 present visualizations for the segmented Gaussian likelihood method.

The accuracy in results by hierarchical and segmented Gaussian likelihood clustering for path clusters indicates the efficacy of the developed approach of spatial clustering for vessel path identification.

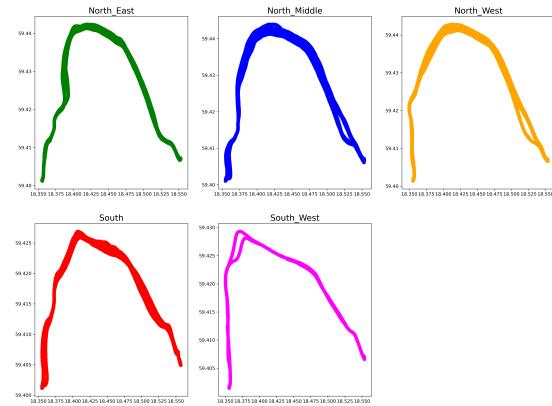


Fig. 17: Display of the five clusters of path for Cinderella II vessel, (ground truth).

TABLE VI: Result of implementing of both k-means and GMM clustering of the paths of Cinderella II vessel to five clusters.

(a) Precision, Recall, and F1-score

Paths	Precision	Recall	F1-score
North-East (NE)	0.7	1	0.824
North-Middle (NM)	1	0.85	0.919
North-West (NW)	1	1	1
South (S)	1	1	1
South-West (SW)	1	1	1

(b) Confusion Matrix

Actual	Predicted					Total
	NE	NM	NW	S	SW	
NE	14	0	0	0	0	14
NM	6	34	0	0	0	40
NW	0	0	16	0	0	16
S	0	0	0	52	0	52
SW	0	0	0	0	2	2
<b>Total</b>	<b>20</b>	<b>34</b>	<b>16</b>	<b>52</b>	<b>2</b>	<b>124</b>

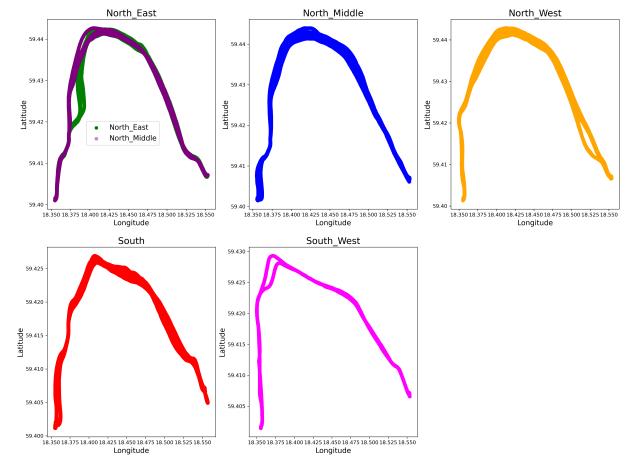


Fig. 18: Results of both k-means and GMM clustering to five path clusters for Cinderella II.

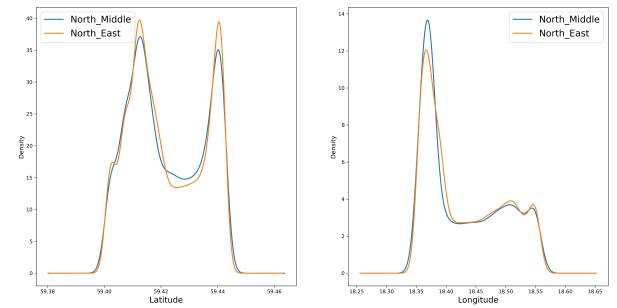


Fig. 19: Probability distribution of location coordinates for mis-clustered paths of Cinderella II by both k-means and GMM.

### 3) Results of Path Clustering for Buro Vessel:

In this part the results from applying the framework of both hierarchical clustering and segmented likelihood Gaussian clustering of Buro vessel paths to three clusters. Following a similar evaluation procedure as in the case of Cinderella II vessel.

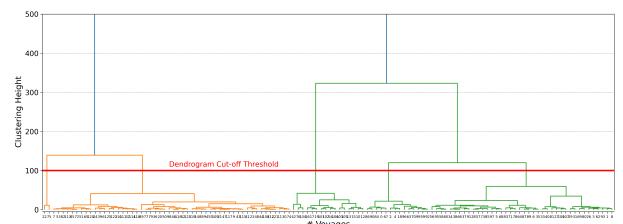


Fig. 20: Results of hierarchical clustering the paths of Cinderella vessel to five clusters.

