11th Hull Performance & Insight Conference

HullPIC'26

Edited by Volker Bertram

Sponsored by

www.jotun.com

tutech.de

albis-mp.com

Accelleron.com

www.cmp-chugoku.com

www.coach-solutions.dk

www.enamor.pl

gitcoatings.com

www.hempel.com

www.hoppe-marine.com

www.idealship.de

i-tech.se

www.akzonobel.com

www.miros-group.com

www.nipponpaint-marine.com

www.sofarocean.com

www.metis.tech

www.vpsolutions.dk

vessely.de

www.bureauveritas.com

www.ppg.com

toqua.ai

www.zeronorth.com

Media Partners

https://hansa-online.de/

https://home.rina.org.uk/

Index

Engage! The Why and How in Crew Involvement

Harriet Hunnisett-Johnson, Signol, London/UK**,** harriet.hunnisett-johnson@signol.io

Abstract

This paper examines how behavioural science techniques can be leveraged to enhance operational efficiency in maritime shipping. Using data from an 18-vessel chemical tanker fleet, this research demonstrates how structured feedback mechanisms, goal-setting frameworks, and judicious and timely deployment of proven behaviour change techniques can lead to measurable reductions in fuel consumption and vessel emissions. The six-month study observed significant improvements in operational behaviours, particularly in vessel trim optimisation, resulting in over 250 t of fuel and 790 t of CO_2 emissions saved.

1. Introduction

Maritime shipping accounts for approximately 3% of global greenhouse gas emissions according to the International Maritime Organisation. Whilst technological solutions are being developed for long-term decarbonisation, substantial near-term efficiency gains can be achieved through operational improvements. This white paper explores how behavioural science principles can be applied to bridge the gap between vessel performance monitoring and actual fuel-saving behaviours.

2. Research Background

The maritime industry has widely adopted sophisticated monitoring technologies that provide detailed operational data. However, a significant challenge remains in translating performance monitoring into consistent behavioural change among crew members. This research investigates whether established behavioural science techniques, including personalised goal-setting, structured feedback, and targeted incentives, can effectively influence maritime operations.

3. Methodology

3.1. Study Design

The study was conducted across an 18-vessel chemical tanker fleet over six months, engaging 70 crew members. Using existing vessel performance data, Signol's behavioural scientists and data analysts established baseline performance metrics, or behaviours, that applied to all vessels.

Four key operational areas were identified as targets for behaviour change:

1. Optimal trim configuration
2. Engine maintenance optimisation
3. Efficient auxiliary engine use
4. Prompt departure procedures

3.2. Implementation Framework

The behavioural intervention consisted of:

1. Personalised Goals: Individual baselines were established for each crew member based on historical performance data
2. Regular Feedback: Structured communications delivered via email and a dedicated web application
3. Dual Incentive Structure: Achievement recognition coupled with charitable contributions

4. Integrated Behavioural Change Techniques: delivered in each communication, including linking the environmental impact of specific operational behaviours

Researchers conducted in-person sessions with crew members before and during the study period to ensure understanding and engagement with the programme.

3.3. Measurement

To assess the indicators of behaviour change, an Econometric analysis was done using a linear regression model, where the primary intent was to obtain values of the coefficients within the equation. This allowed controlling for fixed effects, so any confounding variables were held constant. The fixed effects used were the vessel, the operators, origin/destination/route, weather, month and hull fouling.

4. Results

4.1. Quantitative Findings

The study demonstrated measurable improvements across several key metrics:

- Success rate of the trim behaviour: 16% increase in optimal trim implementation, Fig.1
- Trim Optimisation: Average trim shifted 0.22 m closer to even keel, representing a more efficient configuration for the vessel type, Fig.2
- Fuel Consumption: over 250 t reduction over six months
- Emissions Reduction: over 790 t of CO_2 emissions avoided

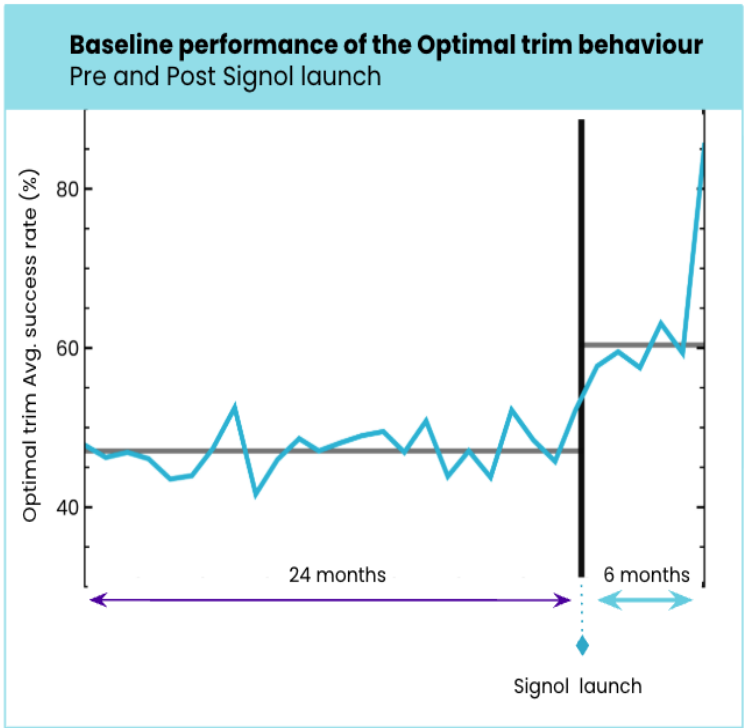


Fig.1: Baseline Performance of Optimal Trim Behaviour Pre and Post Signol Launch

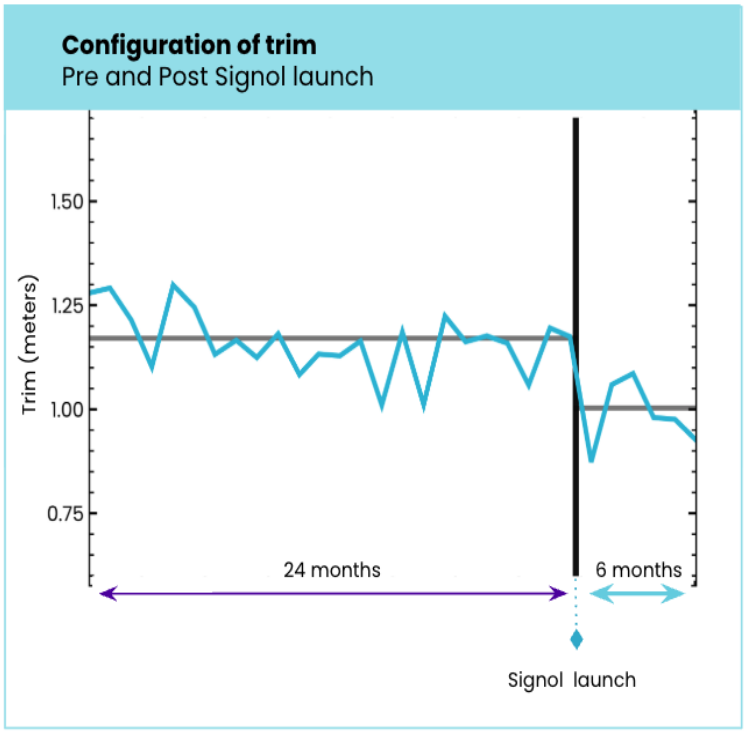


Fig.2: The Configuration of Trim Pre and Post Signol Launch

4.2. Qualitative Outcomes

Interviews with crew members revealed several notable qualitative outcomes:

- Integration of efficiency discussions into daily operational meetings
- Increased awareness of how individual actions impact fuel consumption
- Development of collaborative problem-solving approaches to meet efficiency goals
- Enhanced sense of purpose related to environmental impact reduction

One participant noted: “We discuss Signol with the crew in daily meetings [addressing] how we can save fuel. We also talked about Signol in safety meetings. When we receive emails, if we don’t meet the goals, we discuss with the crew that we should do something to achieve the goals next time.”

5. Discussion

5.1. Behavioural Science Principles in Practice

The research demonstrates how several established behavioural science principles can be effectively applied in maritime operations:

1. Goal Setting Theory: Specific, personalised targets based on historical performance provided clear direction
2. Feedback Loops: Regular, structured feedback facilitated continuous improvement
3. Social Norms: Crew discussions about performance created social reinforcement of desired behaviours
4. Intrinsic Motivation: Environmental impact information connected operational behaviours toa broader purpose

5.2. Implementation Considerations

Organisations seeking to implement similar behavioural programmes should consider several factors:

1. Data Infrastructure: Sufficient operational data is necessary to establish meaningful baselines
2. Leadership Support: Management endorsement facilitates programme acceptance
3. Communication Channels: Reliable communication methods are essential for regular feedback
4. Cultural Factors: Programme design should account for the existing organisational culture
5. Incentive Structure: Alignment with crew values enhances engagement

6. Conclusions

This research demonstrates that behavioural science applications can yield significant operational improvements in maritime shipping. By systematically engaging crew members through personalised goals, regular feedback, and meaningful incentives, organisations can bridge the gap between monitoring capabilities and actual emissions reductions.

The approach detailed in this study offers a complementary strategy to technological solutions, providing immediate efficiency gains whilst longer-term decarbonisation technologies are developed and deployed.

References

THALER, R.H.; SUNSTEIN, C.R. (2021), *Nudge: The final edition*, Penguin Books

WOOLDRIDGE, J.M. (2012), *Introductory econometrics: A modern approach,* South-Western Cengage Learning

Enhancing Hull Performance Analytics with Vision-AI - Introduction of a Hybrid Biofouling Analytics Approach

Michael Stein, Vesselity Maritime Analytics, Hamburg/Germany, stein@vesselity.de
Srusti Ujjini Math, Vesselity Maritime Analytics, Hamburg/Germany
Matthias Brueck, Fraunhofer IAPT, Hamburg/Germany

Abstract

The past decade's advances in machine learning and robotics now enable data-driven biofouling management. AI-based image recognition effectively bridges the gap between underwater inspections and data analytics. This paper introduces a vision-centric evaluation method for performance reporting, linking quantified fouling states to longitudinal hull tracking and maintenance decision support. To address the scarcity of underwater training data, the paper presents a synthetic image pipeline to boost AI recognition. Through a unique cross-domain analysis of aerial test plates and ROV inspections, it establishes a state-of-the-art synthesis workflow. Key findings show that while spatially adaptive normalization improves above-water performance, Pix2PixHD's simpler architecture generalizes better in the low-texture, low-illumination conditions typical of underwater environments.

