Article

Quantitative Assessment of Consistency between IMO DCS and EU MRV Systems Using Large-Scale Operational Data

Hyunju Lee and Hyerim Bae *

Article

Quantitative Assessment of Consistency Between IMO DCS and EU MRV Systems Using Large-Scale Operational Data

Hyunju Lee [1] and Hyerim Bae [2,*]

[1] Korean Register, Busan, Republic of Korea

[2] Department of Data Science, Graduate School of Data Science, Pusan National University, Busan, Republic of Korea

* Correspondence: hrbae@pusan.ac.kr; Tel: +82-51-510-2733

Abstract

This study presents a large-scale empirical comparison of operational efficiency metrics derived from the IMO Data Collection System (DCS) and the EU Monitoring, Reporting and Verification (MRV) framework. Using a matched dataset of 15,755 dual-reported vessels and over 50,000 ship-year observations from 2019 to 2024, paired non-parametric tests, effect size estimation, and agreement diagnostics were applied to assess consistency across monitoring systems. Results indicate that although statistically significant differences are detected ($p < 0.001$), practical differences are negligible (Cohen's $d < 0.025$), with MRV-based values averaging approximately 1.4% lower Annual Efficiency Ratio (AER) and fuel intensity than DCS values. Distributional analysis confirms substantial overlap between datasets, and temporal trends show progressive convergence following the implementation of the Carbon Intensity Indicator (CII) regulation. However, pronounced vessel-type heterogeneity is observed. Flexible cargo vessels exhibit consistent efficiency improvements in EU-related voyages, whereas container ships show minimal variation and LNG carriers demonstrate indicator-dependent patterns. Overall, the findings indicate that DCS and MRV provide broadly comparable representations of operational efficiency, with observed differences primarily reflecting vessel-type-specific operational characteristics rather than structural inconsistencies in reporting systems. The study contributes a scalable statistical validation framework for cross-regulatory monitoring assessment.

Keywords: IMO DCS; EU MRV; carbon intensity indicator (CII); operational efficiency; maritime GHG monitoring; vessel-type heterogeneity; statistical agreement analysis

1. Introduction

International shipping serves as the backbone of global trade, transporting approximately 80% of world merchandise by volume. However, this essential sector contributes around 3% of global anthropogenic greenhouse gas (GHG) emissions, equivalent to over 1 billion tonnes of CO_2 annually [1]. As maritime decarbonization efforts intensify, large-scale monitoring systems have become essential tools for tracking operational efficiency and emissions performance.

Two major data collection frameworks currently govern maritime GHG reporting: the IMO Data Collection System (DCS) and the EU Monitoring, Reporting and Verification (MRV) system. The IMO DCS, operational since 2019, requires ships above 5,000 gross tonnage engaged in international voyages to report annual fuel consumption, distance traveled, and hours underway [2]. The EU Monitoring, Reporting and Verification (MRV) Regulation, in force since 2018, mandates ships calling at EU ports to report CO_2 emissions, fuel consumption, and transport work for voyages involving EU ports, with voyage-level granularity and public data availability [3,4].

While both systems share the common objective of enhancing transparency in maritime emissions, they differ in geographic scope, reporting granularity, and transparency. Such structural differences introduce potential discrepancies in efficiency indicators derived from the two datasets. Ensuring cross-platform consistency is therefore critical for reliable performance assessment, indicator calibration, and monitoring system validation.

Recent regulatory developments further increase reliance on large-scale operational datasets for performance evaluation. The IMO adopted its revised GHG Strategy in 2023, setting a net-zero target by or around 2050 with intermediate checkpoints [5]. The Carbon Intensity Indicator (CII) regulation, effective from 2023, converts DCS-reported fuel consumption and activity data into operational efficiency ratings [6,7]. Concurrently, the EU has implemented FuelEU Maritime [8] and extended the EU Emissions Trading System (EU ETS) to maritime transport [9], both of which utilize MRV-reported emissions and activity data as quantitative compliance inputs. As these regulatory instruments increasingly depend on reported operational datasets, rigorous assessment of cross-system data consistency becomes critical for reliable performance evaluation and monitoring system validation.

Despite the critical importance of data consistency for regulatory harmonization, no comprehensive statistical comparison of efficiency metrics derived from matched DCS-MRV observations has been published in the peer-reviewed literature. Existing studies have primarily focused on institutional comparisons [4] or single-dataset analyses [10,11], without systematically quantifying vessel-level discrepancies using paired statistical methodologies.

This study addresses this gap by developing and applying a reproducible statistical validation framework to matched operational datasets from 15,755 dual-reported vessels over six years (2019–2024), comprising more than 50,000 ship-year observations. The analysis integrates paired non-parametric testing, effect size estimation, and agreement diagnostics to evaluate discrepancies in key efficiency indicators. Temporal convergence patterns following the implementation of CII are also examined.

By establishing a scalable quantitative approach for cross-platform monitoring system validation, this study contributes to applied maritime data analytics and provides methodological tools for evaluating large regulatory datasets.

The remainder of this paper is organized as follows. Section 2 reviews related literature on monitoring system comparisons. Section 3 outlines the regulatory data structures. Section 4 describes the research methodology and statistical framework. Section 5 presents the empirical results. Section 6 discusses methodological implications for monitoring system evaluation. Section 7 concludes.

2. Literature Review

2.1. Structural Characteristics of DCS and MRV Data Reporting Systems

The parallel implementation of the IMO Data Collection System (DCS) and the EU Monitoring, Reporting and Verification (MRV) framework has generated scholarly attention regarding their structural design and reporting characteristics. These studies primarily examine differences in data granularity, transparency, and reporting architecture rather than conducting quantitative validation of reported operational metrics.

Adamowicz [4] conducted a comprehensive analysis of the coexistence between EU MRV and IMO DCS, examining enforcement mechanisms, port state control implications, and penalty structures. The study highlighted that while both systems target ships above 5,000 GT, they differ fundamentally in data transparency—EU MRV mandates public disclosure of individual ship data through the THETIS-MRV platform, whereas IMO DCS data remains confidential at the ship level. The analysis concluded that regulatory alignment would require addressing these structural differences, but did not provide quantitative comparison of reported data.

Zis and Psaraftis [11] examined operational measures to mitigate potential modal shifts resulting from environmental legislation in the maritime sector. While their study focused primarily on the

competitive implications of sulphur regulations for short-sea shipping, it established important methodological precedents for analyzing how regulatory changes affect fleet operational behavior—a framework subsequently applied to MRV-related research.

Psaraftis [12] provided comprehensive discussions on the decarbonization challenges facing maritime transport, including critical assessments of the IMO's regulatory approach and the relationship between EU and IMO monitoring frameworks. These analyses emphasized the importance of data transparency and the limitations of confidential reporting systems for enabling external verification and academic research.

2.2. Quantitative Applications of DCS and MRV Datasets

Since their implementation, the IMO DCS (2019–) and EU MRV (2018–) databases have enabled a growing number of data-driven studies analyzing ship operational performance and emission characteristics. Existing research has primarily utilized either DCS or MRV datasets independently to evaluate efficiency indicators, emission distributions, and fleet-level trends.

Studies based on IMO DCS data have mainly focused on aggregated fuel consumption analysis and carbon intensity assessment. Yeo et al. [13] estimated greenhouse gas emissions from ships registered in South Korea using a bottom-up approach based on activity data. For the first time in South Korea, ship activity data was employed to estimate energy consumption and emissions by ship type, providing a methodological foundation for national-level fleet emission inventories that complement international reporting frameworks such as IMO DCS. Zhang et al. [14] employed IMO DCS fuel consumption data to investigate technical requirements for achieving the 2023 IMO GHG Strategy targets. Their analysis projected efficiency improvement pathways by combining DCS data with regulatory scenarios for EEDI, CII, and alternative fuel adoption through 2050, demonstrating the utility of DCS data for long-term decarbonization planning.

The EU MRV system has been extensively utilized for empirical analysis due to its public accessibility and voyage-level granularity. Panagakos et al. [15] conducted an early assessment of the MRV regulation using 2018 reporting data from 1,041 dry bulk carriers operated by a leading Danish shipping company. By recalculating MRV indicators on a global basis and comparing them with EU-only values, they demonstrated that geographic coverage restrictions introduce significant bias, limiting the intended use of MRV data for global benchmarking. Nevertheless, they acknowledged that the MRV regulation played a role in prompting the IMO to adopt its Data Collection System.

Xing et al. [10] conducted a comprehensive analysis of ship energy efficiency using EU MRV data from 2023, providing insights into fleet-wide efficiency distributions and identifying vessels requiring attention under tightening regulatory requirements. The study concluded that data accumulated in the EU MRV system provides powerful support for ship energy efficiency appraisals but identified data quality and harmonization with IMO DCS as urgent issues requiring resolution.

Luo et al. [16] analyzed MRV data from 2018 to 2021 to investigate temporal trends in ship emissions and evaluate carbon emission performance of shipping companies during the MRV implementation period. Their findings highlighted the impact of COVID-19 on the shipping industry, evidenced by declining numbers of ships recorded in the MRV system between 2019 and 2020, with passenger ships being most affected.

In addition to system-level analyses, indicator-based studies have applied IMO-defined operational efficiency metrics. Park and Cho [17] applied the Energy Efficiency Operational Indicator (EEOI) to analyze operational efficiency changes in large container vessels following bulbous bow modifications for slow steaming. Their study demonstrated fuel consumption reductions of 5.0–9.9% at speeds of 14–18 knots compared to the original design speed of 27 knots, providing empirical validation of EEOI as an operational efficiency metric and demonstrating the practical application of IMO-defined indicators to fleet optimization decisions.

Xing et al. [10] explicitly addressed the co-existence of MRV and DCS systems, noting that with the implementation of the IMO Net-Zero Framework, two parallel mechanisms on ship energy efficiency would continue to operate. They raised critical questions about whether this situation

would weaken Europe's industrial competitiveness and how effectively the systems could be harmonized to avoid fragmented regulation.

Although these studies demonstrate the analytical value of DCS and MRV datasets, they primarily analyze each system independently. Comprehensive statistical validation of consistency between matched DCS and MRV observations at the vessel level remains limited, leaving an important methodological gap in cross-platform monitoring system assessment.

2.3. Simulation and Modeling Studies under Overlapping Maritime Regulations

Recent studies have applied simulation and modeling approaches to evaluate the operational and economic implications of multiple maritime GHG regulations. These analyses primarily investigate compliance strategies and fuel transition pathways under overlapping regulatory constraints. Kwon et al. [18] presented a comparative analysis of EU ETS, FuelEU Maritime, and IMO GFI regulations through scenario-based simulation of a Pure Car and Truck Carrier (PCTC). Four strategies—conventional operation, e-diesel adoption, route optimization, and wind-assisted propulsion—were modeled to assess changes in GHG intensity and regulatory performance metrics. Their results demonstrated that alternative fuel adoption significantly reduces GHG intensity, whereas operational adjustments alone provide limited improvements without technological modification.

Jung [19] examined the characteristics of GHG life cycle assessment for alternative marine fuels, providing foundational understanding of well-to-wake emission calculations that underpin both FuelEU Maritime and the proposed IMO GFI framework.