TABLE VII: Result of implementing both hierarchical clustering and segmented likelihood Gaussian clustering of paths for Cinderella II vessel into five clusters.

(a) Precision, Recall, and F1-score

Paths	Precision	Recall	F1-score
North-East (NE)	1	1	1
North-Middle (NM)	1	1	1
North-West (NW)	1	1	1
South (S)	1	1	1
South-West (SW)	1	1	1

(b) Confusion Matrix

Actual	Predicted					Total
	NE	NM	NW	S	SW	
NE	14	0	0	0	0	14
NM	0	40	0	0	0	40
NW	0	0	16	0	0	16
S	0	0	0	52	0	52
SW	0	0	0	0	2	2
<b>Total</b>	<b>14</b>	<b>40</b>	<b>16</b>	<b>52</b>	<b>2</b>	<b>124</b>

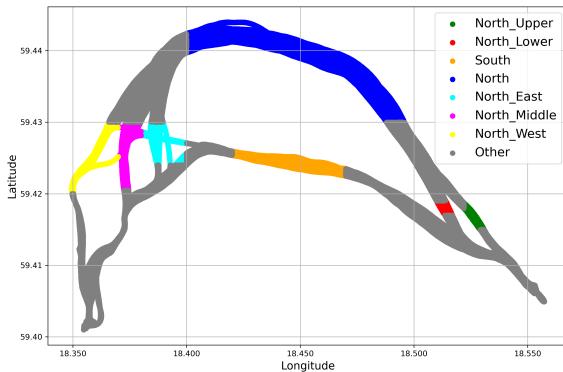


Fig. 21: Distribution of the route of Cinderella II into eight segments.

As it can be seen, the results of likelihood Gaussian clustering also achieved F1-score of 1, as in Cinderella vessel. But for the hierarchical clustering in Buro vessel case has F1-scores of 0.957 and 0.996 for clustering East\_Canal and West\_Canal paths respectively. This results can be considered remarkable

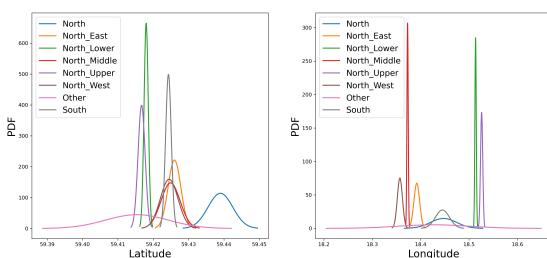


Fig. 22: Probability distributions of location coordinates for the eight segments of the route of Cinderella II.

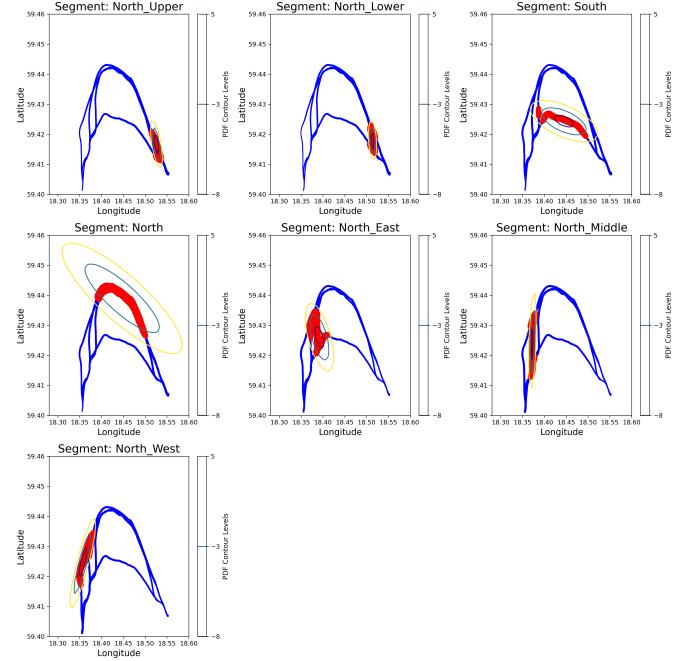


Fig. 23: Gaussian distributions for seven segments of Cinderella II vessel route.

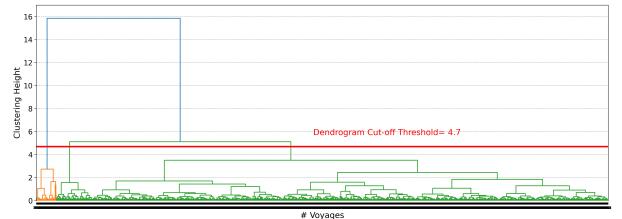


Fig. 24: Results of hierarchical clustering to three path clusters of Buro vessel.

when take into account that these two paths, East\_Canal and West\_Canal, are challenging to be clustered, since they have several paths that are slightly different from each other. Figure 28 shows the heatmap of the distance matrix of 12 sample paths for Buro vessel. Notably, the East\_Canal and West\_Canal paths exhibit high similarity, which make not easy clustering task.