1. Introduction

Marine biofouling remains to be one of the most persistent (hidden) drivers of ship energy demand. Even moderate biological growth increases hull roughness and frictional resistance, which can translate into substantial additional fuel burn and emissions, often dwarfing the direct costs of coating or cleaning interventions, *Schultz et al. (2011), Farkas et al. (2021).* In parallel to its efficiency penalty, biofouling is also recognized as a vector for invasive aquatic species transfer, motivating international guidance on systematic biofouling management and documentation, *IMO (2023).*

Recent advances in compact robotics and computer vision introduce a shift in hull-performance management from periodic, experience-based decisions toward condition-based strategies that allow the combination of operational data streams with objective evidence of the hull's physical state. Micro-ROVs can acquire hull imagery more frequently and with better repeatability than traditional drydock-only assessment − creating the possibility of "longitudinal hull condition" datasets rather than isolated inspection snapshots. The potential of this rising technology for the maritime sector has been highlighted by *Stein (2023)*.

Yet, translating real-world underwater inspection footage into reliable quantitative indicators remains difficult. Underwater imaging suffers from scattering, backscatter, wavelength-dependent attenuation and severe contrast loss, to name only a few of 27 influencing factors, which distort texture cues and suppress boundaries that segmentation models rely on, *Akkaynak and Treibitz (2018)*. In addition, operational hull imagery is highly variable across ports, seasons, coatings, lighting setups, camera optics, and cleaning histories − conditions that routinely produce domain shifts between curated training data and deployment reality. These constraints have driven a growing body of work on ML-supported hull inspection, including semantic segmentation benchmarks for underwater ship inspections and transfer-learning approaches for hull-surface condition recognition, *Waszak et al. (2022), Kim et al. (2022).*

For hull performance analytics, the technical challenge is not only "detect fouling," but "detect fouling in a way that is operationally interpretable." This is where vision-centric inspection can bridge into performance monitoring standards and decision support. ISO 19030 formalizes principles and indicators for tracking changes in hull and propeller performance over time, enabling consistent, within-ship comparisons for maintenance planning, *ISO (2016)*. However, operational performance deltas alone rarely identify why performance changes (e.g., fouling vs. weather routing effects vs. sensor drift). A quantified, class-resolved fouling state derived from ROV imagery can act as an evidence layer to

interpret performance trends − linking the "what we observe on the hull" to "what we observe in propulsion and fuel-time series," thereby strengthening both reporting credibility and maintenance decision quality.

This paper contributes to this integration by presenting a vision-centric method for robust biofouling imagery evaluation under real inspection conditions, with an explicit focus on scalability and downstream utility for performance-oriented reporting. It frames the core bottleneck as data scarcity and generalization: high-quality pixel labels are often disclosed, while real-world hull imagery varies dramatically between aerial (in-air) and underwater domains (turbidity, illumination, reflections, and coating appearance). To address this, the paper evaluates mask-conditioned synthetic data generation − leveraging high-resolution conditional GAN families (e.g., Pix2PixHD and SPADE) and exploring diffusion-based control − to generate realistic, label-aligned fouling imagery at scale.

2. Problem statement

High-quality, well-labelled underwater ship-hull imagery is a critical bottleneck for advancing hull performance analytics. Vision-AI models that quantify biofouling state, coating degradation, corrosion, and structural anomalies depend on pixel-accurate ground truth to learn robust representations and to produce outputs that can be trusted in operational decision support. Without reliable labels, models may appear accurate on narrow test sets yet fail when confronted with the real diversity of vessel types, coatings, lighting conditions, water properties, and inspection practices encountered across the global fleet.

Conventional underwater data acquisition relies heavily on diver-led class or in-water surveys. While divers remain essential for many tasks, diver-based inspection introduces structural constraints for building scalable labelled datasets. Even when conditions allow, diver time-on-task is limited and the work environment can be hazardous; inspections must be performed efficiently, often under schedule pressure, and imagery collection is typically optimized for immediate survey findings rather than for consistent, ML-ready coverage and annotation. More fundamentally, diver inspections do not naturally produce standardized, repeatable scans at the scale required for data-driven hull-condition tracking.

Drydock inspections provide higher-quality access and can support meticulous documentation, but they are episodic. Relying on drydock-only imagery therefore produces an "ex ante" condition snapshot: it explains the hull state at docking, but it does not capture how condition evolves during the years between dockings, when decisions about cleaning, coating performance, routing, speed management, and efficiency interventions are most valuable. For hull performance analytics, the central requirement is longitudinal evidence: frequent, comparable observations that can be linked to time-series performance indicators, as emphasized by in-service monitoring standards, *ISO (2016)*.

Hull-scanning ROV systems offer a practical path through these constraints. ROV-based inspection can be scheduled more frequently than drydock events, can reduce human exposure to hazardous conditions, and can be engineered for systematic coverage (e.g., consistent stand-off distance, lighting, overlap, and georeferenced scan paths). This improves not only inspection repeatability but also the feasibility of building large labelled corpora - especially when paired with semi-automated annotation tools, active learning workflows, and synthetic augmentation pipelines validated against real deployment conditions. In short, ROV scanning shifts underwater imagery from "ad hoc evidence" to "measurement infrastructure."

Despite progress, publicly available datasets are still insufficient to train models that generalize across the global shipping industry. A useful example is the LIACi contribution, *Waszak et al. (2022)*, used in recent work on synthetic marine-growth segmentation: while valuable for benchmarking, it remains limited in scale and scope in the variety of hull geometries, coatings, biofouling regimes, and imaging environments required for cross-sector, cross-region robustness. Surveys of underwater image synthesis and datasets similarly emphasize that data scarcity and domain shift are persistent barriers to trans-

ferable performance as for example seen in *O'Byrne et al. (2018), Mai et al. (2024)* and *Barbosa and Apolinario (2025).* The core problem, therefore, is not simply to "collect more images," but to build well-labelled, operationally representative datasets - at scale, over time, and across conditions - so that hull analytics can support five-year-cycle decision making rather than just post-drydock assessments.

3. Implications to Maritime Economics

Biofouling remains a persistent, operationally "hidden" driver of energy demand for seagoing ships. Even moderate growth elevates hull roughness and frictional resistance, *Schultz et al. (2011)*, with measurable consequences for fuel consumption, emissions, and total voyage cost. Beyond efficiency, biofouling constitutes a biosecurity concern through invasive species transfer, which has increased the regulatory emphasis on systematic biofouling management and documentation, *IMO (2023)*. Within this context, the present work positions vision-based inspection not as an isolated detection task, but as an evidence layer that can be linked to longitudinal performance monitoring and maintenance decision support. A key implication for shipping is the role of micro-ROV systems in shifting hull assessment from episodic snapshots toward continuous, repeatable measurement. Compared with diver-based surveys or drydock-only documentation, micro-ROVs can be deployed more frequently, potentially generating a vast amount of crucial hull condition knowledge. This supports the creation of high-definition, temporally comparable "longitudinal hull condition" datasets, which are essential if hull-state evidence is to be integrated with operational efficiency indicators. In practical terms, frequent imagery closes the evidentiary gap between observed performance deltas and their physical cause, strengthening the credibility of performance reporting.

However, the industry-wide adoption of data-driven biofouling analytics is constrained by underwater data scarcity and domain variability. Real-world hull imagery is difficult to label at scale and is affected by scattering, backscatter, attenuation, and highly heterogeneous operational conditions, which collectively induce domain shift and degrade model generalization. The paper's synthetic, mask-conditioned image generation pipeline has direct implications for overcoming this bottleneck: by producing label-aligned synthetic hull imagery, it can expand training coverage, mitigate class imbalance (including rare fouling states), and reduce dependence on costly manual annotation. Crucially, the reported results emphasize that synthetic realism alone is insufficient; generator choice and hybrid (real + synthetic) training must be validated against underwater imaging physics to yield robust gains in downstream segmentation performance.

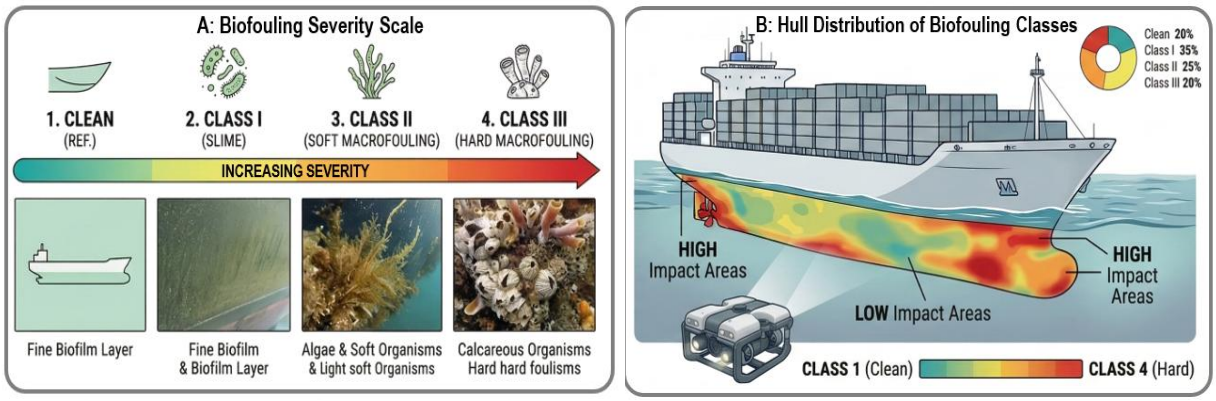


Fig.1: Classification and spatial distribution of marine biofouling on ship hulls. (A) Severity scale illustrating simplified biofouling classes used for classification. (B) Example spatial distribution of biofouling classes across the ship hull.

4. Methodology

This study develops and evaluates a vision-centric pipeline for quantifying hull biofouling from inspection imagery and translating the resulting fouling state into inputs suitable for longitudinal hull-condition tracking and performance reporting. The core methodological idea is data-centric robustness: rather than relying solely on larger real-world labels, the approach systematically expands and diversifies training data with label-preserving synthetic imagery, reducing the sim-to-real gap through controlled variability, Tobin et al. (2017). To further close bespoke gap, this study combines aerial test plate data as well as real-world ROV inspection data in its research. To the best of our knowledge, this mixed-data source approach has not been conducted before and reflects a novelty in biofouling research. Two complementary, manually annotated datasets were constructed to reflect distinct operational inspection domains:

1. Aerial test plate dataset: 216 labelled images spanning 11 fouling categories
2. Underwater ship hull dataset: 225 labelled images spanning 6 fouling categories.