Psaraftis and Kontovas [20] analyzed structural differences between EU MRV and IMO DCS reporting systems, including variations in geographic scope, data aggregation level, and transparency. While their analysis clarified conceptual and institutional distinctions between the systems, it did not provide quantitative validation of consistency using matched vessel-level observations.

Overall, existing modeling and comparative studies focus on scenario evaluation, life-cycle calculation methods, and structural regulatory differences. However, empirical statistical assessment of cross-system consistency using large-scale matched operational datasets remains limited.

2.4. Methodological Gap in Cross-System Validation

The review of existing literature reveals a clear methodological gap: although structural and qualitative comparisons of DCS and MRV have been conducted, no study has systematically evaluated operational efficiency metrics derived from matched ship-year observations across both systems using large-scale statistical techniques. Table 1 summarizes the positioning of existing studies relative to this gap.

Table 1. Comparison of Existing Studies and Research Gap.

Study	Primary Focus	Data Source	Quantitative Comparison
Yeo et al. [13]	Korean GHG emissions estimation	Activity data (Korea)	No
Zhang et al. [14]	Emission pathway modeling	IMO DCS (Global)	No
Panagakos et al. [15]	MRV regional bias assessment	EU MRV only	No
Xing et al. [10]	Ship energy efficiency analysis	EU MRV only	No

Park and Cho [17]	EEOI operational assessment	Operational data	No
Kwon et al. [18]	Scenario-based regulatory modeling	Simulation (1 ship)	No
Psaraftis and Kontovas [20]	Framework comparison	Regulatory documents	No
This study	**Cross-system efficiency validation**	**matched DCS & MRV dataset**	**Yes (15,755 ships, 2019-2024)**

This study addresses the identified gap by developing and applying a large-scale paired statistical validation framework to matched DCS–MRV operational datasets. Using 15,755 dual-reported vessels over six years (2019–2024), the analysis evaluates whether systematically different efficiency profiles emerge when comparing globally aggregated DCS data with voyage-level MRV data for the same vessels. By quantifying cross-platform consistency using reproducible statistical methods, this research contributes a scalable methodological approach for monitoring system validation and large regulatory dataset assessment.

3. Maritime GHG Monitoring and Reporting Systems

This section provides an overview of the regulatory instruments governing maritime GHG emissions monitoring and reporting, establishing the policy context for the subsequent empirical analysis.

3.1. IMO Data Collection System and Carbon Intensity Indicator

The IMO Data Collection System (DCS), operational since January 2019 [2], establishes a standardized framework for collecting annual ship-level operational activity data, including fuel consumption, distance traveled, and hours underway for vessels above 5,000 GT engaged in international voyages. Reported data are aggregated at the flag state level and submitted to the IMO database, with ship-level records remaining confidential.

The DCS dataset serves as the primary data source for computing operational carbon intensity metrics under the Carbon Intensity Indicator (CII), introduced in 2023 [6]. CII transforms reported fuel consumption and transport activity into efficiency indicators expressed as grams of CO_2 per unit transport work (e.g., gCO_2/dwt-nm for cargo vessels). The Annual Efficiency Ratio (AER) is commonly derived from DCS data and functions as a core operational efficiency metric within the CII evaluation framework [7].

3.2. EU Monitoring, Reporting and Verification Regulation

The EU MRV Regulation (Regulation 2015/757) has been in force since January 2018, predating the IMO DCS by one year [3]. It requires ships above 5,000 GT calling at EU/EEA ports to report fuel consumption, CO_2 emissions, distance traveled, time at sea, cargo carried, and transport work. In contrast to the aggregated reporting structure of DCS, MRV data are publicly available at the ship level through the THETIS-MRV platform, providing greater transparency and voyage-level granularity.

From 2024 onward, the scope of MRV has been expanded to include methane (CH_4) and nitrous oxide (N_2O) emissions on a well-to-wake basis, thereby aligning the reporting structure with lifecycle-based emission accounting methodologies adopted in subsequent regulatory instruments.

3.3. Comparison of DCS and MRV Systems

Table 2 summarizes the key differences between IMO DCS and EU MRV regulatory frameworks.

Table 2. Comparison of IMO DCS and EU MRV Monitoring Systems.

Category	IMO DCS	EU MRV
Entry into force	January 2019	January 2018
Applicable ships	>=5,000 GT, international voyages	>5,000 GT, calling at EU/EEA ports
Geographic scope	Global	EU-related voyages
Reporting period	Annual (calendar year)	Annual (calendar year)
GHG scope	CO_2 (Tank-to-Wake)	CO_2, CH_4, N_2O (WtW from 2024)
Reporting Items	Annual fuel consumption, distance, hours underway	Fuel consumption, CO_2 emissions, distance, cargo, time at sea, transport work
Data accessibility	Confidential (IMO GISIS)	Public (THETIS-MRV)
Documentation Required	SEEMP Part II: Fuel Oil Data Collection Plan	Monitoring Plan, annual Emissions Report, Document of Compliance
Verification	Flag state or RO	Third-party verifier
Link to Net Zero Strategy	Provides baseline database for IMO GHG strategy	Integrated with EU ETS and FuelEU Maritime as part of Fit for 55 package

The two systems share common elements - both target large ships, require annual reporting, and employ similar fuel-based emission calculation methodologies. However, they differ significantly in geographic scope (global vs. EU-related), data transparency (confidential vs. public), and enforcement mechanisms (flag state oversight vs. port state expulsion orders).

3.4. Evolution of Intensity-Based and Market-Based Instruments

Beyond monitoring and reporting systems, several instruments introduce performance-based emission intensity targets and market mechanisms. The IMO revised GHG Strategy (2023) established long-term emission reduction pathways and introduced the concept of GHG Fuel Intensity (GFI), expressed in lifecycle-based units (gCO₂eq/MJ) [21]. Complementary methodological guidelines for well-to-wake accounting were adopted in Resolution MEPC.391(81) [22], defining default emission factors and calculation procedures for marine fuels.

At the regional level, FuelEU Maritime establishes progressive reductions in lifecycle GHG intensity relative to a 2020 reference value [8], while the EU ETS applies carbon pricing to reported emissions [9]. Although these instruments differ in implementation mechanisms, they rely on reported operational datasets and intensity metrics derived from DCS and MRV structures. Table 3 provides a comprehensive comparison of these regulatory instruments, highlighting their differences in scope, target metrics, and compliance mechanisms.

3.5. Implications for Cross-System Data Consistency

The increasing reliance on reported operational datasets for intensity-based evaluation and emission accounting highlights the importance of cross-system consistency between DCS and MRV. If efficiency metrics derived from these systems exhibit systematic divergence, analytical comparability and monitoring system reliability may be affected.

Given the methodological alignment toward lifecycle-based intensity indicators, empirical validation of consistency between DCS and MRV datasets becomes essential for ensuring robustness of performance evaluation metrics. This study addresses this requirement through large-scale matched statistical assessment.

4. Methodology

This section describes the data architecture, matching algorithm, efficiency indicator formulation, and statistical procedures employed to evaluate cross-system consistency between IMO DCS and EU MRV datasets.

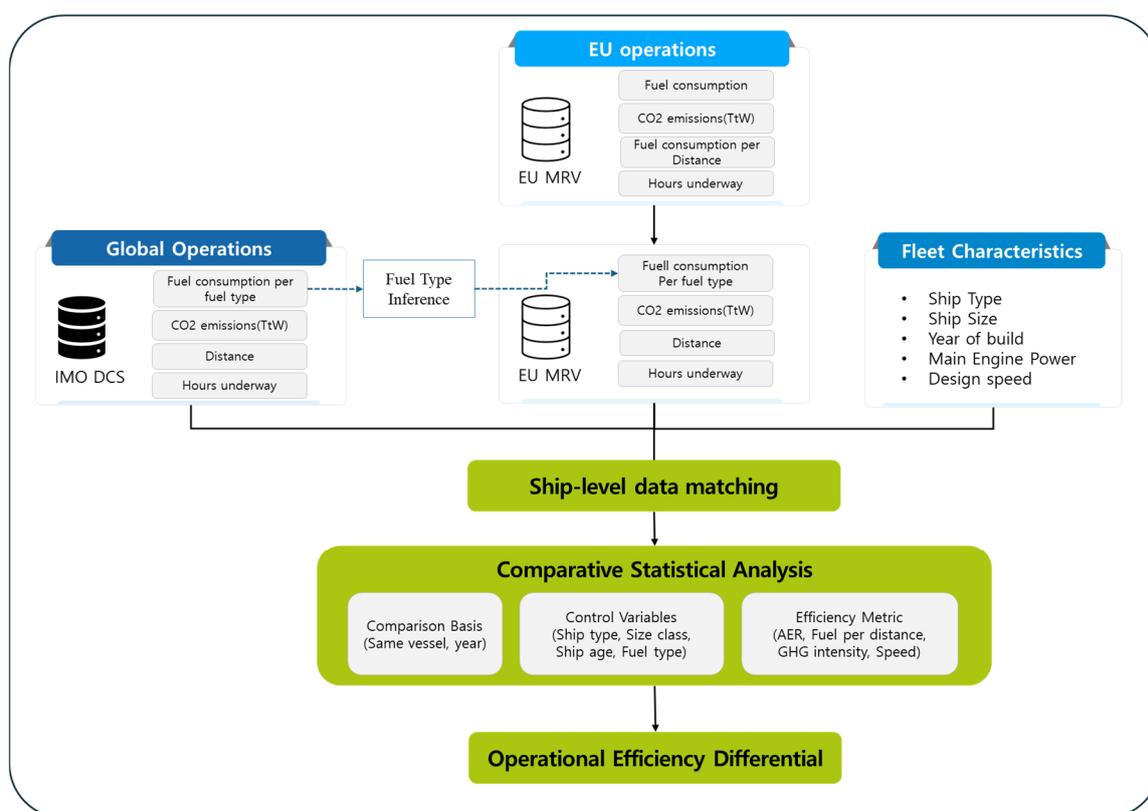Figure 1 illustrates the analytical workflow. The framework consists of four sequential stages:

Figure 1. Analytical workflow for cross-system comparison between IMO DCS and EU MRV datasets. The framework includes data acquisition, ship-level matching, fuel type harmonization, efficiency metric computation, and statistical consistency evaluation.

(1) Data acquisition from IMO DCS (global operations) and EU MRV (EU-related voyages);

(2) Ship-level matching based on IMO number and calendar year;

(3) Fuel type harmonization through inference using DCS-reported fuel composition for matched vessels;

(4) Computation and statistical comparison of operational efficiency metrics.

Because EU MRV does not provide fuel consumption disaggregated by fuel type, fuel type allocation for MRV records was inferred proportionally using corresponding DCS fuel type distributions for each matched vessel-year observation. Fleet characteristics including ship type,

deadweight tonnage, build year, main engine power, and design speed were incorporated as control variables.

Comparative metrics include Annual Efficiency Ratio (AER), fuel consumption per nautical mile, greenhouse gas intensity, and average operating speed. Statistical analysis quantifies the magnitude and direction of efficiency differentials between global (DCS) and EU-related (MRV) operations.

4.1. Data Sources

Two regulatory datasets are employed in this study.

