Rule and Blockchain-based Data Management Framework to Facilitate Ship Efficiency Assessment

Shuai Wang[1], Bingjie Guo[1] and Nikita Karandikar[1]

[1] *Group Research and Development, DNV AS, Veritasveien 1, 1363 Høvik, Norwsay*

Abstract

To meet the International Maritime Organization's (IMO) target of achieving net-zero greenhouse gas (GHG) emissions by 2050, the maritime industry is exploring diverse decarbonization strategies. Among them, enhancing ship energy efficiency through data-driven performance assessment has gained significant attention. The Vessel Technical Index (VTI), introduced by DNV, serves as a hydrodynamic performance indicator to support operational efficiency monitoring and emissions reduction. However, the reliability of VTI is highly dependent on the quality and integrity of its input data. This paper presents an early-stage industrial data management framework that combines rule-based mechanisms, maritime domain knowledge, and blockchain technology to ensure trustworthy input data for VTI calculation. The framework aims to improve data quality (accuracy, completeness and representativeness) and data integrity (data tamper-proof). One use case is further discussed to demonstrate the applicability and potential impact of the proposed framework.

Keywords

Data management framework, data quality, data integrity, ship efficiency assessment, blockchain

1. Introduction

The International Maritime Organization (IMO) has defined a comprehensive strategy for maritime industry to reach the greenhouse gas (GHG) emission net-zero ambition by 2050. Guided by this, a substantial effort has been invested by various maritime industries for shipping decarbonization with the focus on multiple areas, e.g., energy efficiency measures, alternative clean fuels and battery-powered electric vessels [1]. Among these, one of the most promising pathways is to adopt cost-effective solutions to increase ship energy efficiency in operations. The maritime forecast to 2050 indicates that one third of emissions can be reduced through speed reduction and implementation of energy efficiency measures [4]. However, several factors impede the implementation of these measures, including insufficient information, uncertainty regarding hidden costs and benefits, asymmetry of information among stakeholders, conflicts of interest, and split incentives [2][3]. To remove the barrier for application of energy efficiency measure, it is required a trustworthy, transparent and reliable means to assess and monitor the operational ship performance (e.g., fuel consumption and GHG emission) [1].

To achieve this goal, a hydrodynamic performance indicator named vessel technical index (VTI) has been introduced by Det Norske Veritas (DNV) with the aim of assisting ship owners and operators in monitoring the operational ship energy efficiency performance and thus reducing fuel consumption and GHG emissions when ships are on operations over time [2]. DNV is a globally independent assurance and risk management company, which provides certification, classification, advisory, and verification services across industries like maritime, energy, healthcare, and digital technology. While the technical details of calculating VTI are out of scope in this paper (can be consulted in [2][3]), VTI is a data-driven measure, which heavily replies on the input data (e.g., weather, shaft power and calm-water resistance) [2][5]. Poor data quality or incorrect/biased input data could lead to inaccurate calculation of VTI and thereby result in wrong decision-making

RuleML+RR'25: Companion Proceedings of the 9th International Joint Conference on Rules and Reasoning, September 22--24, 2025,Istanbul, Turkiye

shuai.wang@dnv.com (S. Wang); bingjie.guo@dnv.com (B. Guo); nikita.karandikar@dnv.com (N. Karandikar)

processes and potentially bring in economic loss for stakeholders (e.g., ship owners, operators and cargo owners).

To tackle this challenge, this paper introduces an early-stage industrial data management framework based on rules, maritime domain expertise and blockchain to ensure high data quality and integrity for VTI calculation input. Data quality in this context refers to the degree which the input data to calculate VTI (e.g., shaft power) is accurate, complete and representative (Section 2.1 and 2.2). Data integrity in this context refers to the degree which the input data to calculate VTI remains unmanipulated/untampered from its original state collected from the sensors of vessels and thus can be trusted by the subsequent decision-making process (Section 2.3). In addition, one case scenario is discussed to demonstrate the applicability of the proposed data management framework (Section 3) before the conclusion (Section 4).

