

Article

Measurement Adequacy and Parameter Identifiability Before Calibration of a Physics-Informed Ship-Performance Model: A Methodological Case Study on Four Tanker Voyages

Doru Coşofreţ ¹, Alecu Toma ², Florin Postolache ¹, Adrian Popa ¹, Octavian Narcis Volintiru ^{1,*} and Daniel Mărăşescu ¹

¹ Department of Naval Electromechanical Systems, Faculty of Marine Engineering, Naval Academy “Mircea cel Batran”, Fulgerului 1, 900218 Constanta, Romania; doru.cosofret@anmb.ro (D.C.); florin.postolache@anmb.ro (F.P.); adrian.popa@anmb.ro (A.P.); daniel.marasescu@anmb.ro (D.M.)

² Department of Navigation and Naval Transport, Faculty of Navigation and Naval Management, Naval Academy “Mircea cel Batran”, Fulgerului 1, 900218 Constanta, Romania; alecu.toma@anmb.ro

* Correspondence: octavian.volintiru@anmb.ro

Abstract

Calibrating ship-performance models on operational data presupposes traceable, independent and sufficiently excited onboard channels. This methodological case study audits 207 records (0.5–5 h each) from four voyages of one tanker before calibration; the fuel and power instruments are undocumented. A five-parameter fuel observation is structurally limited to two composite directions, of which the available records support one: a two-composite fit returns inadmissible added-power multipliers. The scalar estimated instead κ is the ratio of reconstructed main-engine fuel to reported power at a nominal specific consumption, not a physical efficiency. Frozen after the opening quarter of each voyage, κ predicts remaining-route fuel totals to within -5.67% to $+0.70\%$, against -24.95% to $+0.16\%$ uncorrected; a scalar fitted once on the same window performs equally, so the gain is scaling, not recursion. Across two fuel-conversion baselines, two fuel-field readings and seven freeze-point definitions, the totals stay within -8.5% to $+2.1\%$ and the dispersion of κ within 25–31%; per-voyage RMSE changes and segment-level speed profiles are not stable. Reported and modelled power are different observation operators: on matched references, the V2 cost reduction is $+1.15\%$ with κ and $+0.06\%$ with the scale that modelled power requires. Optimisation results are conditional sensitivity experiments, not realised savings.

Keywords: measurement adequacy; data provenance; structural identifiability; practical identifiability; excitation; offline replay; ship performance; fuel consumption; voyage optimisation; case study



Academic Editors: Amin Nazemian,
Amir Bordbar and Mina Tadros

Received: 24 September 2026

Revised: 11 October 2026

Accepted: 13 October 2026

Published: 17 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Maritime transport accounts for approximately 80–90% of global merchandise trade and remains one of the most energy-efficient freight transport modes [1]. The sector nevertheless contributes nearly 3% of global anthropogenic greenhouse gas (GHG) emissions and is subject to progressively stricter environmental regulation. Recent regulatory developments, including the IMO 2023 Greenhouse Gas Strategy [2], the Carbon Intensity Indicator (CII) framework [3], the extension of the European Union Emissions Trading System to maritime transport [4], and FuelEU Maritime [5], have increased the pressure on ship operators to improve operational efficiency while reducing lifecycle CO₂ emissions.

In previous work [6], a physics-informed optimisation framework was developed for route-level maritime decarbonisation. The framework integrated hydrodynamic resistance prediction, propulsion modelling, fuel-consumption estimation, CO₂-emission accounting, and regulatory compliance assessment within a single optimisation environment. Using operational voyage data from an oil/chemical tanker, the study evaluated speed optimisation under ETA, CII, EU ETS, and FuelEU Maritime constraints. The model used fixed calibrated parameters, which define the starting point for the study reported here. The present work is a direct continuation of that framework on the same vessel and the same four voyages, but it addresses a different question. The earlier study asked what speed profile a calibrated model recommends under economic and regulatory constraints and took the calibration as given. This study asks the prior question: what the available operational channels can identify about that calibration in the first place and what follows for the speed decision when the answer is that they identify less than the model requires. The optimisation layer is therefore no longer the object of study but the instrument through which a measurement-level limitation is expressed in euros, tonnes and knots.

Table 1 separates the physical and economic model inherited from the previous study from the audit, identifiability and decision-sensitivity procedures introduced here.

Table 1. Relationship between the previously published optimisation model and the offline case study reported here.

Component	Previous Framework	Present Study
Physical, fuel, emission and regulatory models; BT, DP and PPO solvers; static parameter calibration	✓	✓
Record-level provenance audit and channel screening	–	✓
Identifiability analysis of the candidate parameters	–	Structural and practical
Offline replay	–	✓ (no live connection)
Implementation example after the audit	–	Scalar Kalman filter on κ ; frozen window scalar as comparator
Posterior uncertainty	–	Retained; not propagated into reported decisions
Transfer between power bases	–	Five applied scales on matched references
Scenario-conditioned optimisation	–	Recorded conditions and deterministic perturbations; dynamic programming
Secondary material	–	Kinematic comparator (Arm 2) and PPO benchmark, Supplementary Section S15

Among the available operational measures, speed optimisation remains one of the most effective [7,8]. It is normally implemented using physics-based propulsion and resistance models, of which the semi-empirical Holtrop–Mennen method [9,10] is the most widely used. Such models are calibrated offline and then applied with fixed parameter values, so agreement deteriorates as the hull, the machinery and the operating environment depart from the calibration condition [11–13]. Section 2.1 reviews both strands in detail.

Digital Twin architectures normally require a maintained connection between a physical asset and its virtual representation [14]. The present work does not implement that connection. It evaluates the estimation chronology by replaying archived records in their original order and is therefore described as an offline, Digital-Twin-oriented calibration workflow rather than an operational Digital Twin.

Maritime studies increasingly combine physics-based models with operational data [15–18], and the adequacy of such data has itself become a subject of study. Data pro-

cessing for ship performance analysis has been reviewed systematically [19]; the effect of the diversity, quality and quantity of real ship data on black-box and grey-box fuel models has been quantified [20]; variable dependence and domain-informed predictor selection have been examined on sensor data from a gas carrier [21]; and the influence of chronological validation on the reported performance of fuel-consumption models has been demonstrated [22]. Section 2.3 reviews this work. These studies assess data adequacy with respect to the predictive accuracy of data-driven or grey-box models. A different question arises when the model is physics-based and its parameters are meant to be interpreted: whether the available observation can separate those parameters at all. A numerically stable estimator can fit a reporting relationship even when the underlying physical parameters are not identifiable.

This paper addresses that question as a methodological case study of one tanker. Its research question is the following: what can the routinely reported fuel and power channels of this vessel identify about the parameters of a physics-informed fuel model, and what does that limit imply for a speed decision that relies on the calibrated model? The channels are audited and screened first. The five-parameter fuel candidate is then examined for identifiability before any estimation. One scalar, κ , relating reconstructed main-engine fuel to reported main-engine power, is estimated over the opening quarter of each voyage, frozen, and used to predict the remaining route. The frozen value is finally transferred to the fuel model used by dynamic programming on the remaining route, where fuel is computed from modelled power and not from reported power.

The study does not propose a new estimator, a new solver or an operational Digital Twin, and the phenomenon it documents is known: a calibration parameter estimated without a discrepancy model absorbs that discrepancy and does not extrapolate (Section 2.4). What it adds is limited to three findings on primary operational records.

- How much of a five-parameter physical correction the reported data support. Three parameter combinations are structurally unidentifiable whatever the data. A fourth, the environmental composite, is structurally identifiable but not practically so on these voyages. On these records, only one scale is practically identifiable.
- That this scale is specific to the power basis on which it is identified. The correction identified against reported power and the scale required by modelled power differ by a factor of up to 1.7 over the same window, and substituting one for the other changes the modelled decision outcome.
- Which conclusions survive the assumptions that cannot be verified. Under two fuel-conversion baselines, two readings of the source fuel field and seven freeze-point definitions, the prospective fuel totals and the between-voyage dispersion of κ are stable; per-voyage error improvements and segment-level speed profiles are not.

The scalar Kalman filter used for κ is an implementation example and not a contribution: a scalar fitted once on the same window gives the same prospective result (Section 5.2.4).

The scope of these findings is one medium-range tanker, four Mediterranean voyages and 207 records of 0.5–5 h. The study does not estimate any physical condition of the hull, propeller or engine. It supports no statement about other vessels, vessel types or operating envelopes. It reports no realised fuel, cost or emission saving, and it does not rank estimators or solvers in general. A reinforcement-learning benchmark and a kinematic speed-over-ground comparator, which do not bear on the research question, are reported in the Supplementary Materials.

The remainder of this paper is organised as follows. Section 2 reviews the literature on ship-performance modelling, estimation in maritime Digital Twins, and the adequacy of operational data, and states the research gap. Section 3 presents the offline replay workflow, the identifiability analysis of the candidate parameters, the reduced estimator and the decision

problem. Section 4 describes the dataset, its provenance, the definition and measurement chain of the reported channels, the five-test screening and the evaluation protocol. Section 5 reports the screening and identifiability results, the scalar correction and its sensitivity to the fuel-field reading and the freeze point, and the transfer and decision-sensitivity tests. Section 6 discusses the results and states the limitations. Section 7 gives the conclusions.

2. Related Work and Research Gap

2.1. Ship Optimisation and Performance Modelling

Because propulsion power depends nonlinearly on speed, moderate speed reductions produce disproportionately large reductions in fuel consumption. Psaraftis and Kontovas [7] reported that a 10% speed reduction may yield fuel savings approaching 27%, and Fagerholt et al. [8] quantified these gains for combined speed and route decisions.

Recent studies have expanded traditional optimisation approaches by incorporating regulatory compliance indicators such as CII, EU ETS costs, and FuelEU Maritime penalties into voyage planning and operational decision-making. Voyage optimisation must weigh economic objectives, environmental performance, and regulatory compliance within one decision-support environment.

Physics-based ship performance models remain the foundation of most maritime energy-efficiency studies. Semi-empirical resistance formulations, particularly the Holtrop–Mennen method [9,10], associate resistance components with hull geometry, displacement, propulsion characteristics, and environmental loading, and therefore support physically interpretable estimates of resistance and propulsion power across a broad range of vessel types and operating conditions. They remain widely adopted in operational optimisation, ship-performance monitoring, voyage planning, and regulatory compliance studies.

However, the predictive accuracy of such models depends on the calibration. Parameters identified under one set of operating conditions progressively lose validity as the hull, the machinery and the operating environment change. Coraddu et al. [11] and Tillig et al. [12] demonstrated that environmental variability and operational uncertainty can affect ship-performance predictions when static model parameters are used over extended periods.

Hybrid models that combine physical knowledge with data-driven techniques have been developed in response. Gkerekos et al. [23] showed that machine-learning models can improve fuel-consumption prediction accuracy when integrated with domain-specific physical information. The Physics-Informed Neural Networks (PINNs) of Raissi et al. [24] embed the governing physical equations in the learning architecture, which improves prediction while keeping it physically consistent. Both approaches underlie the combination of physics-based and data-driven methods in maritime applications.

Data-driven and grey-box models of fuel consumption have developed alongside the physics-based ones. Cai et al. [20] proposed a multi-source data-processing framework, combined ship theory with a long short-term memory network to form a grey-box model, and examined how the diversity, quality and quantity of real ship data affect black-box and grey-box predictions. In their experiments, the grey-box model achieved higher predictive accuracy with less training data than the black-box model. Vorkapić et al. [21] built interpretable linear-regression and decision-tree models of propulsion fuel consumption for a very large gas carrier from engine sensor data, after data exploration, cleaning and verification, and a comparison of correlation-based, genetic and ReliefF feature selection.

2.2. Adaptive Digital Twins and Estimation

Tao et al. [14] defined a Digital Twin as a virtual representation of a physical asset that continuously evolves through the integration of real-time data, physical models, and computational intelligence. Unlike conventional simulation models, a Digital Twin maintains

an ongoing connection with the physical system and can therefore support monitoring, prediction, optimisation, and decision-making throughout the operational lifecycle.

In the maritime sector, early Digital Twin applications addressed ship-performance monitoring and machinery diagnostics. Coraddu et al. [25] developed a data-driven Digital Twin that estimates the speed loss caused by hull fouling. Lee et al. [26] extended the concept to real-time ship operation in waves. Two recent contributions bear directly on the present work: Taghavi and Perera [27] apply adaptive estimation with Kalman filtering to ship fuel-consumption prediction on operational data, and Jeon and Theotokatos [28] propose a trustworthiness framework for physics-based digital twins of marine engines. The present study differs from the first in placing channel screening and an identifiability analysis before calibration and not after it, and complements the second by supplying an empirical case in which the trustworthiness of the reported fuel and power channels cannot be established from the records.

More recent studies have explored the use of Digital Twins for voyage optimisation, autonomous navigation, predictive maintenance, and energy-efficiency management. Wei et al. [16] proposed a Digital Twin-assisted routing framework capable of evaluating alternative operational strategies under decarbonisation constraints, while Kinaci [18] discussed Digital Twin architectures supporting autonomous sailing and intelligent ship operations.

Reviews report a growing use of Digital Twins in maritime decarbonisation and digitalisation. Mauro and Kana [17] concluded that they can improve situational awareness, prediction and operational decision support. Most existing applications nonetheless address monitoring, diagnostics or prediction, and not optimisation or adaptive decision-making.

Because fixed model parameters gradually become inconsistent with the actual vessel state, state- and parameter-estimation techniques are used to maintain model fidelity. Among these methods, the Extended Kalman Filter (EKF) remains one of the most widely used approaches for nonlinear state estimation and system identification [29]. The EKF extends the Kalman filter by linearising the system dynamics around the current operating point, so that states and model parameters can be estimated recursively [30,31]. The properties of this linearisation, including the conditions under which the filter remains stable, are well established in the estimation literature [32,33].

In marine systems, estimation methods based on Kalman filtering have been extensively applied to navigation, motion prediction, fault detection, and control. Fossen [34] gives the theoretical basis for marine craft modelling, estimation and control, including observability analysis and nonlinear state estimation. Perera and Mo [35] further demonstrated how operational ship and engine data can be used to identify performance-relevant operating regimes and detect deviations from nominal vessel behaviour.

Existing Digital Twin applications typically employ estimation for monitoring and diagnostics. The consequences of updating vessel-performance models from routinely reported channels, and of carrying the resulting correction into optimisation decisions, have received less attention.

2.3. Data Quality, Feature Selection and Chronological Validation

A separate body of work addresses whether operational data are fit for performance modelling at all. Kim et al. [19] reviewed data processing for ship performance analysis over the literature from 2014 to 2024, covering the main data sources (AIS, onboard sensors and noon reports), their typical quality problems, and the techniques of synchronisation, missing-value imputation, cleaning and uncertainty management, and noted the absence of a unified processing pipeline. Cai et al. [20] treated the diversity, quality and quantity of real ship data as experimental factors. Vorkapić et al. [21] showed that correlation-based feature selection fails when predictors are strongly inter-correlated and that removing redundant

variables requires domain understanding; in their study fuel flow was measured by Coriolis mass-flow meters with a stated total error of 0.15% and shaft power by a torque meter. Vorkapić [22] isolated the effects of chronological validation, temporal input features and autoregressive persistence on reported accuracy, and showed that values obtained under random partitioning are not comparable with those obtained under strict chronological validation, under which a baseline regression reached $R^2 = 0.788$.

Measurement and data quality, variable dependence and reduction, domain-informed feature selection, and chronological validation on real vessel data are therefore established practice. The present study relies on them: the screening tests of Section 4.3 are data-quality checks, and the evaluation protocol of Section 3.3 is chronological. None is claimed here as a contribution. The present dataset is also far smaller than those used in the studies above; no cleaning, imputation or synchronisation pipeline is applied to it, and no instrument record comparable to that of [21] is available for the vessel (Section 4.2.4).

2.4. Measurement Adequacy and Research Gap

Three limitations persist across this literature.

First, fixed-parameter optimisation can lose agreement when operating conditions or measurement characteristics change. Second, a recursive estimator may reduce numerical error without identifying the physical state assigned to it. Third, prediction gains do not establish that optimisation decisions improve. Measurement screening, identifiability, excitation, prediction assessment and decision assessment must therefore be separated.

These limitations are instances of a result established in the calibration literature outside the maritime field. In the Bayesian calibration framework of Kennedy and O'Hagan [36], a physical model is fitted to observations together with a discrepancy term. Brynjarsdóttir and O'Hagan [37] showed that when the discrepancy is not modelled and constrained, the calibration parameter absorbs it: the fit improves within the observed domain while the estimated parameter loses its physical meaning, and agreement obtained there does not license extrapolation to a different context or to a different model–observation relationship. A calibrated coefficient is therefore interpretable as a physical quantity only if the observation is informative enough to separate it from the discrepancy it might be absorbing. Whether that condition holds is an identifiability question. It can be tested before estimation by the rank of an observability–identifiability matrix, an established method in systems biology [38]; Section 3.2 applies the same rank test to the fuel observation.

The gap addressed here is correspondingly narrow. The studies reviewed in Section 2.3 assess the adequacy of operational data with respect to the predictive accuracy of data-driven or grey-box models. They do not test whether the individual parameters of a physics-based model can be separated from the fuel observation before those parameters are estimated and interpreted, and they do not distinguish the power basis on which a correction is identified from the basis on which a decision model applies it. The present case study examines these two questions on one set of primary records and follows the answer through to a speed decision. It retains the baseline physics-informed optimiser of [6], estimates one declared reporting-channel ratio, and compares static and κ -conditioned solutions after a fixed adaptation window.

Table 2 positions the case study in two panels. Panel (a) compares representative maritime optimisation and Digital Twin studies along the steps followed here. Panel (b) lists the data-adequacy and prediction studies of Section 2.3 along the questions they address. The entries record the stated scope of each study and not its quality: a study that does not test parameter identifiability is not deficient if its model has no physical parameters to identify.

Table 2. Positioning of the present case study. (a) Maritime optimisation and Digital Twin studies. (b) Data-adequacy and prediction studies.

