Chapter 6
Enhanced and Holistic Voyage Planning Using Digital Twins

Dimitris Kaklis
DANAOS, Greece

Antonis Antonopoulos
Konnecta, Greece

ABSTRACT

The chapter explains techniques and approaches to optimize a ship's voyage in terms of environmental and business parameters, utilizing the digital twin (DT) concept. It demonstrates how voyage planning and navigation management, in general, is enhanced by taking into account vessel state in real time as reflected and analyzed by the digital twin ecosystem. The theoretical backbone of voyage planning entails a multitude of state-of-the-art processes from trajectory mining and path finding algorithms to multi constraining optimization by including a variety of parameters to the initial problem, such as weather avoidance, bunkering, Just in Time (JIT) arrival, predictive maintenance, as well as inventory management and charter party compliance. In this chapter, the authors showcase pertinent literature regarding navigation management as well as how the envisaged DT platform can redesign voyage planning incorporating all the aforementioned parameters in a holistic digital replica of the en-route vessel, eventually proposing mitigation solutions to improve operational efficiency in real-time, through simulation, reasoning, and analysis.

TRADITIONAL METHODS AND CHALLENGES IN VOYAGE PLANNING

Voyage planning involves the holistic enhancement and optimization of a vessel voyage by considering various factors such as weather conditions, fuel efficiency, vessel performance, cargo considerations, and safety protocols. It encompasses a broad perspective and aims to create an overall efficient and effective voyage experience. A subset of voyage planning is route optimization that aims to determine the path that minimizes travel time, reduces fuel consumption, and enhances overall operational efficiency,

DOI: 10.4018/978-1-6684-9848-4.ch006

taking into account factors like weather patterns, currents, wind conditions, traffic congestion, and other navigational challenges.

In recent years, there have been several state-of-the-art solutions for route optimization and weather routing in the maritime sector. One popular approach to routing optimization is to use dynamic programming that takes into account various parameters, such as the ship's speed, the sea conditions, and the distance to the destination. These models can be used to generate optimized routes that minimize fuel consumption while ensuring that the ship arrives at its destination on time. One example of such a model is the BunkerOpt system developed by DNV GL, which uses a mathematical model to calculate the optimal speed and route for a ship based on real-time data.

Another approach to routing optimization is to use machine learning algorithms to predict weather patterns and optimize routing accordingly. For example, the Weather Intelligence for Shipping (WIS) system developed by StormGeo uses machine learning to predict weather conditions up to ten days in advance, allowing ships to adjust their routes to avoid adverse weather conditions and reduce fuel consumption.

Pertinent literature regarding smart and efficient transportation in the industry (supply chain optimization, autonomous vehicle, logistics management, voyage planning) as well as regarding societal frameworks (traffic management, public transportation networks, suburban mobility), concerns a variety of multi-constraint optimization methods varying from Genetic Algorithms (Xing Wei and and Huang, 2021), Simulated annealing (Fermani et al.,2021) and Particle Swarm Optimization (Chondrodima et al., 2020) to AI integrated decision making using Reinforcement Learning (Kaklis et al., 2022, Kaklis et al., 2023, Bai et al., 2019) or Natural Language Processing (NLP) based approaches (Garg et al., 2021). The aforementioned practices and methodologies attempt to exploit the proliferation of IoT devices and state-of-the-art communication networks (5G) to build self-contained Information Hubs and provide a sustainable, safer and cost-effective transportation.

In the maritime sector, the optimization criteria adopted in the context of the ship routing problem deal with the minimization of voyage time, fuel consumption (or Fuel Oil Consumption, FOC) and voyage risk. The approaches, which have appeared so far in the literature, can be classified into three broader categories:

- Vessel-based optimization, which aims in optimizing a given route with respect to vessel characteristics, e.g., vessel speed, main-engine rotational speed, draft, trim and sea-keeping behavior: roll, heave and pitch motions, (Roh and Lee, 2018);
- Environmental-based optimization, which aims in optimizing a given route by taking into account environmental conditions, e.g., wind (speed, direction), wave (height, frequency, direction), currents. (Kim and Kim, 2017);
- Holistic optimization that combine the two previous approaches in a common context. (Vettor and Soares, 2015; Varelas et al., 2013).
- Analytical approaches trying to tackle the problem with the use of exact (NP-complete) and/or heuristic algorithms like label-setting algorithms, non-linear integer programming, or simulated annealing (Shin et al., 2020).

In order to incorporate more constraints, several methods split a vessel's voyage into areas of critical interest, involving for example zones of extreme weather conditions, emission control areas (ECAs, SECAs), high-risk zones (piracy), etc. Then, they seek for Pareto optimal solutions from a set of routes that are optimal in terms of Expected Time of Arrival (ETA), FOC, and safety, or they use Genetic Algorithms (Kim et al., 2017) in order to find the best route, as a composition of optimal route segments.

Methods like PSO (Particle Swarm Optimization) (Zhao et al., 2020) are also employed in order to solve the multi-constraint, non-linear optimization problem of optimal route planning.

The techniques employed in the literature for estimating FOC based on vessel characteristics and/or environmental conditions can be grouped into the following categories:

- Data-oriented approaches that combine vessel-trajectory data, gathered from sensors, satellites (AIS data), or Noon Reports, with Machine and Deep-Learning algorithms. These techniques range from simple Regression analysis like Support Vector Regression, Lasso Regression, and Polynomial Regression to ensemble non-parametric schemes like Random Forest (RF) regression, Decision Trees, or AdaBoost. Some studies have also experimented with baseline sequential Artificial Neural Networks (ANN) by tuning a number of hyperparameters (learning rate, number of neurons, number of layers, activation function). (Jeon et al., 2018, Gkerekos et al., 2019).
- Approaches where machine learning (ML) methods (also known as black-box models - BBM), are combined with theoretical models (also known as white-box models - WBM), such as the equations of motion of a freely floating body moving with constant forward speed, in order to increase the prediction accuracy. The proposed models are known as gray-box models (GBM) (Coraddu et al., 2017, Kaklis et al., 2019).

ANNs have been at the center of attention lately in many research areas. As far as vessel FOC is concerned, not many studies utilize the computational power of ANNs to approximate FOC mainly due to the problem of missing historical data. The studies found in pertinent literature dealing with FOC estimation from a deep learning perspective are presented briefly below. Some studies experiment with baseline sequential ANNs by applying a dropout in the weights in order to achieve better generalization error (Gkerekos et al., 2020) or by tuning a number of hyperparameters (learning rate, number of neurons, number of layers, activation function) utilizing brute force methods like randomized grid search (Papandreou et al., 2020, Jeon et al., 2018). In (Yongjie et al., 2020) a Recurrent NN is employed in order to estimate FOC but without further research as far as the architecture, or the generalization capabilities of the neural proposed.

The majority of the approaches found in literature regarding Operational Optimization and Navigation Management in the maritime sector, concern the implementation of standalone services in the sense that they are employing isolated information silos that lack the support mechanisms and enhancement of a centralized Information Hub that exploits the upsurge of IoT and Industry 4.0 advancements, to train, validate and update these services in real-time. Furthermore, they are usually tested on a single vessel and therefore lack the generalization capabilities of models evaluated in a variety of ships that are able to adjust and adapt to the underlying function that describes the relationship between FOC and each specific vessel, continuously, by exploiting the vast amount of data collected by IoT installations. Frameworks and technological advancements regarding the continuous monitoring of the vessel are inextricably linked with the emerging concept of the so-called Digital Twin in the shipping industry, as they employ a digital replica of the en-route vessel that is able to simulate-project and validate in real time the majority of the operational procedures.