The test plate data was extracted from photographs of anti-fouling coating test sites in Scandinavia and the Mediterranean provided by an anonymous stakeholder in coating research. The underwater ship hull dataset has been provided by a service operating company specializing in global ROV ship hull inspections. The sample of images varies across multiple inspections and ships in order to account for small sample biases and provide a variation of lighting, visibility, fouling distribution and other image defining factors.

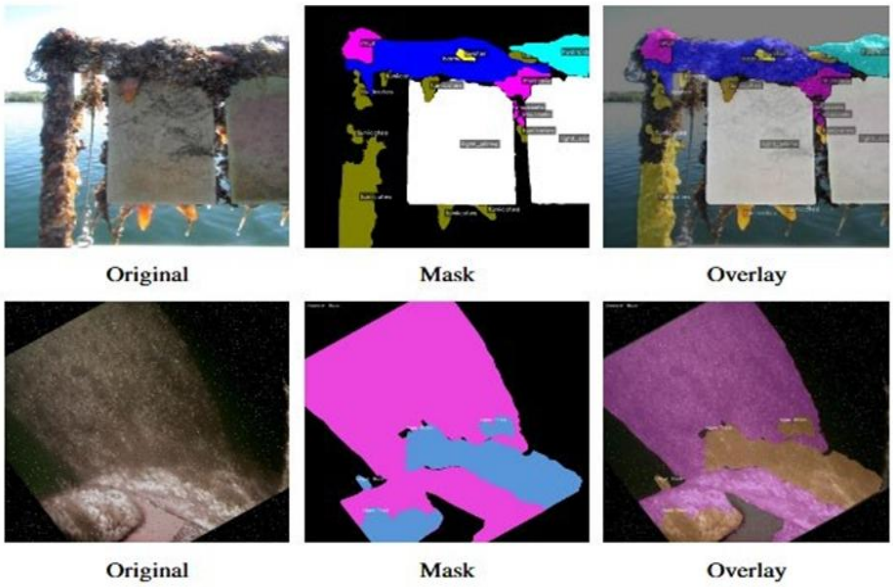


Fig.2: Examples of image–mask alignment used during quality verification

In the first stage, all images and annotations were standardized into a unified training representation. Raw images were automatically oriented and resized to a fixed resolution to support downstream segmentation. The corresponding labels were exported in COCO JSON format, *Lin et al. (2014).* Because the export provides polygon vertices rather than raster masks, we developed custom scripts to (i) rasterize polygons into RGB/ID mask images, (ii) generate mask–image overlay previews for visual quality control, and (iii) compute per-class statistics to audit label distributions and identify potential imbalance. In parallel, routine pre-processing and validation steps were applied to ensure dataset integrity, including mask–image alignment checks, removal of corrupted samples, and any model-specific normalization required by the respective training pipelines.

In the second stage, synthetic image generation was performed using paired mask-to-image translation. This follows established conditional synthesis paradigms in which semantic layouts guide high-

resolution image generation (e.g., Pix2PixHD and SPADE-style conditioning) as already introduced by *Wang et al. (2018)* and *Park et al. (2019).* Both high-resolution conditional GAN gene-rators were trained separately for the aerial and underwater domains using paired (m, x) samples, where m denotes the segmentation mask and x denotes the corresponding RGB image. Both generators were optimized using Adam for 200 epochs per dataset. Training used a conditional adversarial objective augmented with stabilization and perceptual components (adversarial loss, feature-matching loss, and a VGG-based perceptual loss), while SPADE additionally injected the mask through spatially-adaptive normalization layers to strengthen semantic alignment and boundary preservation, *Wang et al. (2018), Park et al. (2019).* For GAN training, images were resized to 512×512 and normalized to [-1,1], while downstream segmentation training was conducted at 640×640.

In the third stage, synthetic datasets were constructed via automatic label propagation, in which each synthetic image inherits its exact conditioning mask as its segmentation label. This eliminates manual annotation for generated data and ensures strict image–label correspondence. This approach enables the isolation of domain-gap effects and measurement of practical gains from synthetic augmentation following *Tobin et al. (2017).* Deterministic file naming and label duplication were used to preserve alignment between generated images and YOLO-format label files. To test whether synthetic data improves real-world performance, the study trains YOLO-family segmentation models under three regimes:

1. Real-only: trained solely on the curated labelled images
2. Synthetic-only: trained solely on generated images with propagated labels
3. Hybrid (real + synthetic): trained on the union of real and synthetic data

The final evaluation stage focussed on multi-level performance assessment of the models. Because visually plausible images do not always improve downstream learning, generator evaluation was conducted on both perceptual/structural similarity and task utility. For perceptual/structural similarity, realism was assessed using widely adopted metrics, including:

- FID for distributional similarity in deep feature space, *Heusel et al. (2017)*
- LPIPS for perceptual similarity aligned with human judgment, *Zhang et al. (2018)*
- SSIM for structural fidelity, *Wang et al. (2004)*

These metrics provide complementary signals: FID emphasizes feature-distribution alignment, LPIPS captures perceptual differences, and SSIM focuses on structural consistency. Importantly, the study treats these as necessary but insufficient, prioritizing downstream segmentation outcomes as the decisive criterion. For task-level utility, segmentation models from the YOLOv8 and YOLOv11 families were trained with consistent hyperparameters (fixed input resolution, fixed epoch budget, and unified augmentation policies) and evaluated on held-out real test images. Performance was reported using precision, recall, F1-score, mIoU, and per-class IoU, with additional attention to robustness in visually challenging cases such as low visibility, occlusion, and texture variability, *Redmon et al. (2016).*

Performance is evaluated on held-out real test data, using mean Intersection-over-Union (mIoU) as primary segmentation metric (consistent with common semantic segmentation practice). The hybrid regime provides the strongest robustness, with reported absolute mIoU gains up to ~0.2 on real test imagery. Improvements are especially evident for minority/rare fouling classes, consistent with the hypothesis that synthetic augmentation can expand coverage of underrepresented visual states and reduce class-imbalance effects. A final methodological element lies in the explicit comparison between aerial and underwater domains. The study reports that generator and training effectiveness can differ by environment, underscoring a practical lesson for ROV-based analytics: synthetic data is not a single universal lever. Generator selection, conditioning strategy, and augmentation scope must be validated against the target deployment domain's imaging physics and operational variability.

5. Results

Synthetic outputs were first evaluated qualitatively, since visual plausibility alone does not ensure effective segmentation training expressed by improved downstream segmentation performance. Beyond realism, generated imagery must preserve conditioning mask boundaries and structural integrity, *Wang et al. (2004),* while also avoiding perceptual artifacts that can mislead learning, *Zhang et al. (2018)*. Representative samples were selected to capture diverse mask complexities and environmental conditions, ensuring balanced coverage of both aerial and underwater domains. Each qualitative comparison is presented as a triplet: (A) input mask, (B) real image, and (C) corresponding synthetic image. This protocol enables consistent evaluation of each generator's ability to produce realistic, mask-aligned outputs across contrasting inspection scenarios, *Heusel et al. (2017).*

1) Aerial Domain: In aerial imagery, both Pix2PixHD and SPADE accurately reproduce the global structure of the conditioning mask, but they differ in texture fidelity and lighting adaptation. Pix2PixHD often over-smooths fine-scale regions, such as thin slime trails or surface micro-irregularities, which can reduce perceived detail. SPADE generates visually sharper outputs with stronger local contrast and more realistic micro-texture variation. Boundaries remain well aligned with the conditioning masks for both models. SPADE shows greater robustness to uneven illumination and reflective hull surfaces. to its spatially adaptive normalization that better preserves local semantic and lighting context. Overall, SPADE produces more convincing aerial results, due to its spatially adaptive normalization that better preserves local semantic and lighting context, Fig.3.

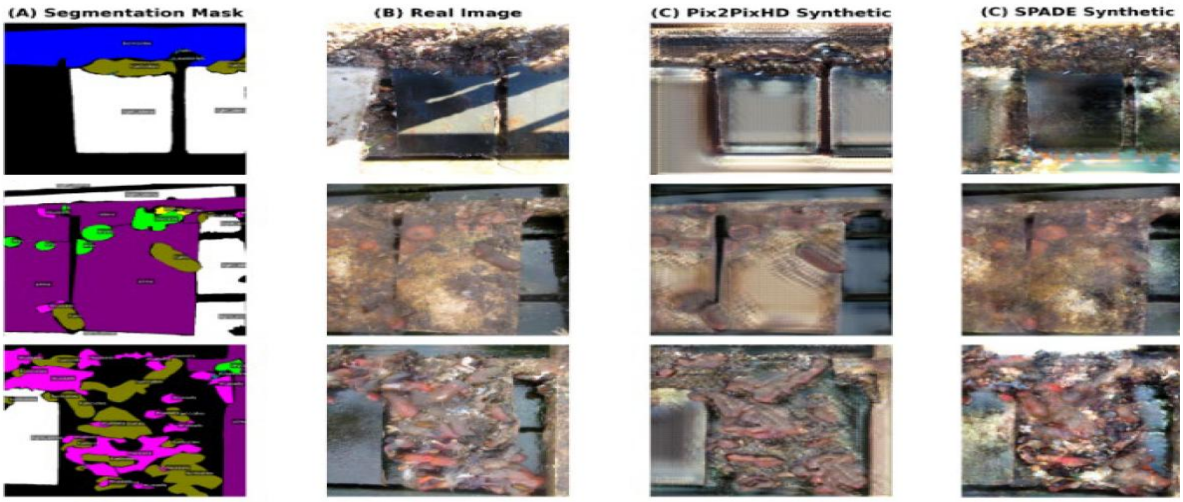


Fig.3: Aerial Training Data Example Results, Pix2PixHD vs. SPADE synthesis

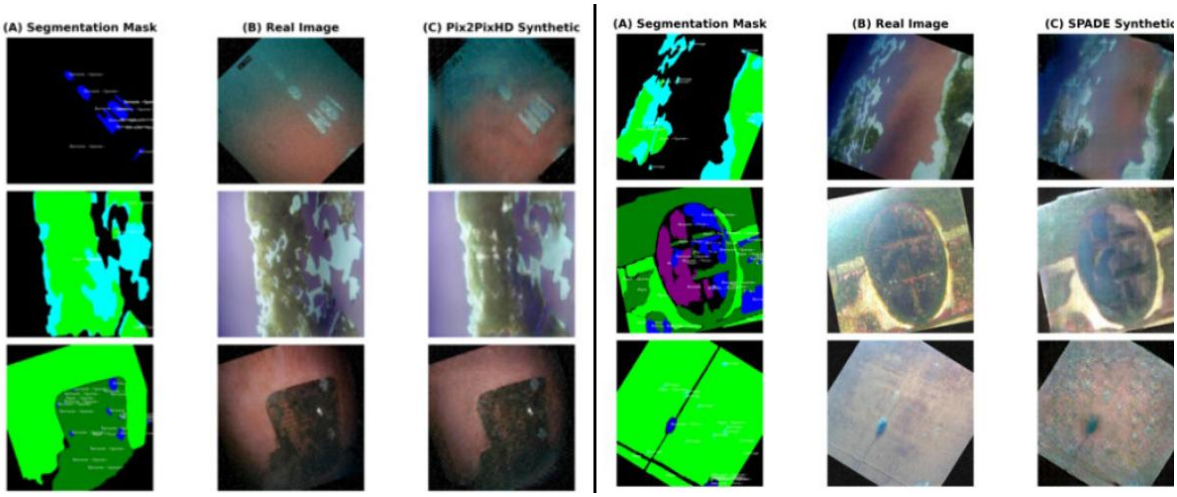


Fig.4: Underwater Training Data Example Results, left Pix2PixHD, right SPADE synthesis