(a)						
Study	Channel Provenance Audit	Identifiability Tested Before Estimation	Recursive Estimation	Reported and Modelled Power Separated	Decision-Level Consequence Quantified	Real Ship Data
Psaraftis and Kontovas [7]	–	–	–	–	Outside scope	–
Fagerholt et al. [8]	–	–	–	–	Outside scope	–
Gkerekos et al. [23]	Partial	–	–	–	Partial	✓
Coraddu et al. [25]	Partial	–	Partial	–	–	✓
Lee et al. [26]	–	–	–	–	–	✓
Wei et al. [16]	–	–	Partial	–	Outside scope	✓
Kinaci [18]	–	–	–	–	Partial	✓
Taghavi and Perera [27]	–	Not applicable	✓	Not applicable	–	✓
Jeon and Theotokatos [28]	Partial	Partial	–	–	–	Test-bed
Present case study	Records traced; instruments undocumented	✓	Scalar; implementation example	✓	Modelled, not realised	One vessel, four voyages
(b)						
Study	Data Quality or Processing	Variable Dependence, Feature Selection	Chronological Validation	Identifiability of Physical Parameters	Power Bases Separated	Real Ship Data
Vorkapić et al. [21]	✓ (exploration, cleaning, verification)	✓	– (random split, stratified by date)	Not applicable	Not applicable	✓
Cai et al. [20]	✓	–	–	Not applicable	Not applicable	✓
Kim et al. [19]	✓ (review)	–	–	Outside scope	Outside scope	Review
Vorkapić [22]	–	✓ (temporal features)	✓	Not applicable	Not applicable	✓
Present case study	Partial: screening tests only; no cleaning pipeline	Partial: algebraic dependence among channels	✓ (adopted from prior practice)	✓	✓	One vessel, four voyages

Note: Entries record the stated scope of each study, not its quality; a dash marks an aspect that is not part of that stated scope. “Outside scope” marks a question the work was not designed to address; “Not applicable” marks one that does not arise in that formulation, as in a data-driven model or a realised state-space model whose quantities are not physical parameters. In the present case study, the estimated quantity is restricted to one reporting-channel scalar, and the decision comparison is performed after a fixed adaptation window without propagation of posterior uncertainty. Source: Authors’ compilation.

Within the studies listed, the individual steps are well represented. Among the entries of panel (a), only [28] touches on the identifiability of physical parameters, through a sensitivity analysis of its engine model, and none separates the power basis on which a correction is identified from the one on which it is applied. In panel (b) these questions do not arise, because the models are not parameterised physically; those studies address data quality, feature selection and chronological validation, on which the present case study is partial and makes no claim. The statement is limited to the studies listed and is not a claim about the literature as a whole.

3. Offline Calibration Workflow and Mathematical Formulation

3.1. Architecture and Information Chronology

Figure 1 shows the information flow from the archived vessel records to the counterfactual speed decision. The estimator receives each observation once, in chronological order. The decision layer is evaluated only after κ is frozen; it receives no fuel observation from the planned portion of the route.

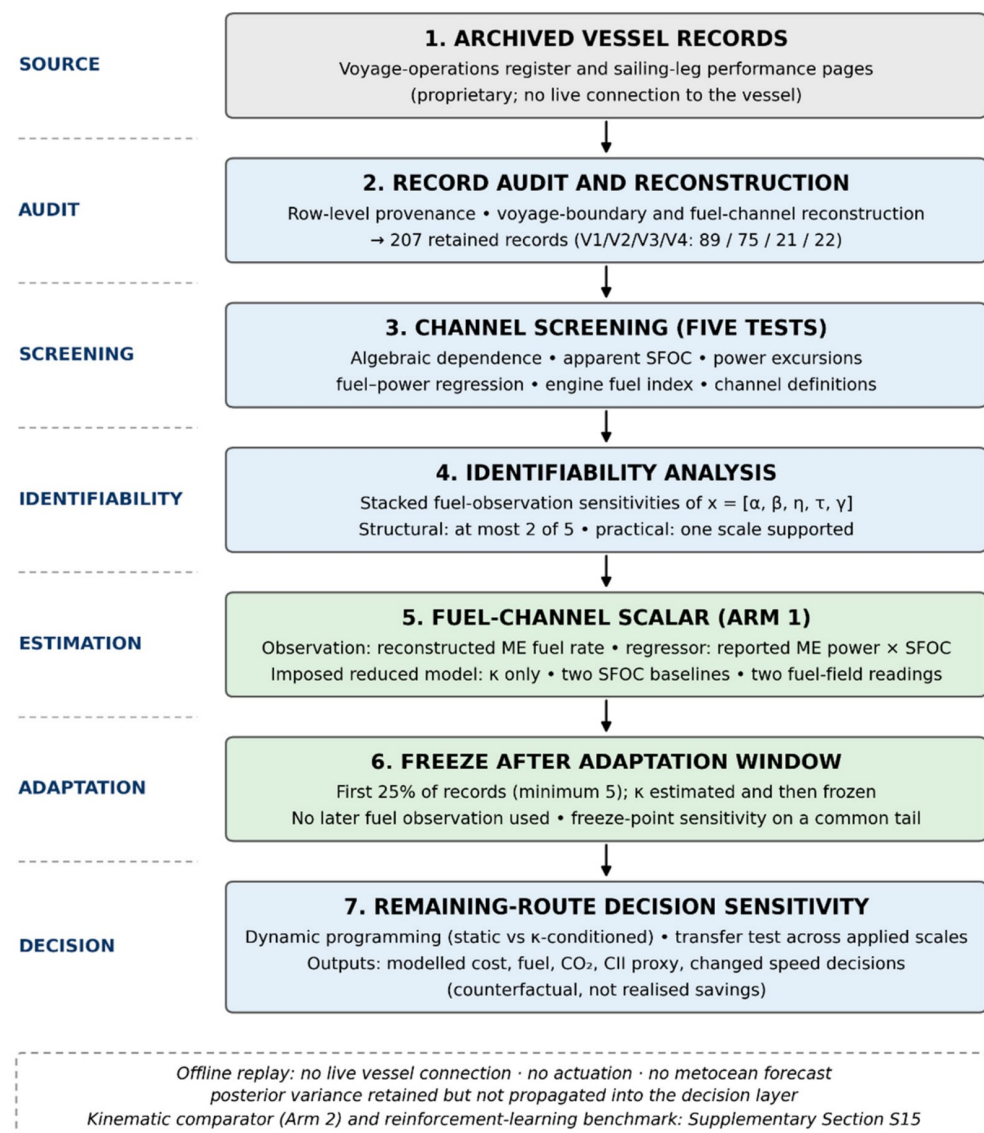


Figure 1. Offline replay workflow: information flow from the archived vessel records to the counterfactual speed decision. The kinematic comparator (Arm 2) and the reinforcement-learning benchmark are reported in Section S15 and do not enter the fuel, power or decision model.

The baseline model uses fixed parameters. The fuel-only candidate is examined first, so that the number of composite directions its observation can support is known before any estimator is defined. Arm 1 then fixes α and β at unity, uses reported main-engine power, and estimates only $\kappa = \tau/\eta$; this reduction is a declared modelling choice and not a consequence of the rank result alone (Section 3.2). A separate kinematic comparator (Arm 2) is described in Supplementary Section S15 and is not propagated into the fuel, power or decision model.

The evaluated workflow links physics-informed prediction of propulsion, fuel, emissions and regulatory indicators; estimation of the reduced reporting-channel correction; and post-adaptation speed optimisation by dynamic programming, with a beam search as a cross-check. Posterior variance is stored but not propagated through the reported optimisation.

The Physical System Layer represents the vessel and operating environment. The primary inputs are vessel-reported channels extracted from the onboard performance-monitoring records. They contain speed, power, fuel, environmental, loading, and route fields; the underlying sensor and acquisition architecture was unavailable. Source and verification status are reported in Supplementary Table S2a.

The Data Acquisition and Processing Layer applies the documented wind-speed conversion, reconstructs the main-engine fuel channel, and prepares the estimator and model inputs.

The State Estimation Layer begins from the five-parameter identifiability analysis. After the imposed reduction, Arm 1 estimates κ from reconstructed main-engine fuel and reported power with a scalar linear Kalman filter, and posterior variance is retained.

The Physics-Informed Modelling Layer combines fixed physical-model inputs with the reduced corrections and produces modelled propulsion, fuel, emission, and regulatory outputs. The estimated corrections are not interpreted as independently identified vessel-condition states.

The Optimisation and Decision Layer evaluates candidate speed profiles under the stated operational, economic, and regulatory conditions. In this study it produces counterfactual recommendations only; no profile is applied onboard.

The information chronology is the same for every chronological record sequence. During an initial adaptation window, κ is updated from the available fuel observations and its predictions are scored one step ahead.

At the freeze point, the terminal adaptation estimate κ_{ad} is retained. All subsequent records form the prospective remaining-route interval, in which κ is held fixed and the solvers are evaluated without access to later fuel observations.

Figure 2 summarises this information chronology independently of any voyage-specific record count or estimated value.

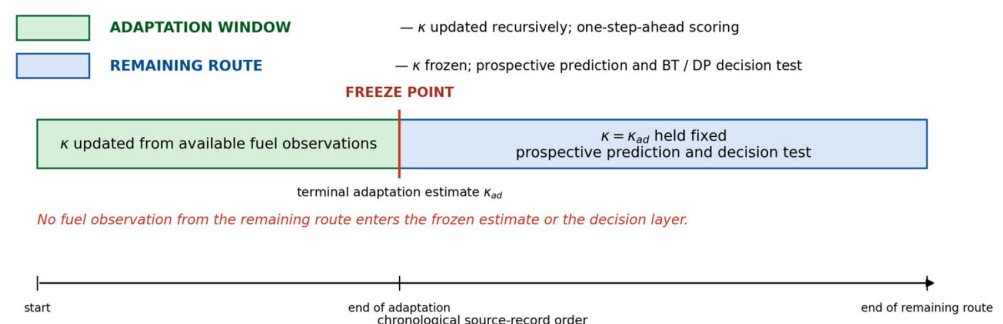


Figure 2. Conceptual information chronology of the prospective adaptation and decision test. The reporting-channel correction κ is updated only within the adaptation window and is then frozen as κ_{ad} for the remaining route; no subsequent fuel observation enters the frozen estimate or the decision layer.

The implementation is an offline replay. It has no live vessel connection, does not actuate the ship, and does not constitute predictive control. Its purpose is to test what the archived channels can support and how an imposed reporting correction affects a model-based decision.

3.2. Candidate Parameter Vector and Identifiability Reduction

Equation (1) states the five-parameter candidate vector used for the identifiability analysis. Its components are candidate correction directions; the measurements available in this study determine which combinations of them can be estimated.

$$x_k = [\alpha_k, \beta_k, \eta_k, \tau_k, \gamma_k]^T \quad (1)$$

where

α is the calm-water power correction in the candidate vector [-];

β is the added environmental-power correction in the candidate vector [-];

η is the propulsive-efficiency correction in the candidate vector [-];

τ is the main-engine fuel-conversion correction in the candidate vector [-];

γ is the auxiliary-demand correction in the candidate vector [-].

The physical meanings listed above define the candidate model directions. They are not treated as identified quantities unless the observation supports them. Three notions are kept separate in what follows.

Structural identifiability. The analysis is a sensitivity-based identifiability test of the parameters of an algebraic observation: the fuel-rate model of Equation (4) is differentiated with respect to the five candidates, and the segment rows are stacked by voyage (Supplementary Section S3). Because Equation (4) is multiplicative, three linear dependencies among the sensitivities hold at every parameter value and for every dataset: γ does not enter the main-engine observation; τ and η enter only through their ratio; and α , β and η admit a common scaling (Equations (S8d), (S9) and (S10)). The rank of the stacked sensitivity matrix therefore cannot exceed two. This bound is a property of the model. It does not depend on the linearisation point or on the records.

Practical identifiability. Whether the two remaining composites, $\tau\alpha/\eta$ and $\tau\beta/\eta$, can actually be estimated is an empirical matter. It requires the ratio of added to calm-water power to vary across records. It is assessed by the conditioning of the two non-zero singular values (Section 4.3) and by fitting both composites directly and examining their standard errors, correlation and sign (Section 5.2.5).

Relation to observability. If the candidate parameters are treated as a constant or random-walk state with identity transition, the observability Gramian accumulated over a voyage is the sum of the outer products of the sensitivity rows, and its rank equals that of the stacked sensitivity matrix [29–31]. In that formulation parameter identifiability and state observability coincide, and the rank test may be read either way. The term identifiability is used throughout this paper, because the object analysed is a set of parameters of a static observation and not the state of a dynamical system.

The stacked sensitivity matrix establishes which parameter combinations no estimator could separate from this observation model. It does not establish that the observation model is correct, that the identifiable combinations are estimated accurately, or anything about a model with a different regressor.

The candidate formulation follows standard nonlinear state-estimation principles [30,31], in which a five-dimensional covariance matrix,

$$P_{k|k} = \text{Cov}(x_k - \hat{x}_{k|k}) \quad (2)$$

would be propagated by an extended Kalman filter. That five-parameter filter is not run. The estimator reported in this study is a scalar linear Kalman filter on a separately defined reduced model (Section 3.3); it acts as a random-walk parameter tracker, and no observability claim is attached to it. The reported optimisation uses the frozen scalar estimate and does not propagate posterior variance into its constraints. The sensitivity matrix, rank analysis, update equations and evaluation order are given in Supplementary Sections S3–S5.

3.2.1. κ -Conditioned Propulsion, Fuel, and Emission Model

The offline model preserves the implemented propulsion, fuel, emission, economic, and regulatory structure of the baseline framework [6]. The complete relationships used in the present calculations are reproduced in Equations (S24)–(S37b) of Supplementary Section S6 so that the current study can be read independently of the earlier article.

For each segment (i), the model evaluates a candidate speed through water together with recorded current, wave, wind, and auxiliary-load inputs. Static optimisation uses $\kappa = 1$; κ -conditioned optimisation uses the frozen κ estimated from the adaptation window.

The total propulsion power demand is expressed as:

$$P_{total,i} = \alpha_i \cdot P_{calm}(V_i) + \beta_i \cdot P_{add,i}(V_i, W_i) \quad (3)$$

where

$P_{calm}(V_i)$ denotes the calm-water propulsion power obtained from the baseline physics-informed model [kW];

$P_{add,i}$ represents the additional propulsion power associated with environmental loading, including wind and wave effects [kW];

α_i is the candidate calm-water resistance correction factor [-];

β_i is the candidate environmental-resistance correction factor [-].

The candidate coefficients α and β parameterise departures from the nominal calm-water and added-power terms. They are included in the identifiability analysis but are fixed at unity in the reported static and κ -conditioned solver runs.

The corresponding main-engine fuel-consumption rate is obtained as:

$$\dot{m}_{ME,i}^{DT} = \left(\frac{\tau_i}{\eta_i} \right) \cdot \frac{P_{total,i} \cdot SFOC_{ME}}{1000} \quad (4)$$

where

$SFOC_{ME}$ denotes the nominal main-engine specific fuel oil consumption [g/kWh];

η_i represents the effective propulsion-efficiency correction factor [-];

τ_i represents the fuel-consumption correction factor [-]. Equation (4) defines a mass flow rate in [kg/h]; the fuel mass consumed over segment (i) is the product of this rate and the segment duration (t_i), in [kg]. The superscript DT is retained from [6]; it marks a modelled quantity and does not denote an operational Digital Twin.

The structural bound of two admits the composites $\theta_1 = \tau\alpha/\eta$ and $\theta_2 = \tau\beta/\eta$, in terms of which Equation (4) reads $\dot{m} = (SFOC/1000)(\theta_1 P_{calm} + \theta_2 P_{add})$. These composites belong to the modelled-power observation. The reported Arm 1 estimator does not use that observation. It regresses the reconstructed main-engine fuel rate on reported main-engine power, which is a different observation operator with a single regressor, and κ is its only parameter. κ is therefore not one of the two composites with the other discarded. It is the parameter of a separately declared reporting-channel model, written $\kappa = \tau/\eta$ only by analogy with the candidate. The choice of reported power as the regressor is a modelling decision and does not follow from the rank result. Section 5.2.5 evaluates the two-composite formulation on the modelled-power basis and shows why it is not adopted.

The fuel-consumption model separates propulsion and auxiliary demand. Fuel-specific emission accounting follows established inventory practice [39], while the separation of main and auxiliary-engine energy use is consistent with ship energy-system analysis [40]. Auxiliary-engine consumption is not included in the Arm 1 observation model.

The corresponding segment-level CO₂ emissions are calculated as:

$$m_{CO_2,i} = EF_{ME} \cdot \dot{m}_{ME,i}^{DT} \cdot t_i + EF_{AE} \cdot \gamma_i \cdot \dot{m}_{AE,i} \cdot t_i \quad (5)$$

where

EF_{ME} and EF_{AE} denote the emission factors of the main-engine and auxiliary-engine fuels [kg CO₂/kg fuel];

$\dot{m}_{AE,i}$ denotes the auxiliary-engine fuel-flow rate [kg/h], reconstructed from the source auxiliary-consumption channel as defined in Section S1;

t_i represents the duration of route segment (i) [h];

γ_i denotes the auxiliary-energy correction factor [-]. It does not enter the observation model of Equation (4), which contains main-engine consumption alone, so the measurement carries no information about it.

The auxiliary-engine contribution is explicitly retained because it influences voyage-level emission inventories and regulatory indicators such as the Carbon Intensity Indicator (CII), EU ETS costs, and FuelEU Maritime compliance metrics [3–5].

3.2.2. Integration with Regulatory Assessment

The segment fuel and emission quantities enter the route-level CII proxy, EU ETS cost, route-level modelled FuelEU compliance-cost proxy, and total-cost objective reproduced in Supplementary Section S6. Static and κ -conditioned evaluations differ only through κ .