Choosing a proper value of the agglomerative threshold for the hierarchical clustering to get the three clusters, as shown in Figure 24, with an agglomerative threshold = 4.7.

## VII. CONCLUSION

In this section, we present briefly the key findings of this project and provide recommendations for future research or practical implications for the maritime industry.

### A. Modeling and Improving Voyage Energy Efficiency

- The modeling approach by using the Efficiency Score, instead of directly working with the EngineFuelRate

TABLE VIII: Result of Buro vessel by implementing hierarchical clustering to three path clusters

(a) Precision, Recall, and F1-score

Paths	Precision	Recall	F1-score
Direct	1	1	1
East_Canal	0.917	1	0.957
West_Canal	1	0.993	0.996

(b) Confusion Matrix

Actual	Predicted			Total
	Direct	E-C	W-C	
Direct	62	0	0	62
E-C	0	122	0	122
W-C	0	11	1560	1571
Total	62	133	1560	1755

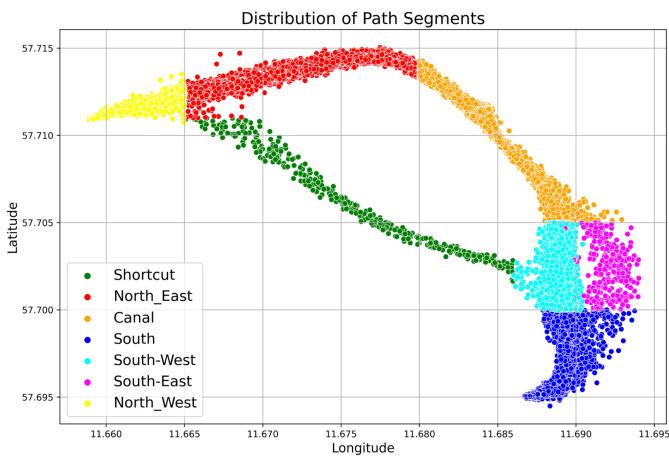


Fig. 25: Distribution of the route of Buro vessel into seven segments.

onboard signal, is more effective in facilitating decision-making.

- The resulting model is based on a more comprehensive understanding of the critical factors that impact fuel consumption, both temporally and spatially, resulting in more dependable counterfactual predictions.
- The quantitative evaluation indicates that estimating the Efficiency Score produces more precise and less biased

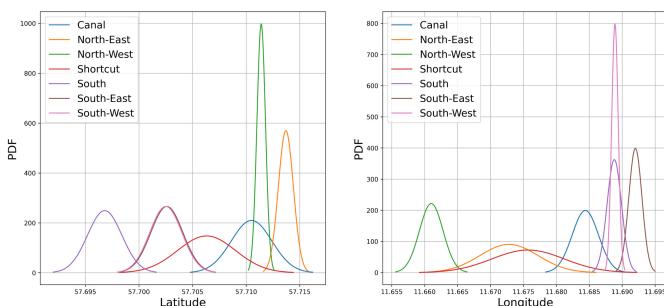


Fig. 26: Probability distributions of location coordinates for the seven segments of the route of Buro vessel.

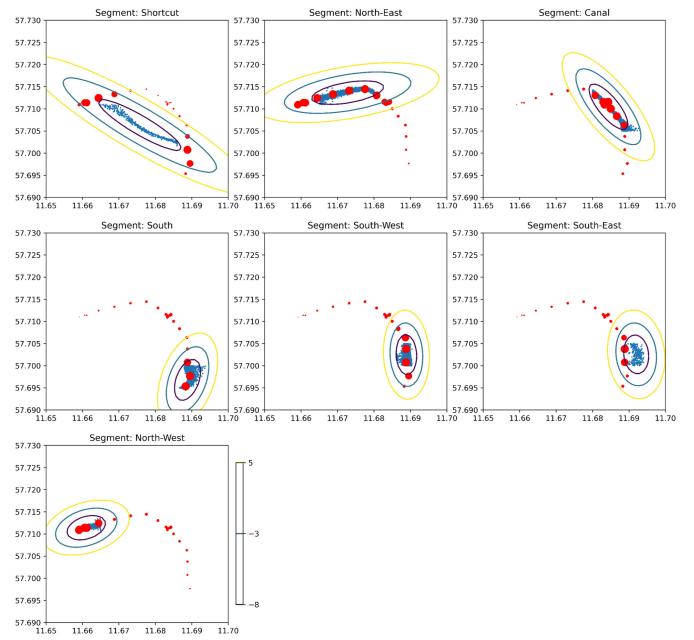


Fig. 27: Gaussian distribution for seven segments of the route of Buro vessel.