2) Underwater Domain: Underwater synthesis is more challenging due to turbidity, backscatter, and pronounced colour distortion to only name a few underwater image influencing factors, *Akkaynak and Treibitz (2018)*. In this context, Pix2PixHD demonstrates greater stability, reproducing global structure and consistent illumination patterns that closely resemble real underwater lighting conditions. SPADE maintains accurate mask adherence and produces clear object boundaries, but some outputs show excessive contrast or colour exaggeration in highly scattered scenes. As a result, Fig.4, Pix2PixHD generates more natural underwater representations, while SPADE remains superior in aerial imagery. These domain-specific behaviours show that spatially adaptive normalization benefits above-water performance, while Pix2PixHD's simpler architecture generalizes more effectively under the low-frequency illumination and reduced texture typical of underwater environments.

Beyond visual confirmation, the generated images were rigorously assessed using five widely recognized metrics: FID, IS, CAS, LPIPS, and SSIM. While qualitative analysis focuses on how realistic and accurate the images appear, these metrics provide a statistical lens on how closely synthetic images mirror real ones in both structure and perception. In this context, lower FID and LPIPS scores, alongside higher IS, CAS, and SSIM values, signal superior generative performance. Separate evaluations for the aerial test plate and underwater marine fouling datasets were conducted. In each case, synthetic images from Pix2PixHD and SPADE were compared under identical preprocessing steps, ensuring a fair match. FID and IS relied on InceptionV3 features, CAS drew from CLIP ViT-L/14, and LPIPS and SSIM were calculated on paired samples with matching masks, Table I. By fixing random seeds, reproduction of every experiment is guaranteed.

Table I: Statistical comparison of data synthesis

Sample	Model	FID ↓	IS ↑	CAS ↑	LPIPS ↓	SSIM ↑
Aerial Test Plates	Pix2PixHD	244.52	3.40	0.6600	0.7073	0.1354
	SPADE	**142.52**	3.40	**0.7700**	**0.6967**	**0.1428**
Underwater ROV Inspection	Pix2PixHD	**140.94**	3.65	**0.8500**	**0.4563**	**0.4432**
	SPADE	176.19	3.65	0.7900	0.5887	0.3891

The statistics echo the visual impressions described above. On the aerial dataset, SPADE stands out with lower FID and LPIPS and higher CAS and SSIM, pointing to images that look more realistic, with sharper textures and stronger structural integrity. Its spatially adaptive normalization helps capture subtle lighting and intricate textures, making the aerial scenes especially convincing. In contrast, Pix2PixHD took the lead underwater, delivering lower FID and LPIPS and higher CAS and SSIM. This suggests Pix2PixHD is more attuned to the gentle lighting and smooth textures found beneath the surface, whereas SPADE tends to boost contrast and color in murkier waters.

In summary, SPADE excels when aerial images demand intricate textures and shifting light, while Pix2PixHD proves more reliable underwater, where preserving overall structure and color is key. These results underscore the importance of matching generative models to their environment and show how domain-aware synthetic data can elevate maritime inspection.

The synthetic data utility was evaluated towards its downstream improvement by training YOLOv8 and YOLO11 under three regimes: real-only, synthetic-only, and hybrid (real + synthetic). All results are reported on held-out real test data. In addition to global metrics, per-class IoU is reported to analyze whether synthetic augmentation benefits rare or visually subtle fouling categories.

As revealed above, hybrid training achieves the most consistent improvement in mIoU across both datasets and generators, indicating that synthetic data is most effective when combined with real supervision.

Table II: Downstream training results of aerial test plates dataset

Pix2PixHD					
Model	Regime	Prec.	Rec.	F1	mIoU
YOLOv8	Real	0.8022	0.7398	0.7697	0.3568
	Synthetic	0.4653	**0.9307**	0.6205	0.4031
	Hybrid	**0.8876**	0.7822	**0.8316**	**0.5756**
YOLO11	Real	0.7861	0.7335	0.7589	0.4073
	Synthetic	0.7089	0.5545	0.6222	0.3221
	Hybrid	**0.8478**	**0.7723**	**0.8083**	**0.5318**

SPADE					
Model	Regime	Prec.	Rec.	F1	mIoU
YOLOv8	Real	0.8022	0.7398	0.7697	0.3568
	Synthetic	0.6000	0.7426	0.6637	0.3999
	Hybrid	**0.8125**	**0.7723**	**0.7919**	**0.5412**
YOLO11	Real	0.7861	0.7335	0.7589	0.4073
	Synthetic	0.6190	**0.7723**	0.6872	0.4163
	Hybrid	**0.8298**	**0.7723**	**0.8000**	**0.5334**

Table III: Downstream training results of underwater ROV inspection dataset

Pix2PixHD					
Model	Regime	Prec.	Rec.	F1	mIoU
YOLOv8	Real	0.7241	0.7778	0.7500	0.4859
	Synthetic	0.7059	**0.8889**	0.7869	0.4923
	Hybrid	**0.7500**	0.8400	**0.7925**	**0.5062**
YOLO11	Real	0.6296	0.6800	0.6538	0.4171
	Synthetic	0.7059	0.8889	0.7869	0.5577
	Hybrid	**0.7667**	**0.9200**	**0.8364**	**0.5833**

SPADE					
Model	Regime	Prec.	Rec.	F1	mIoU
YOLOv8	Real	0.7241	0.7778	0.7500	0.4859
	Synthetic	0.7692	0.3704	0.5000	0.2589
	Hybrid	**0.7931**	**0.8519**	**0.8214**	**0.5153**
YOLO11	Real	0.6296	0.6800	0.6538	0.4171
	Synthetic	0.8333	0.4000	0.5405	0.3125
	Hybrid	**0.8750**	**0.8400**	**0.8571**	**0.6019**

Table IV:

	Pix2PixHD					SPADE			
YOLOv8	Class	Real	Syn-thetic	Hybrid	YOLOv8	Class	Real	Synthetic	Hybrid
	barnacles	0.3302	0.4545	**0.5833**		barnacles	0.3302	0.4615	**0.5000**
	bryozoa	0.1298	0.2273	**0.5714**		bryozoa	0.1298	0.2857	**0.8333**
	heavy slime	0.4444	0.2500	**0.5714**		heavy slime	**0.4444**	0.1667	**0.4444**
	hydroids	0.1318	0.2000	**0.4000**		hydroids	0.1318	0.2000	**0.3750**
	light slime	0.7437	0.7727	**0.9412**		light slime	0.7437	0.7143	**0.8333**
	mussels	0.7333	0.5909	**0.9231**		mussels	0.7333	0.6000	**0.9231**
	oyster	0.0000	0.0000	0.0000		oyster	0.0000	0.0000	0.0000
	slime	0.4731	**0.5455**	0.4667		slime	**0.4731**	0.4615	0.4375
	tubeworms	0.0000	0.1429	**0.5000**		tubeworms	0.0000	0.1250	**0.3333**
	tunicates	0.7711	**1.0000**	**1.0000**		tunicates	0.7711	**1.0000**	**1.0000**
	white tunicates	0.1668	0.2500	**0.3750**		white tunicates	0.1668	**0.3846**	0.2727
	Pix2PixHD					SPADE			
YOLO11	Class	Real	Syn-thetic	Hybrid	YOLO11	Class	Real	Synthetic	Hybrid
	barnacles	0.3777	0.3333	**0.5385**		barnacles	0.3777	0.3846	**0.6923**
	bryozoa	0.3540	0.5000	**0.5714**		bryozoa	0.3540	0.3077	**0.8333**
	heavy slime	0.2862	0.1667	**0.5714**		heavy slime	0.2862	0.1429	**0.4444**
	hydroids	0.2218	**0.2727**	0.0000		hydroids	0.2218	0.2500	**0.2857**
	light slime	0.7425	0.7500	**0.8824**		light slime	0.7425	0.8421	**0.8824**
	mussels	0.7511	0.2667	**0.9231**		mussels	0.7511	0.6190	**0.8571**
	oyster	**0.4041**	0.0000	0.0000		oyster	**0.4041**	0.0000	0.0000
	slime	0.4620	0.1429	**0.5294**		slime	0.4620	**0.5333**	0.4000
	tubeworms	0.0000	0.0000	**0.5000**		tubeworms	0.0000	0.2143	**0.2500**
	tunicates	0.7502	**1.0000**	**1.0000**		tunicates	0.7502	**1.0000**	**1.0000**
	white tunicates	0.1309	0.1111	**0.3333**		white tunicates	0.1309	**0.2857**	0.2222

- Aerial Dataset (Pix2PixHD): Hybrid training substantially improves segmentation quality on real test images for both YOLO variants, with strong gains in mIoU compared to real-only training. Synthetic-only training tends to increase recall but reduces precision, suggesting over-segmentation and weaker boundary alignment under domain shift.
- Aerial Dataset (SPADE): SPADE-based synthetic data improves mIoU under hybrid training for both models, while synthetic-only training achieves moderate performance. The largest improvements are observed for rare categories such as bryozoa and tubeworms, suggesting that label-aligned synthetic examples can strengthen minority-class learning.
- Underwater Dataset (Pix2PixHD): On underwater imagery, Pix2PixHD synthetic data improves performance particularly under hybrid training, with the strongest gains observed for YOLO11. Per-class results indicate notable improvements for algae and barnacle variants, while highly sparse classes such as Damage with Rust and mixed growth remain challenging.
- Underwater Dataset (SPADE): For SPADE, synthetic-only training degrades performance, indicating that the generator alone may not reproduce underwater texture statistics sufficiently for robust generalization. However, hybrid training recovers and improves performance for both YOLO variants, showing that real samples provide essential anchoring for learning stable underwater representations.