The route-level CII value is a diagnostic voyage proxy. EU ETS and FuelEU coverage are applied by voyage according to Supplementary Tables S8 and S9. The FuelEU term is a route-level modelled compliance-cost proxy and not an actual voyage-level compliance obligation; no annual attained CII validation is claimed.

The formulation uses the same physical and regulatory accounting in both configurations and isolates the effect of the transferred reporting-channel correction.

A complete notation reference is provided in Supplementary Section S2.

3.3. Scalar Estimator

The estimator assimilates the reconstructed main-engine fuel-flow rate and estimates the scalar correction κ . The posterior estimate for segment i is obtained after the one-step forecast for that segment has been scored.

For the decision experiment, κ is updated only during the first 25% of the source records of each voyage, with a minimum of five records. The 25% split was fixed before the analysis to balance adaptation against the number of records retained for prospective evaluation, and the five-record minimum preserves an initial sample on the two shorter voyages. The rule is a fraction of records, not of elapsed time: because record durations range from 0.5 to 5 h, the window covers 24.6%, 24.0%, 22.3% and 17.6% of the duration of V1–V4, 27.2%, 23.9%, 22.2% and 18.2% of the distance, and 23.6%, 23.8%, 22.0% and 16.5% of the reconstructed main-engine fuel; its environmental excitation also differs from that of the whole voyage (Table 9). Section 5.2.6 examines the sensitivity of the results to the freeze point, including windows defined by elapsed time. The terminal adaptation-window estimate is then frozen, and the remaining route is optimised without access to later fuel observations. This ordering prevents future observations from entering the decision test.

The candidate observation and identifiability analysis are reported in Supplementary Section S3, the scalar update equations and chronology in Section S4, the statistical pro-

cedures in Section S5, and the software mapping in Sections S9 and S11. A second scalar recursion, a kinematic speed-over-ground comparator (Arm 2), is defined in Supplementary Section S4.2 and reported in Section S15.1; it is not combined with Arm 1 and is not propagated into the fuel or propulsion model.

3.4. Decision Problem and Objectives

The optimisation layer compares the same remaining-route problem under $\kappa = 1$ and under the frozen post-adaptation κ . Each case is evaluated against its own recorded-speed remaining-route reference. The comparison tests model-decision sensitivity to the transferred correction; it is not an estimate of realised onboard savings.

The optimisation problem is formulated over a structured operational scenario space. An operational scenario is represented as:

$$s = (r, W, L) \in S = R \times W \times L \quad (6)$$

where $r \in R$ denotes the route configuration; $W \in W$ represents the metocean state; and $L \in L$ denotes the vessel loading condition.

The reported scenarios are the recorded remaining-route segments of V1–V4. The common decision set, constraints, model-specific references, and solver settings are detailed in Supplementary Section S7.

For a given scenario s and a parameter estimate $x_{k|k}$, the decision variable is the segment-wise speed profile:

$$v = (V_1, V_2, \dots, V_N) \in V^N \quad (7)$$

where N denotes the number of route segments.

The nominal run uses $\kappa = 1$, and the κ -conditioned run uses the value frozen after the adaptation window. Arm 1 identifies κ from reconstructed fuel and reported main-engine power, whereas candidate speeds are evaluated from modelled power, which is a different observation basis.

The comparison records modelled fuel, CO₂, cost, time, CII proxy, solver agreement, and segment-wise profile differences. It does not use later observations to update κ during optimisation.

3.4.1. Objective Functions

Two optimisation objectives are considered.

The first objective minimises voyage fuel expenditure and emission-related compliance costs:

$$J_{fuel}^{DT}(v, x_{k|k}, s) = \sum_{i=1}^N \left[C_i^{fuel}(v, x_{k|k}) + C_i^{ETS}(v, x_{k|k}) \right] + C_{FuelEU}(v, x_{k|k}, s) \quad (8)$$

where fuel costs, EU ETS costs, and the route-level modelled FuelEU compliance-cost proxy, all expressed in [€], are evaluated from the modelled fuel and emission quantities.

The second objective minimises total voyage cost:

$$J_{total}^{DT}(v, x_{k|k}, s) = \sum_{i=1}^N \left(C_i^{fuel} + C_i^{time} + C_i^{ETS} \right) + C_{FuelEU} + C^{ETA} \quad (9)$$

where C_{time} represents time-related operational costs and C_{ETA} denotes schedule-deviation penalties, both expressed in [€].

Equation (9) is the objective of the reported comparison and decision test. Equation (8) provides a grid-constrained fuel-directed comparison on the same remaining routes and under the same latest-arrival limit. Dynamic programming is used because it is exact on

the adopted discrete grid. Section 5.4 reports this comparison and the deterministic-solver cross-checks. The same cost components and constraints are used in the nominal and κ -conditioned runs; only κ differs.

The implemented physical and cost components are reproduced in Supplementary Section S6; the matched-reference definition and solver protocol are given in Supplementary Section S7.

3.4.2. Constraint Formulation

The optimisation problem is subject to operational, technical, and regulatory constraints:

- Voyage completion must satisfy the required arrival time:

$$T_{Route}(v, s) = \sum_i t_i(V_i, s) \leq T_{max} \quad (10)$$

where the voyage duration and the contractual limit are expressed in [h]; the constraint enforces ETA feasibility under the operating conditions considered.

- Propulsion power must remain within the available machinery limits:

$$P_{total}(V_i, x_{k|k}, W) \leq P_{max} \quad (11)$$

for all route segments, with both terms expressed in [kW].

- When regulatory compliance is enforced as a hard constraint, the route-level Carbon Intensity Indicator must satisfy:

$$CII_{Route}(v, x_{k|k}, s) \leq CII_{allow} \quad (12)$$

where the route-level CII proxy is computed from the modelled voyage emissions; both terms are expressed in [g CO₂ / (DWT·NM)].

FuelEU is represented through the route-level modelled compliance-cost proxy in the objective, while the latest-arrival requirement is enforced as a hard constraint.

Posterior uncertainty is retained by the estimators but is not propagated into the reported decisions. The prospective covariance-based formulation is provided only in Supplementary Section S7.9.

3.5. Solver Summary and Sensitivity Analysis

The decision instrument is dynamic programming on a discrete speed set, with the physical model, total-cost objective, ETA limit, propulsion-power limit and regulatory accounting of Section 3.4. The routing structure follows established speed-optimisation formulations [41]; the exact settings are given in Supplementary Section S7.

Dynamic programming uses a 5 min cumulative-time discretisation and returns the minimum-cost path on that grid. Its result is an optimum of the adopted discrete problem, not of the continuous speed-control problem.

A feasibility-aware beam search (BT) is run on the same problem as a cross-check. At each segment it evaluates all admissible speeds, retains the lowest-cost state in each 5 min cumulative-time bin, and limits the active set to 200 states. It is heuristic, because discarded paths cannot re-enter the beam, and it scales as $O(NBM)$ for N remaining-route segments, beam width B and M admissible speeds. A reinforcement-learning benchmark based on proximal policy optimisation [42] was also run against the dynamic-programming solution. Because it does not bear on the research question, its configuration, results and seed variation are reported only in Supplementary Sections S7, S11.2, S12 and S15.2.

The reported sensitivity analysis concerns operational inputs, not posterior-state uncertainty. Significant wave height (H_s) and model-input wind speed (U_w), expressed

in metres and metres per second, are independently perturbed by $\pm 10\%$; displacement is independently perturbed by $\pm 5\%$. Static DP and κ -DP are re-solved for each perturbation.

For model $m \in \{\text{DP}, \kappa\text{-DP}\}$, the optimal-cost sensitivity to operational input (ξ) is defined as:

$$\Phi_m(\xi) = 100 \frac{J_m^*(\xi_{\text{perturbed}}) - J_m^*(\xi_{\text{reference}})}{J_m^*(\xi_{\text{reference}})} \quad (13)$$

The reference term is the unperturbed optimum for the same voyage and the same model. The perturbed term changes one input while all other recorded inputs and solver settings remain fixed.

- A positive value of $\Phi_m(\xi)$ indicates that the perturbation increases optimised total cost; a negative value indicates a reduction.
- The evaluated inputs are $\xi \in \{H_s, U_w, \Delta\}$. Wave height and wind speed are perturbed by $\pm 10\%$, and displacement by $\pm 5\%$.

The analysis uses the same DP discretisation for static and κ -conditioned models, so differences in sensitivity are not caused by a change in optimisation algorithm.

The resulting cost changes and numbers of altered speed decisions are reported in Section 5.3.4 and Supplementary Tables S14 and S15.

3.6. Nomenclature

Table 3 defines the principal symbols used in the main text. The recorded and derived operational variables are listed in Supplementary Table S2, and the additional derivation notation is listed in Supplementary Table S4.

Table 3. Principal symbols used in the main text.

Symbol	Definition	Unit
$x = [\alpha, \beta, \eta, \tau, \gamma]^T$	Candidate five-parameter vector used only for identifiability analysis	–
$\kappa = \tau/\eta$	Scalar reporting-channel fuel-to-power correction estimated in Arm 1	–
$V_{STW,i}$	Speed through water on segment (i)	kn
$V_{SOG,i}$	Speed over ground on segment (i)	kn
$c_{long,i}$	Longitudinal current component, $V_{SOG,i} - V_{STW,i}$	kn
α	Calm-water power correction in the candidate vector	–
β	Added environmental-power correction in the candidate vector	–
η	Propulsive-efficiency correction in the candidate vector	–
τ	Main-engine fuel-conversion correction in the candidate vector	–
γ	Auxiliary-demand correction in the candidate vector	–
P_i^- (prior), P_i (posterior)	Prior and posterior scalar state variance of the reduced estimator	state-dependent
Q, R	Process-noise and observation-noise variance	state ² ; observation ²
P_0	Initial scalar state variance	state-dependent
$\mathcal{H}_i; H_i$	$\mathcal{H}_i$ is the 1×5 sensitivity row of the candidate vector; (H_i) is the scalar observation sensitivity of the reduced estimator, $H_i = g_i$	observation/state
K_i	Scalar Kalman gain	state/observation
$y_i, \hat{y}_i, v_i$	Observation, one-step prediction, and innovation	kg/h or kn
$P_{calm,i}, P_{add,i}, P_{total,i}$	Calm-water, added, and total propulsion power	kW
k_{wave}	Wave-added-resistance coefficient	–
$C_D(\beta_{rel,i})$	Aerodynamic drag coefficient at relative wind direction	–
$\dot{m}_{ME,i}, m_{ME,i}, m_{AE,i}$	Main-engine fuel-flow rate and segment fuel masses	kg/h; kg; kg
$SFOC_{ME}$	Nominal main-engine specific fuel-oil consumption	g/kWh
EF	Fuel-specific CO ₂ emission factor	kg CO ₂ /kg fuel
m_{CO_2}	Route-level CO ₂ mass	kg
v, V_i	Speed-profile vector and segment speed	kn
n	Main-engine shaft speed	rpm
$H_{s,i}, U_{w,i}$	Significant wave height and model-input wind speed	m; m/s
d_i, t_i	Segment distance and duration	NM; h
T_{max}, P_{max}	Voyage-time and propulsion-power limits	h; kW
CII_{route}	Route-level CII proxy	g CO ₂ /(DWT·NM)
$\chi_{ETS,v}$	EU ETS geographic coverage factor	–
$\Phi_m(\xi)$	Optimal-cost sensitivity for input ξ and model m	%
J	Total modelled cost	EUR
$s = (r, W, L)$	Operational scenario: route, metocean state, and loading condition	–

4. Data and Evaluation Protocol

The case study is evaluated on 207 processed operational records from four commercial voyages of one medium-range oil/chemical tanker. The records are irregularly aggregated: 203 are hourly entries and the remaining four cover 0.5, 0.9, 2.42 and 5.0 h. The dataset supports a methodological case study and no population-level inference. The physical, fuel, emission, economic and regulatory equations follow the modelling structure introduced in [6], while the present study applies its own audited record selection, reconstructed main-engine fuel channel and voyage-specific operating-displacement estimates. Static and κ -conditioned cases are compared under this common implementation.

4.1. Vessel and Voyages

Data were obtained from a medium-range oil/chemical tanker operating on Mediterranean and eastern Mediterranean routes, the same vessel used in the baseline study [6]. Table 4 lists the principal particulars and the fixed model inputs.

Table 4. Principal vessel particulars and fixed model assumptions.

Characteristics	Data
Ship type	Oil/chemical tanker
Deadweight, DWT [t]	40,439
Calibration reference displacement, Δ_{ref} [t]	50,029.65
Cargo capacity [m ³]	38,000
Maximum speed [kn]	16.0
Reported service-speed values [kn]	14.5; 12.0
Installed main-engine MCR, P_{max} [kW]	9580
Nominal main-engine SFOC [g/kWh]	194
Nominal propulsive efficiency, η_{prop} [–]	0.60
Projected frontal area, A_{proj} [m ²]	480

Table 4 combines vessel particulars with fixed model assumptions inherited from [6]. The 14.5 kn/5780 kW calibration point, reference displacement, propulsive efficiency, frontal area, wave coefficient, and wind-drag profile are not independently remeasured here. The calm-water model used in the calculations is the calibrated cubic relation reproduced in Supplementary Section S6, not a full Holtrop–Mennen implementation. Because it is fitted at a single point, its use below 14.5 kn is an extrapolation of the assumed power law; Section Limitations states the consequence.

The nominal SFOC of 194 g/kWh is the fixed fuel-conversion reference coefficient inherited from the baseline framework [6] for this engine type [43]. The available records do not identify the exact engine rating and mark, and no shop-test curve, derating record or fuel calorific analysis was available, so the value is not verified for the installed engine at the operating loads of Table 8. The recursive analysis therefore treats it as an assumed baseline coefficient rather than a measured engine map, and Section 5.2.2 carries a second, load-dependent baseline through the whole of Arm 1 for this reason.

The database comprises four Mediterranean voyages with different routes, metocean sequences and loading conditions. V1 and V4 were operated laden, V2 and V3 in ballast. Every value in Table 5 is taken from the vessel's voyage-operations register or from the sailing-leg performance page, not from a summary field; the procedure and its consequences are described in Section 4.2.

The final database contains 207 records. The V4 recursive sequence is restricted to the continuous Messina–Fiumicino operation; records associated with the adjacent Augusta operation are excluded. The single source duration corrected during the boundary audit belongs to row 7, which is itself excluded, so no corrected value enters the retained 22-record

set. The excluded rows and their evidence, including the field-by-field identity of rows 5 and 6, are listed in the traceability workbook and verified by `run_v4_leg_boundary_audit.py`.

Table 5. Operational database and record-selection audit.

Voyage	Analysed Leg	Selection Audit	<i>n</i>	Duration [h]	Distance [NM]	Loading	Operating Δ [t]
V1	Milazzo–Motril	START row removed; rows 2–90 retained	89	88.90	921.54	Laden	41,010
V2	Bar–Sidi Kerir	START row removed; 75 positive-duration rows retained	75	75.00	889.60	Ballast	24,400
V3	Livorno–Sarroch	Rows 1–21 retained	21	22.42	267.13	Ballast	24,400
V4	Messina–Fiumicino	Rows 1–4 (Messina–Augusta) removed; rows 5–6 removed as identical duplicate entries; row 7 removed as the operation-boundary record, its reported 0.5 h having been corrected to 1.5 h during the audit; rows 8–29 retained unaltered	22	25.50	272.00	Laden	41,135

Note: Source-row traceability is reported in Supplementary Tables S1 and S16–S19.

Economic and Regulatory Parameters

The economic and regulatory inputs are fixed before the static and κ -conditioned comparisons and are listed in Table 6. Fuel-price inputs range from 620 to 640 EUR/t. The 800 EUR/h time value and 750 EUR/h ETA-deviation coefficient are scenario assumptions rather than observed charter invoices. The model assigns full value to each hour of earlier arrival and does not include berth availability, waiting time, or an early-arrival cost. Because arrival time is imposed as a hard upper limit, the ETA-deviation coefficient is inactive for feasible solutions. Economic results are therefore conditional on these assumptions.

Table 6. Economic and regulatory parameters per voyage.

Voyage	ME/AE Fuel Price [EUR/t]	Time Value [EUR/h]	ETS Price [EUR/t CO ₂]	ETS Coverage	FuelEU Coverage
V1 Milazzo–Motril	640	800	70 ($\times 0.70$)	1.0	1.0
V2 Bar–Sidi Kerir	625	800	70 ($\times 0.70$)	0.0	0.0
V3 Livorno–Sarroch	630	800	70 ($\times 0.70$)	1.0	1.0
V4 Messina–Fiumicino	620	800	70 ($\times 0.70$)	1.0	1.0

The EU ETS calculation uses an allowance price of 70 EUR/tCO₂ and the 70% surrender requirement applicable to emissions reported for 2025. The geographic coverage factor (χ_{geo}) is 1.0 for V1, V3, and V4 and 0.0 for V2. Bar and Sidi Kerir both lie outside the Union, so no leg of V2 falls within the scope of either instrument. The FuelEU Maritime coverage factor follows the same voyage classification. The physical, fuel, emission, and regulatory equations are reproduced in Supplementary Section S6.

The FuelEU calculation uses a well-to-wake greenhouse-gas intensity of 90.63 gCO₂e/MJ for the adopted MGO/MDO configuration and the 2025 limit of 89.34 gCO₂e/MJ [5]. It is reported as a route-level modelled compliance-cost proxy. The quoted 90.63 gCO₂e/MJ is the complete well-to-wake value, of which about 76 gCO₂e/MJ is the tank-to-wake component alone. The 2025 upper limit of CII rating band C used in the diagnostic assessment is 7.984 g CO₂/(DWT·NM). It follows from the tanker reference line of [44], the 2025 reduction factor of [45] and the rating boundaries of [46]. CO₂ mass is obtained from fuel mass with a factor of 3.206 kg CO₂ per kg of fuel, which is the MDO/MGO value used for both engines [3,39]. Every CII value reported here is a route-level proxy expressed in g CO₂/(DWT·NM).