The IMO DCS dataset was obtained through the Global Integrated Shipping Information System (GISIS), providing annual ship-level records of fuel consumption (disaggregated by fuel type), total distance traveled, and hours underway for global operations.

The EU MRV dataset was accessed via the THETIS-MRV platform operated by the European Maritime Safety Agency (EMSA). It provides verified annual ship-level totals of fuel consumption, CO_2 emissions, time at sea, cargo carried, and transport work for EU-related voyages.

The study period spans 2019–2024, covering six consecutive reporting years during which both monitoring systems operated in parallel. This period includes pre- and post-implementation phases of the CII regulation (effective January 2023), enabling assessment of temporal convergence patterns.

4.2. Ship Matching and Data Harmonization

Ship-level matching was performed using the IMO number as a unique identifier. Only observations appearing in both datasets within the same calendar year were retained. This one-to-one ship-year matching procedure ensures that differences reflect variation in operational context (global versus EU-related voyages) rather than differences in fleet composition.

Observations with inconsistent or duplicate IMO identifiers were removed prior to matching. The matching process resulted in a balanced panel of vessels reported under both monitoring regimes.

Because the EU MRV dataset does not disclose fuel consumption disaggregated by fuel type, fuel-type-specific components for MRV observations were harmonized using proportional fuel mix distributions derived from the corresponding DCS ship-year records. This proportional allocation preserves total reported fuel consumption while ensuring consistency in fuel-based emission calculations across systems.

The harmonization procedure assumes that the annual fuel composition reported under DCS reasonably approximates the fuel mix used during EU-related voyages for the same vessel-year. Given the annual aggregation structure of both systems, this assumption is considered methodologically consistent for comparative intensity analysis.

4.3. Data Validation and Filtering

To ensure robustness and engineering plausibility, identical validation procedures were applied to both datasets prior to statistical comparison. The filtering process comprised three sequential stages.

First, ship-year observations with missing, zero, or non-positive values for key variables (fuel consumption, distance traveled, and deadweight tonnage) were excluded. This step removed approximately 2.3% of matched records.

Second, physically implausible operational profiles were removed using engineering-based thresholds. Observations were excluded if: (i) AER exceeded 50 gCO2/dwt-nm (above the 99.9th percentile of the matched sample), (ii) average sailing speed exceeded 30 knots, or (iii) fuel consumption per nautical mile exceeded 1,000 kg/nm. These thresholds were selected to eliminate extreme reporting anomalies rather than operational variability.

Third, statistical outliers were identified using the Interquartile Range (IQR) method. For each efficiency metric, observations outside the interval [Q1 − 3×IQR, Q3 + 3×IQR] were excluded. The

3×IQR criterion was adopted to minimize excessive trimming while maintaining robustness against extreme distortions.

After filtering, the final analytical sample comprises between 50,055 and 50,079 ship-year observations (depending on metric-specific missing values), representing 15,755 unique vessels across 13 ship-type categories.

4.4. Efficiency Indicator Definitions

Three operational efficiency indicators were constructed to evaluate cross-system consistency. The selection was based on three criteria: (1) methodological compatibility with both monitoring systems, (2) availability of required variables in DCS and MRV datasets, and (3) complementary representation of ship operational performance.

(1) Annual Efficiency Ratio (AER)

The Annual Efficiency Ratio (AER) was adopted as the primary carbon intensity metric. AER normalizes CO_2 emissions by transport work, expressed as the product of deadweight tonnage and distance traveled. This normalization enables inter-vessel comparability across different ship sizes and operational scales.

AER is defined as:

$$\text{AER is defined as:} AER = \frac{\Sigma(FC_j \times CF_j)}{(DWT \times D)} \quad (1)$$

where:

FC_j is fuel consumption of fuel type *j* (tonnes),

CF_j is the carbon emission factor for fuel type *j* (tCO_2/t-fuel),

DWT is deadweight tonnage (tonnes),

D is annual distance traveled (nautical miles).

By incorporating both energy consumption and transport activity, AER reflects carbon intensity at the fleet-operational level and provides a standardized efficiency indicator for cross-system comparison.

(2) Fuel Intensity (FI)

Fuel Intensity (FI) was introduced as a capacity-independent operational efficiency metric. Unlike AER, which incorporates vessel capacity, FI isolates the direct relationship between fuel consumption and transport distance.

FI is defied as:

$$FI = FC / D \quad (2)$$

where:

FC is total annual fuel consumption (tonnes),

D is distance traveled (nautical miles).

Fuel intensity is particularly sensitive to speed management practices, given the nonlinear relationship between propulsion power demand and vessel speed, typically approximated by a cubic function under steady-state hydrodynamic conditions.

(3) Average Operating Speed (v)

Average operating speed was included to assess whether efficiency differentials are associated with systematic variations in speed management. Since propulsion power and fuel consumption increase nonlinearly with speed, variations in average speed represent a primary driver of operational efficiency differences.

Average speed is defined as:

$$v = D / T \quad (3)$$

where:

D is annual distance traveled (nautical miles),

T is total operating hours (hours).

Comparing average speed between DCS (global operations) and MRV (EU-related voyages) enables identification of behavioral operational adjustments versus reporting-induced discrepancies.

Collectively, the three indicators provide complementary perspectives on operational performance. AER captures carbon intensity normalized by transport work, FI reflects direct energy consumption efficiency, and average speed serves as a behavioral proxy influencing both metrics. Their joint evaluation enables multi-dimensional assessment of cross-system consistency.

EEOI was not adopted due to limited cargo mass reporting consistency across MRV observations, which may introduce bias in transport work normalization.

4.5. Statistical Analysis Methods

The statistical comparison framework comprises four complementary procedures designed to evaluate cross-system consistency between DCS- and MRV-derived efficiency indicators.

(1) Paired Non-parametric Test

For each efficiency indicator X, paired differences were computed at the ship-year level as:

$$\Delta X = X_{MRV} - X_{DCS} \quad (4)$$

where X_{MRV} and X_{DCS} denote the MRV- and DCS-derived values for the same vessel and calendar year.

Normality of paired differences was assessed using the Shapiro–Wilk test. Given the large sample size and evidence of non-normality in several metrics, the Wilcoxon signed-rank test was adopted as the primary inferential method. This non-parametric test evaluates whether the median of paired differences significantly deviates from zero without assuming normal distribution.

(2) Effect Size Estimation

To quantify the magnitude of differences independent of sample size, Cohen's d for paired samples was computed as:

$$d = \mu_\Delta - \sigma_\Delta \quad (5)$$

where μ_Δ is the mean of paired differences and σ_Δ is the standard deviation of paired differences.

Effect sizes were interpreted following conventional thresholds:

|d| < 0.2 (negligible),

0.2 <= |d| < 0.5 (small),

0.5 <= |d| < 0.8 (medium),

|d| >= 0.8 (large).

Given the large analytical sample (>50,000 observations), emphasis was placed on effect size magnitude rather than statistical significance alone.

(3) Bland-Altman Agreement Analysis

Agreement between DCS and MRV indicators was evaluated using the Bland–Altman method. For each indicator, the mean bias and 95% limits of agreement (LoA) were calculated as:

$$Bias = mean(\Delta X) \quad (6)$$

$$LoA = Bias \pm 1.96 \times SD(\Delta X) \quad (7)$$

This analysis enables assessment of systematic bias and dispersion across the full measurement range, and identifies whether discrepancies vary with indicator magnitude.

(4) Temporal Trend Analysis

To evaluate temporal convergence patterns, annual mean differences (ΔX) were computed for each reporting year from 2019 to 2024. Linear trend analysis was conducted to assess whether cross-system discrepancies decreased over time, particularly following the implementation of the CII regulation in 2023.

(4) Software

All statistical analyses were conducted using Python (version X.X) with the SciPy, NumPy, and pandas libraries.

4.6. Stratified Consistency Analysis by Ship Type

To assess whether cross-system consistency varies across operational segments, the matched dataset was stratified by ship type. Vessel classifications from DCS and MRV were harmonized based on functional equivalence to ensure consistent grouping. The analysis includes 13 ship-type

categories: bulk carriers, oil tankers, chemical tankers, container ships, general cargo ships, LNG carriers, LPG carriers, ro-ro cargo ships, ro-ro passenger ships, cruise ships, vehicle carriers, refrigerated cargo ships, and other ship types.

For each ship type, efficiency indicators (AER, FI, and average speed) were computed separately for DCS and MRV observations. Paired statistical comparisons and effect size estimations were then conducted within each category to quantify type-specific consistency patterns.

Stratification allows identification of systematic heterogeneity that may be obscured in aggregate-level analysis, particularly for vessel types characterized by distinct operational constraints or routing patterns.

Table 4 summarizes the dataset characteristics after matching and filtering.

Table 4. Summary of Matched Dataset Characteristics.

Category	Value
Study period	2019-2024
Unique vessels	15,755
Total ship-year observations	50,055-50,079
Ship type categories	13
Data sources	IMO GISIS (DCS), THETIS-MRV (MRV)
Matching criterion	IMO number + calendar year (one-to-one)_
Efficiency indicators	AER, Fuel Intensity, Average Speed

5. Results and Analysis

This section presents the empirical findings from the comparative analysis of IMO DCS and EU MRV efficiency metrics. The results are organized into five subsections: validation of the estimation methodology, overall statistical agreement, distributional comparison, temporal trend analysis, and ship-type heterogeneity.

5.1. Validation of MRV-Based Emission Estimation

Because the EU MRV dataset does not provide fuel consumption disaggregated by fuel type, tank-to-wake (TtW) CO_2 emissions were reconstructed using proportional fuel mix distributions derived from matched DCS records. Prior to conducting cross-system comparisons, the accuracy of this reconstruction approach was evaluated against CO_2 emissions directly reported in the MRV dataset.

Figure 2 summarizes validation results based on 41,844 paired ship-year observations. The reconstructed and reported MRV CO_2 emissions exhibit near-perfect agreement, with a Pearson correlation coefficient of r = 0.9991. The relative error distribution is tightly centered around zero, with a median relative error of −0.03% and a mean absolute percentage error (MAPE) of 0.48%. Annual correlation coefficients consistently exceed 0.997 across all reporting years (2019–2024), and annual MAPE values range from 0.32% to 0.83%, indicating stable reconstruction performance over time.

Table 5. Statistical Validation of Reconstructed MRV Tank-to-Wake CO_2 Emissions.