Table V:

Pix2PixHD				
YOLOv8	Class	Real	Synthetic	Hybrid
	Algae_sparse	0.6250	0.7000	**0.7500**
	Algae_thick	0.6667	0.6667	**0.7143**
	Barnacle_sparse	0.6667	0.5714	**0.8000**
	Barnacle_thick	0.5000	**1.0000**	**1.0000**
	Damage	0.4286	**0.5000**	0.2857
	Mussels_thick	**1.0000**	0.5000	0.5000

SPADE				
YOLOv8	Class	Real	Synthetic	Hybrid
	Algae_sparse	**0.6250**	0.5714	0.5556
	Algae_thick	0.6667	0.5000	**1.0000**
	Barnacle_sparse	**0.6667**	0.0000	**0.6667**
	Barnacle_thick	0.5000	0.0000	**1.0000**
	Damage	**0.4286**	0.0000	0.4000
	Mussels_thick	**1.0000**	**1.0000**	0.5000

Pix2PixHD				
YOLO11	Class	Real	Synthetic	Hybrid
	Algae_sparse	0.5714	**0.7000**	0.6250
	Algae_thick	0.5556	0.5000	**0.8750**
	Barnacle_sparse	0.6667	**0.8333**	**0.8333**
	Barnacle_thick	0.4000	**1.0000**	**1.0000**
	Damage	0.1429	**0.4286**	0.3333
	Mussels_thick	**1.0000**	**1.0000**	**1.0000**

SPADE				
YOLO11	Class	Real	Synthetic	Hybrid
	Algae_sparse	0.5714	**0.8333**	0.6250
	Algae_thick	0.5556	0.3333	**0.8571**
	Barnacle_sparse	**0.6667**	0.0000	**0.6667**
	Barnacle_thick	0.4000	0.3333	**1.0000**
	Damage	0.1429	0.0000	**0.6667**
	Mussels_thick	**1.0000**	**1.0000**	**1.0000**

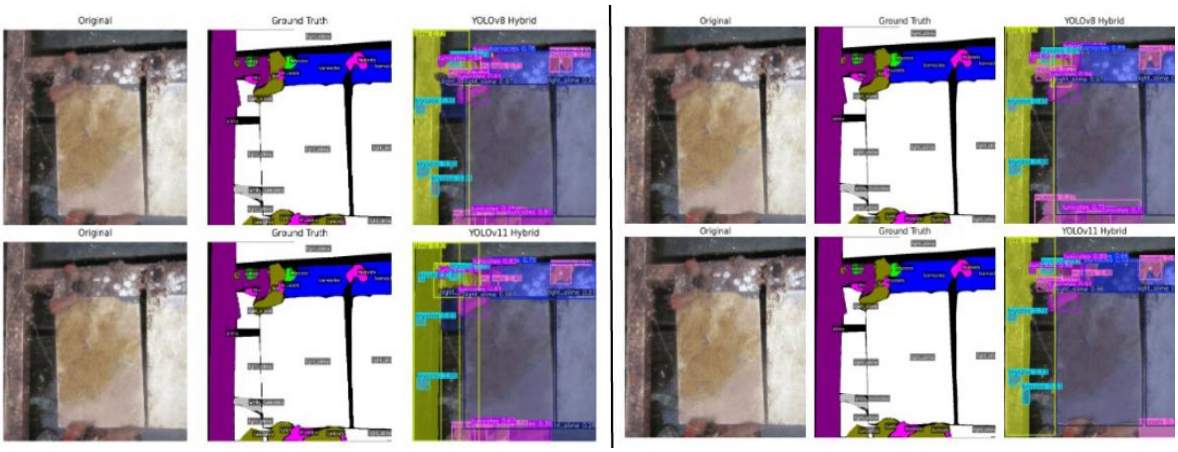


Fig.5: Downstream segmentation aerial data, left Pix2PixHD, right SPADE hybrid training

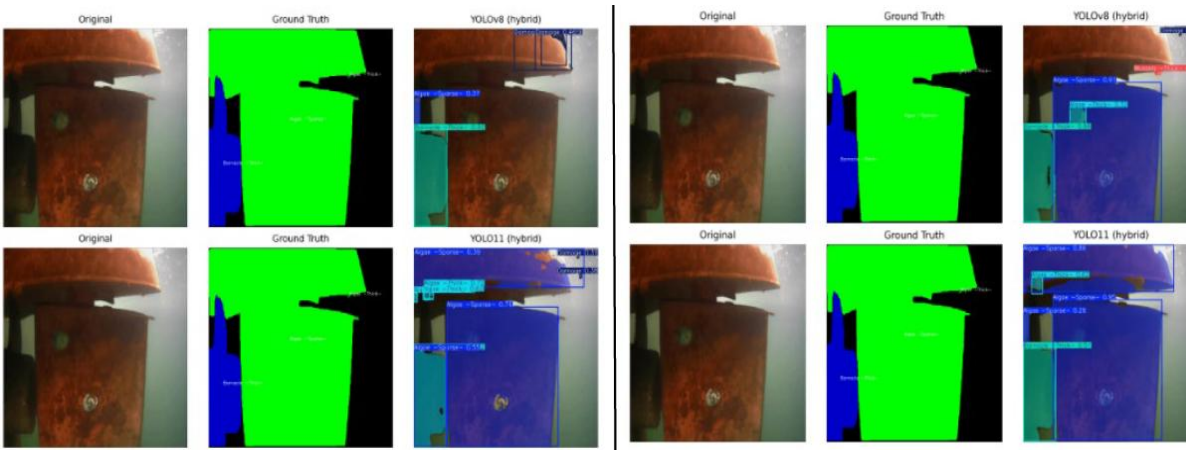


Fig.6: Downstream segmentation underwater data, left Pix2PixHD, right SPADE hybrid training

6. Conclusion

This paper demonstrates that Vision-AI can serve as a practical bridge between biofouling in-water inspection and performance-oriented shipping analytics. By combining aerial test plate imagery with real underwater ROV inspection data, the study shows that biofouling recognition is inherently domain-dependent: generator behaviour and training utility differ between above-water and underwater environments, reflecting the distinct imaging physics and operational variability that characterise each domain.

To address the central bottleneck of underwater label scarcity, the paper evaluates a mask-conditioned synthetic pipeline that preserves exact image-label correspondence through automatic label propagation, enabling scalable data augmentation without additional manual annotation. Critically, the results confirm that perceptual realism metrics alone are insufficient; synthetic outputs must be assessed against downstream task utility. Across experiments, hybrid training (real + synthetic) provides the most consistent robustness on held-out real data, yielding absolute mIoU improvements and notable benefits for minority and rare fouling classes, indicating that synthetic augmentation can expand coverage of underrepresented states when anchored by real supervision.

For current research, the implication is a methodological shift toward domain-aware synthesis selection and evaluation protocols centred on operational generalisation. For future work, priorities include scaling longitudinal micro-ROV datasets, integrating uncertainty and domain-shift monitoring in deployment, and extending controllable generation (including diffusion-based control) to better reproduce underwater texture statistics while remaining aligned with performance-reporting requirements. Overall, these findings are highly relevant for future marine biofouling research, and their implications for AI-based fouling detection are of significant importance.

Acknowledgement

This research has been funded by the Investitions und Förderbank (IFB) Hamburg as part of their Green Potential Screening initiative as part of the project Sustainable ROV Operations for Marine Fouling Identification with AI (SuROMI-AI) project no. 51171204.

References

AKKAYNAK, D.; TREIBITZ, T. (2018), *A revised underwater image formation model*, IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)

BARBOSA, L.A.; APOLINÁRIO, A.L. (2025), *From physically based to generative models: A survey on underwater image synthesis techniques*, J. Imaging, 11(5), p.161

FARKAS, A.; DEGIULI, N.; MARTIĆ, I.; VUJANOVIĆ, M. (2021), *Greenhouse gas emissions reduction potential by using antifouling coatings in a maritime transport industry*, J. Cleaner Production 295

HEUSEL, M.; RAMSAUER, H.; UNTERTHINER, T.; NESSLER, B.; HOCHREITER, S. (2017), *GANs trained by a two time-scale update rule converge to a local Nash equilibrium*, Advances in Neural Information Processing Systems 30

IMO (2023), *2023 guidelines for the control and management of ships' biofouling to minimize the transfer of invasive aquatic species*, Res. MEPC.378(80), Int. Mar. Org., London

ISO (2016), *ISO 19030-1:2016 Ships and marine technology – Measurement of changes in hull and propeller performance*, International Organization for Standardization, Geneva

KIM, B.C.; KIM, H.C.; HAN, S.; PARK, D.K. (2022), *Inspection of underwater hull surface condition using the soft voting ensemble of the transfer-learned models*, Sensors 22(12), 4392

LIN, T.Y.; MAIRE, M.; BELONGIE, S.; BOURDEV, L.; GIRSHICK, R.; HAYS, J.; PERONA, P.; RAMANAN, D.; ZITNICK, C.L.; DOLLAR, P. (2014), *Microsoft COCO: Common objects in context*, Eur. Conf. Computer Vision, pp.740-755

MAI, C.; WIELE, C.; LINIGER, J.; PEDERSEN, S. (2024), *Synthetic subsea imagery for inspection under natural lighting with marine-growth*, Ocean Eng. 313, 119284

O'BYRNE, M.; PAKRASHI, V.; SCHOEFS, F.; GHOSH, B. (2018), *Semantic segmentation of underwater imagery using deep networks trained on synthetic imagery*, J. Marine Science and Eng. 6(3), p.93

PARK, T.; LIU, M.Y.; WANG, T.C.; ZHU, J.Y. (2019), *Semantic image synthesis with spatially-adaptive normalization*, IEEE/CVF Conf. Computer Vision and Pattern Recognition

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. (2016), *You only look once: Unified, real-time object detection*, IEEE Conf. Computer Vision and Pattern Recognition, pp.779-788

SCHULTZ, M.P.; BENDICK, J.A.; HOLM, E.R.; HERTEL, W.M. (2011), *Economic impact of biofouling on a naval surface ship*, Biofouling 27(1), pp.87-98

STEIN, M. (2023), *How the micro ROV class will change the maritime sector: An introductory analysis on ROV, big data and AI*, Autonomous Vehicles: Applications and Perspectives

TOBIN, J.; FONG, R.; RAY, A.; SCHNEIDER, J.; ZAREMBA, W.; ABBEEL, P. (2017), *Domain randomization for transferring deep neural networks from simulation to the real world*, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)

WANG, T.C.; LIU, M.Y.; ZHU, J.Y.; TAO, A.; KAUTZ, J.; CATANZARO, B. (2018), *High-resolution image synthesis and semantic manipulation with conditional GANs*, IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)

WANG, Z.; BOVIK, A.C.; SHEIKH, H.R.; SIMONCELLI, E.P. (2004), *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Processing 13(4), pp.600-612

WASZAK, M.; CARDAILLAC, A.; ELVESÆTER, B.; RØDØLEN, F.; LUDVIGSEN, M. (2022), *Semantic segmentation in underwater ship inspections: Benchmark and dataset*, IEEE J. Oceanic Eng.