4.2. Provenance and Fuel-Channel Definition

The analysed records were obtained from the vessel's onboard energy-efficiency and performance-monitoring terminal and from the voyage-operations register. The available records did not identify the sensor type, manufacturer, calibration status, sampling rate, signal-processing chain, or whether each field was acquired automatically, calculated internally, or entered manually. The variables are therefore treated as vessel-reported operational channels rather than independently verified sensor measurements; their source and verification status are summarised in Supplementary Table S2a.

4.2.1. Reconstruction from the Voyage-Operations Register

The dataset was reconstructed from the voyage-operations register and the corresponding sailing-leg performance pages rather than from voyage summaries. This audit established the meaning of the cargo field, confirmed the loading conditions and departure drafts, defined the V4 operational boundary, and checked the sailed distances against the register.

The voyage-summary Cargo field records the quantity loaded at the destination and is therefore not used to define the cargo carried on the analysed leg. The voyage-operations register and the recorded departure drafts confirm V1 and V4 as laden and V2 and V3 as ballast, consistent with the voyage classification reported in [6]. For the present recursive analysis, V4 is restricted to the Messina–Fiumicino operation from 21/08 at 16:30 to 22/08 at 18:00, and source rows 1–7 are excluded, leaving 22 segments and 25.50 h: rows 1–4 belong to the preceding Messina–Augusta operation, rows 5 and 6 are identical duplicate entries, and row 7 is the operation-boundary record. The retained sequence therefore departs Augusta; it is the continuous sailing operation recorded under the Messina–Fiumicino voyage in the register and is tabulated as Augusta–Fiumicino in the supplementary material of [6]. Table 5 records each exclusion and its evidence. The analysed V4 distance, window and date span describe the retained continuous leg, whose register distance is 277 NM over 21–22 August. They are route-leg rather than route-level quantities and are not intended to reproduce voyage-level values reported elsewhere. The retained record counts of 89, 75, 21 and 22 therefore differ from route-level counts for the same four voyages, because route-level planning treats each route as a single unit whereas the present study admits only records that belong to one continuous sailing operation and carry a complete set of screened channels. On V1, the operational-window rule adopted here excludes the opening START row and retains the closing records of the passage, which is why the present count exceeds the route-level one. All record-level values were re-extracted from the voyage-operations register and the sailing-leg performance pages under the definitions of this section; record-level correspondence with route-level tabulations is therefore not expected in either direction. Recursive replay assimilates records sequentially and cannot span a change in sailing operation. The retained set comprises an opening part-hour entry of 0.50 h, 20 hourly entries, and one final entry of 5.00 h and 53.0 NM that the performance page logs as a single aggregated approach leg at 90 rpm and 3312 kW, consistent with the adjacent records; the final entry is retained because the passage closes against the register duration only when it is included. The retained voyage boundaries, durations and distances were checked against the voyage-operations register; the analysed distances agreed with the register distances of 920.1, 880, 278 and 277 NM to within 0.2%, 1.1%, 3.9% and 1.8% on V1–V4.

4.2.2. Displacement

The lightship follows from Table 4 as $50,029.65 - 40,439 = 9590.65$ t. The laden displacement estimates combine lightship, register cargo, and an assumed 1500 t consumables allowance, giving 41,010 t on V1 and 41,135 t on V4 at recorded drafts of 9.25 m and 9.35 m.

The ballast value of 24,400 t is an operating estimate associated with the recorded drafts and checked for internal consistency against the cubic relation in Section 5.3.5; it is not read from a departure-condition report. With the same consumables allowance, the implied ballast-water quantity is 13,309 t, or 32.9% of deadweight. The estimate is associated with the 6.10 m mean draft of V3 and is applied unchanged to V2, whose mean departure draft is 6.50 m.

The calm-water coefficient is calibrated at the service point and reference displacement of Table 4 and then held fixed. Voyage-specific operating-displacement estimates enter only through the two-thirds scaling. This defines the record-informed model evaluated here and is not used to re-estimate the integrated optimisation results reported in [6].

4.2.3. Fuel-Channel Definition

The source “Fuel flow” field is treated as total hourly fuel consumption. The reported daily auxiliary consumption of 2.0 t/day is converted to 83.33 kg/h and subtracted from every segment; main-engine segment mass then follows from segment duration. This constant allocation is an assumption, not a measured segment-level auxiliary load, and Supplementary Table S28 repeats the analysis at 1.5 and 2.5 t/day.

The source field admits a second reading, in which it is taken as main-engine consumption alone and no auxiliary rate is subtracted. The records themselves bear on the choice. The performance pages carry a separate main-engine consumption field (“ME FOC”), and on 202 of the 207 retained records that field agrees with “Fuel flow” minus the auxiliary rate to within 5 kg/h and on 200 of them to within about 2 kg/h, which is the rounding of the daily entries; the five exceptions are listed in Supplementary Section S1.2. The reporting system therefore treats “Fuel flow” as total consumption, which is the convention adopted here. This is evidence about the system’s own convention and not an independent verification, because the main-engine field may be computed by the same system from the same entries.

The alternative reading is nevertheless carried through the whole chain and not only through the screening. It changes reconstructed main-engine mass by roughly 10%. The screening result of Section 4.3 is unchanged: with no auxiliary subtraction the apparent specific consumptions are 217.4, 266.8, 283.7 and 245.7 g/kWh, a spread of 66.3 g/kWh against the 3.3 g/kWh attributable to the observed load range. Its effect on the frozen correction, the prospective prediction and the transfer test is reported in Sections 5.2.1 and 5.3.3 and in Supplementary Table S37.

Voyages are replayed sequentially. Static and κ -conditioned calculations use the same physical relations, constraints, and solvers. The sole difference in the decision test is the transferred κ factor.

4.2.4. Channel Definitions and Measurement Chain

κ relates two vessel-reported channels, neither of which is independently validated. Table 7 states, for these and for the speed channels, what the records establish about each quantity and its relation to other fields, and what remains undocumented. The records identify no instrument, so the table is built from the records alone.

Fuel. “Fuel flow” is an hourly rate in kg/h, with one value per record. Its relation to the main-engine and auxiliary fields is set out in Section 4.2.3; the daily total field equals 24 times the hourly rate, rounded to 0.1 t/day, on 203 of 207 records, so the two are one channel. The records do not identify the flowmeter, its principle, or whether a temperature or density correction is applied. No instrument record of the kind reported for the instrumented vessel of [21] is available.

Table 7. Channel definitions and measurement chain as established from the records.

Channel	Quantity and Unit	Relation to Other Fields Established from the Records	Not Documented
Fuel flow	Total fuel consumption rate, kg/h; one value per record	Agrees with ME FOC plus the auxiliary rate to within 5 kg/h on 202 of 207 records; the daily total field equals $24 \times$ the hourly rate, rounded to 0.1 t/day, on 203 records	Flowmeter type; temperature or density correction; averaging
ME FOC	Main-engine consumption, t/day	Equals the daily total minus the daily auxiliary entry on 202 records	Whether measured or computed by the reporting system
AE consumption	Auxiliary consumption, t/day; one generator	Constant entry of 2.0 t/day on all 207 records	Segment-level auxiliary load
ME power	Main-engine power, kW	Ratio to fuel index $\times$ shaft speed: 0.7205–0.7454 on four comparison legs, 0.8440 on V1 (Test E)	Shaft torsion meter or computed value; calibration
ME load	Main-engine load, %, integer values	Equals reported power divided by 9480 kW, rounded, on 199 of 207 records	Derived field; no independent information
Shaft speed	rpm, integer values	89–102 rpm; constant at 90 rpm on V4	Source of the signal
Speed through water	kn, 0.1 kn resolution	Not a function of shaft speed alone; falls with wave height at constant shaft speed on V1 and V4	Speed-log type; whether the field is its direct output
Speed over ground	kn, 0.1 kn resolution	Equals segment distance divided by duration on all 207 records	Position source
Longitudinal current	kn	Equals speed over ground minus speed through water exactly (Test A)	Which of the three fields is primary

Power. Reported main-engine power is given in kW at each record. The records do not state whether it is the output of a shaft torsion meter or a value computed by the monitoring system from engine-side signals. Test E of Section 4.3 does not resolve this: the product of fuel index and shaft speed is itself approximately proportional to power, so a near-constant ratio of reported power to that product is expected whichever way the power is obtained. Both possibilities are therefore retained. If the power is computed from engine-side signals, the power and fuel channels are not independent and κ relates two quantities with a common origin. If it is measured at the shaft, the two channels are independent but, as far as the records show, uncalibrated. The load field of the performance page equals reported power divided by 9480 kW, rounded to the nearest per cent, on 199 of 207 records; it is derived from the power channel, on a reference about 1% below the installed MCR of Table 4, and adds no independent information.

Speed through water. The field is logged at a resolution of 0.1 kn. It is not a function of shaft speed alone: on V4 the shaft speed is constant at 90 rpm while speed through water ranges from 9.7 to 10.7 kn, and its correlation with shaft speed on V1–V3 is 0.86, 0.12 and -0.02 . The records therefore do not support a simple derivation of speed through water from shaft speed alone. At constant shaft speed it falls as significant wave height rises on V1 ($r = -0.49$, -0.85 and -0.98 at 91, 92 and 93 rpm; 13, 44 and 23 records) and on V4 ($r = -0.89$, 22 records), as a speed responding to added resistance would; on V2 the association is positive ($r = 0.46$ to 0.84) and on V3 the field varies by only 0.7 kn. The records do not identify the speed log, and the provenance of this channel is not established. Because speed through water enters the modelled calm-water power to the third power, a 2% bias in it changes that power by about 6% and is carried into the modelled-power basis.

Speed over ground, distance and current. In the released dataset, speed over ground equals segment distance divided by segment duration on all 207 records, so these two fields are not independent of each other. The longitudinal-current field equals the difference between speed over ground and speed through water exactly (Test A), so the three fields carry two independent quantities and not three.

Whether logged values are averages over the record or instantaneous readings, and whether they are acquired automatically or entered manually, are not documented. These gaps are carried into the limitations of Section Limitations. Because neither channel that defines κ is independently validated, κ is used below only as a characterisation of the relationship between two reported quantities.

4.3. Five-Test Screening and Identifiability

Five checks precede calibration: exact dependence among the speed and current channels, apparent specific consumption, response to large power excursions, linear fuel–power consistency, and consistency of reported power with the engine fuel index. They identify information limitations and do not by themselves assign a sensor fault. The reported channels cannot be shown to meet the input conditions presumed by ISO 19030-2 [13], which requires an independent speed measurement, a power measurement traceable to shaft torque, and rejection of transient records; ISO 19030 normalisation is therefore not applied anywhere in this study, and the checks below and Section 4.2.4 establish why. Agreement of route-integrated fuel totals with the acceptance levels quoted for ISO 19030 is a tolerance on an aggregate quantity, not the set of input conditions the standard imposes for hull- and propeller-performance normalisation; Section 5.1 shows that route-level totals can close while the underlying segment-level relationships remain non-identifiable. No ISO 19030 conformity claim is made here in either direction. Table 8 gives the voyage-level results.

Test A. On all four voyages the reported longitudinal current reproduces the difference between speed over ground and speed through water exactly, to machine precision. The

three reported quantities therefore contain only two independent numerical degrees of freedom and must not be counted as three independent information channels. This algebraic identity alone does not establish which field is primary and which is derived. The longitudinal-current field is consequently retained only as a kinematic descriptor and is not used as an additional independent predictor. The test says nothing about the provenance or measurement independence of speed through water itself, which is addressed separately in Section 4.2.4. Mean differences are -1.229 , -0.071 , -0.071 and $+0.364$ kn on V1–V4.

Table 8. Screening of the reported measurement channels.

Voyage	Loading	Apparent SFOC [g/kWh]	Mean Load [% MCR]	Shaft Speed [rpm]	CV [%]	Fuel–Power Slope [g/kWh]	R ²	Max ΔP /Matching Δ Fuel [%]	ME Fuel per Mile [kg/NM]	Current Identity
V1	Laden	193.6	36.6	90–93	0.98	118.5	0.316	$-29.0/+7.9$	65.5	exact
V2	Ballast	242.0	35.0	90–102	2.46	148.9	0.716	$+51.3/+35.6$	68.5	exact
V3	Ballast	256.7	32.3	89–91	0.49	-299.6	0.129	$+2.3/-12.1$	66.6	exact
V4	Laden	220.0	33.9	90	0.00	728.8	0.767	$+1.1/+3.3$	67.0	exact

Note: the current identity concerns the longitudinal-current field only (Test A); it does not characterise the provenance of speed through water.

Test B. Apparent specific consumption ranges from 193.6 to 256.7 g/kWh at mean loads of 32.3–36.6% MCR and shaft speeds of 89–102 rpm. Part-load operation raises specific consumption above the project-guide reference on all four voyages, but it cannot produce this dispersion: over the observed 4.3-percentage-point span of mean load, a representative two-stroke part-load correction accounts for 3.3 g/kWh of the observed 63.1 g/kWh spread. Section 5.2.2 carries both baselines through Arm 1. The remaining dispersion is evidence of voyage-dependent reporting or unrecorded operating effects, not proof of an engine or sensor fault.

Test C. The channels disagree in sign on V1 and V3. On V4, using the retained Messina–Fiumicino records, the largest consecutive power change is +1.1%, with a matching fuel change of +3.3%.

Test D. Voyage-level fuel-on-power slopes are 118.5, 148.9, -299.6 , and 728.8 g/kWh. Moving-block 95% intervals are [43.6, 209.6], [111.3, 211.8], $[-718.1, 259.9]$, and [443.4, 963.8] g/kWh. The V3 interval crosses zero, while the V4 slope is far above the nominal static relation. These regressions are channel diagnostics with omitted variables and limited excitation; they are not engine maps.

Test E. Four comparison sailing legs give ratios of reported main-engine power to the product of fuel index and shaft speed between 0.7205 and 0.7454, whereas V1 gives 0.8440. On V1 records containing fuel index, the cross-leg relationship implies 2796–3329 kW against reported values of 3421–3731 kW. The separation supports a V1-specific inconsistency in the reported-power relationship. The test cannot say how the power channel is obtained: the product of fuel index and shaft speed is itself approximately proportional to power, so a near-constant ratio is expected whether the power is measured at the shaft or computed from engine-side signals, and the fields may share upstream calculations. The comparison is therefore not an independent sensor validation. It is used only as corroborating channel evidence; it neither identifies which field is incorrect nor quantifies an effect on the earlier route-level study.

Reconstructed main-engine consumption per mile ranges from 65.5 to 68.5 kg/NM. Plausible voyage totals can coexist with inconsistent segment-level relationships because aggregation removes part of the timing and scaling information. An independent mass-flow meter and shaft-torque measurement would be required to resolve the discrepancy. Figure 3 plots the reconstructed fuel rate against reported power.

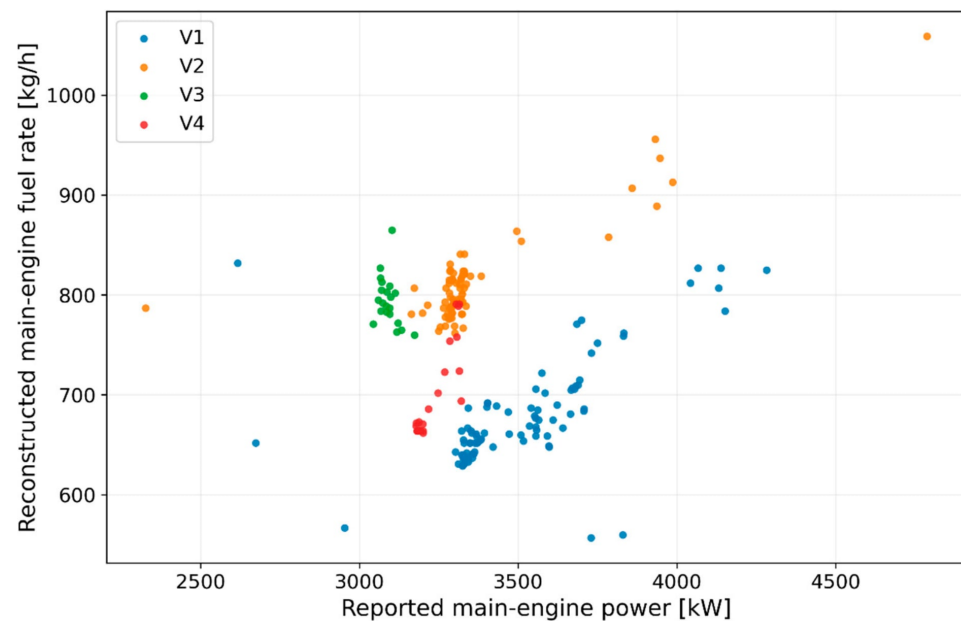


Figure 3. Reconstructed main-engine fuel rate against reported main-engine power. The figure is a channel-consistency diagnostic, not an engine-map validation.

The identifiability analysis has a structural part and a data-dependent part. Structurally, the stacked sensitivity matrix of the five-parameter candidate has at most rank two on any dataset (Section 3.2). Numerically, singular-value decomposition of the voyage-stacked matrices returns rank two on every voyage, before and after column normalisation, under the stated machine-precision tolerances; Supplementary Table S32 reports the ranks, tolerances and condition numbers. The bound is thus attained, and the fuel observation admits at most the two composites $\tau\alpha/\eta$ and $\tau\beta/\eta$, not five physical corrections. How well those two composites are separated depends on the data. Table 9 gives the excitation diagnostics by voyage.

Table 9. Voyage-level identifiability and excitation diagnostics. The structural rank is 2 of 5 on every voyage and for any data; all columns shown are data-dependent.