In recent years, there has been a growing interest in using Digital Twins to optimize routing in the maritime sector. Digital twins serve as virtual replicas of physical systems, offering a dynamic platform

Table 1. Navigation management and operational optimization approaches in the literature

Category	Sub-Categories	Approach/Methodology	Input Data	Example Models or Systems
Smart & Efficient Transportation (Beyond Maritime)		Multi-constraint optimization methods: Genetic Algorithms, Simulated Annealing, Particle Swarm Optimization, AI integrated decision making, NLP.	Real-time traffic data through various sensor installments on EDGE	Xing Wei and Huang, 2021 - Fermani et al., 2021 - Chondrodima et al., 2020 - Kaklis et al., 2022, 2023 - Bai et al., 2019 - Garg et al., 2021
Ship Routing Optimization	Vessel-based optimization	Multi-constraint optimization methods	Weather data/ Operational Data/ Charter Party contracts	Roh and Lee, 2018 - Kim and Kim, 2017 - Vettor and Soares, 2015 - Varelas et al., 2013
	Environmental-based optimization	Dynamic Programming		
	Holistic optimization	RL, Dynamic Programming		

to simulate, analyze, and control real-world conditions. Unlike traditional methods that depend on historical data or pre-set conditions, digital twins make use of real-time analytics. This live data feed can include everything from a vessel's hull condition and load status to the available fuels and engine performance. The implication is a quantum leap in the level of granularity and customization that voyage planning algorithms can achieve.

By embracing this real-time, vessel-specific approach, maritime operators can leverage sophisticated algorithms designed to interpret and act upon a cascade of live data points. As a result, voyage planning becomes a dynamic process, continually updated to reflect the vessel's current actual condition. This increased level of detail not only allows for more precise route optimization but also leads to safer and more efficient voyages. Whether accounting for the wear and tear on a hull over a journey or optimizing fuel consumption based on real-time metrics, digital twins provide a holistic and up-to-the-minute view that is revolutionizing voyage planning.

In the table below we summarize the approaches found in pertinent literature regarding Navigation Management and Operational Optimization and provide the reader with a comprehensive-consolidated breakdown analysis based on the specific methodology adopted as well as input data utilized on each category.

Towards this direction, a novel approach that aims to extend a Digital Twin framework with a Voyage Planning module in the context of a multimodal-adaptive Digital Twin ecosystem will be proposed. This extension enables stakeholders to simulate diverse scenarios and optimize routes using real-time data, considering factors like weather, traffic, fuel consumption, hull conditions, and commercial considerations. Through the integration of various data sources and the application of machine learning algorithms to forecast future conditions, stakeholders can dynamically optimize routes, leading to cost reduction, enhanced safety, and improved efficiency.

Section 2 outlines the proposed perspectives to enhance voyage planning beyond route optimization and weather routing, leveraging the opportunities provided by the digital twin platform.

Section 3 provides a brief overview of the digital twin framework proposed for accommodating the voyage planning application.

Section 4, we delve into the methodologies for estimating fuel oil consumption and weather routing, both fundamental components of any voyage planning tool. We explore how these methodologies can be extended to leverage the benefits of the digital twin.

Section 5 offers considerations for potential advancements that could enhance the value of the proposed solution, as well as acknowledging certain limitations.

The chapter concludes Section 6 with a summary of voyage planning using digital twins.

INTEGRATION OF VESSEL OPERATIONAL AND COMMERCIAL ASPECTS IN VOYAGE PLANNING

In the traditional paradigm of voyage planning, strategies often rely on static data models and broad generalizations. These generalizations, while useful for general navigational purposes, fall short of accounting for the dynamic and ever-changing conditions of individual vessels. In this section, we propose specific vessel operational aspects that hold value for integration into the digital twin, aimed at enhancing voyage planning. It's notable that some aspects discussed are innovative for voyage planning, and we'll demonstrate how this innovation is further enhanced by the application of digital twins.

Trim Optimization

Trim is defined as the draft at the stern (or aft of the ship), minus the draft at the bow (or forward). Trim optimization is one of the approaches considered by the industry to improve the energy efficiency of ships, having a potential in both reducing operational costs and to decrease the emissions of the ship. Trim optimization is the selection of trim with the goal of fuel consumption reduction, by ballast water management and load distribution, which can be done without significant changes to the ship structure. The application has been investigated by Gao (2019) and Islam (2019).

The MEPC (2008) has estimated that optimizing a vessel's trim and draft can result in fuel consumption reductions ranging from 0.5% to 3% for most vessel categories. In the case of ships operating with partial loads, these savings can soar to as high as 5%. Coraddu (2017), in their research, have even demonstrated the potential for surpassing a 2% improvement in fuel consumption for handymax chemical tankers. A case study conducted by DNV-GL in 2013, which assessed the effectiveness of the commercial optimization tool known as the ECO Assistant, revealed impressive fuel savings ranging from 2% to an astonishing 14% across various draft and speed combinations for handymax bulk carriers. Furthermore, research by Yuan (2018) indicates that trim optimization for Very Large Crude Carriers (VLCCs) can lead to a notable 1.8% reduction in fuel consumption. The work of Du (2019), revealed that trim optimization has the potential to save between 5% and 6% of bunker fuel for 9000 TEU container ships. Similarly, Gao (2019) reported significant bunker fuel savings of 3% to 7% for Pure Car and Truck Carriers (PCTCs). In conclusion, it is evident that the impact of trim optimization on reducing fuel consumption can be substantial, with the extent of savings contingent upon factors such as the vessel's type, operational profile, and maneuverability in achieving the desired trim.

Typically, the optimum trim is often ascertained through reference trim tables. These tables are derived either from model-scale towing experiments or, in certain instances, from computational fluid dynamics (CFD) simulations. In the past years, alternatives to trim tables in the field of trim optimization have emerged. A range of commercial trim optimization solutions have been introduced to the market, offering diverse pricing options, levels of user-friendliness, underlying methodologies, and performance capabilities, as per MEPC (2008)

Integrating trim optimization into voyage planning with the assistance of a digital twining platform offers numerous advantages. On one hand, incorporating trim information into the fuel oil estimation process enhances the precision of voyage planning predictions, contributing to a well-informed, dynamic digital twin. On the other hand, since the primary goal of voyage planning is to minimize fuel oil consumption, integrating trim optimization aligns seamlessly with this objective.

Within the digital twin framework, calculation algorithms are continuously supplied with real-time data, including parameters essential in describing the relationship between trim and fuel oil consumption. Notably, hull condition, a critical factor often overlooked in trim optimization scenarios without access to real-time data, is meticulously addressed in digital twin applications. Furthermore, the temporal and spatial variations in weather conditions can be effectively taken into account through digital twin-powered trim optimization. Additionally, the intricacies of cargo management, particularly in cases of partial loading and unloading in port calls, can be seamlessly integrated into a digital twin application, providing trim optimization without requiring extensive crew involvement.

Route/Port Congestion and JTI

Reducing speed is the most essential method to reduce fuel oil consumption. ABS (2021) investigated the potential efficiency improvements created by just-in-time shipping by presuming an average 5 percent reduction in speed, assuming no impact on cargo-carrying capacity and no adjustment to the size of the fleet. Based on that basic analysis, the CO2 emissions savings are around 10-11 percent annually.

For a number of years, the use of data from automated identification systems (AIS) has made it possible for the industry to operationally benefit from knowing details such as the estimated time of arrival, the arrival port, draught and navigational speeds, etc.

Those data points are being used today to track vessels and to support limited adjustments to voyage planning. However, deeper analyses of the data on vessel positions and berth availability are revealing the type of information that soon could make 'just-in-time' shipping a reality.

Improving communication efficiency between vessels and ports regarding berth availability and services like tug operators holds the promise of optimizing marine traffic, delivering both commercial and environmental advantages. In the context of the discussion in this chapter about the digital twin as a dataspace actor, this concept can play a pivotal role in realizing these improvements.

Digital twins inherently exist in the digital realm, and they can offer a more seamless interface with other IT systems compared to AIS (Automatic Identification Systems). Furthermore, they can provide enriched information for a ship's voyage plan and conditions, enhancing operational effectiveness.

Conversely, data regarding shore-side infrastructure availability and readiness can be automatically accessed and factored into real-time voyage planning. This two-way flow of information, facilitated by digital twins, has the potential to significantly enhance the efficiency and sustainability of maritime operations.