TABLE IX: Result of Buro vessel by implementing segmented likelihood Gaussian clustering to three path clusters

(a) Precision, Recall, and F1-score

Paths	Precision	Recall	F1-score
Direct	1	1	1
East_Canal	1	1	1
West_Canal	1	1	1

(b) Confusion Matrix

Actual	Predicted			Total
	Direct	E-C	W-C	
Direct	62	0	0	62
E-C	0	122	0	122
W-C	0	0	1571	1571
Total	62	122	1571	1755

outcomes than estimating the measured EngineFuelRate.

- The study employs four distinct models: LSTM, kNN, 1NN-DTW, and HMM, to optimize vessel speed profiles with the objective of enhancing energy efficiency in short sea voyages. The key observation is that model performance varies significantly across these algorithms. However, the performance of the models varies depending on the data cluster used to train the model and the weather conditions.
- We developed a data-driven framework for optimizing vessel speed profiles to improve energy efficiency in SSS. The framework integrates a data-driven modeling approach to energy efficiency with the 1NN-DTW algorithm. We evaluated the added value of the framework using a real-world dataset and found that it can effectively improve vessel energy efficiency, especially with limited options, which are common in short-sea shipping.

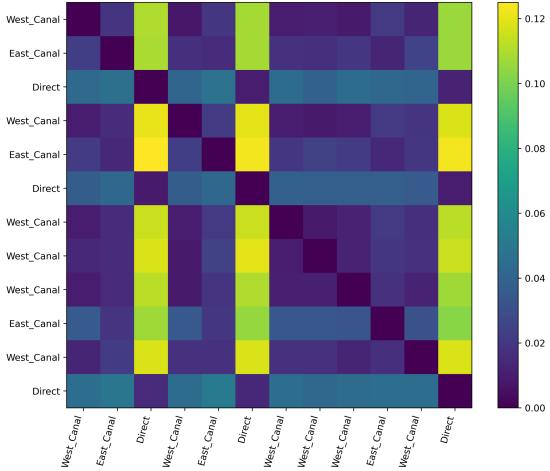


Fig. 28: Heatmap of the distance matrix for 12 paths of Buro vessel.

- 1NN-DTW exhibits the ability to capture temporal dependencies within speed profiles, especially within the constraints of short-sea shipping where opportunities for actively controlling the vessel to enhance its energy efficiency are restricted.
- Although the kNN can handle multivariate data and incorporating additional features like weather conditions, in this case study, the 1NN-DTW performs better due to its specialized handling of time-dependent data and inherent patterns.
- The result findings emphasize that in terms of searching the best behavior of vessel from the observed data, the 1NN-DTW exhibits superior performance compared to LSTM and kNN, since the 1NN-DTW selects the best measured speed profiles. On the other hand, the HMM is the most effective approach in our study, where the HMM optimizes these measured speed profiles further by offering strategies to them informed by their weather states.
- The study also reveals that the HMM model exhibits notable adaptability to different weather states (Calm, Moderate, and Rough). In each weather state, the HMM consistently delivers efficiency gains, indicating its ability to adapt speed profiles according to varying environmental conditions. This adaptability is crucial for real-world maritime applications, where weather can change rapidly.

#### B. Vessel Path Identification

- The approach is able to identify the vessel paths with partially defined or unknown paths.
- In the distance-based method, the hierarchical clustering used in the approach outperforms k-means and GMM clustering techniques.
- The approach of hierarchical clustering includes a user-customized, a cut-off threshold, which allows desired

control for the number of path clusters, enhancing the flexibility and adaptability of the proposed approach.

- In the distance-based method, adopting ANND as a measure of similarity makes path clustering less affected by noise or outliers and provides a more intuitive interpretation of path similarity, ultimately enhancing the robustness and interpretability of our approach.
- The segmented Gaussian likelihood method is particularly useful for identifying and analyzing the vessel path alterations at different segments of the vessel route.
- The proposed approach is computationally efficient and has the potential to be a valuable tool for planning vessel paths. Accurate path identification can contribute to safer and more efficient maritime transportation practices, aiding in route planning, collision avoidance, and navigation optimization.
- Nevertheless, the framework has some potential limitations, such as the segmented Gaussian likelihood method exhibiting sensitivity to segment definition which could affect its salable performance, particularly in complex maritime scenarios. Moreover, while the study case demonstrates that the framework is computationally efficient, it is essential to discuss any potential scalability issues, especially when dealing with large datasets, since the computational efficiency may vary depending on the dataset size and the nature of paths.