Voyage	Loading	Condition Number	Added-Power Share, Whole Voyage [%]	Added-Power Share, Adaptation Window [%]	Required Overall Scale [–]	Shaft-Speed Range [rpm]	Shaft-Speed CV [%]
V1	Laden	10.18	15.63	2.60	1.089	90–93	0.98
V2	Ballast	25.97	8.58	2.76	1.909	90–102	2.46
V3	Ballast	86.92	4.26	5.81	1.922	89–91	0.49
V4	Laden	21.95	12.73	7.78	1.783	90	0.00

Note: The diagnostics were evaluated using the voyage-specific displacements reported in Table 5. The condition number refers to the non-zero singular values of the stacked fuel-observation sensitivity matrix. Added-power shares are duration-weighted. The required overall scale is the multiplicative factor needed to reconcile the nominal physical model with the reconstructed voyage fuel mass.

Over whole voyages, added power represents 4.3% to 15.6% of total propulsion power, so the environmental direction is appreciably excited on V1 and V4 only. Within the adaptation windows the share is 2.6% to 7.8%, and on V1 it is lowest, because most of the weather on that voyage occurs after the window. The two ballast voyages require nearly identical overall scales of 1.909 and 1.922, and V4 requires 1.783; V1 requires 1.089, which is the same outlier identified by Test E.

RPM excitation is insufficient for an adaptive shaft-speed term. V4 has constant RPM; V1 and V3 have coefficients of variation below 1%; only V2 exceeds the prespecified

1% threshold. Power–RPM fits for constant or weakly varying RPM are reported as not estimable and are not used to validate the power channel.

4.4. Scoring, Statistics, Sensitivity and Software

The scalar κ model is imposed after the identifiability analysis. κ is the ratio between two vessel-reported quantities, the reconstructed main-engine fuel rate and reported main-engine power multiplied by a nominal specific consumption. It characterises that reporting relationship and is not assigned a physical efficiency interpretation.

Arm 1 estimates κ from reconstructed main-engine fuel and reported main-engine power. The decision layer applies the frozen value to fuel calculated from modelled power. This change in power basis is tested explicitly and is not assumed to be physically valid.

The estimator is scored one step ahead: segment (i) is predicted from the posterior available after segment (i – 1). The opening 25% of each voyage, with a minimum of five records, is reserved for adaptation. A second test freezes κ at the end of that window and scores only the remaining records, with no further update. Table 10 summarises the configuration; that of the kinematic comparator is given in Supplementary Table S6.

Table 10. Configuration of the scalar estimator (Arm 1).

Quantity	Arm 1
Estimated state	$\kappa = \tau/\eta$
Initial value	1.000
Initial covariance	0.05
Process noise per record	1×10^{-5}
Measurement variance	Adaptation-window residual variance: 278.3, 432.3, 1866.4, 28.7 (kg/h) ²
Comparator	Reported ME power $\times$ 194 g/kWh
Bounds	None
Adaptation window	First 25% of records, minimum 5
Record weighting	Equal per source record

The same frozen κ values are used in the decision experiment. This aligns the prediction and optimisation chronology: all observations used for calibration precede the planned route.

Arm 1 uses an initial value of unity, a prior variance of 0.05, and a process noise of 1×10^{-5} per record. The observation-noise variance is the sample variance of the static residual over the adaptation window, floored at 1.0. No state bounds are imposed. Each source record is treated as one update and one scored observation regardless of segment duration; the implications of the 0.5–5 h record durations are stated in Section Limitations. No scored residual contributes to R, and Supplementary Table S25 reports the response to tenfold changes in Q and R.

Tenfold variations in Q and R leave the qualitative one-step result unchanged. This tuning check does not address observation uncertainty, which is examined separately through the auxiliary-fuel allocation and transfer-scale tests.

The primary deterministic grid is 6.5–13.5 kn in 0.1 kn increments with 5 min cumulative-time resolution, which contains every recorded speed exactly. A 10.0–13.5 kn grid in 0.5 kn increments is solved separately as a discretisation check. The reinforcement-learning benchmark reported in the Supplementary Materials uses the same admissible speed set and physical constraints.

“Recorded-speed static” and “recorded-speed κ -conditioned” are model evaluations of the same remaining-route speeds; they are not measured fuel totals. Each optimiser is compared only with the recorded-speed evaluation under the same model.

Segment performance is reported with RMSE, MAPE, and signed bias. Moving-block bootstrap intervals use 10,000 within-voyage resamples [47]. They describe conditional un-

certainty within each recorded sequence and do not represent between-voyage or between-vessel generalisation [48]. The four voyages, not the 207 segments, are the experimental units, so the across-voyage direction is assessed by an exact one-sided sign test [49] rather than by a segment-level test.

Voyage-level solver metrics are total main- plus auxiliary-engine fuel, CO₂, total cost, route duration, a route-level Carbon Intensity Indicator proxy, cost change relative to the corresponding recorded-speed reference, and the number of segments on which the static and κ -conditioned profiles differ.

Robustness checks comprise: $\pm 10\%$ wave and wind perturbations; $\pm 5\%$ displacement; auxiliary consumption of 1.5, 2.0 and 2.5 t/day; the alternative reading of the source fuel field, with no auxiliary subtraction, and the source main-engine field used directly, both propagated through the frozen estimator and the transfer test; seven definitions of the adaptation window, by record fraction and by elapsed time, scored on a common tail; a two-composite fit on the modelled-power basis; five κ -transfer scales, from unity through the reported-power correction and the part-load-baseline correction to the window model-required scale, each scored against a matched recorded-speed reference; the adopted 0.1 kn decision grid against the 0.5 kn check grid; moving-block intervals for the fuel–power slopes; raw and column-normalised identifiability ranks; economic coefficients; filter tuning; a no-intercept scalar fitted only on the adaptation window and then frozen; and calm-water speed-exponent sensitivity at $n = 2.8, 3.0$ and 3.2 . The fuel-conversion baseline, the fuel-field reading and the adaptation window are varied one at a time; no combination of them is evaluated. Seed sensitivity of the reinforcement-learning benchmark is reported in the Supplementary Materials.

The verified environment is Python 3.11.9 with the pinned package versions listed in requirements.txt: NumPy 2.3.5, pandas 2.2.3, Matplotlib 3.10.8, openpyxl 3.1.5 and PyTorch 2.10.0+cpu. Deterministic reconstruction does not require PyTorch.

The master script regenerates the estimator, identifiability, robustness, deterministic solver, economic-sensitivity, and figure outputs from the processed 207-record dataset. Archived PPO outputs are used by default; retraining is optional. The analyses added in revision are generated by run_reviewer_analyses.py.

The current package passes 21/21 application checks and 383/383 published-value checks.

The package contains the processed CSV and traceability workbook, all analysis modules, archived PPO outputs, result files, figures, checksums, and verification scripts. It reproduces the reported offline calculations and is not onboard software.

5. Results

5.1. Screening and Identifiability

The database contains 207 records and closes against register duration and distance within the limits reported in Section 4.2.1. The channel audit distinguishes traceable source values, derived fields, and quantities that lack independent measurement.

The screening tests reveal algebraic dependence, weak excitation, and inconsistent segment-level fuel–power behaviour. They do not establish that four sensors failed, and they do not revise the earlier route-level calculations.

The fuel-only sensitivity matrix attains its structural bound of rank two on every voyage, before and after column scaling. Condition numbers of the two non-zero raw singular values are 10.18, 25.97, 86.92 and 21.95. The observation therefore admits two composite directions at most; whether the second is usable is examined in Section 5.2.5.

The full-voyage model-required scales are 1.089, 1.909, 1.922, and 1.783. They reconcile reconstructed fuel with power computed from speed and displacement over all records

of each voyage. By contrast, κ reconciles fuel with reported power. The two quantities measure different relationships and are not interchangeable. A third quantity appears in the transfer test of Section 5.3.3: the scale required by the modelled-power basis over the adaptation window alone, which is what would be knowable at the moment κ is frozen. Its values are 0.914, 2.144, 1.884 and 1.617 on V1–V4; they differ from the full-voyage scales because the adaptation window covers a different operating segment, and on V1 they fall on the opposite side of unity.

5.2. Arm 1: Fuel-Referenced Scalar Correction

5.2.1. One-Step Prediction Performance

Table 11 reports one-step RMSE, MAPE, bias, filtered diagnostics, and terminal κ .

Table 11. One-step Arm 1 prediction results.

Voyage	Loading	<i>n</i>	Adaptation Records	R [(kg/h) ²]	RMSE Static [kg/h]	RMSE One-Step [kg/h]	MAPE [%]	Bias [kg/h]	Δ RMSE One-Step [%]	Δ RMSE Filtered [%]	κ Terminal
V1	Laden	89	22	278.3	56.73	58.27	4.09	+2.74	−2.7	+6.8	1.018
V2	Ballast	75	18	432.3	163.15	39.30	2.57	+1.98	+75.9	+77.7	1.253
V3	Ballast	21	5	1866.4	199.96	21.55	2.05	+12.97	+89.2	+90.4	1.323
V4	Laden	22	5	28.7	88.35	29.54	2.79	+13.80	+66.6	+77.1	1.196
Mean									+57.2	+63.0	

Segment RMSE improves on three voyages, by 66.6% to 89.2%, and degrades by 2.7% on V1. The unweighted mean across the four voyages is 57.2% under one-step evaluation and 63.0% under filtered evaluation; the 5.8 percentage-point difference reflects evaluation order rather than a model change.

A record-duration sensitivity check gave the same result when process noise was scaled as $Q_i = Q_{t_i}$ (changes below 0.001 percentage points). Duration-weighted RMSE changes were −2.7%, +75.9%, +89.6%, and +71.0% on V1–V4, compared with −2.7%, +75.9%, +89.2%, and +66.6% under equal record weighting. The direction is unchanged; the largest difference is 4.4 percentage points on V4.

The prospective frozen-parameter test gives remaining-route RMSE changes of −0.5%, 75.9%, 83.9%, and 45.7% on V1–V4. Aggregate remaining-route errors after freezing are 0.62%, 0.70%, −3.50%, and −5.67%. These values are predictions made after the calibration window; they replace the earlier whole-voyage aggregate, which continued to update κ during the scored sequence.

Figure 4 instantiates the generic chronology of Figure 2 for the four retained voyage sequences. The adaptation windows contain 22, 18, 5, and 5 records for V1–V4, leaving 67, 57, 16, and 17 records for prospective evaluation; the corresponding frozen adaptation-window corrections are $\kappa_{ad} = 1.0046, 1.2542, 1.2858, \text{ and } 1.0804$.

Figure 5 reports the prospective remaining-route RMSE and aggregate errors after κ is frozen.

The auxiliary-fuel sensitivity confirms that κ depends on the subtraction used to construct the main-engine observation. Across 1.5–2.5 t/day, frozen κ spans 0.9728–1.0363 on V1, 1.2219–1.2865 on V2, 1.2522–1.3195 on V3, and 1.0467–1.1141 on V4. The prospective RMSE improvement is stable and large on V2–V3, but changes from −0.5% at the baseline to 3.7–7.6% on V1 and from 31.1% to 55.8% on V4. Supplementary Table S28 gives the complete results.

The alternative reading of the source fuel field (Section 4.2.3) has been propagated through the same analysis (Supplementary Table S37). Using the source main-engine field directly reproduces the adopted construction: the frozen κ agrees to within 0.002 on every voyage. With no auxiliary subtraction, the frozen κ rises by about 0.13, to 1.1316, 1.3834,

1.4204 and 1.2151 on V1–V4, as expected from a roughly constant additive change in the observation. The prospective remaining-route totals are essentially unchanged, at +1.10%, +0.76%, −3.50% and −4.85% against +0.62%, +0.70%, −3.50% and −5.67%, and the between-voyage dispersion of the frozen κ moves from 28.0% to 25.5%. The remaining-route RMSE change is not invariant: on V1 it moves from −0.5% to +39.2%, and on V4 from 45.7% to 72.4%. The level of κ and the per-voyage RMSE change are therefore reading-dependent, in the same way as they are baseline-dependent (Section 5.2.2), and are not used to support any conclusion. The prospective totals and the dispersion of κ are stable under both readings.

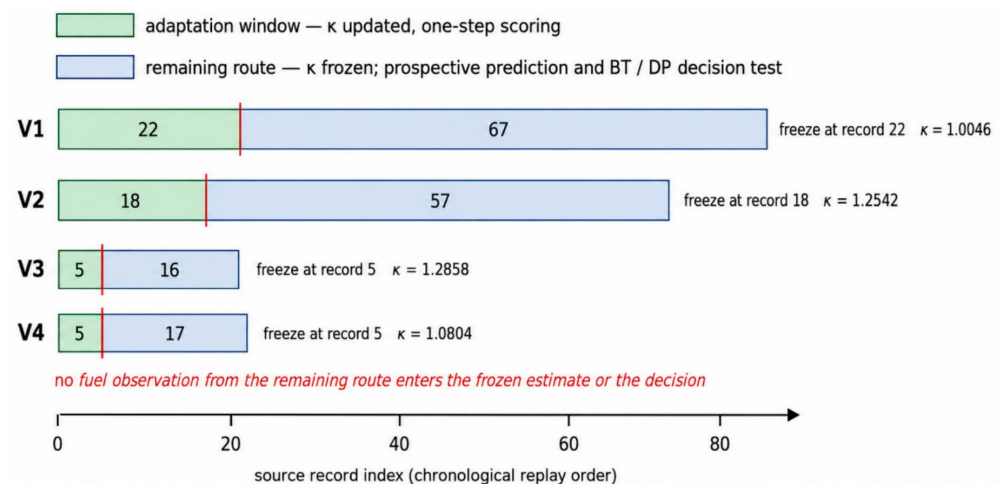


Figure 4. Voyage-specific implementation of the prospective adaptation protocol. Green segments denote the adaptation windows; blue segments denote the remaining-route records used for prospective prediction and decision testing. The displayed κ values are the frozen adaptation-window corrections.

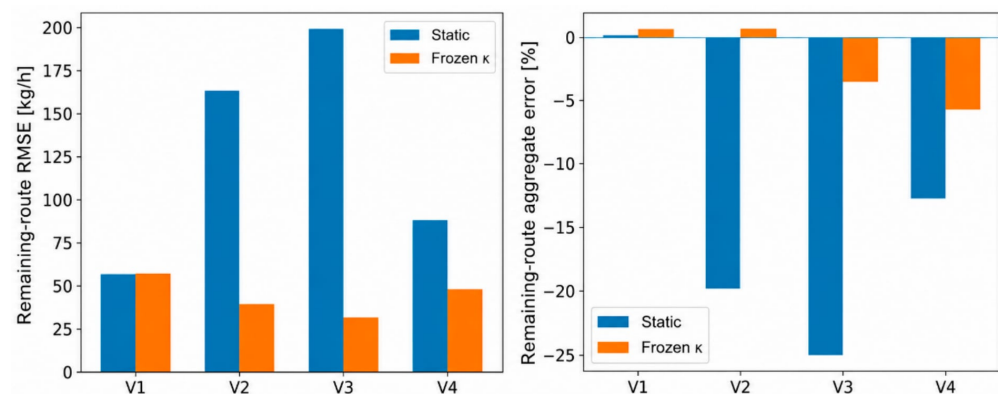


Figure 5. Prospective remaining-route performance after κ is frozen at the end of the adaptation window.

5.2.2. Identifiability of the Fuel-Conversion Baseline

The Arm 1 static baseline is the product of reported main-engine power and a nominal specific consumption. Table 8 shows that the four voyages are operated at 32.3% to 36.6% MCR, far from the reference point at which the project-guide value is quoted, so the choice of baseline is not neutral. Two baselines are therefore carried through the whole of Arm 1. Arm A uses the documented 194 g/kWh, constant with load. Arm B scales the same value with the load-dependent SFC correction used in the Third IMO Greenhouse Gas Study [50]; its quadratic factor has a minimum at 78.0% MCR and gives an effective 209.6 to 213.1 g/kWh on these voyages. Neither is verified for the installed engine: no shop-test curve, derating record or fuel calorific analysis was available. κ is therefore identified only up to the assumed baseline. This is a further identifiability limit of the same kind as the

rank deficiency of Section 4.3, not a defect of the estimator. Table 12 reports Arm 1 under both baselines.

Table 12. Arm 1 results for the two fuel-conversion baselines.

Quantity	V1	V2	V3	V4	Mean
Mean load [% MCR]	36.6	35.0	32.3	33.9	34.5
Effective baseline SFOC, Arm A [g/kWh]	194.0	194.0	194.0	194.0	194.0
Effective baseline SFOC, Arm B [g/kWh]	209.6	210.8	213.1	211.8	211.3
Terminal κ , Arm A [–]	1.018	1.253	1.323	1.196	–
Terminal κ , Arm B [–]	0.954	1.157	1.205	1.101	–
One-step RMSE change, Arm A [%]	–2.7	+75.9	+89.2	+66.6	+57.2
One-step RMSE change, Arm B [%]	+28.8	+69.1	+85.6	+34.6	+54.5
Remaining-route total error, static, Arm A [%]	+0.16	–19.71	–24.95	–12.69	14.38
Remaining-route total error, static, Arm B [%]	+8.04	–12.81	–17.54	–4.74	10.78
Remaining-route total error, frozen κ , Arm A [%]	+0.62	+0.70	–3.50	–5.67	2.62
Remaining-route total error, frozen κ , Arm B [%]	+0.05	+0.49	–3.12	–5.93	2.40

Note: Arm A uses the documented project-guide SFOC value, assumed constant with engine load. Arm B applies the IMO load-dependent correction defined in Supplementary Section S14.6. The Mean column reports the unweighted mean across the four voyages; for the remaining-route total-error rows, it reports the mean absolute error.

Between about 8 and 11 percentage points of the correction estimated over the adaptation window are attributable to part-load operation rather than to reporting. Under Arm B the terminal κ falls below unity on V1, which strengthens rather than weakens the reported-power finding of Section 4.3 for that voyage. Three results are invariant to the baseline. The between-voyage dispersion of κ , defined as $(\kappa_{max} - \kappa_{min})/\kappa_{min}$ across the four terminal values, changes only from 29.9% to 26.3%, because the mean loads span 4.3 percentage points of MCR, over which the part-load correction varies by 1.6%. The prospective remaining-route prediction is effectively unchanged, at 2.62% against 2.40% mean absolute error, whereas the static baseline error moves from 14.38% to 10.78%; a common multiplicative baseline error is absorbed by the scalar correction whatever its origin. The apparent specific consumptions of Table 8 span 63.1 g/kWh, while the observed load range explains only 3.3 g/kWh, so the dispersion identified by Test B is not a part-load artefact. The per-voyage one-step RMSE change is not invariant: it changes sign on V1 and approximately halves on V4. Per-voyage improvement figures are therefore reported as observations under a stated baseline and are not used to support any conclusion.