2. Data Management Framework

This section presents the data management framework in detail for ensuring data quality and integrity (as shown in Figure 1), which consists of three key components: a) rule-based data cleaner with the aim at cleaning and filtering the input data (e.g., outliers for shaft power from broken sensors) with pre-defined metrics and rules (Section 2.1), b) a domain expertise based data validator with the aim at further validating the input data against the domain expertise based on different independent data sources and physical models (Section 2.2) and 3) a blockchain-based data checker with the aim at assessing whether the input data to VTI preserves the same compared with the original data collected from sensors (Section 2.3). As illustrated in the left-top of Figure 1, a data listener microservice is first designed to continuously collect data either directly from customer data streams or via a data batch file (i.e., .csv files) and store the collected data into database (DB) of our industrial data cloud platform Veracity. The three components (discussed below) will interact with the DB and maritime domain experts correspondingly. Note that there is no sequential order practically to apply these three components and each component can be enabled and disabled based on VTI calculations of specific vessels. However, it is usually recommended to first apply the data checker for assessing the data is not tampered/manipulated before the data cleaning and validation.

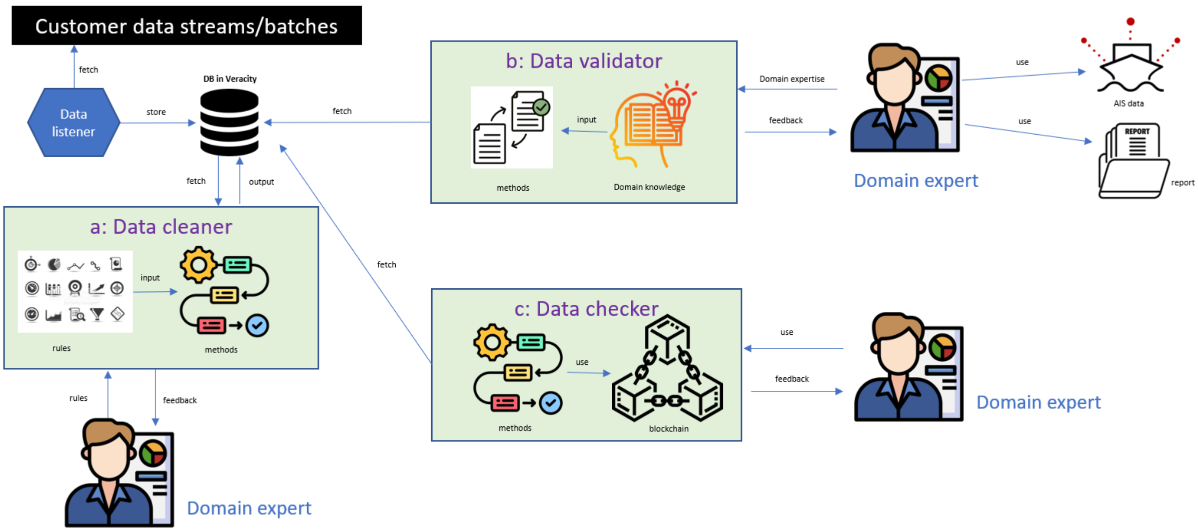

Figure 1: An overview of data management framework

2.1. Rule-based data cleaner

The first component is a rule-based data cleaner (in Figure 1) to systematically assess and clean raw data by applying a set of pre-defined rules. The cleaner currently focuses on three dimensions of data quality: accuracy, completeness, and representativeness. Note that additional dimensions could be introduced in other contexts.

Accuracy. To ensure accuracy, the cleaner utilizes logical constraints based on the expected behavior of onboard systems and sensors, which includes checking ranges, checking rate-of-change and checking sensor correlation.