Bunkering Optimization

In engines that run on oil-based fuels, different types of fuel are necessary to ensure vessels comply with regulations. These fuel types include Heavy Fuel Oil (HFO), Low Sulfur Fuel Oil (LSFO), Very

Low Sulfur Fuel Oil (VLSFO), and Marine Gas Oil (MGO). Furthermore, globally, alternative fuels are emerging as environmentally viable options to oil-based fuels. As reported by DVC GL (2019), among the fuel alternatives to marine bunker oil, LNG is the most prolific with more than 300 ships in operation. In addition to LNG, biofuels and methanol are available in certain ports. Several types of dual-fuel engines have been introduced in the market which increases fuel flexibility significantly. Beyond LNG, fuels such as methanol, ethanol, LPG (liquid petroleum gas) and soon likely also ammonia can be burned in different types of dual fuel engines, in addition to HFO/MGO. Promising steam- and gas- turbine concepts are also being considered.

The accessibility of fuels in close proximity to vessels for bunkering purposes significantly influences voyage planning. Additionally, varying fuel prices at different bunkering locations play a crucial role from a commercial standpoint. It's worth noting that storing fuels in vessel tanks contributes to increased displacement, subsequently impacting fuel oil consumption.

Integrating digital twin technology in real-time to monitor fuel type availability and prices at different locations, on one hand, and tracking fuel tanks' arrangement, capacity, and current levels, on the other, can introduce bunkering optimization as an innovative and transformative element in voyage planning. This approach enhances both operational efficiency and cost-effectiveness by making data-driven decisions about fuel sourcing and consumption during voyages.

Supplies On-Boarding Location and Frequency

In addition to applying digital twins for monitoring vessel seagoing performance and cargo operations, there is a growing focus on enhancing internal operations. One significant area of improvement is condition-based maintenance, which can be effectively facilitated through the capabilities of digital twins. In a broader perspective, the digital twin can proactively anticipate and track the spare parts required over time, streamlining maintenance processes and improving overall operational efficiency.

Integrating spare part supply into voyage planning can have a profound impact, enhancing onboard availability while concurrently reducing freight costs. By optimizing the supply location and timing, vessels can ensure that necessary spare parts are on hand when needed, minimizing downtime and maintenance delays. This proactive approach not only improves operational efficiency but also contributes to cost savings in the long run, making it a valuable component of comprehensive voyage planning strategies.

The realization of this approach is enhanced by the digital twin's predictive maintenance component that estimates in real time machinery condition and thus forecast spare parts needs

Hull Degradation and Hull Cleaning Effectiveness

Biofouling, namely the undesirable accumulation of microorganisms, algae, and animals on artificial surfaces immersed in seawater results in performance decay of hull and propellers, as discussed in XXX.

The substantial impact of biofouling on fuel oil consumption is widely acknowledged. For instance, in studies conducted by Watanabe (1969), which involved rotor and towing tank experiments, they identified a frictional resistance increase ranging from 8% to 15%. Similarly, Farkas (2020), employing Computational Fluid Dynamics (CFD) simulations, found that the total resistance increased by 50% to 120%. Moreover, Schultz's (2015) research, which included laboratory-scale drag measurements and analysis based on boundary layer similarity laws, revealed a total resistance increase of approximately 10% for a light slime film and around 20% for a heavy one. In certain cases, the total resistance increase was even more pronounced, ranging from 35% to 86%. These findings underscore the significant impact of biofouling on vessel performance and the need for effective anti-fouling measures to mitigate fuel consumption increases.

Techniques to reduce the impact of biofouling are underwater cleaning of the hull, propeller polish and application of antifouling coatings. However, effective maintenance can be responsible for up to 20% of total operational costs and, therefore, is a perfect candidate for optimisation and improvement.

Methods to mitigate the impact of biofouling include underwater cleaning of the hull, propeller polishing, and the application of antifouling coatings. However, effective maintenance practices can account for up to 20% of the total operational costs, as reported by Valchev (2022). Hence, optimizing and improving these maintenance processes present excellent opportunities for cost-saving measures in maritime operations.

Voyage planning is significantly influenced by hull degradation, and a digital twin can play a pivotal role in ensuring that the voyage planning model and its associated attributes remain updated with the actual state of the vessel. Furthermore, the digital twin can provide estimates of the benefits derived from corrective actions, such as underwater cleaning, and their impact on both operational and commercial performance. For instance, if underwater cleaning is conducted before a voyage, it may introduce operational costs and time delays into the journey. However, these factors can be offset by the improved efficiency and reduced fuel oil consumption during the voyage. Unlike dedicated systems for hull performance assessment, fuel oil consumption estimation, and voyage planning, the digital twin serves as a central system that maintains all the necessary real-time information required to assess and execute such actions effectively. This centralization of data and insights enhances decision-making in maritime operations.

AN OPEN DIGITAL TWIN FRAMEWORK FOR VOYAGE PLANNING

A Digital Twin, adapted to the needs of the maritime sector, constitutes a virtual holistic representation of the vessel that spans its life cycle and is updated from near to real-time data, utilizing simulation, machine learning and reasoning to help in decision-making, sensing and control actuation. By combining core structural properties of traditional MIS and digital twins, organizations can gain a better insight of their internal operations and pave the way for a fully automated and fault tolerant decision-making procedure, substantially improving their efficiency and effectiveness.

The framework developed within the scope of the DT4GS Horizon Europe research project, parts of it are described in the preset work, consists of a variety of SOTA tools and services that aim to vastly automate the majority of the procedures concerning the existing Voyage Planning in several ways, incorporating a variety of state-of-the-art streaming tools for real-time analysis of vessel data as well as tools for continuous integration/deployment (CI/CD) of ML/DL models regarding operational optimization, causal analysis, and event recognition. The resulting platform constitutes a prototype version of a virtual replica of a vessel that aims to assist shipowners to achieve efficiency in fleet management with tangible benefits in terms of emission reduction, environmental compliance and protection of crew safety onboard. In addition, DT4GS framework can also facilitate the refinement of the existing Voyage Planning framework by providing a platform to test different scenarios and strategies in the form of an adaptive Decision Support System. This can include simulating different routes, adjusting speeds, or optimizing the vessel's trim to reduce fuel consumption. By testing these strategies in a virtual environment, companies can identify the most efficient and cost-effective approach to the voyage and define tailor made mitigation strategy towards a carbon neutral operational blueprint.

Specification of the Digital Twin Framework

The present section discusses the specification and requirements relevant to a Digital Twin capable for voyage planning, to ensure a consistent and effective implementation, promoting seamless integration and interoperability among the various required components.

Data Acquisition and Integration

In the past years the main source of information for vessel operation resulted from the noon report, which was compiled manually by vessel crew and was sent to the Head Quarters (HQs) with the available means of communication. Increasingly, ship data used for various purposes including building the digital twin are collected automatically, from sensors. To keep track of metadata and to structure the storage and processing of sensor data, there is a need for a unique identification of sensors as well as the components and systems subject to monitoring by sensors, as suggested by LÅG (2017).

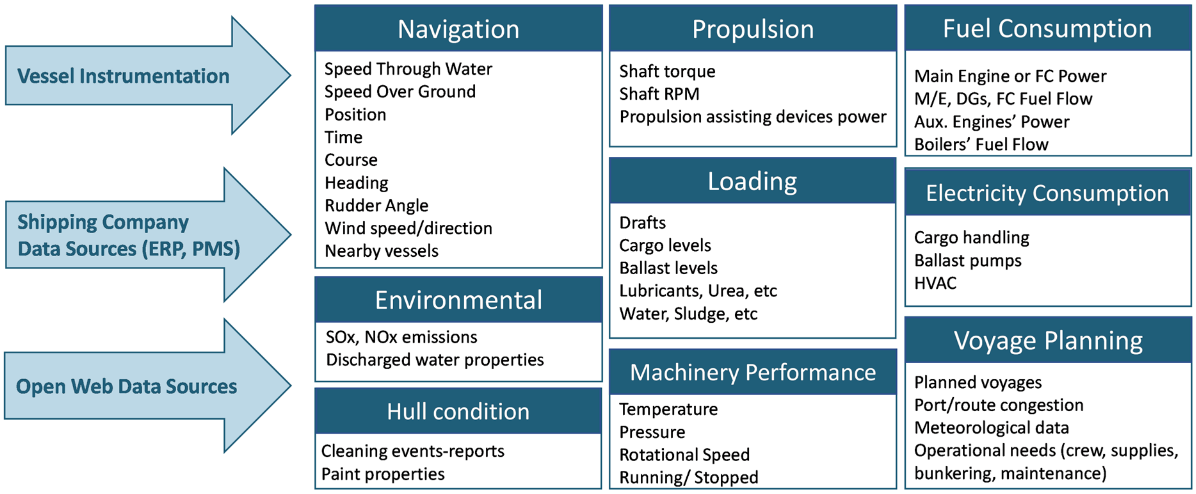
Within the context of voyage planning, data is required for the following purposes:

- To compose a digital shadow, mirroring actual vessel condition, to compare against the digital twin (simulated values) for integrity verification.
- To feed the models that reproduce vessel operation and compose the digital twin, such as training models.
- To provide the required information, such as green fuel availability and cost, port congestion, weather forecast for weather routing, etc.
- There exist three sources to acquire the required data.
- Vessel instrumentation,
- Shipping company's core systems that provide information about voyage planning, spares/stores deliver, crew on-boarding, planned maintenance, etc.
- Web data sources: weather forecast, etc

In Figure 1 are listed the most important parameters required for the deployment of a Shipping Digital Twin.