ZHANG, R.; ISOLA, P.; EFROS, A.A.; SHECHTMAN, E.; WANG, O. (2018), *The unreasonable effectiveness of deep features as a perceptual metric*, IEEE Conf. Computer Vision and Pattern Recognition, pp.586-595

Performance Verification Methodologies for Vessels with Wind Assisted Propulsion

Josef Camilleri, Anemoi Marine Technologies, London/UK, josef.camilleri@anemoimarine.com
Luke McEwen, Anemoi Marine Technologies, London/UK, luke.mcewen@anemoimarine.com
Santiago Suarez de la Fuente, Lloyd's Register, London/UK, santiago.suarezdelafuente@lr.org

Abstract

The paper compares three methods for verifying wind-assist performance: ANEMOI in-service ON/OFF verification and model calibration, ITTC sea-trial and performance-prediction guidelines (7.5-04-01-02; 7.5-02-03-01.9), and DNV in-service ON/OFF practice (DNV-RP-0686). Although all use ON/OFF comparisons, they differ in objective, scope, treatment of environmental variability, and approaches to uncertainty and extrapolation. The analysis finds the methods complementary: ITTC offers a controlled baseline, DNV a statistically rigorous long-term framework, and ANEMOI shows how measurements plus modelling yield actionable fuel-saving predictions. It is intended to inform, and provoke, debate on evolving verification practice for wind assistance installations and standards.

1. Introduction

Maritime transport is a critical enabler of global trade, carrying over 80% of goods traded worldwide, *UNCTAD (2024)*, but it is also a significant and growing source of emissions, with shipping's share of global anthropogenic greenhouse gas (GHG) emissions increasing from 2.76% in 2012 to 2.89% in 2018, *IMO (2020)*. In response, increasingly stringent decarbonisation targets and regulatory frameworks are stimulating interest and accelerating the adoption of wind-assisted propulsion as a practical and compelling fuel-saving measure for energy efficiency. A growing number of wind assisted propulsion systems (WAPS) are now available on the market, including rotor sails, wing sails, and suction sails, which generate aerodynamic thrust that reduce net propulsion power demand and/or to supplement propulsion power at a given power setting, *LR (2024)*.

As installations of WAPS increase, there is a growing requirement for performance verification methodologies that are technically credible, operationally practical, and comparable across vessels, routes and operating profiles. Unlike conventional speed-power trials, WAPS performance is strongly dependent on apparent wind conditions (speed and angle), and verification approaches must address wind measurement difficulties, coupling between aerodynamic and hydrodynamic response, and the challenge of isolating small performance benefits from environmental variability.

1.1. Purpose of Paper

This paper evaluates three representative methodologies of performance verification of wind assistance installations:

- ANEMOI, Performance Verification of Wind-Assisted Ship Propulsion Systems by ON/OFF Testing, *Anemoi (2025)*, verified by Lloyd's Register, hereinafter; 'ANEMOI verification'
- ITTC, Recommended Procedures and Guidelines 7.5-04-01-02 - Sea trials for assessing the power saving from wind assisted propulsion, *ITTC (2024a)*, together with ITTC, Recommended Procedures and Guidelines 7.5-02-03-01.9 - Predicting the Power Saving of Wind Powered Ships, *ITTC (2024b)*, hereinafter; 'ITTC guidelines'
- DNV, Recommended Practice DNV-RP-0686 - Performance of wind assisted propulsion systems, *DNV (2025)*, hereinafter; 'DNV practice'.

The purpose of this paper is to compare and clarify differences, identify gaps and limitations, of each methodology and provide recommendations to support a possible need for a future industry-wide standardisation of performance verification of wind assistance installations.

Alternative approaches not addressed in this paper are briefly discussed in Appendix 2.

"Performance verification" is used in a broad sense, covering controlled verification (e.g., sea trials), statistically defensible in-service assessment, and where applicable the calibration/validation of a prediction method for voyage- or route-level savings. This paper assumes wind assistance installation is fully available (i.e., no downtime affecting delivered savings); the impact of system availability is outside the scope of this paper.

2. Challenges with Performance Verification

Changes in vessel performance with a wind assistance installation cannot be interpreted as a simple "added thrust" problem without a clearly defined analytical framework where:

- Wind-assisted propulsion alters the coupled force and moment balance of a vessel, and
- Aerodynamic thrust and side force influence propeller loading, propulsive efficiency, leeway, rudder demand, and hydrodynamic resistance.

The key challenges with performance verification are presented in Fig.1.

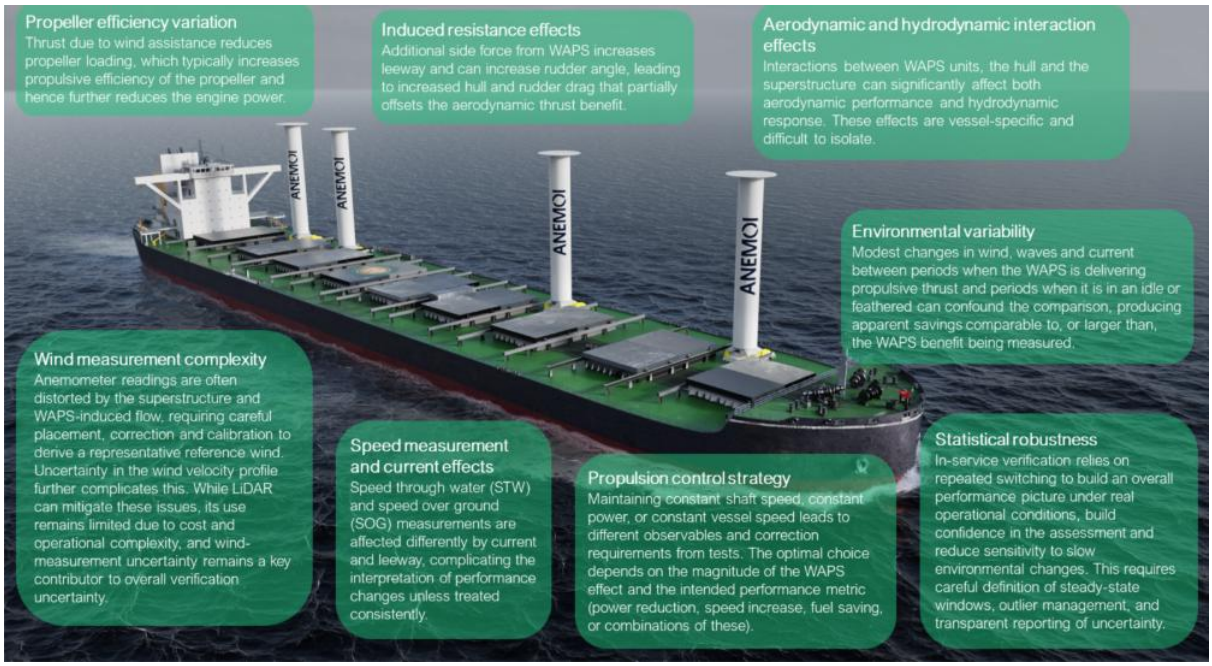


Fig.1: Key challenges with performance verification of vessels with wind assistance installations

3. Overview of Test Methodologies for Performance Verification

Fig.2 illustrates a generic performance verification workflow applicable to all three methodologies of performance verification evaluated. All three methodologies seek to quantify the performance impact of WAPS by comparing vessel behaviour with the system active (ON) and inactive (OFF), under broadly comparable environmental and operational conditions. However, they differ significantly in how tests are executed, how environmental variability is treated, how uncertainty is quantified, and whether measured results are extrapolated beyond the test conditions.

Fig.2: Flowchart showing the performance verification workflow

The differences between methodologies are discussed in more detail in Section 4 along with recommendations to draw together the most effective elements to inform a possible debate on consolidated methodologies.

3.1. Anemoi Verification Methodology

The ANEMOI verification methodology is designed for in-service performance verification using repeated ON/OFF tests during normal vessel operation, with the explicit objective of both measuring performance changes and developing and validating a vessel-specific Performance Prediction Program (PPP), *Anemoi (2025)*.

Data quality assurance is central to the approach. ON/OFF comparisons are structured as paired baseline (OFF) and test (ON) windows, with the OFF condition represented by the time-weighted average of two baselines taken before and after the test window. Fig.3 illustrates an example OFF-ON-OFF test sequence in which measured power and speed exhibit a slow underlying drift, showing how a single "OFF" snapshot can misrepresent the true baseline during the ON window. Using two baselines (before and after) enables a time-weighted interpolation of the expected OFF condition at the test time, materially reducing scatter of results and inaccuracy due to slow environmental changes.

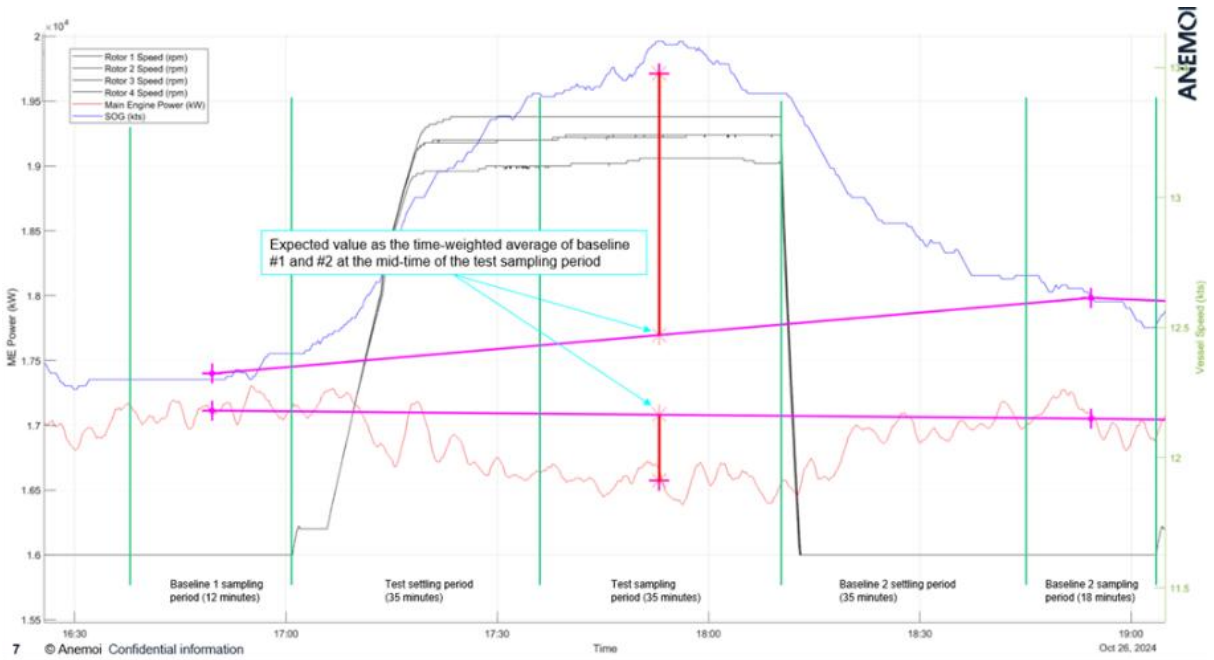


Fig 3: Example OFF-ON-OFF test sequence illustrating slow baseline drift and the use of pre- and post-OFF baselines to interpolate the expected OFF condition during the ON window (pink markers and connecting lines), with red vertical lines indicating the deviation of the ON condition from the interpolated OFF reference for vessel speed and main engine power.