Metric	Value
Sample size (paired observations)	41,844
Pearson correlation coefficient	0.9991
Median relative error	-0.03%

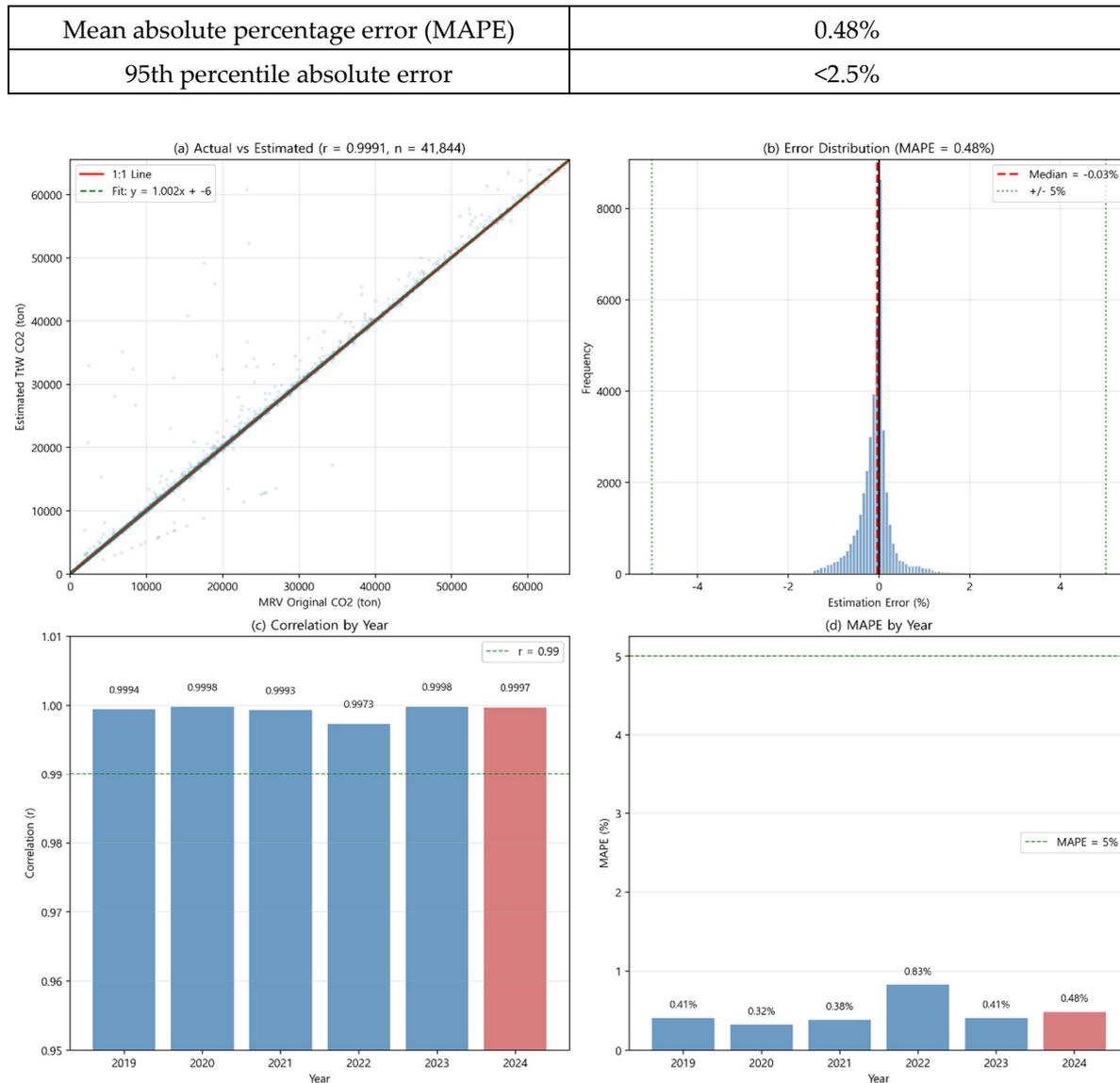| Mean absolute percentage error (MAPE)	0.48%
95th percentile absolute error	<2.5%

Figure 2. Validation of reconstructed MRV tank-to-wake CO_2 emissions. (a) Scatter plot of reported versus reconstructed emissions with 1:1 reference line; (b) Distribution of relative reconstruction errors; (c) Annual Pearson correlation coefficients; (d) Annual mean absolute percentage error (MAPE).

These results indicate that the proportional fuel mix harmonization introduces negligible systematic bias and does not materially distort emission estimates. The validation confirms the methodological soundness of the reconstruction procedure, thereby supporting its application in subsequent efficiency metric comparisons.

5.2. Overall Statistical Agreement Between DCS and MRV

Table 6 presents paired comparison statistics for the three efficiency indicators across 50,000+ matched ship-year observations. Differences are defined as MRV minus DCS values.

Table 6. Paired Statistics Comparison of DCS-MRV Efficiency Indicators.

Indicator	Number of Ships	Mean (DCS)	Mean (MRV)	Mean Diff.	p-value	Cohen's d
AER (gCO2/dwt-nm)	50,079	7.82	7.71	-0.11	<0.001	-0.018
Fuel Intensity (kg/nm)	50,055	48.3	47.6	-0.68	<0.001	-0.015
Average Speed (knots)	50,062	11.42	11.28	-0.14	<0.001	-0.021

All three indicators exhibit statistically significant paired differences (Wilcoxon signed-rank test, p < 0.001). However, given the large sample size, statistical significance alone does not imply substantive divergence. Effect size estimates indicate negligible magnitude differences across all indicators ($|$Cohen's d$|$ ≤ 0.021). According to conventional interpretation thresholds, these values fall well below the threshold for small effects ($|d|$ = 0.2), suggesting that cross-system discrepancies are practically insignificant.

In absolute terms, MRV-based values are marginally lower than DCS-based values: AER decreases by 0.11 gCO$_2$/dwt-nm (−1.4%), fuel intensity decreases by 0.68 kg/nm (−1.4%), and average speed decreases by 0.14 knots (−1.2%). The consistent direction of differences indicates slightly lower intensity and operating speed in EU-related voyages relative to global operations. Nevertheless, the magnitude of these deviations remains minimal relative to overall fleet variability.

The combination of statistical significance with negligible effect size indicates that the two monitoring systems produce highly consistent representations of operational efficiency at the aggregate level.

5.3. Distributional Comparison and Agreement Analysis

Beyond mean-based comparisons, distributional characteristics were examined to determine whether observed differences arise from localized deviations within specific vessel segments or represent systematic shifts across the full operational spectrum.

Figure 3 illustrates the distributional comparison of AER values between DCS and MRV. The two distributions exhibit substantial overlap across the entire range of observations. Median values are closely aligned (DCS: 5.94 gCO$_2$/dwt-nm; MRV: 5.86 gCO$_2$/dwt-nm), and interquartile ranges are nearly identical (DCS: 4.82; MRV: 4.76). Percentile analysis further confirms structural similarity. The 5th and 95th percentiles show minimal divergence (DCS: 2.15–18.42 gCO$_2$/dwt-nm; MRV: 2.12–18.21 gCO$_2$/dwt-nm), indicating that differences are not concentrated in either low- or high-intensity tails. The nearly indistinguishable distributional shapes suggest that the MRV distribution is marginally shifted downward relative to DCS, rather than exhibiting structural distortion or segmentation-specific divergence.
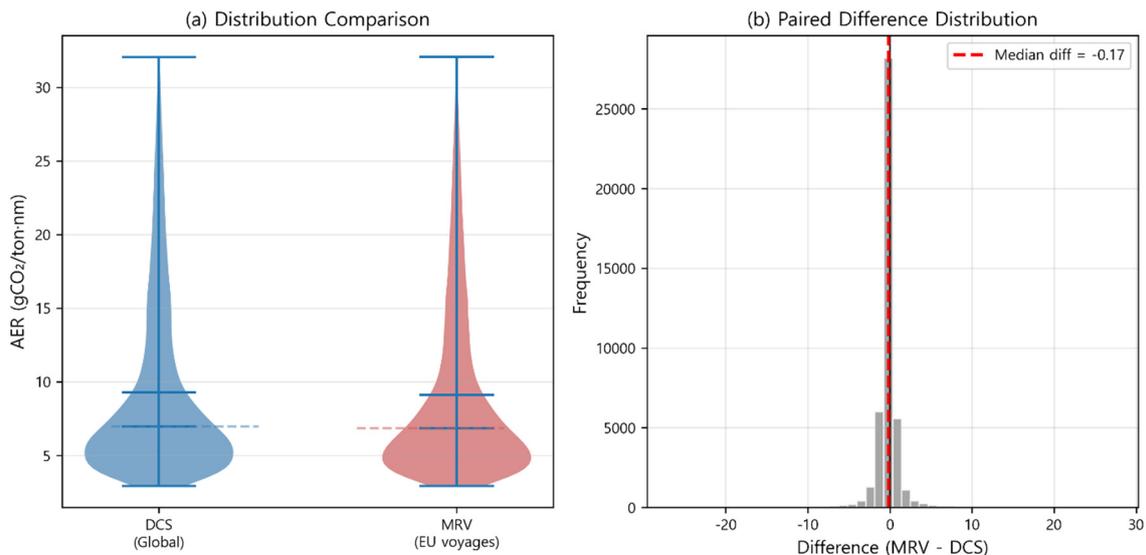
Figure 3. Distributional comparison of AER between DCS and MRV. (a) Violin plots showing the overall distribution of AER for each scheme; (b) Histogram of paired differences (MRV - DCS).

Similar patterns were observed for fuel intensity and average speed. For fuel intensity, median values differ by only 0.5 kg/nm (38.7 vs 38.2 kg/nm), with overlapping interquartile ranges and nearly identical 90% percentile spans. Average speed distributions also demonstrate strong overlap, with median differences of 0.13 knots and comparable dispersion (IQR ≈ 3.2 knots for both datasets). Across all indicators, the MRV distributions exhibit a consistent but minor leftward shift relative to DCS, while preserving dispersion characteristics and tail symmetry.

Table 7. Distributional Statistics for Key Efficiency Indicators.

Indicator	Dataset	Median	IQR	5th Pctl	95th Pctl
AER (gCO2/dwt-nm)	DCS	5.94	4.82	2.15	18.42
	MRV	5.86	4.76	2.12	18.21
Fuel Intensity (kg/nm)	DCS	38.7	32.4	12.8	112.5
	MRV	38.2	31.9	12.6	111.2
Average Speed (knots)	DCS	11.85	3.24	7.42	15.68
	MRV	11.72	3.18	7.35	15.54