The methods for integrating vessel instrumentation depend on the available technologies, often influenced by the vessel's build year. Typically, data acquisition on ships involves various protocols such as NMEA (standard for navigational instruments), Modbus serial or TCP, HART, UDP broadcasts, analog signals (e.g., pulse, 4-20mA), SNMP, or REST APIs. Centralized sources of combined data, like the AMS, collect operational data from various subsystems of the vessel and provide it through serial communication or over LAN using protocols like NMEA, proprietary, or Modbus. Other sources of combined data include the VDR, engine control system for electronic engines, tanks level and drafts monitoring system (connected to the loading computer), power management system, and fuel conditioning system. Additionally, individual instruments like flowmeters, level sensors, torque meters, temperature, and pressure transmitters can be integrated separately.

Figure 1. Data required for the formulation of digital twin



Data Processing and Analytics

To depict a unified vessel condition, simultaneous sampling of all signals is required. A reference procedure is described in ISO 19030-2 (2016). However, in most cases, the sampling rate is not configurable, except for analog data acquisition and communication with raw sensors using request-reply protocols. To address this matter, data temporal bucketing can be applied for data synchronization in the context of a digital twin IT system. As data streams from various sources arrive at irregular intervals, temporal bucketing organizes the data into fixed time intervals, typically referred to as buckets. This process enables aligning the data to a common time scale, ensuring that all data points are associated with a specific time frame, which is crucial for generating a unified and coherent view of the vessel's condition. However in the decision of the bucketing interval length it should be taken in to account that the data sampling rate may coincide with the frequency of a natural phenomenon for the vessel in question (e.g. wave encounter frequency) and thereby influencing the accuracy of associated data point.

In addition to synchronization, data temporal bucketing allows for the creation of new variables based on existing data. By aggregating or summarizing data within each bucket, it becomes possible to compute derived quantities or perform statistical analyses. Some examples are the following:

- Fuel flows calculations such as M/E FOC = M/E INLET FOC – M/E OUTLET FOC
- Fuel mass flow estimation using volumetric flow, nominal density (at 15oC) and actual temperature
- Estimation of absolute wind speed from relative wind speed, wind direction and heading
- Unit conversions form instrument specific unit to platform standard
- Estimation of system and operational performance indicators

As the digital twin's accuracy directly relies on the data quality used in its construction, identifying data-related issues and defining specific requirements becomes critical. Poor data quality, such as missing values, outliers, or inaccuracies, can lead to erroneous simulations and unreliable predictions.

Therefore, it is essential to implement robust data validation mechanisms to identify and rectify data anomalies. Data quality checks should be performed at various stages, from data acquisition to temporal bucketing and variable creation. In cases where data discrepancies are detected, appropriate corrective actions or data cleansing processes should be applied to maintain the integrity of the digital twin system. Statistical method for outlier detection, such as Chauvenet's criterion, and rule based mechanisms can provide a means for data integrity. Custom data cleansing methods as the one described in the following paragraphs are also applicable.

Data Storage

The choice of data storage solution for the digital twin system depends on the specific implementation and requirements. There are various options available, each with its own strengths and use cases. Some common data storage solutions include:

- Relational Databases (e.g., PostgreSQL, MariaDB): Relational databases are well-established and widely used for structured data storage. They offer robust data integrity, support for complex queries, and data consistency through ACID (Atomicity, Consistency, Isolation, Durability) properties. Relational databases are suitable when data relationships and schema are well-defined and stable.
- NoSQL Databases (e.g., MongoDB): NoSQL databases are designed to handle unstructured or semi-structured data. They provide greater flexibility and scalability compared to relational databases, making them suitable for scenarios where data schemas may evolve over time or data volume is large and continuously changing.
- Time Series Databases (e.g., influxDB): Time series databases are optimized for handling time-stamped data, making them particularly suitable for storing sensor readings, historical performance data, and other time-related data. They provide efficient data retrieval based on timestamps and support data compression techniques for storage optimization.

From the perspective of the platform requirements, adherence to a strict definition of variables registry is crucial. This means that each measurement or data point should have a solid reference or metadata that precisely defines what it represents. Having a well-defined variable registry ensures consistency in data representation, avoids ambiguity, and facilitates effective data analysis and interpretation.

By adhering to a strict variable registry, it can be achieved a clear mapping between the data stored in the database and the physical or virtual variables it represents, across the different DT implementations on different applications (different ships, etc) which result in interoperability. An application would be the ability to transfer knowledge between dt instances and also reuse resources, such as simulation models.

Regardless of the specific data storage solution chosen, data retention policies, and backup strategies which may vary between implementations, ensuring a solid variable registry and well-defined metadata for each measurement is fundamental for maintaining data quality and facilitating the seamless functioning of the digital twin system. This approach promotes consistency, accuracy, and reliability in data-driven simulations and decision-making processes.

Simulation and Modelling

Indicatively, simulation is applied for the following purposes:

- Forecast future conditions: Simulation is used to predict and model the future behavior of the system or process based on current conditions and known parameters.
- Estimate parameters required for generating the digital twin but not directly available: Sometimes, certain parameters essential for building the digital twin may not be directly obtainable from data. Simulation helps estimate these parameters through modeling and analysis.
- Estimate the effect of a change in performance: Simulation allows for testing different scenarios and changes in the system to understand their potential impact on performance and outcomes.

Assist in the acquired data validation or event detection (failure) by detecting patterns that deviate from the simulation results: The digital twin can aid in data validation by identifying discrepancies between real-world data and simulated results. These deviations can indicate data quality issues or reveal anomalies that require investigation. On the other side the same approach can be utilized for the detection of deterioration events assisting in their early detection.

The modeling framework for the digital twin system must have the following main characteristics:

- Support for single moment simulation, time steps, and events based: The framework should be capable of conducting simulations at specific time points, discrete time intervals, or in response to specific events or triggers.
- Support for dockerized code, Python, Java, FMI, R: The framework should be versatile and support different programming languages and technologies, allowing flexibility in implementing and integrating simulation models.
- Integration with data-driven model instantiation and versioning systems (such as MLflow): Seamless integration with data-driven models and version control systems enhances reproducibility, transparency, and management of different model versions.
- Description of model source (e.g. binary) and versioning: The framework should provide mechanisms for storing and managing simulation models, including version control to track model changes over time.

An important aspect of modeling is its utilization for optimization purposes. Optimization is a crucial requirement for DT systems, and the following approaches are considered:

- Repetitive procedure for determining the best solution based on predefined scenarios and variable ranges: This approach involves iteratively testing different combinations of variables within specified ranges to identify the optimal solution.
- Similar approach based on Monte Carlo principles instead of using predefined scenarios: In this approach, random sampling is used to explore a wide range of possible scenarios, allowing for a more comprehensive optimization analysis.
- Utilization of genetic algorithms, such as NSGA II: Genetic algorithms mimic the process of natural selection to iteratively evolve and refine potential solutions, making them well-suited for multi-objective optimization problems.