#### C. Future work and Recommendations

Future work could include the following:

- For modeling of vessel energy efficiency, considering spatial dimension, such as distance variable, might be worth of investigation.
- Including the use of heuristic algorithms such as genetic programming, for optimizing voyage efficiency.
- Exploring the scalability and real-world applicability of the proposed path clustering approach, as well as its integration with related systems of maritime transportation.
- Applying incremental map-matching algorithms for real-time vessel path identification.
- Future research in path detection may involve the study of graph theory and the application of evolutionary algorithms, including ant colony optimization.

#### ACKNOWLEDGMENT

This research work has been funded and supported by Vinnova. We would like to thank CetaSol AB for their support and for providing the resources necessary to conduct this research.

We also wish to thank the diverse group at the Center for Applied Intelligent Systems Research (CAISR), Halmstad University, for helpful discussions.

#### SUPPLEMENTARY MATERIALS

The source codes that are implemented on Python 3.9.7 to produce the results of case studies are available as following:

- For modeling energy efficiency at <https://github.com/MohamedAbuella/ST4EESSS>;
- For voyage optimization at <https://github.com/MohamedAbuella/TSA4EESSS>;
- For path identification at [https://github.com/MohamedAbuella/Path\\_Clustering](https://github.com/MohamedAbuella/Path_Clustering).