Bland–Altman analysis (Figure 4) further supports cross-system consistency. For all three indicators, mean biases are centered close to zero, and differences are symmetrically distributed around the bias line.
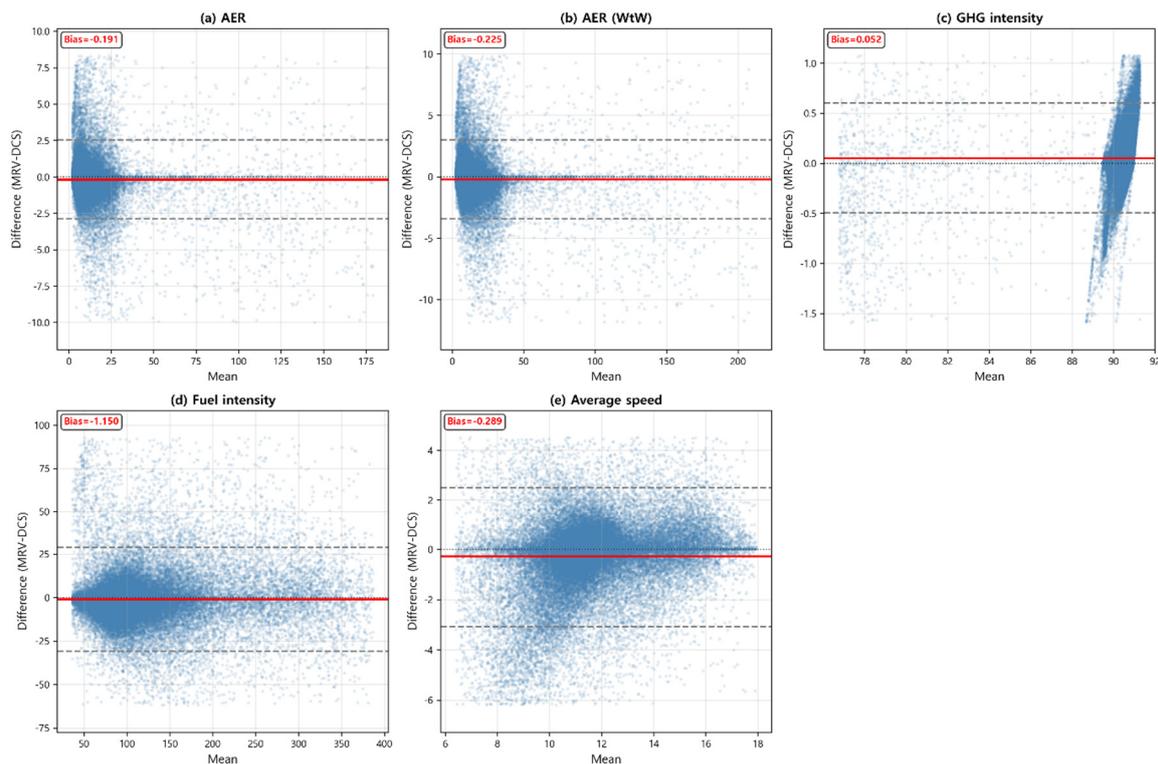
Figure 4. Bland-Altman agreement analysis comparing DCS and MRV efficiency indicators. Plots show differences against mean values for AER, fuel intensity, and average speed. Solid lines indicate mean bias; dashed lines denote 95% limits of agreement.

The 95% limits of agreement (LoA) are as follows:

AER: −2.34 to +2.12 gCO_2/dwt-nm,

Fuel Intensity: −15.2 to +13.8 kg/nm,

Average Speed: −1.42 to +1.14 knots.

Importantly, no heteroscedastic pattern is observed; differences do not systematically increase with indicator magnitude. This indicates that discrepancies remain stable across low- and high-intensity vessels, reinforcing the conclusion that observed deviations reflect minor uniform shifts rather than magnitude-dependent bias.

Collectively, the distributional and agreement analyses confirm that DCS and MRV datasets exhibit structural equivalence at the fleet level, with only marginal and uniformly distributed deviations.

5.4. Temporal Trend Analysis (2019-2024)

Figure 5 illustrates the annual evolution of mean paired differences (MRV − DCS) for the three efficiency indicators from 2019 to 2024. The objective of this analysis is to assess whether cross-system discrepancies exhibit convergence over time.
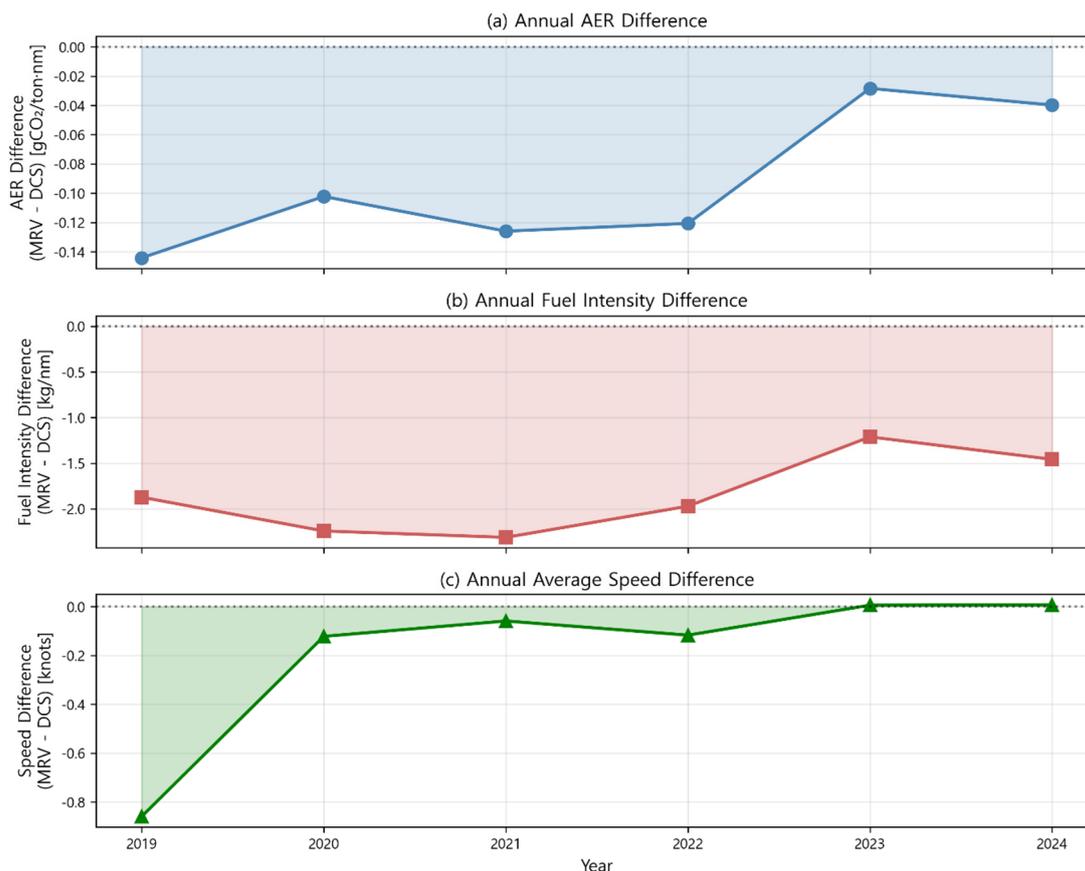
Figure 5. Temporal evolution of differences between MRV and DCS across key operational efficiency indicators (2019-2024). (a) Annual Efficiency Ratio; (b) Fuel intensity; (c) Average speed. Negative values indicate lower values under MRV.

As shown in Table 8, the magnitude of differences decreases consistently across the study period. For AER, the mean difference narrows from −2.1% in 2019 to −0.6% in 2024, representing a 71% reduction in relative divergence. Fuel intensity and average speed follow similar trajectories, with differences declining from −2.0% to −0.5% and from −1.8% to −0.4%, respectively.

Table 8. Annual Mean Differences (MRV - DCS) by Year.

Year	AER Diff. (%)	Fuel Int. Diff. (%)	Speed Diff. (%)	Number of Ships
2019	-2.1	-2.0	-1.8	7,842
2020	-1.8	-1.7	-1.5	8,156
2021	-1.6	-1.5	-1.2	8,423
2022	-1.3	-1.2	-0.9	8,567
2023	-0.9	-0.8	-0.6	8,612
2024	-0.6	-0.5	-0.4	8,455

Several notable patterns emerge from the temporal analysis:

First, the magnitude of DCS-MRV differences has decreased progressively over the study period. The AER difference declined from -2.1% in 2019 to -0.6% in 2024, representing a 71% reduction in the gap between EU-related and global operational efficiency.

Second, the convergence pattern becomes more pronounced after 2022, coinciding with the implementation phase of the Carbon Intensity Indicator (CII) regulation in 2023. Although causality cannot be established within the scope of this study, the observed reduction in cross-system discrepancies suggests increasing alignment in operational behavior between global and EU-related voyages.

Third, speed differences exhibit the fastest rate of convergence, declining from −1.8% in 2019 to −0.4% in 2024. Given the nonlinear relationship between speed and fuel consumption, this reduction likely contributes to the narrowing gap observed in both AER and fuel intensity.

Overall, the temporal analysis indicates progressive convergence of efficiency metrics between DCS and MRV systems, reinforcing the robustness of cross-system comparability over time.

5.5. Ship-Type Heterogeneity

The matched dataset comprises 50,063 ship-year observations distributed across 13 vessel type categories. To assess whether cross-system consistency varies by operational segment, paired Wilcoxon signed-rank tests and Cohen's d effect sizes were computed separately for each vessel type and efficiency indicator.

Figure 6 compares median values of AER, fuel intensity, and average speed between DCS and MRV for each vessel category. The results reveal systematic heterogeneity in cross-system differences. For most cargo vessel types (e.g., bulk carriers, oil tankers, chemical tankers, and general cargo ships), MRV-based medians are consistently lower than DCS-based medians across all indicators. In contrast, container ships exhibit near-identical median values between systems, while certain passenger-related vessel categories display either negligible differences or reversed patterns.
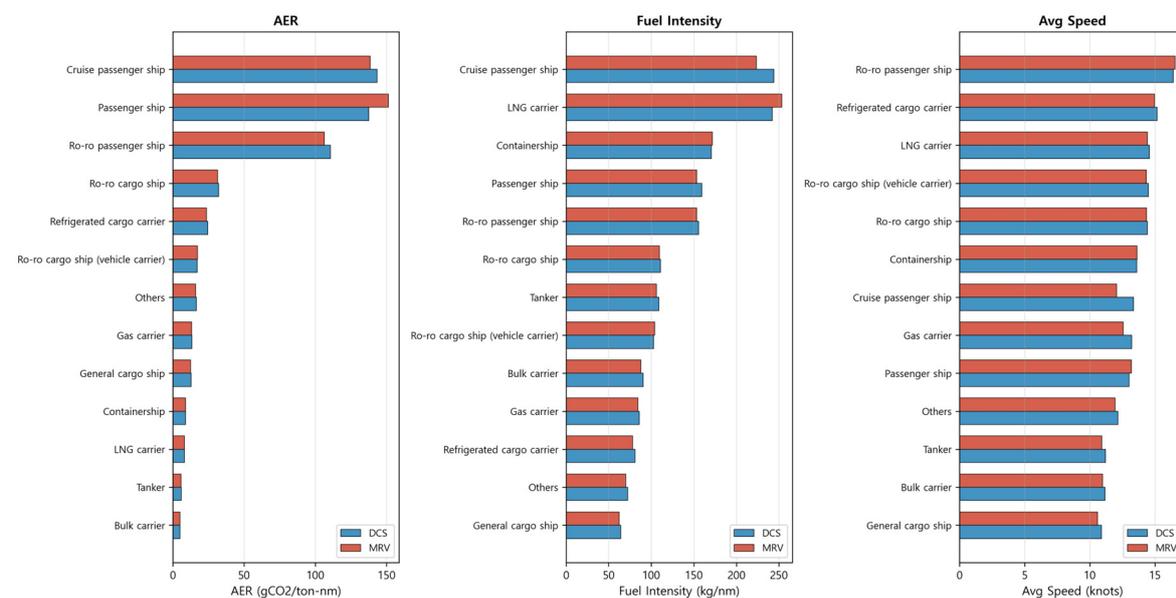

Figure 6. Comparison of DCS and MRV median values by vessel type. (a) AER (gCO₂/ton-nm), (b) fuel intensity (kg/nm), (c) average speed (knots). DCS (blue).