Multiple anemometers are used, typically located forward and aft, with directional selection logic applied to identify the least disturbed sensor for a given apparent wind direction. A defined correction and calibration workflow is then applied to the on-board apparent wind measurements to derive a consistent upstream reference wind, including bias-correction of the measured wind time series against hindcast datasets (e.g., ERA5), Fig.4, and scaling to sail mid-height using a power-law vertical profile. This reference-wind calibration anchors the wind input used to derive non-dimensional forward force coefficients and allows the calibrated performance model to be applied in voyage simulation against standard hindcast data. Current effects are estimated and compensated so that comparisons are made on a speed-through-water basis (assuming the current is steady or varies approximately linearly over the test window).

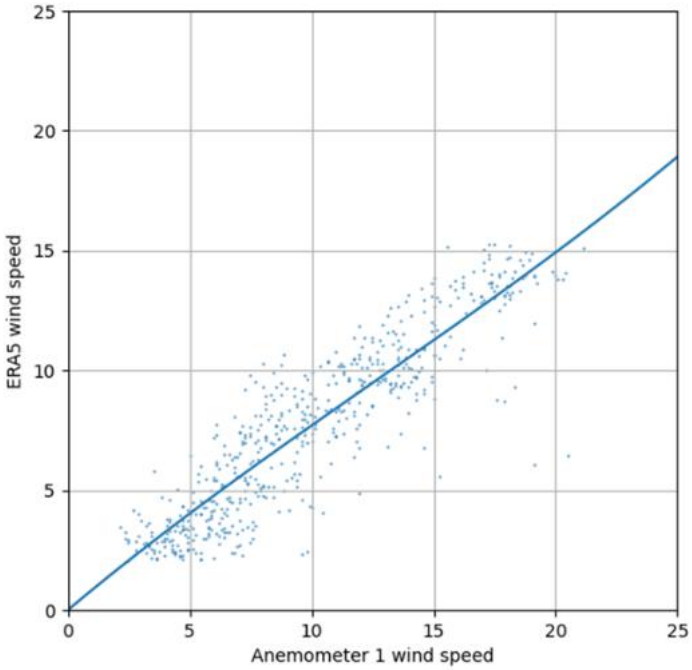


Fig.4: Example reference-wind calibration using ERA5 hindcast data: anemometer wind speed compared against ERA5 to derive the upstream wind correction applied in subsequent analysis

The change in net forward force due to WAPS is inferred from measured changes in the propulsion operating point. Propeller open-water characteristics, Fig.5, together with measured shaft speed and torque, and thrust deduction are used to estimate effective thrust and net forward force. Where ON/OFF speeds differ, the methodology estimates the WAPS-OFF resistance at the WAPS-ON speed using a quadratic resistance-speed relationship ($R \propto V^2$), with an exponent of two generally appropriate at low Froude numbers or using the resistance-speed relationship from model tests.

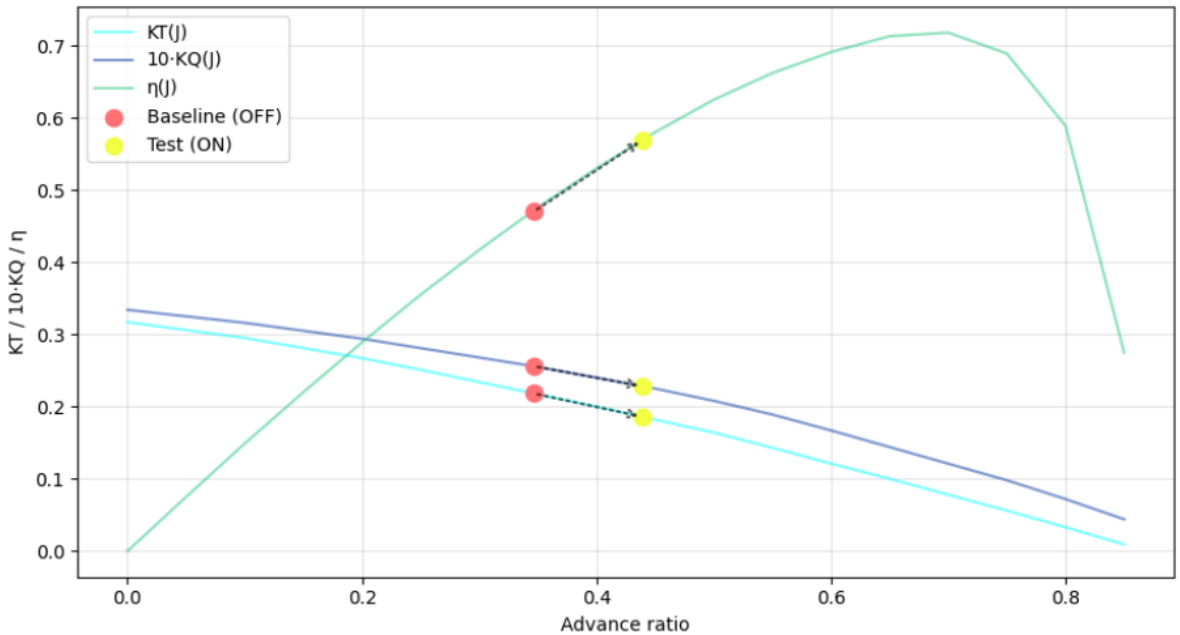


Fig.5: Typical propeller open-water characteristics (KT, KQ and efficiency) with representative ON/OFF operating points overlaid. The arrows indicate the shift in propeller operating point from baseline (rotors OFF) to test condition (rotors ON), illustrating the increase in advance ratio and associated decrease in torque coefficient and thrust coefficient and the significant increase in propeller efficiency.

Using the open-water curves captures the change in advance ratio and open-water propeller efficiency as the WAPS thrust unloads the propeller, improving the fidelity of the inferred force balance. As thrust is derived directly from measured shaft torque and speed (rather than from a local speed-power relationship), the method remains applicable even when the ON/OFF response produces a material change in vessel speed and is less sensitive to any change in the local speed-power slope between OFF and ON conditions caused by altered propeller loading. The measured change in net forward force between OFF and ON conditions is then compared with PPP predictions. Discrepancies are used to calibrate and adjust the PPP model, producing a performance-verified model of the vessel with WAPS.

Once calibrated, the PPP is used to extrapolate performance beyond the tested conditions and to generate route- and voyage-specific predictions of fuel and emissions savings. This enables gaps in the measured dataset (e.g. apparent wind angles between tested values) to be addressed through model-based extrapolation, while retaining traceability to full-scale verification data.

3.2. ITTC 7.5-04-01-02 Methodology

The ITTC guidelines define a standardised sea-trial procedure intended to verify the power savings delivered by a WAPS under controlled conditions (7.5-04-01-02), *ITTC (2024a)*, and for predicting the power savings of wind-powered ships (7.5-02-03-01.9), *ITTC (2024b)*. Verification is achieved through paired runs with and without the WAPS active, typically conducted over a limited number of wind directions dependent on the test day wind conditions but can be extended to include a larger number of conditions conducted over a longer period during operation. The approach allows for early confirmation of the WAPS performance under the observed conditions.

A baseline speed-power relationship without WAPS is used as a counterfactual baseline mapping to convert observed differences in speed or power between paired runs into a consistent “power difference due to WAPS” at the trial condition. This baseline may be derived from a conventional speed trial, model tests, or a speed-variation test conducted in conjunction with the wind-assisted sea trial.

The ITTC procedure prioritises control of test conditions and measurement quality over post-test correction. It does not define a formal correction methodology for local wind disturbance caused by superstructure or WAPS, instead recommending mitigation measures such as multiple anemometers, LiDAR measurements, or CFD-informed sensor placement. Similarly, current correction is deliberately limited; the preferred speed reference (speed over ground or speed through water) is selected based on local current variability and instrument reliability rather than applying analytical current corrections. Corrections are applied narrowly to maintain run-pair comparability, including air-resistance alignment when wind varies within a pair of runs and accounting for idling/non-retractable WAPS resistance in the OFF state.

Post-processing follows a structured sequence. The power difference attributable to WAPS is first derived at the sea-trial conditions and compared to the model predictions at the same conditions to validate the prediction method. The prediction method should be selected and implemented in accordance with ITTC 7.5-02-03-01.9, *ITTC (2024b)*, which outlines approaches for predicting wind-assisted power savings across design stages and links standard indicators to procedures of increasing fidelity (and computational cost). However, it is primarily a framework and does not provide detailed procedures for performance predictions or model calibration. If agreement is satisfactory, the validated model may be used to extrapolate performance to broader wind conditions through a power-saving matrix and voyage simulation. The ITTC procedure explicitly recognises that uncertainty in both the sea-trial procedure and performance predictions is not yet well established and therefore advises against using sea trials alone to confirm contractual performance guarantees. Nonetheless, it strongly recommends standardised sea trials for each installation to provide early, comparable verification.

3.3. DNV RP-0686 Methodology

The DNV Recommended Practice (DNV-RP-0686) is designed primarily for repeated in-service ON/OFF switching tests conducted during normal vessel operation, *DNV (2025)*. Its objective is to quantify WAPS performance during the encountered environmental conditions with explicit treatment of statistical uncertainty and confidence intervals, rather than to validate or calibrate a predictive performance model. The DNV practice defines a framework structured around specific "use cases" ranging from confirmation of favourable-condition performance to long-term operational evaluation. Two alternative test procedures are specified, with the choice determined by the magnitude of the vessel speed change when switching the WAPS ON or OFF.