1. *Checking ranges*: Shaft power values are checked to ensure they fall within the reasonable engine power range for specific vessel types and configurations. Values that fall outside reasonable operational ranges (e.g., negative shaft power) are discarded.
2. *Checking rate-of-change*: Relation consistency is checked by assessing the rate of change between consecutive data points. For instance, sudden spikes or drops for speed and shaft power that exceed acceptable thresholds are marked as anomalies.
3. *Checking sensor correlations*: Correlations between parameters are checked (e.g., shaft power and vessel speed). Unrealistic deviations such as high power but abnormally low speed may indicate faulty sensors with data anomalies or adverse environmental conditions.

Completeness. To ensure the completeness of time-series data, especially for VTI, the data cleaner performs the following tasks:

1. *Detecting missing data*: Missing values (e.g., shaft power, wind speed) are detected using methods such as timestamp-based continuity checks. Gaps longer than a predefined duration (e.g., 10 minutes) will be discarded.
2. *Generating missing data*: For short- duration data loss, data generation methods (e.g., linear models) are tried to produce consistent and reliable augmented data to bridge the data gap. For long-duration data loss, methods such as generating data based on historical medians/deviations for similar operational profiles are tried to assess whether realistic data could be generated. If not, the data will be excluded.
3. *Checking thresholds*: A completeness score is calculated for given time windows. Data segments falling below a minimum completeness threshold (e.g., <80%) are discarded to preserve data completeness.

Representativeness. To maintain data representativeness, the cleaner filters out data that does not reflect realistic vessel performance with the following actions:

1. *Filtering weather conditions*: Data collected during extreme sea states (e.g., wave height > 5m) are excluded as they are not typical sea state of ship operation.
2. *Removing outliers*: Outliers detection methods (e.g., local outlier factor) are applied to detect and remove outliers that are not representative compared with overall data distributions.

The rule-based component focuses on filtering and removing data that are not accurate, incomplete and not representative for ensuring the reliability of VTI evaluations and decision-making.

2.2. Domain knowledge-based data validator

The second key component is a domain knowledge-based data validator (in Figure 1) that validates the input data against various independent sources (such as Automatic Identification System (AIS), noon report and hindcast data) and physical models. This aims to ensure the plausibility and consistency of data used for VTI. It is worth mentioning that measurement data from vessels often faces sensor drift and zero-adjustment issues, which standard analysis can't detect [13]. Therefore, cross-checking independent data and physics correlations is essential to ensure data validation. In practice, the crosschecking validator divides the measured parameters into cross-checking of navigation data, cross-checking of ship performance data and cross-checking of weather data (presented as below).

Cross-checking of navigation data. For ships over 300 gross tons on international voyages, cargo ships over 500 gross tons not on international voyages, and all passenger ships, navigation information is collected by AIS, which uses independent measurement systems. The cleaned and postprocessed AIS data is then used to cross-check high-frequency navigation data for VTI calculations.

Cross-checking of ship performance data. IMO introduced certain regulations that are mandatory for ships larger than 5000 GT in international voyages since 2019. Fuel oil consumption, distance travel, hours underway, and other operational parameters are reported and can verify measured performance data like ship shaft power and fuel consumption.

Cross-checking of weather data. There are several numerical models for estimating global sea state, for example, ECMWF Reanalysis v5 (ERA5) [1], which represents the fifth generation of the atmospheric reanalysis of the global climate. ERA5 amalgamates extensive historical observations into comprehensive global estimates through advanced modeling and data assimilation systems. It delivers hourly estimates of numerous atmospheric, land, and oceanic variables. Using ship navigation information, the ship operation environment can be interpolated in the ERA 5 hindcast data set. More details refer to hindcast weather interpolation can refer to [13]. The interpolated hindcast data can be used to verify the measured weather data.

For a given ship, performance parameters correlate with navigation and environment. The correlation can also be utilized to check the plausibility of measured data. The proposed domain knowledge-based data validator aims at checking relevant parameter validity before VTI calculation. Significant discrepancies will result in data being discarded, with a warning sent to domain experts. Note that this data validator component is highly coupled with the maritime context for ship energy efficiency, but the idea behind it could be employed in other contexts by incorporating corresponding domain knowledge.