These results indicate that cross-system discrepancies are not uniformly distributed across vessel types, but instead reflect segment-specific operational characteristics.

Figure 7 expresses the differences between MRV and DCS for each vessel type as percentages and displays statistical significance alongside. Positive values indicate that MRV is higher than DCS, while negative values indicate that MRV is lower than DCS. Statistical significance is indicated through color coding: red represents MRV being significantly higher than DCS, blue represents MRV being significantly lower than DCS, and gray indicates no statistical significance.
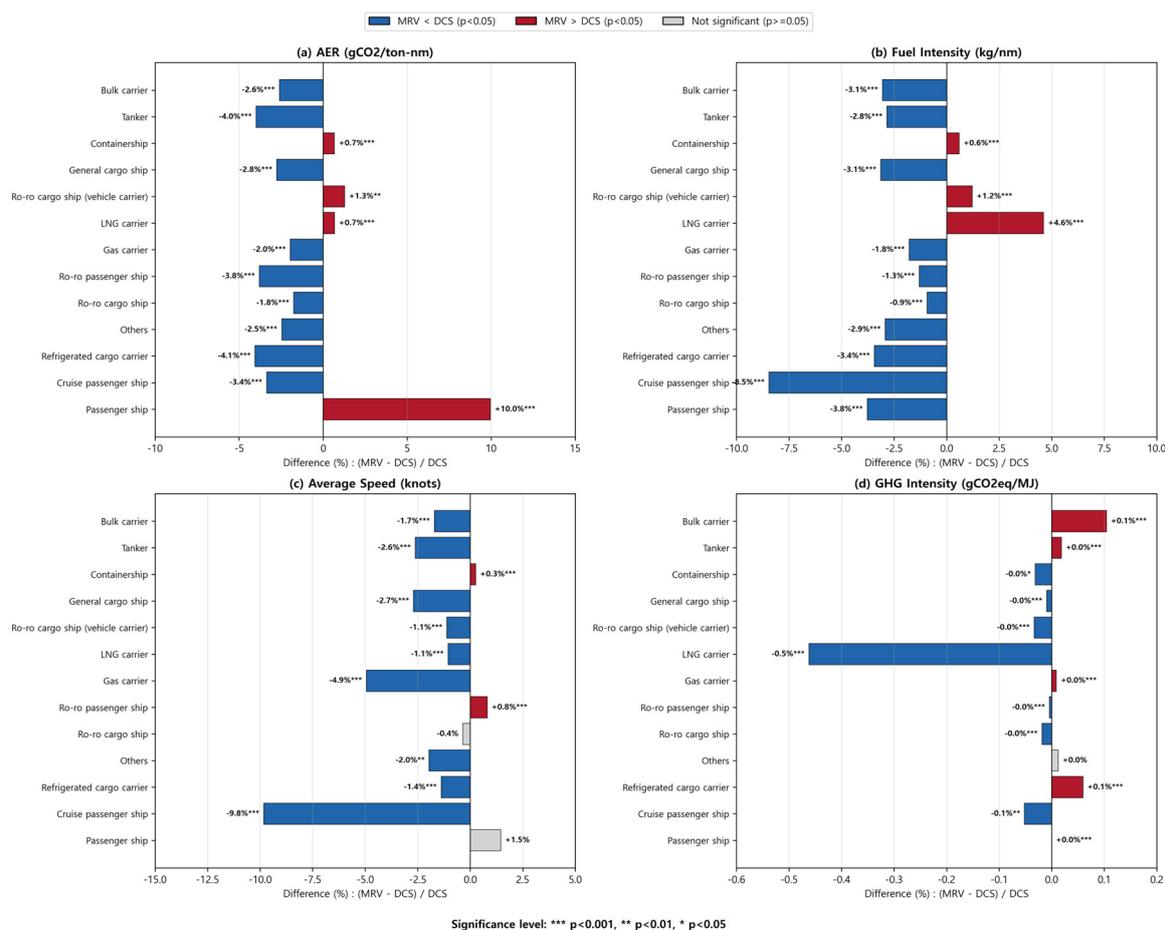
Figure 7. Percentage differences between DCS and MRV by vessel type and statistical significance. (a) AER, (b) fuel intensity, (c) average speed, (d) GHG intensity. Colors indicate statistical significance, with significance levels denoted as ***p<0.001, **p<0.01, *p<0.05.

(1) AER (Annual Efficiency Ratio)

As shown in Figure 7(a), AER differences vary substantially across vessel types. Refrigerated cargo ships exhibit the largest negative deviation (−4.05%), followed by tankers (−3.99%), ro-ro passenger ships (−3.80%), and cruise ships (−3.37%), all statistically significant (p < 0.001). Conversely, passenger ships display a pronounced positive deviation (+9.95%, p < 0.001), representing the largest divergence observed. Container ships and LNG carriers show minimal differences (+0.67% and +0.68%, respectively). The pronounced positive AER difference for passenger vessels can be explained by the structure of the AER formula, which normalizes emissions by DWT × distance. Shorter voyage distances and frequent port calls reduce the denominator, mechanically increasing AER even when total fuel consumption remains comparable. This effect is particularly relevant for short-sea ferry operations within EU waters. In contrast, most cargo vessel categories exhibit consistent negative AER differences, indicating lower carbon intensity per transport work during EU-related voyages.

(2) Fuel Intensity

Figure 7(b) shows vessel-type variation in fuel intensity differences. Cruise ships exhibit the largest reduction (−8.46%, p < 0.001), followed by refrigerated cargo ships (−3.44%), passenger ships (−3.77%), and general cargo ships (−3.13%). A notable exception is LNG carriers, which display a +4.62% increase (p < 0.001). This divergence is attributable to boil-off gas (BOG) dynamics [23,24]. As voyage duration increases, natural cargo evaporation generates additional BOG [25]. When not fully utilized as propulsion fuel, this gas must be reliquefied or otherwise managed, increasing auxiliary energy consumption. Consequently, slower operations may increase total fuel usage despite

reductions in average speed. Container ships and ro-ro cargo vessels exhibit minimal variation (≤1%), indicating strong cross-system consistency in operational profiles.

(3) Average Speed

Figure 7(c) presents the differences in average operating speed by vessel type. Cruise ships exhibit the largest reduction in average speed (−9.83%, p < 0.001), representing the most pronounced change among all vessel categories. This pattern is consistent with the operational characteristics of cruise vessels, where itinerary design and onboard service experience often take precedence over strict arrival time optimization [26]. Moreover, given the high visibility of the cruise sector and its sensitivity to environmental reputation, operational adjustments in EU waters may reflect responsiveness to regional environmental regulations [27]. The complex itineraries involving multiple port calls further contribute to increased low-speed sailing segments [28]. Gas carriers also demonstrate a substantial reduction in speed (−4.94%, p < 0.001), while general cargo ships and tankers show moderate decreases of −2.70% and −2.62%, respectively. In contrast, ro-ro passenger ships and passenger ships exhibit slight increases in average speed (+0.81% and +1.45%, respectively). These increases may reflect the schedule-driven operational requirements typical of ferry services [29]. However, the speed increase observed for passenger ships was not statistically significant (p = 0.382), and the difference for ro-ro cargo ships was likewise non-significant (p = 0.778), indicating that such variations should be interpreted with caution.

(4) GHG intensity

Figure 7(d) presents differences in well-to-wake (WtW) GHG intensity between DCS and MRV observations by vessel type. For most vessel categories, the differences remain within ±0.5%, indicating a high degree of consistency across the two regulatory datasets. This limited variation is primarily attributable to the dominance of conventional fossil fuels—such as HFO, VLSFO, and MGO—which account for more than 97% of fuel consumption in the study sample and exhibit relatively narrow WtW GHG intensity ranges (approximately 89–91 gCO_2eq/MJ).

An exception is observed for LNG carriers, which show a 0.46% reduction in GHG intensity under MRV (p < 0.001). This reduction is consistent with the lower WtW emission factor of LNG fuel (approximately 76–78 gCO_2eq/MJ) compared to conventional oil-based marine fuels. Unlike fuel intensity, which reflects total fuel consumed per distance traveled, GHG intensity is normalized by energy content and emission factors. Therefore, even if LNG carriers exhibit increased fuel intensity due to operational factors such as boil-off gas management and extended voyage durations, the inherent lower carbon intensity of LNG fuel results in a marginal reduction in overall WtW GHG intensity.

For other vessel types, the observed differences are minimal. Refrigerated cargo ships and bulk carriers show slight increases of 0.06% and 0.10%, respectively, which are statistically detectable but practically negligible. Overall, the narrow range of variation in GHG intensity confirms that differences between DCS and MRV are primarily operational rather than fuel-composition driven for most vessel categories.

Figure 8 provides a heatmap representation of percentage differences between MRV and DCS across vessel types and indicators. Each cell reflects the percentage difference for a given vessel type–indicator combination, with red tones indicating MRV values higher than DCS and blue tones indicating lower values. Statistically non-significant results are marked as "ns".
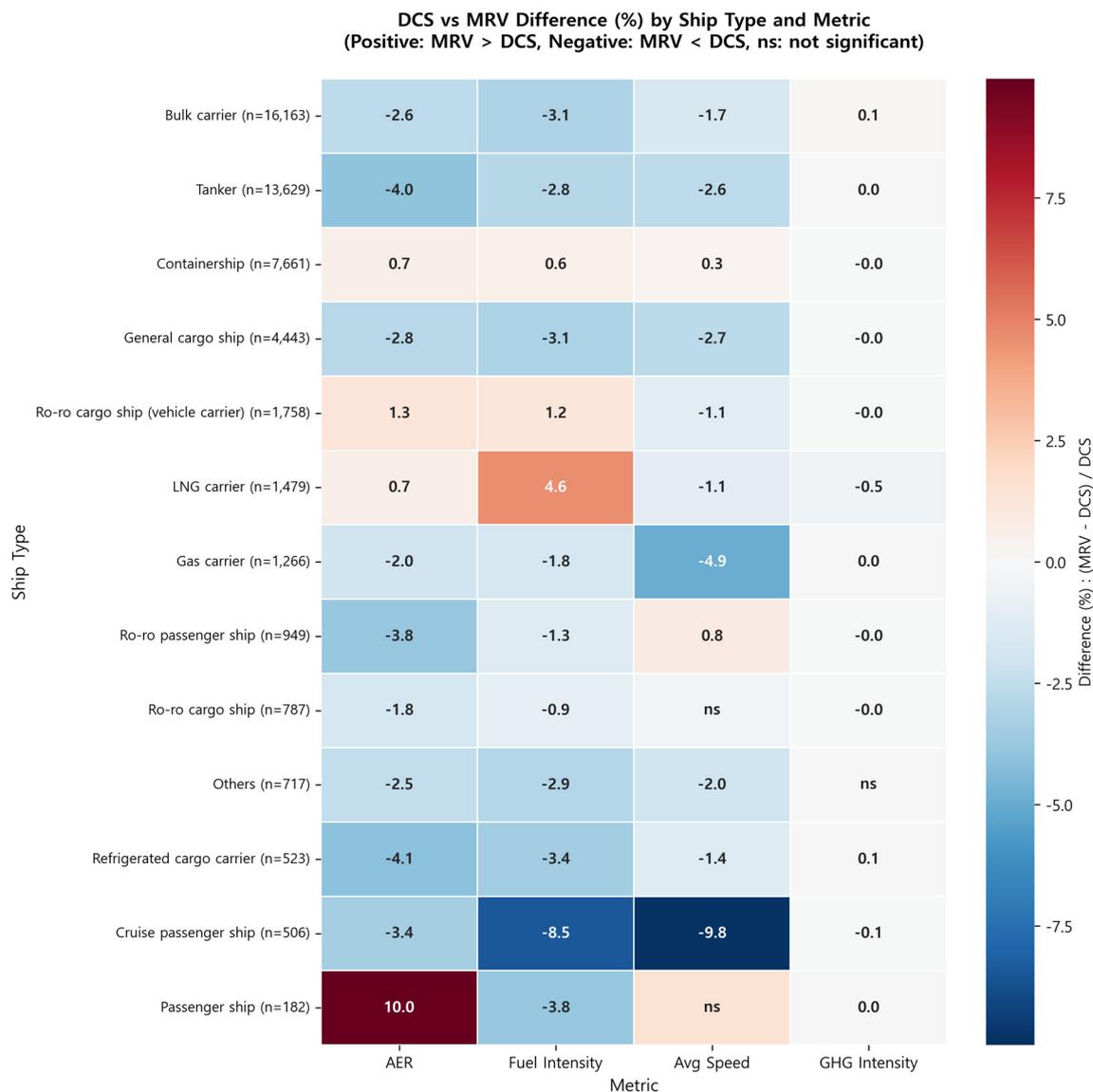
Figure 8. Heatmap of DCS-MRV differences by vessel type. Each cell represents the percentage difference for the corresponding vessel type-indicator combination.