## REFERENCES

- [1] "Development of short sea shipping," [Online]. Available: <https://www7.transportation.gov/testimony/development-short-sea-shipping>, Feb. 2007.
- [2] P. Donner and T. Johansson, "Sulphur directive, short sea shipping and corporate social responsibility in a eu context," *corporate social responsibility in the maritime industry*, pp. 149–166, 2018.
- [3] J. D. Ampah, A. A. Yusuf, S. Afrane, C. Jin, and H. Liu, "Reviewing two decades of cleaner alternative marine fuels: Towards imo's decarbonization of the maritime transport sector," *Journal of Cleaner Production*, vol. 320, p. 128871, 2021.
- [4] Eurostat, "Short sea shipping - country level - gross weight of goods transported to/from main ports," [Online]. Available: [https://ec.europa.eu/eurostat/databrowser/view/mar\\_sg\\_am\\_cw/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/mar_sg_am_cw/default/table?lang=en), 2023.
- [5] W. Bank, *Accelerating Digitalization: Critical Actions to Strengthen the Resilience of the Maritime Supply Chain*. Washington: World Bank, 2020.
- [6] C. IMO, "Fourth imo ghg study 2020," 2020.
- [7] T. P. Zis, H. N. Psaraftis, and L. Ding, "Ship weather routing: A taxonomy and survey," *Ocean Engineering*, vol. 213, p. 107697, 2020.
- [8] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, 2017.
- [9] Y. Zheng, "Trajectory data mining: an overview," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 3, pp. 1–41, 2015.
- [10] L. M. Petry, C. A. Ferrero, L. O. Alvares, C. Renso, and V. Bogorny, "Towards semantic-aware multiple-aspect trajectory similarity measuring," *Transactions in GIS*, vol. 23, no. 5, pp. 960–975, 2019.
- [11] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis *et al.*, "Semantic trajectories modeling and analysis," *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, pp. 1–32, 2013.
- [12] J.-G. Lee, J. Han, and K.-Y. Whang, "Trajectory clustering: a partition-and-group framework," in *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 2007, pp. 593–604.
- [13] Y. Zheng and X. Zhou, *Computing with spatial trajectories*. Springer Science & Business Media, 2011.
- [14] R. Yan, S. Wang, L. Zhen, and G. Laporte, "Emerging approaches applied to maritime transport research: Past and future," *Communications in Transportation Research*, vol. 1, p. 100011, 2021.
- [15] P. R. Bellingmo, A. Pobitzer, U. Jørgensen, and S. P. Berge, "Energy efficient and safe ship routing using machine learning techniques on operational and weather data," in *20th International Conference on Computer Applications and Information Technology in the Maritime Industries*, 2021.
- [16] U. Jørgensen, P. R. Bellingmo, B. Murray, S. P. Berge, and A. Pobitzer, "Ship route optimization using hybrid physics-guided machine learning," in *Journal of Physics: Conference Series*, vol. 2311, no. 1. IOP Publishing, 2022, p. 012037.
- [17] K. Sugimoto, "Digital twin for monitoring remaining fatigue life of critical hull structures."
- [18] M. Haranen, S. Myöhänen, and D. S. Cristea, "The role of accurate now-cast data in ship efficiency analysis," in *2nd Hull Performance & Insight Conference*, 2017, pp. 25–38.
- [19] A. Zakaria, A. Md Arof, and A. Khabir, "Instruments utilized in short sea shipping research: A review," *Design in Maritime Engineering*, pp. 83–108, 2022.
- [20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
- [21] X. Lang, D. Wu, and W. Mao, "Comparison of supervised machine learning methods to predict ship propulsion power at sea," *Ocean Engineering*, vol. 245, p. 110387, 2022.
- [22] D. Kim, G. Antaraksa, M. P. Handayani, S. Lee, and J. Lee, "Explainable anomaly detection framework for maritime main engine sensor data," *Sensors*, vol. 21, no. 15, p. 5200, 2021.
- [23] M. Veerappa, M. Anneken, N. Burkart, and M. F. Huber, "Validation of xai explanations for multivariate time series classification in the maritime domain," *Journal of Computational Science*, vol. 58, p. 101539, 2022.
- [24] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.
- [25] H. Chen and P. Ballou, "Art and science of ship voyage optimization: A critical review," 2021.
- [26] L. Walther, A. Rizvanoli, M. Wendebourg, and C. Jahn, "Modeling and optimization algorithms in ship weather routing," *International Journal of e-Navigation and Maritime Economy*, vol. 4, pp. 31–45, 2016.
- [27] A. Fan, J. Yang, L. Yang, D. Wu, and N. Vladimir, "A review of ship fuel consumption models," *Ocean Engineering*, vol. 264, p. 112405, 2022.
- [28] M. H. Moradi, M. Brutsche, M. Wenig, U. Wagner, and T. Koch, "Marine route optimization using reinforcement learning approach to reduce fuel consumption and consequently minimize co2 emissions," *Ocean Engineering*, vol. 259, p. 111882, 2022.
- [29] M. Grifoll, F. X. Martínez de Osés, and M. Castells, "Potential economic benefits of using a weather ship routing system at short sea shipping," *WMU Journal of Maritime Affairs*, vol. 17, no. 2, pp. 195–211, 2018.
- [30] L. Walther, A. Rizvanoli, M. Wendebourg, and C. Jahn, "Modeling and optimization algorithms in ship weather routing," *International Journal of e-Navigation and Maritime Economy*, vol. 4, pp. 31–45, 2016.
- [31] T. W. Smith, J. Jalkanen, B. Anderson, J. Corbett, J. Faber, S. Hanayama, E. O'keeffe, S. Parker, L. Johansson, L. Aldous *et al.*, "Third imo greenhouse gas study 2014," 2015.
- [32] M. Wingrove, "Owners cut fuel costs by 10 per cent," vol. 10, no. 4, p. 42–43, 2016.
- [33] J. Huotari, T. Manderbacka, A. Ritari, and K. Tammi, "Convex optimisation model for ship speed profile: Optimisation under fixed schedule," *Journal of Marine Science and Engineering*, vol. 9, no. 7, p. 730, 2021.
- [34] K. Wang, J. Wang, L. Huang, Y. Yuan, G. Wu, H. Xing, Z. Wang, Z. Wang, and X. Jiang, "A comprehensive review on the prediction of ship energy consumption and pollution gas emissions," *Ocean Engineering*, vol. 266, p. 112826, 2022.
- [35] L. Walther, *Development of a Weather Routing System for Analysis and Optimization of Ship Voyages*. Fraunhofer Verlag, 2021.
- [36] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artificial Intelligence Review*, vol. 47, pp. 123–144, 2017.
- [37] G. Pallotta, M. Vespe, and K. Bryan, "Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, 2013.
- [38] Z. Yan, Y. Xiao, L. Cheng, R. He, X. Ruan, X. Zhou, M. Li, and R. Bin, "Exploring ais data for intelligent maritime routes extraction," *Applied Ocean Research*, vol. 101, p. 102271, 2020.
- [39] L. Zhao and G. Shi, "A novel similarity measure for clustering vessel trajectories based on dynamic time warping," *The Journal of Navigation*, vol. 72, no. 2, pp. 290–306, 2019.
- [40] G. K. D. De Vries and M. Van Someren, "Machine learning for vessel trajectories using compression, alignments and domain knowledge," *Expert Systems with Applications*, vol. 39, no. 18, pp. 13 426–13 439, 2012.
- [41] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, "Deep learning methods for vessel trajectory prediction based on recurrent neural networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4329–4346, 2021.
- [42] H. Li and Z. Yang, "Incorporation of ais data-based machine learning into unsupervised route planning for maritime autonomous surface ships," *Transportation Research Part E: Logistics and Transportation Review*, vol. 176, p. 103171, 2023.
- [43] H. Li, H. Jiao, and Z. Yang, "Ais data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods," *Transportation Research Part E: Logistics and Transportation Review*, vol. 175, p. 103152, 2023.
- [44] N. Zygouras, A. Troupiotis-Kapeliaris, and D. Zissis, "Envclus\*: Extracting common pathways for effective vessel trajectory forecasting," *IEEE Access*, 2024.