5.2.3. Residual Dependence and Interval Estimates

Table 13 reports residual dependence and moving-block bootstrap intervals.

Table 13. Residual dependence and moving-block bootstrap intervals for Arm 1.

Voyage	Lag-1 Residual Autocorrelation	Block Length	RMSE Change [%]	95% Interval [%]
V1	0.372	5	–2.7	[–5.5, 6.4]
V2	0.205	4	+75.9	[71.6, 89.3]
V3	0.397	3	+89.2	[84.8, 93.8]
V4	0.349	3	+66.6	[59.6, 75.5]

Three conditional within-voyage intervals exclude zero, and V1 spans zero. The exact one-sided sign test over four voyages gives $p = 0.3125$. This is descriptive case-study evidence; it does not support population-level significance.

5.2.4. State Trajectory, Scalar Baselines and the Role of the Recursion

Figure 6 shows the sequential trajectory of the composite correction κ .

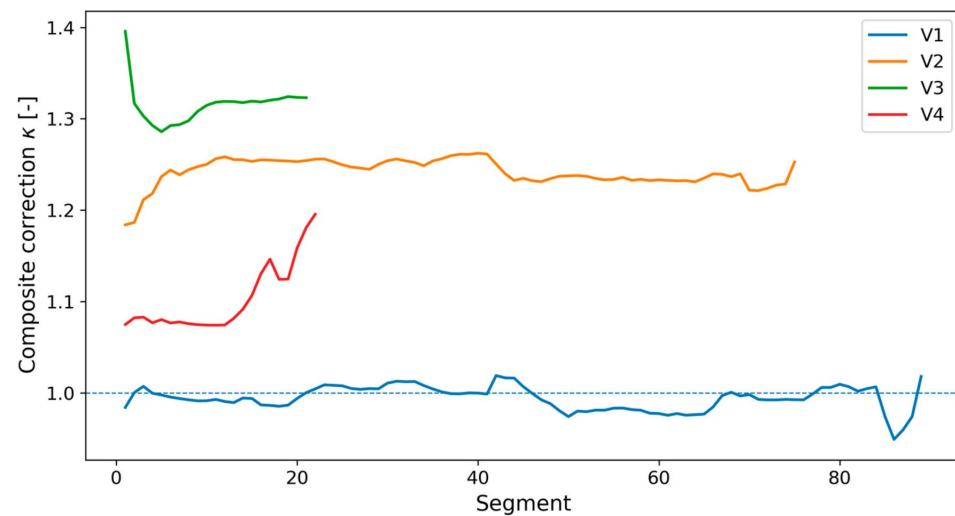


Figure 6. Sequential trajectory of the composite fuel correction κ . The state is a channel-level correction and does not separate SFOC from propulsive efficiency. The dashed line marks $\kappa = 1$, the uncorrected nominal fuel-conversion baseline.

The terminal κ values correspond to the voyage-specific apparent fuel-to-reported-power ratios. The estimator therefore tracks a reporting relationship. It does not separate SFOC, propulsive efficiency, resistance, or degradation.

Table 14 compares the scalar Kalman filter with recursive least squares (RLS) and an exponentially weighted moving average (EWMA).

Table 14. Segment-level RMSE change for the scalar Kalman filter and two baselines carrying no covariance machinery.

Voyage	Scalar Kalman Filter	RLS	EWMA
V1	−2.7	−0.6	−4.4
V2	+75.9	+76.1	+76.0
V3	+89.2	+89.8	+80.1
V4	+66.6	+53.3	+61.7
Mean	+57.2	+54.7	+53.4

RLS gives nearly the same result on three voyages and is less accurate on V4. In this one-state problem the gain follows from scaling and not from a mechanism specific to the Kalman filter. A no-intercept scalar fitted once on the adaptation window and then frozen gives remaining-route RMSE changes of +0.1%, +76.0%, +85.4% and +45.7% on V1–V4, compared with −0.5%, +75.9%, +83.9% and +45.7% for the frozen Kalman estimate (Supplementary Table S33). The prospective gain is a voyage-specific scaling effect, and the scalar Kalman filter is to be read as one implementation of that scaling, not as a contribution of this study. It is retained in the reported results for two practical reasons that predictive accuracy does not measure: it returns a posterior variance, and it respects the causal ordering of the records, so it can be run without knowing in advance where the adaptation window ends.

5.2.5. Two-Composite-Parameter Formulation

The structural analysis leaves two composites, $\theta_1 = \tau\alpha/\eta$ and $\theta_2 = \tau\beta/\eta$, on the modelled-power basis (Section 3.2). To test whether the second is usable, both were fitted by least squares on the adaptation window of each voyage, frozen, and scored on the remaining route in the same way as κ .

Within the adaptation windows the added-power term contributes between 0.9% and 8.4% of modelled power in any single record, and 2.6% to 7.8% over a window as a whole. The two estimates are correlated at -0.86 to -0.99 . The added-power composite θ_2 takes the values 5.38, -1.37 , -18.59 and -0.42 on V1–V4. It is negative on three voyages, which is inadmissible for an added-resistance multiplier, and on V2 and V4 its standard error exceeds its magnitude. Using all records of a voyage does not repair this: θ_2 is then 2.47 and 3.77 on the laden voyages V1 and V4 and -1.45 and -2.02 on the ballast voyages V2 and V3. Its ordinary-least-squares standard error is below its magnitude on every voyage, yet on V1, V3 and V4 its value differs from the window estimate of the same voyage, and its sign follows the loading condition. This is the behaviour of a parameter that absorbs discrepancy of the modelled-power basis [36,37], not that of an identified added-resistance multiplier.

Prospectively, the frozen two-composite model predicts remaining-route fuel totals with errors of +50.9%, +1.1%, +19.7% and -18.5% , and with segment RMSE of 414, 101, 182 and 156 kg/h. A single scale on the same modelled-power basis gives 222, 186, 41 and 95 kg/h, and κ on the reported-power basis gives 57, 39, 32 and 48 kg/h. The two-composite model is therefore worse than a single modelled-power scale on V1, V3 and V4, improves on it only on V2 and then with a negative (θ_2), and is far less accurate than κ on every voyage. Supplementary Tables S35 and S36 give the complete results. The standard errors are ordinary-least-squares values and understate the uncertainty, because the residuals are serially correlated.

The second composite is thus structurally identifiable, but no admissible and stable value of it can be estimated from these records; in that sense it is not practically identifiable here, and one scale is all the fuel observation supports. The comparison also shows that a single scale on the modelled-power basis is a much weaker predictor than κ on the reported-power basis, which anticipates the transfer result of Section 5.3.3.

5.2.6. Sensitivity to the Freeze Point

The adaptation window of Section 3.3 is a fraction of records. Its influence was examined with seven window definitions: the first 10%, 20%, 25%, 33% and 50% of records, and the first 25% and 33% of voyage duration, all subject to the five-record minimum. On V1 and V2, where every record is hourly except the first of V1, the two duration windows select the same records as the 25% and 33% record windows; on V3 and V4 they differ from them by at most one record. The seven definitions thus give five distinct windows on V1, V2 and V4 and four on V3. Because the remaining route changes with the window, each frozen κ was scored on a common tail, the last 50% of the records of each voyage, and the dynamic-programming problem was re-solved on that tail for each window. Supplementary Table S38 gives the complete results.

The frozen correction is stable. Across all windows, κ spans 0.985–1.013 on V1, 1.238–1.260 on V2, 1.286–1.316 on V3 and 1.074–1.080 on V4, at most 3% of its value, while the between-voyage dispersion remains at 28–31%. The common-tail total error stays within -0.8% to $+2.0\%$ on V1, -0.2% to $+1.6\%$ on V2 and -3.0% to -0.8% on V3, against $+0.7\%$, -19.3% and -24.6% uncorrected. On V4, it is -7.9% to -8.5% , against -14.8% uncorrected. The RMSE change stays near zero on V1 and at 70–72%, 86–93% and 40–43% on V2–V4. When scored instead on each window's own remaining route, the total error lies within -1.5% to $+2.1\%$, -0.8% to $+1.6\%$, -3.5% to -0.9% and -8.5% to -5.7% on V1–V4.

Two limitations follow. On V3 and V4 the five-record minimum binds for every record window up to 25%, so the window cannot be shortened on those voyages. On V4 an underprediction of about 8% remains on the common tail whatever the window, because the fuel-to-reported-power ratio rises late in the passage (Figure 6) and a frozen scalar cannot follow it.

5.3. Transfer and Decision Sensitivity

5.3.1. Configuration

The reporting-channel correction is estimated over the adaptation window and frozen at 1.0046, 1.2542, 1.2858, and 1.0804 on V1–V4. The remaining-route problems contain 67, 57, 16, and 17 segments. No observation from the remaining route enters the decision calculation.

5.3.2. Matched-Reference Solver Results

Table 15 summarises the matched-reference remaining-route results. Supplementary Table S12a–d give fuel, CO₂, cost, time and the CII proxy for all configurations.

Table 15. Summary of the remaining-route dynamic-programming results.

Voyage	Model	Recorded-Speed Cost [EUR]	Recorded-Speed Fuel [t]	Δ Cost, DP [%]	Δ Fuel, DP [%]
V1	Static	91,470	45.48	+2.21	−8.76
	κ -cond.	91,623	45.66	+2.19	−8.60
V2	Static	64,283	29.89	+2.65	−17.99
	κ -cond.	68,277	36.28	+1.15	−16.56
V3	Static	21,033	8.63	+0.78	−14.53
	κ -cond.	22,720	10.68	+0.02	−0.74
V4	Static	25,019	10.11	+1.24	−17.79
	κ -cond.	25,566	10.79	+0.77	−13.51

Note: Each configuration is evaluated against the recorded-speed reference calculated with the same model. Positive Δ values indicate reductions relative to that reference. Beam search attained the dynamic-programming optimum in every evaluated case, assessed on achieved fuel, duration and cost because the 0.1 kn grid admits ties, and is therefore omitted. Complete results for fuel consumption, CO₂ emissions, cost, voyage time, the CII proxy and pairwise comparisons are reported in Supplementary Sections S7 and S12.

The beam search reaches the dynamic-programming optimum in all eight primary runs; on the 0.1 kn grid the two solvers occasionally return tied profiles that exchange the speeds of two segments at identical modelled fuel and duration, so agreement is assessed on achieved fuel, duration and cost and not on the profile. The reinforcement-learning benchmark, its seed variation, its one infeasible rollout, and the basis on which its ranking can and cannot be compared with that reported in [6] are given in Supplementary Section S15.2.

At 800 EUR/h, total-cost minimisation reduces modelled cost by 0.78–2.65% under the static model and 0.02–2.19% under the κ -conditioned model, while increasing modelled fuel by 8.76–17.99% and 0.74–16.56%, respectively. The objective assigns full value to early arrival and includes no berth-waiting or early-arrival cost. The result is conditional on that scenario.

5.3.3. Pairwise Static- κ -Conditioned Profile Comparison

Supplementary Table S13 reports the direct static- κ -conditioned solver differences, including the number of changed segments, cost and fuel differences, and the speed sets selected.

Figure 7 shows the dynamic-programming static and κ -conditioned speed profiles. The reported power κ changes the profile on every voyage: 10 of 67 segments on V1, 38 of 57 on V2, 16 of 16 on V3, and 17 of 17 on V4. This establishes decision sensitivity to the transferred factor; it does not establish that either counterfactual profile would have been cheaper onboard.

The transfer-scale test is decisive, and it must be read on matched references: absolute costs computed under different applied scales are not comparable, because the scale multiplies modelled fuel directly. Each case is therefore scored against the recorded-speed

profile evaluated under the same applied scale. Moving from the reported-power κ to the scale required by the modelled-power basis over the adaptation window changes the modelled cost reduction from +2.19% to +2.73% on V1, from +1.15% to +0.06% on V2, from +0.02% to 0.00% on V3, and from +0.77% to +0.02% on V4. The accompanying modelled fuel penalty disappears on three voyages: the 16.56% increase on V2 becomes a 0.13% reduction, the 0.74% increase on V3 falls to 0.05%, and the 13.51% increase on V4 becomes a 0.02% reduction. On V1 the increase grows from 8.60% to 12.65%. Adopting the part-load fuel-conversion baseline of Section 5.2.2 instead of the documented one moves the result by a comparable amount and alters every segment of the profile on V1. The optimisation outcome is therefore not robust either to the assumption that κ transfers unchanged between reported and modelled power, or to the assumed fuel-conversion baseline. Supplementary Table S29 reports all five scale cases.

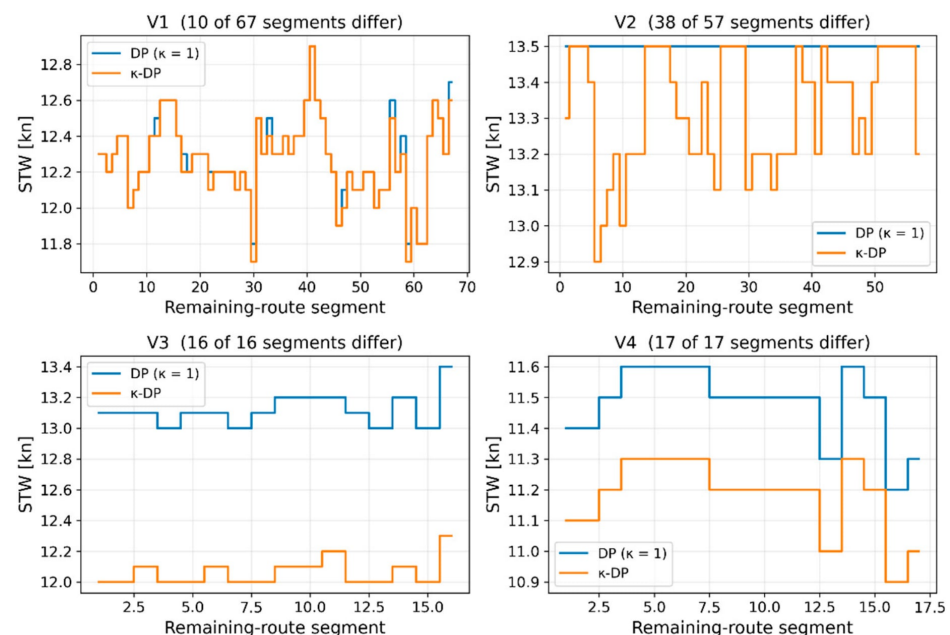


Figure 7. Static and κ -conditioned dynamic-programming speed profiles on the remaining route. The beam search coincides with dynamic programming and is omitted.

The transfer test was repeated under the alternative reading of the source fuel field (Supplementary Table S37). With no auxiliary subtraction, moving from the reported-power κ to the window model-required scale changes the matched cost reduction from +1.74% to +2.07% on V1, from +0.62% to +0.06% on V2, from 0.00% to −0.01% on V3 and from +0.28% to +0.02% on V4. The transfer conclusion therefore holds under both readings. The discrete profile does not: the κ -conditioned optimum under the alternative reading differs from that under the adopted reading on 67 of 67, 55 of 57, 12 of 16 and 17 of 17 segments.

The freeze point has a similar effect (Supplementary Table S38). On the common tail of Section 5.2.6, the matched cost reduction obtained with the transferred κ is 3.3–3.5%, 1.15–1.27%, 0.02–0.03% and 0.96–0.99% on V1–V4 across the seven windows, while the optimised profile differs from that of the 25% window on up to 34 of 44 segments on V1, 20 of 37 on V2 and all 10 on V3. These values refer to the common tail and are not comparable with those of Table 15, which refer to the route remaining after the 25% window.

5.3.4. Operational-Input Sensitivity

Supplementary Tables S14 and S15 report optimal-cost sensitivity to wave, wind and displacement perturbations and the number of speed decisions changed, evaluated in 48 exact re-optimisations.

A $\pm 5\%$ displacement perturbation changes optimised cost by 0.94–1.14% and alters speed decisions on every voyage under at least one model. Wave effects are largest on V1, at up to 1.34%, and fall below 0.02% on V3; wind effects remain below 0.41%. Displacement is the only tested operational input that affects every voyage. On the finer decision grid the counts of altered decisions are substantially larger than on a 0.5 kn grid, because the optimum can move where the coarse grid held it fixed.

The decision grid runs from 6.5 to 13.5 kn in 0.1 kn steps. Every recorded speed lies on that grid exactly, so the recorded profile is itself a feasible member of the decision set and the recorded-versus-optimised comparison is like for like. On a 10.0–13.5 kn grid at 0.5 kn, by contrast, 144 of the 207 recorded speeds are not representable: 127 fall between grid points and 17 lie below the lower bound, including six of the 22 V4 records, whose recorded speeds span 9.70–10.70 kn. Refining that grid from 0.50 to 0.25 kn does not repair the coverage, because the recorded speeds are logged on a 0.1 kn scale. Measured against the 0.5 kn check grid, the adopted grid changes optimal cost by at most 0.018% and modelled fuel by at most 0.87%. Supplementary Table S30 gives the results.

5.3.5. Operational Cross-Check of the Calm-Water Scaling

Table 16 provides an internal check of the cubic calm-water scaling using mean reported power and speed for one ballast and one laden leg. Because the calculation uses voyage means and an estimated ballast displacement, it is approximate and is not an independent resistance or displacement validation.

Table 16. Internal operational cross-check of the two-thirds displacement scaling for two recorded loading conditions.