2.3. Blockchain-based data checker

The third component is a blockchain-based data checker (in Figure 1) based on our previous work [6] [7] to ensure the integrity of input data used for calculating VTI. Recall that data integrity in this context refers to the input data is tamper-proof (not manipulated intentionally or unintentionally) compared with the original data collected from customers' sensor data streams or batches. While the first two components focus on data quality (e.g., accuracy and completeness), blockchain technology [9][10] offers a decentralized and tamper-evident mechanism to verify that the input data remains unaltered from the point of collection to its final use in VTI analysis.

This component builds on the architecture introduced in our earlier work [7], which leverages blockchain (and containerized infrastructure to securely hash and verify maritime data. The checker operates by hashing the input data (e.g., shaft power, weather, vessel resistance) using the hashing algorithm SHA-256 [11], then storing the hash along with metadata (e.g., timestamp and transaction ID) on the blockchain. This immutable record allows any stakeholder (e.g., analysts, charterers) to later verify that the dataset used in the VTI calculation is the same as the original data collected onboard. In practice, this blockchain-based checker functions in three key phases:

Data Hashing and Registration. During data collection onboard, cleaned and representative data segments from the first two components (Section 2.1 and 2.2) are hashed using the algorithm SHA-256. These hashes, along with associated metadata, are registered on the blockchain. The hash is a cryptographic fingerprint of the data—any alteration, however minor, would produce a completely different hash.

Verification Prior to VTI Calculation. Before VTI computation begins, the data checker hashes the input dataset again and queries the blockchain to retrieve the registered hash. If the new hash matches the stored hash, integrity is confirmed, and the data is marked green for trusted use. If there is no

[1] https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5

match or if the hash is missing, a red status is returned, indicating possible tampering or data replacement.

Tamper Alert and Traceability. In the case of red status, the checker logs a detailed alert, enabling analysts or auditors to trace the potential point of compromise—whether due to system errors, unauthorized access, or manual manipulations. This provides a basis for trust not only in the data itself but also in the processes built upon it.

Integrating this blockchain-based checker ensures that the input data integrity is verifiable and transparent, which is particularly important for scenarios where ship efficiency metrics affect economic or regulatory decisions. It also increases confidence in VTI calculation, especially when analysis results are shared across organizations with different data governance practices. Further, it allows a data producer to provide an evidence based timestamped body of evidence for VTI and for the data consumer to verify data integrity without having to be onboarded into the data producer's data infrastructure.

Looking ahead, the checker can be extended to support data provenance tracking, integration with smart contracts for automated auditing, and real-time verification of streaming data, further enhancing the trust ecosystem around ship performance monitoring.

3. One user case scenario

To illustrate the practical application of the proposed data management framework, the use case scenario considers a bulk carrier, and apply VTI as a real-time efficiency indicator to monitor vessel technical performance. During a certain operational period, the vessel's onboard monitoring system collected continuous sensor data, including shaft power, vessel speed, fuel flow, weather conditions, and navigational status. However, an initial check revealed that a set of shaft power values were either missing or fluctuating beyond expected limits. Additionally, several data points seem not representative of typical vessel speed ranges and therefore probably not suitable for VTI calculation.

The blockchain-based data checker was first called via an API to assess if the data is tampered compared with the original data collected from the sensors. A green light was returned in this scenario and thus indicating the input data is trustworthy to be used for cleaning and validating. The rule-based data cleaner was then applied to identify those data anomalies (e.g., unrealistic vessel speed) by either correcting or excluding them from the input dataset. Moreover, the domain knowledge-based validator was utilized to further validate the input dataset. For example, a data segment shows a spike in shaft power while AIS indicates the vessel is at anchor status. Such data segment should be either automatically discarded or reported to data owners/ship operators for clarifications. Once data quality and integrity were assessed, VTI can be calculated based on the methods in [2][3] and the results show the ship technical performance changed over time.

In such way, ship technical performance and efficiency of energy efficiency measures can be evaluated in a transparent way, which can help ship operators to optimize ship maintenance and remove the barriers of utilizing energy efficiency measures. This case demonstrated that high-quality, trustworthy data is essential not only for calculating a trustworthy VTI, but for translating performance insights into actionable decisions.