- [45] C. Huang, X. Qi, J. Zheng, R. Zhu, and J. Shen, "A maritime traffic route extraction method based on density-based spatial clustering of applications with noise for multi-dimensional data," *Ocean Engineering*, vol. 268, p. 113036, 2023.
- [46] L. Wang, P. Chen, L. Chen, and J. Mou, "Ship ais trajectory clustering: An hdbscan-based approach," *Journal of Marine Science and Engineering*, vol. 9, no. 6, p. 566, 2021.
- [47] L. Eljabu, M. Etemad, and S. Matwin, "Spatial clustering model of vessel trajectory to extract sailing routes based on ais data," *International Journal of Computer and Systems Engineering*, vol. 16, no. 10, pp. 491–501, 2022.
- [48] ———, "Spatial clustering method of historical ais data for maritime traffic routes extraction," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 893–902.
- [49] X. Han, C. Armenakis, and M. Jadidi, "Modeling vessel behaviours by clustering ais data using optimized dbSCAN," *Sustainability*, vol. 13, no. 15, p. 8162, 2021.
- [50] F. Farahnakian, F. Nicolas, F. Farahnakian, P. Nevalainen, J. Sheikh, J. Heikkonen, and C. Raduly-Baka, "A comprehensive study of clustering-based techniques for detecting abnormal vessel behavior," *Remote Sensing*, vol. 15, no. 6, p. 1477, 2023.
- [51] A. Moavinis, A. Gounaris, and I. Constantinou, "Detection of anomalous trajectories for vehicle traffic data," in *Proceedings of the Workshops of the EDBT/ICDT 2023 Joint Conference, Ioannina, Greece*, 2023.
- [52] I. Varlamis, I. Kontopoulos, K. Tserpes, M. Etemad, A. Soares, and S. Matwin, "Building navigation networks from multi-vessel trajectory data," *GeoInformatica*, vol. 25, pp. 69–97, 2021.
- [53] H. Liu, X. Chen, Y. Wang, B. Zhang, Y. Chen, Y. Zhao, and F. Zhou, "Visualization and visual analysis of vessel trajectory data: A survey," *Visual Informatics*, vol. 5, no. 4, pp. 1–10, 2021.
- [54] X. Zhang, X. Fu, Z. Xiao, H. Xu, and Z. Qin, "Vessel trajectory prediction in maritime transportation: Current approaches and beyond," *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- [55] "CetaSol AB ,," [Online]. Available: <https://cetasol.com>.
- [56] "Marine Traffic," [Online]. Available: <https://www.marinetraffic.com/en/ais/details/ships/shipid:1088282/mmsi:265513810/imo:8602713/vessel:BURO>, Jul. 2022.
- [57] "Copernicus Marine Service ,," [Online]. Available: <https://marine.copernicus.eu>.
- [58] "StormGlass API ,," [Online]. Available: <https://stormglass.io>.
- [59] J. Carlton, *Marine propellers and propulsion*, 2nd ed. Oxford: Butterworth-Heinemann, January 2007.
- [60] M. Abuella, M. A. Atoui, S. Nowaczyk, S. Johansson, and E. Faghani, "Data-driven explainable artificial intelligence for energy efficiency in short-sea shipping," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2023, pp. 226–241.
- [61] M. Abuella, H. Fanaee, S. Nowaczyk, S. Johansson, and E. Faghani, "Time-series analysis approach for improving energy efficiency of a fixed-route vessel in short-sea shipping," *arXiv preprint arXiv:2402.00698*, 2024.
- [62] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [63] R. C. Staudemeyer and E. R. Morris, "Understanding lstm—a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.
- [64] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [65] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd international conference on knowledge discovery and data mining*, 1994, pp. 359–370.
- [66] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and information systems*, vol. 7, pp. 358–386, 2005.
- [67] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *Proceedings of the VLDB Endowment*, vol. 1, no. 2, pp. 1542–1552, 2008.
- [68] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 262–270.
- [69] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data mining and knowledge discovery*, vol. 31, pp. 606–660, 2017.
- [70] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [71] M. Abuella, M. Atoui, S. Nowaczyk, S. Johansson, and E. Faghani, "Spatial clustering approach for vessel path identification," *arXiv preprint arXiv:2403.05778*, 2024.
- [72] "Marine Traffic," [Online]. Available: [https://www.marinetraffic.com/en/ais/details/ships/shipid:322946/mmsi:265609540/imo:8619821/vessel:CINDERELLA\\_II](https://www.marinetraffic.com/en/ais/details/ships/shipid:322946/mmsi:265609540/imo:8619821/vessel:CINDERELLA_II), Nov. 2023.
- [73] R. Yan, S. Wang, and C. Peng, "An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities," *Journal of Computational Science*, vol. 48, p. 101257, 2021.