By employing these optimization approaches, the digital twin system can identify optimal configurations, settings, or decisions that lead to improved performance, efficiency, or other desired outcomes for the modeled system or process.

System Monitoring and Diagnostics

A Digital Twin (DT) system requires robust functionality for self-monitoring and diagnostics to ensure its own health and performance. The system should continuously monitor critical metrics, such as resource utilization, response times, and event logs. Real-time performance monitoring, resource tracking, and health checks are essential for identifying anomalies and potential issues within the system. The DT system should generate alarms and alerts for abnormal conditions and include an auto-recovery mechanism to handle failures autonomously. It must offer diagnostic tools and historical performance data for troubleshooting and trend analysis. Scalability and security monitoring are crucial to ensure the system can handle increasing workloads and maintain data integrity. Integration with IT operations tools facilitates centralized management and analysis. By implementing these functionalities, the DT system can maintain its reliability, optimize performance, and provide a stable foundation for effective digital twin operations in diverse scenarios.

Compliance and Standards

According to the degree a DT interacts with the physical vessel, the corresponding shipping industry standards, elicited by the IMO, the Institute of Electrical and Electronics Engineer (IEEE) and rules of Classification Societies, should be adopted. The lowest level of interaction is observed in the case of simple data acquisition, while at the other end of the spectrum, there are aspects related to autonomy, which significantly increase the requirements related to compliance to regulations.

The Digital Twin as a Dataspace Actor

As per Nagel (2021), dataspace can be defined as "A data ecosystem, specified by a sector or application, whereby decentralized infrastructure enables trustworthy data sharing with commonly agreed capabilities". Thus, a dataspace is produced by an ecosystem of actors that interact through the sharing of data.

A Shipping Dataspace provides the data infrastructure, specifically data connectors, for creating, managing and interacting with actors, such as weather data providers, fuel availability and pricing data providers, route and port congestion data providers, commercial brokers, and supplies availability data providers. Integration of DT in the dataspace can assist in Information exchange between them, in a uniform and secure way, providing them with collective knowledge. Another aspect cohering to the DT functionality and the dataspace is model sharing and co-simulation on an open or commercial basis. Furthermore development of applications interacting with DTs is enhanced by the standardization that the dataspace offers.

Setting up a dataspace with weather data providers, fuel availability and pricing data providers, route and port congestion data providers, commercial brokers, and supplies availability data providers provides seamless integration of these sources of information.

Dataspace components provide the appropriate infrastructure to collect, store, and analyze data from various sources in real-time. It uses a distributed architecture that enables the processing of large volumes of data while ensuring scalability and reliability. Dataspace provides a variety of processing capabilities, such as filtering, aggregation, and data augmentation, that can be used to further optimize the Routing Optimization module.

THE APPLICATION OF THE DIGITAL TWIN FRAMEWORK IN VOYAGE PLANNING

The following paragraphs focus on the realization of the specific case of FOC estimation and Routing Optimization, by consolidating the aforementioned components of the broader DT4GS frame, towards a holistic Operational Optimization Digital Twin suite that aims to improve voyage efficiency and environmental compliance.

Reference Implementation

In the present section is described a DT implementation conducted within the scope of the DT4GS research project, which is in progress at the time that this chapter is authored. In Figure 2 is provided the implementation of the framework in terms of building blocks. A messaging system is engaged to distribute data, events, and triggers to the intended peripheral components. Storage, permanent and temporary, for both configuration and data, is achieved by the combination of a time series database, a nosql database, a knowledge graph for metadata storage and a cache. The ingestion is achieved using internal connectors for data originating from the vessel, external connectors for external data sources such as CRM systems or the internet (sources not possible to be integrated into the dataspace). Data is also exchanged with the dataspace via an IDSA compliant connector. Peripheral components exist also for the following purposes:

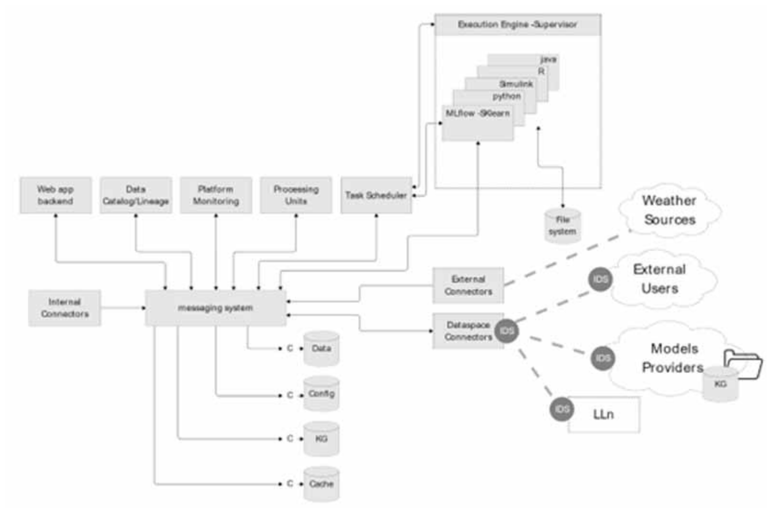
- Web app backend
- Data catalog/lineage
- Platform monitoring tools
- Processing units, assisting ingestion, producing composite variables and simple calculations Task Scheduler
- Model Execution Engine

Fuel Oil Consumption Estimation Methodology

Feature Selection

In order to unveil the relationships between the independent variables as well as their importance and role in estimating FOC, we conduct an initial exploratory analysis with Random Forest regression as the feature ranking algorithm. Then calculate the correlations between the most important features and conclude to an ideal feature set that consists of independent variables that will be utilized accordingly in the context of FOC approximation.

Figure 2. Overview of an implementation



Decision Trees (DT) is a popular classification or regression algorithm that takes into account the importance of features. More specifically, the feature importance defines the order in which features are selected for splitting the initial set of samples to subsets, from the tree root to the leaves. It is defined by the decrease in (tree) node impurity, which is weighted by the node probability. This probability is the number of samples that reach the node, divided by the total number of samples. Higher decreases in impurity denote more important features.

Assuming only two child nodes (left, right) for each node, the node importance is given by the following equation:

$$ni_j = w_j C_j - w_{left(j)} - w_{left(j)} C_{right(j)}$$

where nij is the importance of node j for feature i, wj is the weighted number of samples reaching node j and Cj is the impurity of node j. Impurity is measured using Gini Index or Entropy.

The Random Forest (RF) algorithm extends the concept of Decision Trees, for high-dimensional data, by constructing many individual decision trees during training, using each time a different random subset of the initial set of features. It then collectively examines the predictions of trees in order to make the final prediction. Respectively, RF can be used to evaluate the importance of each feature across all the trees and provide a more comprehensive ranking of feature importance.

In Table 2 we depict the experimental results from conducting regression analysis utilizing RF regression in order to rank the importance of the aforementioned features in estimating FOC.

Besides selecting the most important (i.e. informative) features, we also aim to avoid selecting highly correlated ones. For this purpose, we utilize the Spearman's Rank Correlation (SRC) coefficient. The Spearman's rank-order correlation is the non-parametric equivalent of the Pearson product-moment correlation (ρ) and assesses the strength and direction of the monotonic relationship between two ranked variables R(Xi), R(Yi) using covariance and standard deviation σ, and is calculated as follows:

$$\rho R_{(X),R(Y)=} \mathrm{cov}(R(X), R(Y)) / \sigma R(_{X)\sigma R}(Y_{)}$$

Table 2. Feature ranking using RF

Ranking	Feature	Importance
1	STW	0.94
2	WS	0.13
3	DRAFT	0.011
4	VSL_H	0.005
5	COMBH	0.0058
6	SWH	0.0054
7	CS	0.004
8	$WAVE_H$	0.0039
9	SWP	0.0036
10	COMBD	0.0032
11	SWD	0.0028

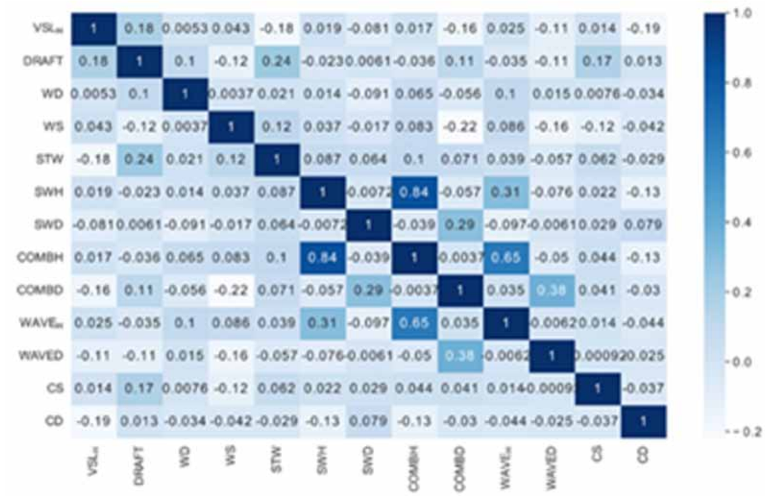
Assembling the ranking of features depicted in Table 2 and the correlation coefficients calculated, depicted in Figure 3 using Algorithm 1, we conclude with a subset of the initial feature set that combines feature importance and independence.