Quantity	Ballast Leg, 6.10 m	Laden Leg, 9.35 m
Mean speed through water [kn]	11.583	10.387
Mean reported main-engine power [kW]	3169.5	3232.9
Displacement ratio implied by cubic relation	0.5944	—
Adopted displacement ratio, 24,400/41,135	0.5932	—
Relative difference	0.2%	—

Note: the cubic-law ratio is calculated from voyage-mean power and speed. Owing to nonlinearity, it is not equivalent to a segment-level ratio. The comparison tests internal consistency only.

The cubic relation gives a displacement ratio of 0.5944, while the adopted operating displacements give 0.5932, a relative difference of 0.2%. Because the ballast displacement is itself an estimate, this agreement supports internal consistency of the implemented scaling but does not validate the absolute resistance level or the displacement values.

Applying the V3-based ballast estimate to V2 remains a model-input approximation. The displacement sensitivity quantifies its effect on cost and decisions; a voyage-specific departure condition is needed to remove it. A separate exponent sensitivity anchors $P_{calm} \propto V^n$ at the same 14.5 kn/5780 kW calibration point for $n = 2.8, 3.0$, and 3.2 . Matched DP cost reductions remain positive in every tested case, but the discrete speed profiles can change substantially: relative to $n = 3.0$, 12–67 remaining-route decisions change in several voyage/model combinations, while static V2 is unchanged. The economic direction is therefore more stable than the segment-level profile; complete results are reported in Supplementary Table S34. Read together, the power-basis transfer test, the auxiliary-allocation and fuel-reading sensitivities, the freeze-point sensitivity and this exponent sensitivity converge on one conclusion: the segment-level speed recommendation is not supported by these data under any assumption tested. The modelled cost reduction is non-negative under all of them, to within 0.01%, as it must be on a grid that contains the recorded profile (Section 5.4); only its size changes.

5.4. Objective-Function Comparison

Table 17 separates the economic effect of early arrival from fuel-directed operation. Under the total-cost objective, early arrival is valued at 800 EUR/h and no waiting cost is applied. Under the fuel-and-compliance objective, the fuel change ranges from a 3.09% reduction on V1 to a 0.05% increase on V3, with the arrival constraint binding on all four voyages.

Table 17. Remaining-route comparison of the two optimisation objectives.

Voyage	Profile	Fuel [t]	Δ Fuel [%]	Total Cost [€]	Δ Cost [%]	Time [h]
V1	Recorded-speed	45.48	—	91,470	—	67.00
	Minimum total cost, Equation (9)	49.46	−8.76	89,447	+2.21	60.33
	Minimum fuel and compliance cost, Equation (8)	44.07	+3.09	90,299	+1.28	67.00
V2	Recorded-speed	29.89	—	64,283	—	57.00
	Minimum total cost, Equation (9)	35.27	−17.99	62,580	+2.65	50.67
	Minimum fuel and compliance cost, Equation (8)	29.86	+0.12	64,260	+0.04	57.00
V3	Recorded-speed	8.63	—	21,033	—	17.42
	Minimum total cost, Equation (9)	9.88	−14.53	20,868	+0.78	15.92
	Minimum fuel and compliance cost, Equation (8)	8.63	−0.05	21,031	+0.01	17.41
V4	Recorded-speed	10.11	—	25,019	—	21.00
	Minimum total cost, Equation (9)	11.91	−17.79	24,710	+1.24	18.78
	Minimum fuel and compliance cost, Equation (8)	10.11	+0.02	25,013	+0.03	20.99

Note: All profiles for a given voyage are evaluated under the same static model and total-cost accounting. Positive Δ values indicate reductions relative to the recorded-speed profile, whereas negative values indicate increases.

Only V1 shows a material fuel reduction. On V2–V4 the fuel-directed optimum differs from the recorded profile by no more than 0.12%, which means that on those three routes the recorded speeds were already at or very near the fuel-optimal profile achievable under the recorded arrival time.

Equation (8) is evaluated with dynamic programming and cross-checked with beam search under both static and κ -conditioned models. Agreement is assessed on achieved fuel, duration and cost under common accounting rather than on the profile itself, because the 0.1 kn grid admits ties. The beam search reaches the dynamic-programming optimum in every case. On V1 the two profiles differ on two segments under both models while returning identical modelled fuel to six decimal places and identical duration, which is a tie and not a disagreement; on V2–V4 the profiles are identical. The κ -conditioned fuel-directed optimum coincides with the static one on all four voyages.

On the adopted grid the recorded profile is representable exactly, so the fuel-directed optimum can be no worse than the recorded profile except through the 5 min cumulative-time discretisation. That residual is visible only on V3, where the optimum uses 0.05% more fuel because the recorded 17.42 h cannot be reproduced exactly on the time grid.

5.5. Evidence Supported by Each Component

A second, kinematic observation arm was run alongside Arm 1 and is reported in Supplementary Section S15.1. It estimates an additive speed-over-ground bias against a leave-one-voyage-out comparator and reduces one-step speed-over-ground RMSE by 55.3–77.7% on the four voyages. It absorbs residual current and comparator error, carries no propulsion information, and is not propagated into the fuel, power or decision model. It is mentioned here only because it shows that a recursive correction can improve prediction markedly while saying nothing about vessel condition.

Table 18 distinguishes the conclusions supported by each evaluation component from interpretations not supported by the available measurements or experimental design.

Beam search reproduces the dynamic-programming solution on the adopted grid. The number of static- κ profile changes depends on correction magnitude, time margin,

segment count and proximity to grid thresholds; it is not a monotone measure of physical importance. No counterfactual route was sailed, so the solver comparison demonstrates numerical and decision sensitivity only.

Table 18. Evidence supported by each evaluation component.

Component	Supported Conclusion	Conclusion Not Supported
Provenance reconstruction	Primary records confirm the loading conditions and provide the departure drafts and voyage boundaries used in the present analysis; the resulting loading classification is consistent with [6].	That [6] used incorrect loading conditions, or that its route-level results are corrected in the present study.
Speed-channel provenance	The longitudinal current and the two speed fields are algebraically linked and carry two independent quantities, not three. Speed through water is not a function of shaft speed alone.	That speed through water is, or is not, the direct output of an independent speed log.
Fuel-channel screening	The reconstructed fuel channel is not uniformly consistent with reported power at segment resolution, under either reading of the source fuel field.	A specific sensor fault or physical degradation mechanism.
Fuel-index test	The V1 ratio of reported power to the product of fuel index and shaft speed is about 15% above the range observed across four comparison sailing legs.	A confirmed sensor fault, the origin of the power channel (measured or computed), or a quantified error in the integrated results reported in [6].
Identifiability analysis	Three parameter combinations are structurally unidentifiable. The environmental composite is not practically identifiable on these records.	That the fuel observation model is correct, or that the same practical limit holds for data with stronger environmental excitation.
Arm 1 scalar correction	A one-dimensional scaling predicts remaining-route fuel totals prospectively; the totals stay within -8.5% to $+2.1\%$ across the two fuel-conversion baselines, the two fuel-field readings and the seven freeze-point definitions.	Independent estimates of hull, propeller, engine or auxiliary condition; per-voyage RMSE improvements, which depend on the baseline and the reading.
Calm-water scaling cross-check	The adopted displacement ratio is internally consistent with the cubic power–speed relation for two recorded loading conditions to within 0.2%.	Independent validation of displacement or the absolute resistance level.
BT–DP comparison	The feasibility-aware beam retains the dynamic-programming solution on these four problems.	A general proof that the beam search is globally optimal.
Post-adaptation comparison	A frozen reporting-channel correction can change discrete speed choices; the modelled cost reduction stays non-negative under the tests applied, to within 0.01%.	Realised onboard fuel or emission savings; the segment-level speed recommendation.
Reconciliation with [6]	The record selection, the main-engine fuel convention, the operating-displacement model and the V2 regulatory coverage differ from [6]; each difference is stated in the Economic and Regulatory Parameters section and Sections 4.2 and 4.3.	That the route-level results of [6] are re-estimated, superseded or corrected here; the two studies are not evaluated on a common record set.
Supplementary components (Section S15)	A recursive offset improves speed-over-ground prediction on all voyages. The fixed-budget reinforcement-learning policy reaches the dynamic-programming cost in one of 39 feasible runs.	Propulsion-condition tracking or independent current estimation; a general ranking of reinforcement learning against mathematical programming, or a contradiction of the separate fuel and time reductions reported in [6].

6. Discussion and Limitations

The evidence has three levels. The released package establishes numerical reproducibility from the processed dataset. The provenance audit establishes how those processed values relate to available vessel records. Neither establishes independent physical validation of power, fuel, or displacement. The internal cross-check of Section 5.3.5 is consistent with that boundary: the cubic power–speed relation reproduces the adopted two-thirds displacement ratio to within 0.2%, which shows the implemented scaling to be self-consistent but, because the ballast displacement is itself reconstructed, does not establish the absolute resistance level.

The principal result is an identifiability limit. A five-parameter physical interpretation is not supported by the available fuel observation: three combinations cannot be separated on any data, and the environmental composite cannot be separated on these data (Section 5.2.5). On these records, only one scale is practically identifiable. The imposed scalar estimator is useful for tracking the voyage-specific fuel-to-reported-power ratio, but κ cannot be assigned to efficiency, resistance or degradation.

Prospective freezing separates adaptation from prediction. It retains large gains on V2–V4 and no gain on V1. The auxiliary-fuel test shows that this conclusion is robust on V2–V3 but less stable on V1 and V4, where the constructed main-engine observation is more influential. The identifiability limit extends to the fuel-conversion baseline itself. Because all four voyages operate between 32.3% and 36.6% MCR, the observation cannot separate a genuine departure from the project-guide specific consumption from the part-load behaviour of the engine, and the two baselines in Section 5.2.2 differ by 8 to 11 percentage points of κ for that reason alone. The study relies on the two results that survive the choice: the prospective remaining-route error moves only from 2.62% to 2.40% mean absolute error, and the between-voyage dispersion of κ from 29.9% to 26.3%. The per-voyage one-step improvement does not survive it and changes sign on V1. The same separation holds for the reading of the source fuel field and for the freeze point: the prospective totals and the dispersion of κ are stable under both, whereas the level of κ and the per-voyage RMSE change depend on the reading alone, and the discrete speed profile depends on both (Sections 5.2.1, 5.2.6 and 5.3.3). The interval estimates of Section 5.2.3 point the same way: the moving-block intervals exclude zero on V2–V4 but span it on V1, and with four voyages the exact sign test reaches $p = 0.3125$, so the per-voyage pattern is descriptive case-study evidence and is not offered as population-level inference. Carrying both baselines is what separates the conclusions the observation supports from those it does not.

The decision layer introduces a second distinction: κ is estimated on reported power and applied to modelled power. The transfer-scale results show that this assumption can dominate the selected profile and its cost. The κ -conditioned optimisation must therefore remain a sensitivity analysis until the two power bases are reconciled by independent measurements.

The economic result is driven by the value assigned to early arrival. At lower time values, the higher-fuel optimum weakens or disappears. These scenarios illustrate objective-function behaviour; they are not recommendations without berth windows, waiting costs, charter terms, and current fuel prices for the intended operation.

The beam search agrees with dynamic programming on the finite grid, which provides the numerical benchmark. The reinforcement-learning benchmark is reported in the Supplementary Materials and supports no general claim about reinforcement learning.

The kinematic comparator of Supplementary Section S15.1 gives a second illustration of the distinction the calibration literature draws between predictive fit and parameter meaning [36,37]: its error reduction is larger and more uniform than that of Arm 1, yet it carries no propulsion information. The case study therefore supports a measurement-first workflow: establish provenance, test independence and excitation, determine the identifiable parameter combinations, and only then select an estimator and a decision model.

Both principal results are the behaviour predicted by the calibration literature. They are offered as an instance of that known result on primary operational records, within a methodological case study, and not as a new theoretical result. Brynjarsdóttir and O'Hagan [37] showed that a calibration parameter estimated without an explicit discrepancy model absorbs that discrepancy, fits well inside the observed domain, and loses its physical interpretation outside it. The two findings reported here are the two halves of that statement, obtained on primary operational records rather than on a constructed example. κ fits well in the domain in which it is identified: frozen after the adaptation window, it predicts remaining-route fuel totals to within -5.67% to $+0.70\%$, against -24.95% to $+0.16\%$ for the uncorrected model. It does not transfer outside that domain: moving from the reported-power basis on which it is identified to the modelled-power basis used by the decision layer changes the modelled cost reduction on V2 from $+1.15\%$ to $+0.06\%$ and removes almost all of the modelled fuel penalty. The observation operators differ, so κ is not transportable between them, and the empirical result is consistent with the theoretical expectation.

The remedy proposed in that literature is to model the discrepancy explicitly and to constrain it with prior information [36,37]. That remedy presupposes an observation informative enough to support both the parameter and the discrepancy function. The identifiability analysis of Sections 5.1 and 5.2.5 shows that this condition does not hold here: the fuel observation admits two composite directions structurally and supports one in practice, and adding a discrepancy term to an already deficient observation would relocate the indeterminacy and not remove it. Reporting a scalar correction that is declared as a reporting-channel ratio is therefore a defensible option under these measurements, and modelling the discrepancy becomes feasible only once an independent power or torque channel enlarges the observation.

These findings position the study within the wider literature. Work on data adequacy has examined processing pipelines [19], the effect of data diversity, quality and quantity on black-box and grey-box models [20], variable dependence and interpretable predictor selection [21], and chronological validation [22]; in each case adequacy is judged by predictive accuracy, whereas the present case study asks which parameters of a physics-based model the data can separate. Digital-twin work on ship performance has concentrated on model architecture and data-driven condition tracking [51,52], and on voyage-level routing services that assume the underlying channels are adequate [53]. Learning-based speed optimisation has been assessed mainly against simulated or aggregated data [54]. Comparative assessments of the EU ETS, FuelEU Maritime and IMO carbon-pricing instruments similarly take the reported operational quantities as given [55]. The present result is complementary rather than contradictory, but the complementarity is narrower than it may appear. Route-level totals of the kind reported for an oil/chemical tanker in [6] and for a container vessel in [56] can be accurate while the underlying segment-level relationships remain non-identifiable, because aggregation removes part of the timing and scaling information that the identifiability test exercises, as Section 4.3 notes. The separation between an accurate aggregate and identifiable components is also visible in the uncertainty analysis of [12], which quantifies the achievable accuracy of fuel-consumption prediction for a RoRo ship and a tanker by propagating the separate uncertainties of the contributing model components across the phases of a vessel's life. The present study therefore bounds what may be inferred about physical states from these channels at segment resolution; it does not test, and does not bear on, the accuracy of route-level accounting. Hull-condition and fouling studies reach a similar conclusion from a different direction, in that resistance degradation can only be separated from operating condition when independent power or thrust measurements are available [57]. Finally, the economic behaviour observed here is the expected behaviour of classical speed models, in which optimal speed rises with the value of time and falls with bunker price [58,59]; the contribution is not the direction of that response but the demonstration that its magnitude is governed by an unverified transfer of scale between two power bases.

Limitations

The results are those of a methodological case study. They are not offered as population-level inference or as validated for another vessel, vessel type or operating envelope.

The instruments behind the fuel, power and speed-through-water channels are not documented. Whether reported power is measured at the shaft or computed from engine-side signals is unknown, so the independence of the two channels that define κ cannot be asserted. The speed log is not identified, and any bias in speed through water enters the modelled-power basis to the third power.

The source fuel field admits two readings. The level of κ , the per-voyage RMSE change and the segment-level speed profile depend on the reading; the prospective totals, the dispersion of κ and the transfer result do not.

The adaptation window is a fraction of records. On V3 and V4 it cannot be shorter than five records, which is 22% and 18% of elapsed time, and on V4 an underprediction of about 8% remains on the common tail whatever the freeze point.

The dataset covers one tanker and four Mediterranean voyages. Two laden and two ballast cases do not support population-level inference; the exact sign test cannot reach conventional significance with this design.

Operating displacement is reconstructed rather than taken from complete departure-condition reports. Laden cases include a 1500 t consumables assumption, and one ballast estimate is applied to both ballast voyages. Displacement is the dominant operational input in the decision sensitivity.

Main-engine fuel is obtained by subtracting a constant auxiliary rate from reported total fuel, and reported power is not verified by shaft torque. Sensor type, calibration, sampling, filtering, and manual-entry status are unavailable. The observed inconsistencies cannot be assigned to a specific sensor or processing step.

The calm-water relation is a single-point calibrated power-law surrogate, not a validated vessel-specific speed–power curve [60]. Every evaluated speed below the 14.5 kn calibration point is therefore an extrapolation of the assumed cubic exponent, and the reported decision sensitivity is conditional on that representation. Environmental and shaft-speed excitation are uneven. V1 contains most of the wave variation, while V4 has constant RPM. Additional records collected under the same narrow operating envelope would not resolve the unidentifiable directions.

Posterior covariance is not propagated into optimisation. Operational-input perturbations, economic coefficients, transfer scales, and grid refinements are deterministic sensitivities rather than probability models of future conditions.

Record durations range from 0.5 to 5 h. Time-scaled process noise leaves the result unchanged, but the record-based adaptation window remains conditional on the source discretisation.

The transfer from reported-power κ to modelled-power fuel is not physically validated. Until independent mass-flow and shaft-torque measurements are available, the resulting profiles should not be labelled as calibrated-vessel optima.

The optimisation omits berth availability, waiting, early-arrival penalties, and continuous speed control. Counterfactual profiles retain the recorded segment-wise current, wave, and wind sequence; changes in segment duration do not re-time the environmental series. The optimisation is therefore a frozen-environment decision-sensitivity experiment rather than a time-consistent weather-routing replay. Cost changes below a few per cent and fuel changes near one per cent should be interpreted within these modelling limits.

7. Conclusions

This case study continues the physics-informed optimisation framework of [6] on the same vessel and voyages, approached from the measurement side and not the decision side. It asks what the routinely reported fuel and power channels can identify about the parameters of that model and what the answer implies for a speed decision.

The provenance audit produces a reproducible 207-record dataset and a defensible continuous Messina–Fiumicino sequence. It does not replace independent calibration records for the underlying fuel, power, speed or draft channels, whose instruments are not documented.