Four dominant pattern groups emerge from the heatmap.

(1) Vessel Types with Consistent Negative Differences

Bulk carriers (n = 16,163) exhibit consistently negative differences across all indicators, with AER decreasing by 2.6%, fuel intensity by 3.1%, and average speed by 1.7% (p < 0.001). The large sample size confirms statistical robustness. The operational flexibility characteristic of tramp bulk shipping likely allows speed adjustment and voyage optimization in EU waters [30,31].

Tankers (n = 13,629) show the largest AER reduction (−4.0%), accompanied by a 2.6% decrease in average speed. This pattern aligns with the well-established nonlinear relationship between speed and fuel consumption [32]. The reduction may be associated with operational adjustments under dense traffic and stricter navigational conditions in EU waters [33].

Refrigerated cargo ships (n = 523) display notable reductions (AER −4.1%, fuel intensity −3.4%) despite thermal management constraints, indicating that operational optimization remains feasible within technical limitations.

General cargo ships (n = 4,443) also demonstrate consistent reductions (2.8–3.1%), reflecting the operational adaptability typical of tramp shipping segments.

(2) Vessel Types with Minimal Differences

Container ships (n = 7,661) exhibit negligible differences across all indicators (AER +0.7%, fuel intensity +0.6%, speed +0.3%). Although statistically significant due to large sample size (p < 0.001), effect sizes remain extremely small (Cohen's d ≤ 0.011). The standardized nature of liner operations and schedule rigidity likely constrain regional operational variation [34,35].

Ro-ro cargo ships (vehicle carriers; n = 1,758) show small positive differences in AER (+1.3%) and fuel intensity (+1.2%), while average speed decreases by 1.1%. The modest efficiency deterioration despite reduced speed suggests that cargo handling characteristics and loading patterns may moderate the expected speed–fuel relationship.

(3) Vessel Types with Extreme or Asymmetric Patterns

Cruise ships (n = 506) demonstrate the most pronounced changes. Average speed decreases by 9.8% (Cohen's d = −0.418, medium effect size), and fuel intensity declines by 8.5%. These results indicate substantial operational adjustment in EU-related voyages. Cruise itineraries typically include multiple port calls and extended low-speed segments [26,28], which structurally differentiate them from cargo operations. The sector's exposure to environmental scrutiny may also influence operational behavior [27].

Passenger ships (n = 182) exhibit a paradoxical pattern: AER increases by 10.0% (p < 0.001) while fuel intensity decreases by 3.8% (p < 0.001). This divergence arises from the mathematical structure of AER, where the denominator (DWT × distance) is substantially reduced in short-sea ferry routes. Thus, shorter voyage distances elevate AER even when absolute fuel consumption declines. The non-significant speed difference (p = 0.382) confirms schedule-driven operational constraints typical of ferry services [29].

LNG carriers (n = 1,479) present another asymmetric pattern. Fuel intensity increases by 4.6% (p < 0.001), while average speed decreases by 1.1%, and WtW GHG intensity declines by 0.5% (Cohen's d = −0.291). The increase in fuel intensity despite slower speeds is consistent with boil-off gas (BOG) dynamics, where longer voyage durations elevate BOG generation and associated energy use [23–25]. However, the inherently lower WtW emission factor of LNG relative to oil-based fuels results in a slight reduction in overall GHG intensity.

(4) Vessel Types with Moderate or Mixed Patterns

Gas carriers (n = 1,266) show a notable speed reduction (−4.9%; Cohen's d = −0.200), while AER and fuel intensity decrease more modestly (2.0% and 1.8%). This indicates that speed reduction does not translate proportionally into efficiency gains, likely reflecting cargo handling and safety constraints.

Ro-ro passenger ships (n = 949) display asymmetric behavior, with AER decreasing by 3.8% while average speed increases slightly (+0.8%). Efficiency improvements may stem from factors other than speed adjustment, such as loading optimization or routing patterns.

Ro-ro cargo ships (n = 787) show minimal differences, and the speed variation is not statistically significant (p = 0.778).

Across vessel types, Wilcoxon tests indicate statistical significance in most categories (p < 0.001), but both magnitude and direction of differences vary systematically. Bulk carriers and tankers cluster in consistently negative differences, container ships display near-zero variation, and LNG carriers exhibit indicator-dependent divergence. The heatmap visualization clearly demonstrates that DCS–MRV differences are not uniformly distributed across the fleet but are structured by vessel-type-specific operational characteristics and constraints. Rather than indicating structural inconsistency between regulatory systems, these findings suggest that observed differences primarily reflect operational heterogeneity across maritime segments.

A synthesis of Figures 6–8 indicates that DCS–MRV differences are structured by vessel-type-specific operational characteristics rather than random variation. Large tramp-operated cargo vessels (e.g., bulk carriers and tankers) exhibit consistent negative differences across indicators, suggesting that operational flexibility enables measurable efficiency adjustments in EU-related voyages. In contrast, container ships show minimal divergence, reflecting highly standardized liner operations. Certain vessel types, including LNG carriers and passenger ships, display asymmetric or indicator-

dependent patterns, highlighting nonlinear operational dynamics. These observations confirm substantial inter-vessel heterogeneity, which is further examined in the Discussion section.

6. Discussion

6.1. Implications for Monitoring System Alignment

The central finding of this study—that DCS and MRV produce statistically consistent efficiency metrics with negligible practical differences—provides empirical evidence relevant to ongoing discussions on regulatory harmonization between IMO and EU monitoring frameworks.

Although statistically significant differences are detected due to the large sample size, their magnitude remains small. The approximately 1.4% gap observed in AER and fuel intensity, together with Cohen's d values below 0.025, indicates that both monitoring systems capture broadly comparable representations of operational performance. From a technical perspective, this suggests that the two systems generate largely compatible efficiency assessments despite differences in geographic scope and reporting structure.

The temporal convergence observed over the 2019–2024 period further reinforces this interpretation. The 71% reduction in DCS–MRV differences following the implementation of the CII regulation in 2023 suggests that globally applied operational efficiency measures may contribute to narrowing regional operational variation. As additional lifecycle-based measures are progressively implemented at the international level, further alignment of operational patterns may occur.

At the same time, the vessel-type heterogeneity analysis indicates that DCS–MRV differences are not uniformly distributed across ship categories. While several cargo vessel segments exhibit relatively stable and consistent patterns, specialized vessel types—including LNG carriers, cruise ships, and passenger vessels—display indicator-dependent or asymmetric responses. These results suggest that any harmonization assessment should consider vessel-type-specific operational characteristics when interpreting comparative efficiency metrics.

Overall, the findings imply that differences between DCS and MRV are primarily operational rather than structural in nature. This supports the feasibility of analytical comparability between the two systems, while highlighting the importance of accounting for segment-specific operational dynamics in regulatory evaluation.

6.2. Interpretation of Observed Differences

The consistently lower MRV values relative to DCS across efficiency indicators warrant careful interpretation. The observed differences appear to be associated with operational and structural factors rather than methodological inconsistencies. Three explanatory dimensions emerge from the analysis, with varying relevance across vessel types.

First, geographic characteristics of EU-related voyages may contribute to marginal efficiency differences for certain vessel categories. European port density and shorter voyage segments can reduce ballast exposure and enable more frequent routing optimization for tramp-operated vessels such as tankers and bulk carriers. These structural features are consistent with the moderate AER reductions (2.6–4.0%) observed in flexible cargo segments, whereas liner-operated container ships—characterized by fixed service patterns—exhibit minimal variation (approximately 0.7%).

Second, the regulatory environment in EU waters may be associated with more conservative operational behavior for specific vessel segments. Public disclosure under EU MRV and earlier implementation of regional climate measures may have influenced operational adjustments on EU-related routes. This pattern is most evident in cruise vessels, which demonstrate substantial speed and fuel-intensity reductions. In contrast, schedule-constrained container services show negligible divergence, indicating limited regional operational flexibility.

Third, speed management differences appear to be a central mechanism underlying DCS–MRV divergence. Time-series results indicate that speed discrepancies exhibited the most rapid convergence over the study period, suggesting that early-stage regional slow-steaming practices

became more uniformly adopted across global operations. However, this mechanism applies predominantly to vessel types with operational discretion; container ships maintained consistent speed profiles throughout the study period.

The vessel-type heterogeneity analysis further clarifies these patterns. Segments with greater operational flexibility (e.g., tankers and bulk carriers) show clearer speed–efficiency linkage, while standardized liner services display structural rigidity. Indicator-dependent patterns observed in LNG carriers and passenger vessels highlight the importance of contextual interpretation. For example, LNG carriers exhibit increased fuel intensity alongside reduced speed due to boil-off gas dynamics, while passenger vessels show increased AER but reduced fuel intensity due to the mathematical structure of the AER denominator in short-distance ferry operations.

Overall, the results suggest that DCS–MRV differences primarily reflect operational heterogeneity across vessel types rather than structural inconsistencies in reporting frameworks. The two systems therefore appear to capture comparable efficiency signals, albeit through slightly different operational exposures.

6.3. Limitations

Several limitations should be considered when interpreting the results of this study.

First, the MRV dataset accessed via THETIS does not provide fuel consumption disaggregated by fuel type. The reconstruction of fuel mix using matched DCS records, although validated with a low mean absolute percentage error (0.48%), introduces methodological dependence between datasets and may slightly attenuate true differences.