Algorithm 1 Feature selection based on RF regression importance and Spearman Correlation.

Require: `featureSet` $\mathcal{F} \leftarrow$ top 10 from RF
Require: `featureSet` $\mathcal{F}_r \leftarrow$ rest of features from RF
Require: `correlations` $Corr \leftarrow$ from SRC
Require: `importances` $\mathcal{I}mp \leftarrow$ from RF

for each $f_i \in \mathcal{F}$ **do**
  set $\bar{\mathcal{F}} = \mathcal{F} \setminus \{f_i\}$
  for each $f_k \in \bar{\mathcal{F}}$ **do**
    if $Corr(f_i, f_k) > 0.5$ **then**
      if $Imp[f_i] < Imp[f_k]$ **then**
        delete f_i From $\mathcal{F}$
        set $f_{temp} = f_k$
      else
        delete f_k From $\mathcal{F}$
        set $f_{temp} = f_i$
    for each $f_r \in \mathcal{F}_r$ **do**
      if $Corr(f_{temp}, f_r) < 0.5$ **then**
        add f_r to $\mathcal{F}$ and **break**
    if $f_{temp} = f_k$ **then**
      break
Return $\mathcal{F}$

Figure 3. Spearman correlation heatmap



Data Cleaning

Raw data, collected from the sensors of the vessel, are in time-series (minutely) form and tend to be ``noisy'' (high variance, high standard deviation from the mean) and in some cases even erroneous. In order to remove noise, we employed a fit\filter technique that effectively ``cleaned'' the data but at the same time kept the bulk of information needed for training robust predictive models.

Data filtering was implemented in two stages. First, assuming that the dataset follows a normal - like distribution, we keep the data points that lie within the 99\% confidence interval around the mean.

Then we apply an appropriately designed Decision Tree based algorithm in order to further cancel the noise in FOC target distribution caused by the flowmeter sensor on the vessel. Then, we proceed to transform our dataset into 15-min rolling window averages in order to further smooth out any spikes and outliers that occur in the feature set from sensor installments.

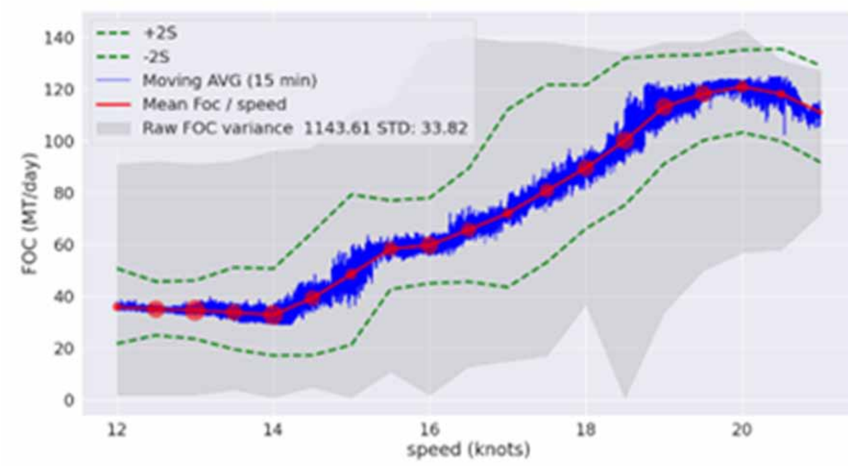
Note that the use of rolling window averages is consistent with the use of the FOC prediction model within a WR algorithm, in which decisions are based upon average values of FOC and not momentary consumption. The raw data of the vessel's speed and corresponding FOC collected from the sensors, versus the mean values per speed range (+/-0.25 V$) and the 15 min rolling window averages are depicted in Figure 4. Red circles are indicative of the number of observations found for a particular range of speed.

Model Implementation

The dynamic estimation of FOC based on vessel state and environmental conditions can be examined as a multivariate time-series prediction problem that takes into account the actual values as well as their recent history, and captures the information hidden in the values' evolution over time.

Based on the superiority of Long Short-Term Memory Neural Network (LSTM) models over traditional time-series prediction methods (e.g., ARIMA) as suggested by Siami (2014), LTSMs are chosen as the basis of our solution.

Figure 4. Raw data values vs. mean values vs. rolling window average values



The initial feature set, collected by sensor installments on-board the vessel, comprises the vessel speed through water, draft and heading and some basic weather features, mapped from external services (i.e. NOOA), such as wind speed and direction. The sampling rate of the sensor based operational data corresponds to minutely measurements.

In order to take maximum advantage of this feature set, we employ a LSTM architecture, using a pre-training step that extracts information from the original features, using spline-based regression (Friedman J. 1999). In what follows, we describe how LSTM is used for FOC estimation and detail the proposed LSTM model and its novel aspects.

LSTM is a variation of traditional Recurrent Neural Network (RNN) architecture, which has been extensively used for time-series prediction tasks.

Unlike standard feed forward neural networks, LSTM also contains feedback connections and can process single data points (e.g., images) as well as entire sequences of data (e.g., speech, video or object trajectories). Compared to RNNs, Hidden Markov Models and other sequence learning methods, LSTMs are not so sensitive to the length of gaps between important events in a time series, which makes them more preferable in numerous applications. To this end, we adopt an LSTM architecture for the prediction of FOC values from the consecutive observation, corresponding to the aforementioned features, in a time window, as described in the following paragraphs.

The input of the LSTM network at timestep t_u comprises N time-series, one for each feature of interest (speed through water, wind speed, wind angle etc) and in order to use the recent history of values in each feature, we employ a fixed-length time-window (time-lag of length m). As a consequence, the window contains the values for each time step for the weather and vessel state features that are used for the estimation of FOC at time t_u, resulting in N time-series, of length m+1, of the form $[F_{N(u-m)}, \ldots, F_{N(u-1)}, F_{N(u)}]$, for each feature F_N. Given a sequence of consecutive time-steps, and a multivariate feature set, we get the following correspondence between the input and the output of the LSTM:

$$\begin{bmatrix} \left[F_{1(u-m)} \ldots F_{j(u-m)} \ldots F_{N(u-m)} \right] \\ \vdots \\ \left[F_{1(i)} \ldots F_{j(i)} \ldots F_{N(i)} \right] \\ \vdots \\ \left[F_{1(u)} \ldots F_{j(u)} \ldots F_{N(u)} \right] \end{bmatrix} \rightarrow \begin{bmatrix} FOC_{u-m} \\ \vdots \\ FOC_{i} \\ \vdots \\ FOC_{u} \end{bmatrix}$$

Weather Routing Aspects and Integration of the Solution

In order to validate the approach in the context of a real-world application, the data driven FOC LSTM model has been coupled with a WR algorithm to support vessel routing decisions towards the reduction of FOC. The WR algorithm that has been utilized is based on the isochrone principle (Hanssen et al., 1960). It builds upon a predetermined basic route; this route can be the original route planned by the vessel's master or provided by a basic routing algorithm. In the context of this work an initial route was employed on the basis of shortest path principles. The original (initial) route is then broken into segments, with respect to a given time step (indicating the master's routing decision horizon, e.g., every 6 hours), and a graph is built around it that enables course and speed deviations, while ``following'' the direction of the vessel's original course. To this end, for each node of the original route, a set of nodes is added in a ``parallel'' fashion on both sides of the route (i.e., parallel to the direction of the original route). Edges are added between all nodes of subsequent sets. Note that nodes that are identified to be on land as well as edges that go above land segments are naturally excluded from the graph.