Of the five candidate corrections of the fuel observation, three combinations are structurally unidentifiable whatever the data, and the environmental composite is not practically identifiable on these voyages: a two-composite fit returns a negative added-power multiplier on three of the four. Separate resistance, efficiency and fuel-conversion

corrections cannot be obtained from the available observation. On these records, only one scale is practically identifiable.

The scalar estimated instead (κ) is the ratio of reconstructed main-engine fuel to reported power at a nominal specific consumption; it is a declared modelling choice and does not follow from the identifiability analysis. Frozen after the opening quarter of each voyage, it predicts remaining-route fuel totals to within -5.67% to $+0.70\%$, against -24.95% to $+0.16\%$ uncorrected. A scalar fitted once on the same window does the same; the Kalman filter is one implementation of the scaling and not a contribution.

κ is specific to the power basis on which it is identified. Transferred to the modelled-power basis of the decision layer, it changes all four optimised profiles; when the scale required by that basis is used instead, the matched cost reduction on V2 falls from $+1.15\%$ to $+0.06\%$ and the modelled fuel penalty disappears on three voyages.

The prospective totals and the between-voyage dispersion of κ are stable under two fuel-conversion baselines, two readings of the source fuel field and seven freeze-point definitions, and the transfer result holds under both readings. The per-voyage RMSE improvement depends on the baseline and the reading, and the segment-level speed profile on all three. On these records, a route-level correction is usable, and a segment-level speed recommendation is not supported.

At 800 EUR/h, earlier arrival lowers modelled cost while increasing fuel. Lower time values and the fuel-directed objective produce different choices. These are conditional counterfactual results, not realised savings.

These findings are an instance, in primary ship records and for one vessel, of the known result that a calibration parameter absorbs unmodelled discrepancy. They are offered as a methodological case study and not as a general framework.

Further validation requires synchronised mass-flow, shaft-torque, speed-log, draft or displacement and weather measurements across more vessels and operating conditions, followed by prospective onboard trials.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jmse14201940/s1>. The supporting file contains operational inputs and channel definitions (S1–S2), identifiability and recursive-estimation details (S3–S5), the complete implemented model (S6), solver methods, complete route outputs, pairwise results and cost/fuel figures (S7), machine-readable dataset manifests (S8), reproducibility mapping, the CII regulatory trajectory and software verification (S9–S11), tuning and PPO seed results (S12), the fuel-index comparison (S13), robustness analyses for frozen prediction, auxiliary-fuel allocation and the reading of the source fuel field, κ transfer, speed-grid resolution, regression intervals, the part-load SFOC baseline, scaled identifiability ranks, an adaptation-window scalar comparator, speed–power exponent sensitivity, the two-composite-parameter formulation and the freeze-point sensitivity (S14), and the kinematic comparator and reinforcement-learning benchmark (S15).

Author Contributions: Conceptualization, D.C. and F.P.; methodology, D.C. and F.P.; software, F.P. and A.P.; validation, D.C. and O.N.V.; formal analysis, O.N.V. and D.M.; investigation, D.C. and A.T.; resources, D.C. and A.P.; data curation, O.N.V.; writing—original draft preparation, O.N.V. and D.M.; writing—review and editing, A.P.; visualization, A.T.; supervision, F.P.; project administration, O.N.V., D.M. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The processed 207-record CSV, traceability workbook, analysis code, archived PPO outputs, generated result files, figures, checksums and verification scripts are supplied with this submission as a reproduction package and are deposited in the Zenodo repository under a Creative Commons Attribution 4.0 International licence at <https://doi.org/10.5281/zenodo.23235788>. The original commercial workbook and the sailing-leg performance pages are proprietary and cannot be distributed.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Maritime Organization (IMO). *Fourth IMO Greenhouse Gas Study 2020*; IMO: London, UK, 2020. Available online: <https://www.imo.org/en/ourwork/environment/pages/fourth-imo-greenhouse-gas-study-2020.aspx> (accessed on 11 March 2026).
2. International Maritime Organization (IMO). *2023 IMO Strategy on Reduction of GHG Emissions from Ships, Resolution MEPC.377(80)*; IMO: London, UK, 2023. Available online: [https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.377\(80\).pdf](https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.377(80).pdf) (accessed on 11 March 2026).
3. International Maritime Organization (IMO). *Resolution MEPC.352(78): 2022 Guidelines on Operational Carbon Intensity Indicators and the Calculation Methods (CII Guidelines, G1)*; IMO: London, UK, 2022. Available online: [https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.352\(78\).pdf](https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.352(78).pdf) (accessed on 1 September 2026).
4. European Parliament and Council of the European Union. Directive (EU) 2023/959 amending Directive 2003/87/EC Establishing a System for Greenhouse Gas Emission Allowance Trading Within the Union. *Off. J. Eur. Union* **2023**, L 130, 134–202.
5. European Parliament and Council of the European Union. Regulation (EU) 2023/1805 on the Use of Renewable and Low-Carbon Fuels in Maritime Transport. *Off. J. Eur. Union* **2023**, L 234, 48–100. Available online: <https://eur-lex.europa.eu/eli/reg/2023/1805/oj/eng> (accessed on 11 March 2026).
6. Coşofreţ, D.; Postolache, F.; Popa, A.; Volintiru, O.N.; Mărăşescu, D. A Hybrid Digital CO₂ Emission-Control Technology for Maritime Transport: Physics-Informed Adaptive Speed Optimization on Fixed Routes. *Fire* **2026**, 9, 136. [\[CrossRef\]](#)
7. Psaraftis, H.N.; Kontovas, C.A. Ship Speed Optimization: Concepts, Models and Combined Speed-Routing Scenarios. *Transp. Res. Part C Emerg. Technol.* **2014**, 44, 52–69. [\[CrossRef\]](#)
8. Fagerholt, K.; Laporte, G.; Norstad, I. Reducing Fuel Emissions by Optimizing Speed on Shipping Routes. *J. Oper. Res. Soc.* **2010**, 61, 523–529. [\[CrossRef\]](#)
9. Holtrop, J.; Mennen, G.G.J. An Approximate Power Prediction Method. *Int. Shipbuild. Prog.* **1982**, 29, 166–170. [\[CrossRef\]](#)
10. Holtrop, J. A Statistical Re-Analysis of Resistance and Propulsion Data. *Int. Shipbuild. Prog.* **1984**, 31, 272–276.
11. Coraddu, A.; Oneto, L.; Ghio, A.; Savio, S.; Anguita, D.; Figari, M. Machine Learning Approaches for Improving Condition-Based Maintenance of Naval Propulsion Plants. *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.* **2016**, 230, 136–153. [\[CrossRef\]](#)
12. Tillig, F.; Ringsberg, J.W.; Mao, W.; Ramne, B. Analysis of Uncertainties in the Prediction of Ships' Fuel Consumption—From Early Design to Operation Conditions. *Ships Offshore Struct.* **2018**, 13, 13–24. [\[CrossRef\]](#)
13. ISO 19030-2:2016; Ships and Marine Technology—Measurement of Changes in Hull and Propeller Performance—Part 2: Default Method. International Organization for Standardization (ISO): Geneva, Switzerland, 2016.
14. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, 15, 2405–2415. [\[CrossRef\]](#)
15. Madusanka, N.S.; Fan, Y.; Yang, S.; Xiang, X. Digital Twin in the Maritime Domain: A Review and Emerging Trends. *J. Mar. Sci. Eng.* **2023**, 11, 1021. [\[CrossRef\]](#)
16. Wei, Q.; Liu, Y.; Dong, Y.; Li, T.; Li, W. A Digital Twin Framework for Real-Time Ship Routing Considering Decarbonization Regulatory Compliance. *Ocean Eng.* **2023**, 278, 114407. [\[CrossRef\]](#)
17. Mauro, F.; Kana, A.A. Digital Twin for Ship Life-Cycle: A Critical Systematic Review. *Ocean Eng.* **2023**, 269, 113479. [\[CrossRef\]](#)
18. Kinaci, O.K. Ship Digital Twin Architecture for Optimizing Sailing Automation. *Ocean Eng.* **2023**, 275, 114128. [\[CrossRef\]](#)
19. Kim, Y.; Gupta, P.; Steen, S. A Comprehensive Review of Data Processing for Ship Performance Analysis. *Appl. Ocean Res.* **2025**, 162, 104737. [\[CrossRef\]](#)
20. Cai, Z.; Li, L.; Yu, L.; Li, C.; Sun, M. Diversity, Quality, and Quantity of Real Ship Data on the Black-Box and Gray-Box Prediction Models of Ship Fuel Consumption. *Ocean Eng.* **2024**, 291, 116434. [\[CrossRef\]](#)
21. Vorkapić, A.; Martinčić-Ipšić, S.; Piltaver, R. Interpretable Machine Learning: A Case Study on Predicting Fuel Consumption in VLGC Ship Propulsion. *J. Mar. Sci. Eng.* **2024**, 12, 1849. [\[CrossRef\]](#)
22. Vorkapić, A. Toward Realistic Ship Fuel Consumption Prediction Under Chronological Validation. *J. Mar. Sci. Eng.* **2026**, 14, 538. [\[CrossRef\]](#)
23. Gkerekos, C.; Lazakis, I.; Theotokatos, G. Machine Learning Models for Predicting Ship Main Engine Fuel Oil Consumption: A Comparative Study. *Ocean Eng.* **2019**, 188, 106282. [\[CrossRef\]](#)
24. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **2019**, 378, 686–707. [\[CrossRef\]](#)
25. Coraddu, A.; Oneto, L.; Baldi, F.; Cipollini, F.; Atlar, M.; Savio, S. Data-Driven Ship Digital Twin for Estimating the Speed Loss Caused by the Marine Fouling. *Ocean Eng.* **2019**, 186, 106063. [\[CrossRef\]](#)
26. Lee, J.-H.; Nam, Y.-S.; Kim, Y.; Liu, Y.; Lee, J.; Yang, H. Real-Time Digital Twin for Ship Operation in Waves. *Ocean Eng.* **2022**, 266, 112867. [\[CrossRef\]](#)

27. Taghavi, M.; Perera, L.P. Digital Twin Development for Ship Fuel Consumption Predictions Using Multiple Model Adaptive Estimation with Eigensystem Realization Algorithm and Observer/Kalman Filter Identification. *Ocean Eng.* **2026**, *345*, 123743. [CrossRef]
28. Jeon, J.; Theotokatos, G. A Framework to Assure the Trustworthiness of Physical Model-Based Digital Twins for Marine Engines. *J. Mar. Sci. Eng.* **2024**, *12*, 595. [CrossRef]
29. Ljung, L. *System Identification: Theory for the User*, 2nd ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1999.
30. Jazwinski, A.H. *Stochastic Processes and Filtering Theory*; Academic Press: New York, NY, USA, 1970.
31. Simon, D. *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*; Wiley: Hoboken, NJ, USA, 2006.
32. *Applied Optimal Estimation*; Gelb, A., Ed.; MIT Press: Cambridge, MA, USA, 1974.
33. Reif, K.; Günther, S.; Yaz, E.; Unbehauen, R. Stochastic Stability of the Discrete-Time Extended Kalman Filter. *IEEE Trans. Autom. Control* **1999**, *44*, 714–728. [CrossRef]
34. Fossen, T.I. *Handbook of Marine Craft Hydrodynamics and Motion Control*; Wiley: Chichester, UK, 2011. [CrossRef]
35. Perera, L.P.; Mo, B. Data Analytics for Capturing Marine Engine Operating Regions for Ship Performance Monitoring. In Proceedings of the ASME 2016 35th International Conference on Ocean, Offshore and Arctic Engineering (OMAE), Busan, Republic of Korea, 19–24 June 2016; Volume 7. [CrossRef]
36. Kennedy, M.C.; O'Hagan, A. Bayesian Calibration of Computer Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2001**, *63*, 425–464. [CrossRef]
37. Brynjarsdóttir, J.; O'Hagan, A. Learning about Physical Parameters: The Importance of Model Discrepancy. *Inverse Probl.* **2014**, *30*, 114007. [CrossRef]
38. Villaverde, A.F.; Barreiro, A.; Papachristodoulou, A. Structural Identifiability of Dynamic Systems Biology Models. *PLoS Comput. Biol.* **2016**, *12*, e1005153. [CrossRef] [PubMed]
39. Intergovernmental Panel on Climate Change (IPCC). *2006 IPCC Guidelines for National Greenhouse Gas Inventories*; IGES: Hayama, Japan, 2006. Available online: <https://www.ipcc-nggip.iges.or.jp/public/2006gl/> (accessed on 11 March 2026).
40. Baldi, F.; Johnson, H.; Gabriellii, C.; Andersson, K. Energy and Exergy Analysis of Ship Energy Systems—The Case Study of a Chemical Tanker. *Int. J. Thermodyn.* **2015**, *18*, 82–93. [CrossRef]
41. Fagerholt, K.; Gausel, N.T.; Rakke, J.G.; Psaraftis, H.N. Maritime Routing and Speed Optimization with Emission Control Areas. *Transp. Res. Part C Emerg. Technol.* **2015**, *52*, 57–73. [CrossRef]
42. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347.
43. MAN Diesel & Turbo. *MAN B&W S50MC-C8.2-TII Project Guide: Camshaft Controlled Two-Stroke Engines, Edition 0.5*; MAN Diesel & Turbo: Copenhagen, Denmark, 2014.
44. International Maritime Organization (IMO). *Resolution MEPC.353(78): 2022 Guidelines on the Reference Lines for Use with Operational Carbon Intensity Indicators (CII Reference Lines Guidelines, G2)*; IMO: London, UK, 2022. Available online: [https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.353\(78\).pdf](https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.353(78).pdf) (accessed on 1 September 2026).
45. International Maritime Organization (IMO). *Resolution MEPC.400(83): Amendments to the 2021 Guidelines on the Operational Carbon Intensity Reduction Factors Relative to Reference Lines (CII Reduction Factors Guidelines, G3)*; IMO: London, UK, 2025.
46. International Maritime Organization (IMO). *Resolution MEPC.354(78): 2022 Guidelines on the Operational Carbon Intensity Rating of Ships (CII Rating Guidelines, G4)*; IMO: London, UK, 2022. Available online: [https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.354\(78\).pdf](https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.354(78).pdf) (accessed on 11 March 2026).
47. Efron, B.; Hastie, T. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*; Cambridge University Press: Cambridge, UK, 2016. [CrossRef]
48. Wasserman, L. *All of Statistics: A Concise Course in Statistical Inference*; Springer: New York, NY, USA, 2004.
49. Hollander, M.; Wolfe, D.A.; Chicken, E. *Nonparametric Statistical Methods*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013.
50. International Maritime Organization (IMO). *Third IMO Greenhouse Gas Study 2014*; IMO: London, UK, 2014. Available online: <https://greenvoyage2050.imo.org/wp-content/uploads/2021/01/third-imo-ghg-study-2014-executive-summary-and-final-report.pdf> (accessed on 6 August 2026).
51. Taghavi, M.; Perera, L.P. Advanced Data Cluster Analyses in Digital Twin Development for Marine Engines towards Ship Performance Quantification. *Ocean Eng.* **2024**, *298*, 117098. [CrossRef]
52. Sánchez-Fernández, Á.; Vlad-Voinea, E.-D.; Pernas-Álvarez, J.; Crespo-Pereira, D.; Sañudo-Costoya, B.; Lamas-Rodríguez, A. Framework for the Development of a Process Digital Twin in Shipbuilding: A Case Study in a Robotized Minor Pre-Assembly Workstation. *J. Mar. Sci. Eng.* **2026**, *14*, 106. [CrossRef]
53. Latinopoulos, C.; Zavvos, E.; Kaklis, D.; Leemen, V.; Halatsis, A. Marine Voyage Optimization and Weather Routing with Deep Reinforcement Learning. *J. Mar. Sci. Eng.* **2025**, *13*, 902. [CrossRef]
54. Zhou, T.; Bao, T.; Liu, Y.; Zhang, C. Q-Learning-Based Sailing Speed Optimization for Ocean-Going Liners Under the EU ETS: Considering Shipper Satisfaction. *J. Mar. Sci. Eng.* **2026**, *14*, 848. [CrossRef]

55. Kwon, B.-J.; Jeong, B.-U.; Lee, S.-B.; Park, Y.-M.; Oh, S.-J.; Shin, S.-C. Comparative Analysis of EU ETS, FuelEU Maritime and IMO Carbon Pricing Regulations: Strategic and Economic Implications for the Shipping Industry. *J. Adv. Mar. Eng. Technol.* **2025**, *49*, 224–238. [[CrossRef](#)]
56. Coşofreţ, D.; Volintiru, O.-N.; Mărăşescu, D.; Deliu, F.; Popa, C. Voyage-Level Assessment of Slow Steaming and B30 Biofuel Strategies for Container Vessel Decarbonisation Under EU ETS and FuelEU Maritime. *J. Mar. Sci. Eng.* **2026**, *14*, 1296. [[CrossRef](#)]
57. Grigoropoulos, G.; Plessas, T. Speed Dependence of Hull Fouling on Ship Performance. *J. Mar. Sci. Eng.* **2026**, *14*, 1107. [[CrossRef](#)]
58. Ronen, D. The Effect of Oil Price on the Optimal Speed of Ships. *J. Oper. Res. Soc.* **1982**, *33*, 1035–1040. [[CrossRef](#)]
59. Psaraftis, H.N.; Kontovas, C.A. Speed Models for Energy-Efficient Maritime Transportation: A Taxonomy and Survey. *Transp. Res. Part C Emerg. Technol.* **2013**, *26*, 331–351. [[CrossRef](#)]
60. International Towing Tank Conference (ITTC). *Recommended Procedures and Guidelines 7.5-04-01-01.1: Preparation, Conduct and Analysis of Speed/Power Trials, Revision 06*; ITTC: Zürich, Switzerland, 2021. Available online: <https://www.ittc.info/media/9874/75-04-01-011.pdf> (accessed on 2 August 2026).

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.