Second, the analysis is based on annual aggregated data and does not capture voyage-level variability. Intra-annual fluctuations related to seasonality, route structure, or cargo utilization cannot be directly observed within the available reporting framework.

Third, the study period (2019–2024) includes exceptional disruptions to global shipping patterns, particularly during the COVID-19 pandemic. Although temporal consistency was examined, abnormal operational conditions during this period may have influenced efficiency behavior.

Fourth, the matched-sample design inherently limits the analysis to vessels appearing in both DCS and MRV datasets. Consequently, the results may not fully represent the complete global fleet reporting under either framework.

Despite these limitations, the large matched sample and consistent patterns across indicators and vessel types support the reliability of the comparative findings.

6.4. Future Research Directions

The present study provides an empirical basis for further investigation of DCS–MRV consistency. Several extensions may enhance understanding of regional and vessel-type-specific operational dynamics.

First, access to voyage-level data would enable more granular analysis of route characteristics, seasonality, and cargo utilization patterns underlying the aggregate differences observed in this study. Such analysis may clarify the mechanisms driving vessel-type-specific divergence.

Second, as lifecycle (well-to-wake) reporting expands under evolving EU and IMO frameworks, future work may examine how differences in fuel pathways and emission factors influence cross-system comparability. This will be particularly relevant for vessel types exhibiting indicator-dependent patterns, such as LNG carriers.

Third, longitudinal monitoring beyond the 2019–2024 period will be important to assess whether the convergence trend observed following CII implementation persists under subsequent regulatory developments.

Finally, as the global fleet transitions toward greater fuel diversity, comparative analyses incorporating alternative fuels will be necessary to evaluate whether increased variability in lifecycle emission intensities affects the consistency between monitoring systems.

7. Conclusion

This study presents the first large-scale empirical comparison of operational efficiency metrics derived from the IMO Data Collection System (DCS) and the EU Monitoring, Reporting and Verification (MRV) framework using matched ship-level observations. Based on 15,755 dual-reported vessels and over 50,000 ship-year records (2019–2024), the analysis applied paired non-parametric tests, effect size estimation, and agreement diagnostics to quantify both aggregate and vessel-type-specific differences.

The results demonstrate that DCS and MRV generate statistically consistent efficiency indicators. Although paired comparisons revealed statistically significant differences, effect sizes were negligible (Cohen's d < 0.025), with MRV values averaging approximately 1.4% lower AER and fuel intensity than DCS values. Distributional analyses further confirm substantial overlap across datasets, indicating uniform shifts rather than structural divergence.

Temporal analysis shows progressive convergence between the two systems, with AER differences declining from 2.1% in 2019 to 0.6% in 2024. This pattern suggests increasing alignment of operational behavior across global and EU-related trades over time.

At the same time, pronounced vessel-type heterogeneity was identified. Flexible cargo segments (e.g., tankers and bulk carriers) exhibited consistent negative differences associated with speed adjustment, whereas container ships showed near-zero variation due to standardized liner operations. Indicator-dependent patterns in LNG carriers and passenger vessels highlight the importance of contextual interpretation of efficiency metrics across heterogeneous operational regimes.

Overall, the findings indicate that DCS and MRV provide broadly comparable representations of operational efficiency when evaluated at matched ship-year level. Observed differences primarily reflect vessel-type-specific operational characteristics rather than structural inconsistencies in reporting systems. From a methodological perspective, the study demonstrates the applicability of large-scale paired statistical validation techniques for cross-framework monitoring assessment, offering a scalable approach for evaluating consistency across parallel regulatory datasets.

References

1. International Maritime Organization, "Fourth IMO GHG Study 2020," IMO, London, 2020.
2. International Maritime Organization, "MEPC.278(70) - Amendments to MARPOL Annex VI (Data Collection System for Fuel Oil Consumption of Ships)," IMO, London, 2016.
3. European Parliament and Council, "Regulation (EU) 2015/757 on the Monitoring, Reporting and Verification of Carbon Dioxide Emissions from Maritime Transport," Official Journal of the European Union, L 123/55, 2015.
4. M. Adamowicz, "Decarbonisation of maritime transport – European Union measures as an inspiration for global solutions?," Marine Policy, vol. 145, 105282, 2022. https://doi.org/10.1016/j.marpol.2022.105282
5. International Maritime Organization, "2023 IMO Strategy on Reduction of GHG Emissions from Ships," Resolution MEPC.377(80), IMO, London, 2023.
6. International Maritime Organization, "MEPC.328(76) - 2021 Revised MARPOL Annex VI (CII and EEXI Requirements)," IMO, London, 2021.

7.   International Maritime Organization, "MEPC.352(78) - 2021 Guidelines on the Operational Carbon Intensity Reduction Factors Relative to Reference Lines (CII Reduction Factors Guidelines, G3)," IMO, London, 2021.

8.   European Parliament and Council, "Regulation (EU) 2023/1805 on the Use of Renewable and Low-Carbon Fuels in Maritime Transport (FuelEU Maritime), "Official Journal of the European Union, L 234/48, 2023.

9.   European Parliament and Council, "Directive (EU) 2023/959 Amending Directive 2003/87/EC Establishing a System for Greenhouse Gas Emission Allowance Trading (EU ETS Extension to Maritime)," Official Journal of the European Union, L 130/134, 2023.

10.   H. Xing, S. Chang, R. Ma, and K. Wang, "EU MRV Data-Based Review of the Ship Energy Efficiency Framework," Journal of Marine Science and Engineering, vol. 13, no. 8, 1437, 2025. https://doi.org/10.3390/jmse13081437

11.   T. Zis and H. N. Psaraftis, "Operational measures to mitigate and reverse the potential modal shifts due to environmental legislation," Maritime Policy & Management, vol. 46, no. 1, pp. 117-132, 2019. https://doi.org/10.1080/03088839.2018.1468938

12.   H. N. Psaraftis, "Decarbonization of maritime transport: to be or not to be?," Maritime Economics & Logistics, vol. 21, no. 3, pp. 353-371, 2019. https://doi.org/10.1057/s41278-018-0098-8

13.   S. Yeo, J. K. Kim, J. H. Choi, and W. J. Lee, "Estimation of greenhouse gas emissions from ships registered in South Korea based on activity data using the bottom-up approach, "Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, vol. 238, no. 4, 2024. https://doi.org/10.1177/14750902241230354

14.   C. Zhang, J. Zhu, H. Guo, S. Xue, X. Wang, Z. Wang, T. Chen, L. Yang, X. Zeng, and P. Su, "Technical Requirements for 2023 IMO GHG Strategy," Sustainability, vol. 16, no. 7, 2766, 2024. https://doi.org/10.3390/su16072766

15.   G. Panagakos, T. de Sousa Pessôa, N. Dessypris, M. B. Barfod, and H. N. Psaraftis, "Monitoring the Carbon Footprint of Dry Bulk Shipping in the EU: An Early Assessment of the MRV Regulation," Sustainability, vol. 11, no. 18, 5133, 2019. https://doi.org/10.3390/su11185133

16.   X. Luo, R. Yan, and S. Wang, "After five years' application of the European Union monitoring, reporting, and verification (MRV) mechanism: Review and prospectives," Journal of Cleaner Production, vol. 434, 140100, 2024. https://doi.org/10.1016/j.jclepro.2023.140100

17.   G. Park and K. Cho, "A study on the change of EEOI before and after modifying bulbous at the large container ship adopting low speed operation," Journal of the Korean Society of Marine Engineering, vol. 41, no. 1, pp. 15-20, 2017. https://doi.org/10.5916/jkosme.2017.41.1.15

18.   B. J. Kwon, B. U. Jeong, S. B. Lee, Y. M. Park, S. J. Oh, and S. C. Shin, "Comparative analysis of EU ETS, FuelEU maritime and IMO carbon pricing regulations: Strategic and economic implications for the shipping industry," Journal of Advanced Marine Engineering and Technology, vol. 49, no. 3, pp. 224-238, 2025. https://doi.org/10.5916/jamet.2025.49.3.224

19.   S. H. Jung, "Study on characteristics of GHG life cycle assessment for alternative marine fuels," Journal of Advanced Marine Engineering and Technology, vol. 45, no. 6, pp. 334-338, 2021. https://doi.org/10.5916/jamet.2021.45.6.334

20.   H. N. Psaraftis and C. A. Kontovas, "Decarbonization of Maritime Transport: Is There Light at the End of the Tunnel?," Sustainability, vol. 13, no. 1, 237, 2021. https://doi.org/10.3390/su13010237

21.   International Maritime Organization, "Report of the Marine Environment Protection Committee on Its Eighty-Third Session (MEPC 83)," MEPC 83/15, IMO, London, 2025.

22.   International Maritime Organization, "2024 Guidelines on Life Cycle GHG Intensity of Marine Fuels," Resolution MEPC.391(81), IMO, London, 2024.

23.   Kulitsa, M., Wood, D.A., 2020. Boil-off gas balanced method of cool down for liquefied natural gas tanks at sea. Advances in Geo-Energy Research 4(2), 199-206.

24.   Jeong, S., Jeong, D., Park, J., Kim, S., Kim, B., 2019. A voyage optimization model of LNG carriers considering boil-off gas. Proceedings of OCEANS 2019 MTS/IEEE SEATTLE, Seattle, WA, USA, 1-7.

25.   Fernández, I.A., Gómez, M.R., Gómez, J.R., Insua, Á.B., 2017. Review of propulsion systems on LNG carriers. Renewable and Sustainable Energy Reviews 67, 1395-1411.

26. Klein, R.A., 2011. Responsible cruise tourism: issues of cruise tourism and sustainability. Journal of Hospitality and Tourism Management 18(1), 107-116.

27. Transport & Environment, 2023. One Corporation to Pollute Them All: Luxury cruise giant emits 10 times more SOx than all of Europe's cars. T&E Campaign Report, Brussels.

28. Psaraftis, H.N., Kontovas, C.A., 2010. Balancing the economic and environmental performance of maritime transportation. Transportation Research Part D 15(8), 458-462.

29. Zis, T.P.V., Psaraftis, H.N., 2017. The implications of the new sulphur limits on the European Ro-Ro sector. Transportation Research Part D 52, 185-201.

30. Stopford, M., 2009. Maritime Economics, 3rd edition. Routledge, London.

31. Christiansen, M., Fagerholt, K., Nygreen, B., Ronen, D., 2013. Ship routing and scheduling in the new millennium. European Journal of Operational Research 228(3), 467-483.

32. Adland, R., Cariou, P., Wolff, F.C., 2020. Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet. Transportation Research Part E: Logistics and Transportation Review 140, 101972.

33. Fagerholt, K., Gausel, N.T., Rakke, J.G., Psaraftis, H.N., 2015. Maritime routing and speed optimization with emission control areas. Transportation Research Part C 52, 57-73.

34. Pasha, J., Dulebenets, M.A., Kavoosi, M., Abioye, O.F., Theophilus, O., Wang, H., Kampmann, R., Guo, W., 2020. Holistic tactical-level planning in liner shipping: an exact optimization approach. Journal of Shipping and Trade 5, 8.

35. Notteboom, T.E., 2006. The time factor in liner shipping services. Maritime Economics & Logistics 8(1), 19-39.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.