Once the graph is created (Figure 5), LSTM NN is used to obtain the FOC of each edge of the graph, i.e., of each corresponding sea route, given the vessel's STW, draft and corresponding weather conditions along that sea route. After scoring each sea route (i.e., graph edge), a variation of Dijkstra's algorithm for the shortest path problem is utilized to obtain the route that minimizes the total route FOC (i.e., considering the calculated FOC of each edge as its corresponding ``edge weight'' or ``distance'').

Note that since the algorithm is isochrone, the produced route also satisfies any constraints concerning the time of arrival (if any).

Note also that the decision variables for the WR algorithm are only the STW and the vessel's direction, since these are the aspects that the vessel's master can control. Obviously, any change in the vessel's speed affects FOC directly (since STW is a basic feature of the corresponding model). However, changes in speed and direction also affect FOC indirectly, since they alter the spatio-temporal state of the vessel and hence the corresponding weather conditions.

Preliminary Experimental Results

We continue by demonstrating the results of the WR optimization algorithm demonstrated briefly above. We compare the total FOC of an initial transatlantic voyage conducted by the vessel's master, with the suggested optimized route produced from the WR algorithm by utilizing the aforementioned LSTM FOC

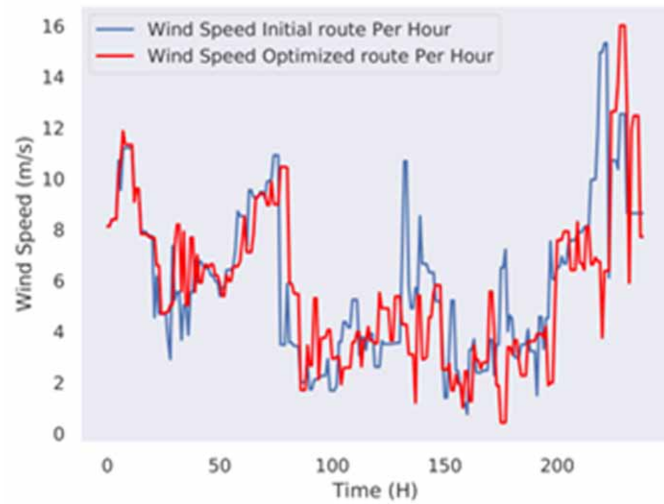
Figure 5. Graph construction comprised of alternative waypoints (red circles) for an example route

Figure 6. Initial (blue) and Optimised (red) route for one leg TAMPA (FLORIDA U.S) - TANGER MED (MOROCCO)

Table 3. Estimation based on weather service (NOAA)

Voyage	Date	Latitude	Longitude
Departure	2019-09-21	27.7°N	82.5°W
Arrival	2019-10-03	35.8°N	6°W
Basic Comparison		**Actual Route Estimation**	**Optimized Route Estimation**
Distance		4787.4	4369.36
Time (hours)		289.997	264.63
Avg. Speed (kt)		16.52	16.51
Total FOC (MT)		774.53	759.97
CO_2 (MT)		2411.88	2366.54

Figure 7. Weather comparison (wind speed) for the initial and optimised route



model. Furthermore, we calculate the total distance traveled, the estimated time of arrival, the average speed and the emissions emitted for the two alternative routes and we exhibit the results in Table 3.

Consecutively we demonstrate the weather (wind speed (m/s)) of the initial and the optimized route per hour, in Figure 7 We can clearly see that the optimized route attempted to avoid ambient weather conditions on average while at the same time complied with ETA constraints.

The accuracy of the estimations we depict in the table below rely heavily on the accuracy of the FOC model we have demonstrated in previous sections. The model showcases promising approximation capabilities, and we can therefore incorporate its prediction to the heuristic function of a path finding algorithm, being confident that the simulation results correspond to the reaction of the physical system with minimal margin of error.

FUTURE TRENDS AND IMPLICATIONS

As we look ahead, the convergence of digital twins with other cutting-edge technologies promises to revolutionize voyage planning.

Further advancements are expected in the near future in data integration. The digital twin ecosystem is extending beyond the vessel itself. Ports, fuels, and spares suppliers will increasingly become integrated into the digital dataspace. This interconnectedness will enable smoother transitions during port calls, efficient refueling, and timely maintenance. Real-time data feeds from satellites, ocean sensors, and on-board IoT devices are increasingly made available and can be exploited by the digital twin improving the on-the-fly adaption on current sea conditions, weather patterns, or other unexpected challenges, ensuring optimal navigation and safety. As standards develop, digital twins will become more interoperable across different platforms and systems. This means that a digital twin created by one shipping company could be easily used by another, leading to a more collaborative and efficient maritime industry. Additionally, the rise of blockchain and decentralized technologies can enhance the security and integrity of the data used in digital twins. This ensures that the information is tamper-proof and comes from verified sources, increasing the reliability of voyage plans.

Apart from data integration advancements in AI technologies are improving the prediction of potential risks or disruptions. By analyzing historical data, they can forecast mechanical failures, piracy hotspots, or hazardous weather long before they pose a threat, allowing for proactive route adjustments. Apart from prediction purpose AI technics are being used in combination with VR and AR technologies, allowing ship crews, planners, and stakeholders to virtually navigate routes, identify obstacles, or even conduct training exercises within a digital replica of the real-world environment.

The advent of autonomous ships will heavily rely on digital twins for navigation, decision-making, and onboard operations. Digital twins will serve as the 'brain' for these vessels, ensuring they can autonomously adapt to changing conditions, detect obstacles, and make split-second decisions. As autonomous vessels become more prevalent, the sophistication and accuracy of digital twins will be paramount in ensuring safe and efficient maritime transport.

As with any advancement in technology, digital twins will also raise ethical and regulatory questions. Who owns the data and how the dataspace is regulated? What happens in the event of a digital twin's malfunction or misinterpretation of data? How are privacy concerns addressed, especially when personal data, such as passenger preferences on cruise ships, are integrated? The maritime industry will need to work closely with regulators and ethicists to ensure that digital twins are used responsibly and that there are clear guidelines and accountability mechanisms in place.

CONCLUSION

The use case of voyage planning demonstrates how concepts and technologies that are already known can be integrated in digital Twin providing holistic and spherical solutions for domain problems.

The voyage planning can be expanded incorporating various concepts and solutions when integrated as a digital twin. Trim optimization can add a benefit of up to 5% in fuel oil consumption and the respective benefit in emissions, being enhanced by the data availability within a digital twin and on the other side the mechanism can be perfectly integrated in the digital twin. The same applies when considering detection and assessment of hull degradation due to biofouling. Interaction with ports in an automated way, facilitated by the digital from vessel side, is an important voyage planning aspect, resulting in reduced voyage cost and more efficient operation, considering that vessel speed has almost a cubic effect on fuel consumption. Further increase in efficiency within voyage planning can be achieved when considering proper position and quantity of fuel and spares supply.

Integration of these features demands an adaptable digital twin architecture with unique characteristics. We have explored these requirements and outlined specifications for a framework encompassing data, modeling, and system management. Additionally, we have proposed a specific implementation.

Finally, the application of this implementation has demonstrated a typical approach to voyage planning, which includes predicting vessel performance, determining the shortest route, and utilizing weather routing. The methodology applied is described, and the results indicate a 2% reduction in fuel oil consumption during a transoceanic -voyage.

In sum, the future of digital twins in voyage planning is one of greater accuracy, predictive capability, and dynamic responsiveness. As these systems evolve, they'll not only optimize routes for efficiency and safety but will also consider factors like passenger experience and environmental impact. The implications are vast, signaling a transformative shift in how the maritime industry approaches and executes voyage planning.