**Mohamed Abuella** received his M.S. and PhD degrees in Electrical and Computer Engineering from Southern Illinois University at Carbondale and University of North Carolina at Charlotte, in 2012 and 2018 respectively. He is a postdoctoral researcher at Halmstad University since 2022. His research interests include energy analytics and AI for sustainability.



**Hadi Fanaee-T** is an Associate Professor at Halmstad University, Sweden. He completed his PhD (with distinction) in Computer Science, under supervision of Professor Joao Gama at the Faculty of Science of University of Porto, Portugal in November 2015. He was the finalist in ERCIM Cor Baayen Young Researcher Award 2017. Prior to this position he was a postdoctoral fellow at the department of biostatistics, University of Oslo, Norway, and also worked as a postdoctoral researcher in European FP7 Project "MAESTRA" at INESC TEC research institute, Portugal. His main research interests are interdisciplinary applications of tensor decompositions, data fusion, anomaly/event detection and spatiotemporal data mining. He is the first-author of several journal and conference papers. He has served as a PC member to over 20 prestigious conferences (e.g. IJCAI, AAAI, ECML-PKDD, ISMIS, ACM SAC, IEEE DSAA, etc.) and also reviewer to several high-impact journals (e.g. TDKE, TKDD, DAMI, ML, KAIS, KBS, CSUR).



**M. AMINE ATOUI** obtained his Ph.D. degree from LARIS, Polytech' Angers, France, in 2015. Currently, he is affiliated with the Center for Applied Intelligent Systems Research at Halmstad University, Sweden. His research interests encompass probabilistic and explainable Machine Learning, causal and Bayesian Inference, transmission/communication, and automatic control.



**Slawomir Nowaczyk** is a Professor in Machine Learning at the Center for Applied Intelligent Systems Research, Halmstad University, Sweden. He received his MSc degree from Poznan University of Technology in 2002 and his PhD from the Lund University of Technology in 2008. During the last decades, his research has focused on machine learning, knowledge representation, and self-organising systems. The majority of his work concerns industrial data streams, often with predictive maintenance as the goal. Given that accurate and relevant labels

are usually impossible to obtain in such settings, Slawomir's contributions primarily take advantage of proxy labels, such as transfer learning and multi-task learning, or concern semi-supervised and unsupervised modelling. He is a board member of the Swedish AI Society and a research leader for the School of Information Technology at Halmstad University. Slawomir has led multiple research projects on applying Artificial Intelligence and Machine Learning in different domains, such as transport and automotive, energy, smart cities, and healthcare. In most cases, this research was done in collaboration with industry and public administration organisations – inspired by practical challenges and leading to tangible results and deployed solutions.



**Simon Johansson** is a MSc graduate of Chalmers University of Technology's program in Engineering Mathematics and Computational Science in 2020, currently works in Cetasol, a marine company specialising in CO<sub>2</sub> reduction and energy optimisation for vessels. His practical application of computational methods and dedication to environmental sustainability align with his role, contributing to global efforts to mitigate climate change. Simon's commitment to advancing eco-friendly solutions in the marine industry reflects a seamless integration of academic excellence and real-world impact.



**Ethan Faghani** is the CEO and founder of Cetasol. Before Cetasol, Ethan was Chief Engineer of Automation and AI at Volvo Penta. Ethan has experience working with cutting-edge technologies in other transportation segments in both big enterprises and his own founded startup. Ethan obtained his Ph.D. in mechatronics from UBC and Innovation and Entrepreneurship from Stanford Business School.