4. Conclusion

This paper presented an early-stage industrial data management framework for ensuring data quality and integrity in support of accurate and trustworthy Vessel Technical Index (VTI) calculation and thereby facilitating ship technical efficiency assessment in the maritime sector. To address the challenges posed by data quality and integrity, the framework is comprised of a rule-based data cleaner, a domain knowledge-based validator and a blockchain-based data checker. Through a realistic case study involving operational voyage data from a bulk carrier, we illustrated how this framework can be employed to enhance the reliability and trustworthiness of VTI calculations.

Looking forward, several future work will be envisioned. First, the data cleaner will be extended with more (domain-specific) rules to predict complex patterns and identify data anomalies. Second, the domain knowledge-based data validator will be further refined based on the feedback from domain experts and further evaluated with the in-service data from vessels. Moreover, the framework will be integrated into the business service of VTI as the first layer for ensuring high data quality and integrity for VTI calculations. This work contributes to ship performance monitoring systems in the context of maritime digitalization and supports the industry's transition toward energy-efficient and sustainable operations. The proposed framework can also be utilized in other industrial contexts for data-driven services with data quality and integrity challenges.

Acknowledgements

This work is supported by the Sea-Prime project (no. 352964) funded by the Research Council of Norway.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly to check grammar and spelling. After using it, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

[1] Det Norsk Veritas (DNV), Energy Efficiency report. 2025.

[2] B. Guo, O. Rognebakke, H.A. Tvete, C. Adal, G. Storhaug, M. Schmidt, T. Bruset, & G. Prytz, Setting the Standard for Evaluation of in-service technical ship performance. HullPic, 2021.

[3] DNV, Recommended Practice for Technical Ship Performance (DNV-RP-0675), 2023.

[4] DNV, Maritime Forecast to 2050, 2024.

[5] H.A. Tvete, B. Guo, C. Agrell, C. Ferreira, S. Eldevik, M. Schmidt, G. Storhaug, Uncertainty analysis on vessel technical index for technical ship performance, 15th International Symposium on Practical Design of Ships and Other Floating Structures, pp. 1107-1123, 2022.

[6] K.E. Knutsen, Q. Liang, N. Karandikar et al, Containerized immutable maritime data sharing utilizing Distributed Ledger Technologies, Journal of Physics, Conference Series, pp. 1-11, 2022.

[7] S. Wang, N. Karandikar, K.E. Knutsen, X. Tong, T. Edseth, Z. Xu, Enhancing Maritime Data Standardization and Integrity using Docker and Blockchain, IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings, pp. 370-374, 2023.

[8] B. Guo, H.A. Tvete, B.J. Vartdal, S. Wang, Piloting of Vessel Technical Index, Proceedings of the ASME 43rd International Conference on Ocean, Offshore and Arctic Engineering. Volume 5A: Ocean Engineering, June 9–14, 2024.

[9] Z. Zheng, S. Xie, H. Dai, X. Chen and H. Wang, An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends, 2017 IEEE International Congress on Big Data (BigData Congress), pp. 557-564, 2017.

[10] M. Sharples and J. Domingue, The blockchain and kudos: A distributed system for educational record, reputation and reward, in Proceedings of 11 th European Conference on Technology Enhanced Learning, pp. 490–496, 2015.

[11] S. Gueron, S. Johnson and J. Walker, SHA-512/256, 2011 Eighth International Conference on Information Technology: New Generations, pp. 354-358, 2011.

[12] Cattaneo, C., 2019, “Internal and external barriers to energy efficiency: which role for policy interventions?,” Energy Efficiency 12, 1293–1311 (2019).

[13] B. Guo, P. Gupta, S. Steen, H.A. Tvete, B.J. Vartdal, Comparison of Ship Technical Performance Analysis Based on Public Hindcast Data and Measured Weather Data, ISOPE 2025, pp.2730 – 2737, 2025.