REFERENCES

Ahlgren, Fredrik, M.E.M., Thern, M. (2018). Predicting dynamic fuel oil consumption on ships with automated machine learning. *Journal of Mechanical Science and Technology*, *32*, 5785–5796.

Andrea, C., Oneto, L., Baldi, F., & Anguita, D. (2017). Vessels fuel consumption forecast and trim optimisation: A data analytics perspective. *Ocean Engineering*, *130*, 351–370. doi:10.1016/j.oceaneng.2016.11.058

Avgouleas, K. & Sclavounos, P. (2008). *Optimal ship routing*. Research Gate.

Bai, Z., Shangguan, W., Cai, B. & Chai, L. (2019). *Deep reinforcement learning based high-level driving behavior decision-making model in heterogeneous traffic*. IEEE.

Bernitsas, M., Ray, D., & Kinley, P. (1981). *KT, KQ and Efficiency Curves for the Wageningen B-series propellers. Dep*artment of Naval Architecture and Marine Engineering College of Engineering The University of Michigan .

Bohlin, T. (1992). Interactive system identification: Prospects and pitfalls. J*ournal of Mechanical Science and Technology.*

Campolucci, P., Capperelli, F., Guarnieri, S., Piazza, F., & Uncini, A. (1996). Neural networks with adaptive spline activation function. *Proceedings of 8th Mediterranean Electrotechnical Conference on Industrial Applications in Power Systems, Computer Science and Telecommunications (MELECON 96).* IEEE. 10.1109/MELCON.1996.551220

Chen, S., Hong, X., Khalaf, E., Morfeq, A., & Alotaibi, N. D. (2015). Adaptive B-spline neural network based nonlinear equalization for high-order qam systems with nonlinear transmit high power amplifier. *Digital Signal Processing*, *40*, 238–249. doi:10.1016/j.dsp.2015.02.006

Chen, X., Xu, X., Yang, Y., Wu, H., Tang, J., & Zhao, J. (2020). Augmented ship tracking under occlusion conditions from maritime surveillance videos. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 42884–42897. doi:10.1109/ACCESS.2020.2978054

Chondrodima, E., Georgiou, H., Pelekis, N., & Theodoridis, Y. (2020). *Particle swarm optimization and rbf neural networks for public transport arrival time prediction using gtfs data*.

Coraddu, A., Oneto, L., Baldi, F., & Anguita, D. (2017). Vessels fuel consumption forecast and trim optimisation: A data analytics perspective. *Ocean Engineering*, *130*, 351–370. doi:10.1016/j.oceaneng.2016.11.058

Deepu, T. & Ravi, V. (2021). Supply chain digitalization: An integrated mcdm approach for inter-organizational information systems selection in an electronic supply chain. *International Journal of Information Management Data Insights, 1*.

Du Meng, Q. (2019). Y. and, S. Wang, and H. Kuang. Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. *Transportation Research Part B: Methodological*, *122*, 88–114. doi:10.1016/j.trb.2019.02.004

Eldershaw, C. & Hegland, M. (1997). Cluster analysis using triangulation. *Computational Techniques and Applications*, 201–208. Research Gate.

Farkas, A., Song, S., Degiuli, N., Martić, I., & Demirel, Y. K. (2020b). Impact of biofilm on the ship propulsion characteristics and the speed reduction. *Ocean Engineering*, *199*, 107033. doi:10.1016/j.oceaneng.2020.107033

Fermani, M., Rossit, D.G., & Toncovich, A. (2021). *A simulated annealing algorithm for solving a routing problem in the context of municipal solid waste collection*. Springer.

Fey, M., Eric Lenssen, J., Weichert, F., & Müller, H. (2018). Splinecnn: Fast geometric deep learning with continuous B-spline kernels. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 869–877). IEEE. 10.1109/CVPR.2018.00097

Filippopoulos, I., Panagiotakopoulos, T., Skiadas, C., Triantafyllou, S. M., Violaris, A., & Kiouvrekis, Y. (2022). Live vessels' monitoring using geographic information and internet of things. *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA).* IEEE.

Folgheraiter, M. (2016). A combined B-spline-neural-network and ARX model for online identification of nonlinear dynamic actuation systems. *Neurocomputing*, *175*, 433–442. doi:10.1016/j.neucom.2015.10.077

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 1–67.

Gao, X., Sun, K., Shi, S., Wu, B., & Zuo, Z. Research on influence of trim on a container ship's resistance performance. In *3rd International Conference on Fluid Mechanics and Industrial Applications*, volume 1300. IOP Publishing, 2019. 10.1088/1742-6596/1300/1/012105

Geneva L. S. & With, S.B. (2017). *Standardization as an enabler of digitalization in the maritime industry.* DNV GL Group Technology & Research Position Paper.

Hanssen, G., & James, R. (1960). Optimum ship routing. *Journal of Navigation*, *13*(3), 253–272. doi:10.1017/S0373463300033580

Islam, H., & Guedes Soares, C. (2019). Effect of trim on container ship resistance at different ship speeds and drafts. *Ocean Engineering*, *183*, 106–115. doi:10.1016/j.oceaneng.2019.03.058

Jeon, M., Noh, Y., Shin, Y., Lim, O. K., Lee, I., & Cho, D. (2018). Prediction of ship fuel consumption by using an artificial neural network. *Journal of Mechanical Science and Technology*, *32*(12), 5785–5796. doi:10.1007/s12206-018-1126-4

Kaklis, D., Eirinakis, P., Giannakopoulos, G., Spyropoulos, C., Varelas, T. J., & Varlamis, I. 2022a. A big data approach for fuel oil consumption estimation in the maritime industry. *2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService).* IEEE. 10.1109/BigDataService55688.2022.00014

Kaklis, D., Giannakopoulos, G., Varlamis, I., Spyropoulos, C. D., & Varelas, T. J. (2019). A data mining approach for predicting main-engine rotational speed from vessel-data measurements. *Proceedings of the 23rd International Database Applications & Engineering Symposium*, (pp. 1–10). ACM. 10.1145/3331076.3331123

Kaklis, D., Varelas, T. J., Varlamis, I., Giannakopoulos, G., Eirinakis, P., & Spyropoulos, C. (2023). From STEAM to Machine: Emissions control in the shipping 4.0 era. *The 8th International Symposium on Ship Operations, Management & Economics (SOME)*, (pp. 9-12). IEEE.

Kaklis, D., Varlamis, I., Giannakopoulos, G., Spyropoulos, C., & Varelas, T.J. (2022b). *Online training for fuel oil consumption estimation: A data driven approach*. Research Gate.

Schultz, M. P., Walker, J. M., Steppe, C. N., & Flack, K. A. (2015). Impact of diatomaceous biofilms on the frictional drag of fouling-release coatings. *Biofouling*, *31*(9–10), 759–773. doi:10.1080/08927014.2015.1108407 PMID:26652667

Shin, Y. (2020). Near-optimal weather routing by using improved A* algorithm." *Applied Sciences*.

Valchev, I., Coraddu, A., Kalikatzarakis, M., Geertsma, R., & Oneto, L. (2022). Numerical methods for monitoring and evaluating the biofouling state and effects on vessels' hull and propeller performance: A review. *Ocean Engineering*, *251*, 110883. doi:10.1016/j.oceaneng.2022.110883

Vettor, R., & Soares, C. G. (2015). Multi-objective route optimization for onboard decision support system. A. Weintrit & T. Neumann (Eds.), Information, Communication and Environment: Marine Navigation and Safety of Sea Transportation. CRC Press, Taylor & Francis Group.

Watanabe, S., Nagamatsu, N., Yokoo, K., & Kawakami, Y. (1969). The augmentation in frictional resistance due to slime. *J. Kansai Soc. Nav. Archit.*, *131*, 45–51. doi:10.2534/jjasnaoe1968.1969.126_45

Yuan, J., & Nian, V. (2018). Ship energy consumption prediction with gaussian process metamodel. *Energy Procedia*, *152*, 655–660. doi:10.1016/j.egypro.2018.09.226

Zhu, Y., Zuo, Y., & Li, T. (2020). Predicting ship fuel consumption based on lstm neural network. *2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*. IEEE. 10.1109/ICCSS52145.2020.9336914