- Procedure I (benefit by reduced power) applies when the ON/OFF speed difference is less than approximately one knot. In this case, performance benefit is quantified by comparing shaft power at near-constant vessel speed. Residual speed differences are corrected using a representative speed-power curve; the curve should be based on WAPS ON but may be based on WAPS OFF if the speed change is less than one knot and the relative ΔP-ΔV behaviour is expected to be similar. The curve may be derived from sea trials, model tests, CFD, or supporting ON-ON and OFF-OFF tests conducted as part of the campaign. Shaft-speed control strategies are prescribed to limit speed deviations and improve comparability between ON and OFF periods.
- Procedure II (benefit by increased speed) applies when the ON/OFF speed difference exceeds one knot. Performance is quantified using energy consumption per distance, with shaft speed set to representative operating points for conventional and WAPS operation, based on the baseline speed-power relationship.

Environmental conditions (wind, waves, current), draft and trim, and hull/propeller condition (fouling and roughening) are assumed to be effectively equal between ON and OFF periods. The analysis does not apply analytical corrections for variation in these factors, because such corrections can introduce additional uncertainty and potential bias (e.g., when normalising results to a reference condition). This is to reduce the uncertainty associated with such corrections, including avoiding bias in the correction to a normalised condition. Instead, it relies on many tests, robust statistical filtering to validate steady-state conditions, and statistical aggregation to reduce the influence of uncontrolled variability. Uncertainty quantification is integral to the methodology, with confidence intervals derived using Student's t-distribution and minimum observation requirements linked to the intended use case and coverage of tested conditions.

DNV practice focuses on verification of in-service WAPS performance rather than supporting the development of predictive performance models. As a result, the performance benefits demonstrated are limited to conditions that are represented in the switching dataset. Where favourable wind conditions occur infrequently during the test period, the assessed savings may be conservative compared to the WAPS's broader operational potential. Conversely, where favourable wind conditions occur often during testing, the assessed savings may be optimistic. Extended verification periods can mitigate this by increasing condition coverage, although increased time spent with the WAPS inactive or in idle mode may influence overall utilisation and hence savings during the verification phase.

3.4. Simplified Comparison of Anemoi, ITTC & DNV Methodologies

A simplified and consolidated comparison of Anemoi verification, ITTC guidelines and DNV practice is shown in Table I. A detailed comparison of the three methodologies is provided in Appendix 1.

Table I: Simplified & consolidated comparison of the three methodologies

Area	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686
Purpose and Scope			
Verification Objectives	Long-term in-service verification (+ model calibration)	Short-term sea trial verification	Long-term in-service evaluation (+ statistical confidence)
Intended Use	Calibrate fuel-saving PPP model (FSAM)	Verify WAPS power saving	Quantify WAPS performance (agreed use case)
Location	In-service trading routes	Dedicated trial area or normal-service voyage (boundary conditions controlled)	In-service routes (may include sea trials)
Core Methodology and Performance Definition			
Logic (ON/OFF)	ON/OFF comparison	ON/OFF comparison	ON/OFF comparison
Performance Metric	Net change in forward force	Power difference (ΔP)	Procedure I: power saving (ΔV <1 kn) Procedure II: energy per distance (ΔV>1 kn)
Speed Handling	Allows speed changes; infers thrust from shaft signals + prop curves; corrects for resistance change with speed squared or model test resistance	Correct speed change using speed-power curve; minimise ΔV	Proc I speed correction using speed-power curve; Proc II uses energy/distance
Test Design and Operational Envelope			
Conditions	Quasi-steady conditions	Defined trial boundary conditions (wind/sea state/water depth)	Many tests across operational profile (filtered for stationarity)
Wind speed	AWS ≥10 kn	TWS ≥ 8 m/s (or sufficient for measurable effect)	Coverage agreed per use case
Wind angle	AWA from ahead to astern	Min. 5 wind angles (>10° apart)	Broad polar coverage
Wind Measurement	At least two anemometers	Ship anemometer + LiDAR recommended (measure and/or calibrate); CFD/multiple anemometers to minimise distortion	At least one anemometer; LiDAR preferred (particularly if multiple WAPS installed)
Control Logic	Constant RPM & heading	Steady course & min. rudder movement; constant power, shaft speed or vessel speed	Formalised control logic to minimise speed-correction uncertainty
Corrections and Uncertainty			
Wind Correction	Select least-disturbed anemometer; interference/height corrections; calibration using hindcast data	LiDAR/calibration emphasis; placement via CFD/multiple anemometers	Consistent wind measurement; LiDAR/anemometer quality focus
Variability Management	Accept in-service variability; corrections + OFF-ON-OFF (averaging of two OFF baselines)	Minimise variability via site + boundary-control; limited corrections	Assumes environment effectively equal; many tests + statistical filtering/aggregation
Uncertainty and Model Use	Diagnostic + PPP calibration; supports extrapolation	Uncertainty not yet established; not for contractual guarantees; validates external matrix/PPP	Formal uncertainty framework; report statistical bounds; prediction optional

4. Three Methodologies Compared: Findings & Recommendations

This section summarises the key findings from comparing the three methodologies, focusing on practical implications for performance verification and standardisation. The recommendations draw together the most effective elements across the three approaches for a consolidated methodology.

4.1. Scope and Application

All three methodologies are founded on ON/OFF comparison logic, but they diverge in objective and scope. The ITTC procedure is explicitly framed as a verification tool, intended to confirm wind propulsion performance in controlled conditions and to validate pre-existing performance prediction matrices. However, it also includes a pathway to estimating the power saving for all weather conditions and supporting voyage simulation, provided that a power-saving matrix has been derived (e.g. in accordance with ITTC 7.5-02-03-01.9, *ITTC (2024b)*) and that the comparison between prediction and sea-trial results is satisfactory. In this sense, ITTC acts as a verification and a validation gate for an external prediction model, but it does not prescribe how that model (or matrix) should be formulated or calibrated; if agreement is not satisfactory, the prediction provider is expected to update the complete matrix, without a defined calibration procedure within the ITTC guidelines itself.

In contrast, both the ANEMOI and DNV approaches are oriented toward in-service application, recognising that representative coverage of operational and environmental conditions cannot realistically be achieved through short-duration trials alone. The ANEMOI methodology further extends beyond verification by explicitly using ON/OFF test results to calibrate a vessel-specific performance prediction model, enabling extrapolation to voyage- and period-level fuel savings.

- **Recommendation:** Where verification is intended to reflect real-world operation, in-service testing under representative conditions is recommended to provide a realistic picture of WAPS performance. If a quick and controlled verification of performance is required, but limited to the conditions encountered during the trials, then the ITTC approach provides the steps needed to quantify the WAPS performance for the trial conditions and to validate the performance predictions at those conditions. Where results are expected to inform operational or commercial decisions beyond the tested conditions, verification should be paired with an accompanying performance prediction methodology that supports model development and calibration for prediction beyond the measured envelope. Given that detailed modelling procedures are difficult to standardise across organisations, such a methodology should require a clearly documented and auditable approach (including defined scope and assumptions, data quality/coverage expectations, calibration and change-control procedures, and stated limits on extrapolation) and include model evaluation and governance requirements including validation/acceptance metrics, uncertainty reporting and benchmarking against shared reference cases and/or curated datasets.

4.2. Performance Metrics

The ANEMOI verification methodology adopts a force-based analysis, deriving the net change in forward force from measured propulsion response. ITTC guidelines express performance primarily as a power difference at the trial condition, optionally normalised to a reference speed consistent with conventional speed-power trial practice. The DNV practice quantifies benefit either as power saving or energy per distance (depending on the operational strategy of the vessel with WAPS) and reports results as statistically bounded estimates rather than deterministic point values. These differences reflect distinct priorities: diagnostic understanding and prediction (ANEMOI), controlled verification at installation (ITTC), and statistically robust in-service assessment (DNV).

A practical advantage of force-based metrics is reduced dependence on speed-power curve use when ON/OFF speeds differ, because thrust is inferred from shaft signals using propeller open-water characteristics, making the propeller unloading effect and associated change in operating point and efficiency explicit. For power- and energy-based metrics, the speed-power curve introduces two separable sensitivities that should be controlled in the verification protocol. First, when ON and OFF speeds differ and results are referenced to a common speed, the corrected power saving depends on the speed-power curve used and on the magnitude of the ON/OFF speed difference. Second (implication), where a WAPS-OFF speed-power curve is used to evaluate the baseline power at the

ON speed, the baseline point generally corresponds to the OFF-propulsion setting required to achieve the ON speed (often a different shaft rate and therefore a different propeller operating point than the paired constant-setting trial point). The reported power saving therefore combines (i) the measured change in shaft power at the trial setting with (ii) a curve-based estimate of the additional power required to reach the ON speed in OFF operation and should not be interpreted as a decomposition at a single propeller operating point. Power-based metrics may nevertheless be more readily understood by stakeholders who are more familiar with kW than kN.

- **Recommendation:** Select performance metrics consistent with the intended use. Force-based metrics are preferred where diagnostic insight, model calibration, and reliable extrapolation are required. Where power- or energy-based metrics are used for reporting, the verification protocol should define (i) how ON/OFF speed differences are handled, (ii) how the speed-power curve is derived and applied (including whether it is based on ON or OFF operation), (iii) acceptable limits on ON/OFF speed difference to control sensitivity of corrected results to curve choice and local speed-power behaviour, and (iv) how changes in propulsive efficiency are accounted for due to WAPS unloading the propeller, both in the tests and in the PPP for voyage simulation.

4.3. Environmental Variability

The ITTC methodology seeks to minimise variability through test-site selection and boundary-condition control, deliberately limiting analytical corrections. The DNV practice takes the opposite stance, treating environmental conditions as effectively equal between ON and OFF periods and relying on large numbers of tests, robust statistical filtering to validate steady-state conditions, and statistical aggregation to manage residual variability. The ANEMOI verification methodology occupies an intermediate position, accepting the variability of in-service conditions but applying correction and normalisation for wind distortion and current effects to improve comparability between test pairs, and using two OFF periods for each ON period (OFF-ON-OFF) to reduce noise from approximately linearly varying environmental effects.

These choices have direct implications for data requirements: ITTC guidelines depend on carefully planned trials (with uncertainty driven by wind conditions on the test day); DNV practice depends on large sample sizes; and ANEMOI depends on structured data processing and auxiliary information such as hindcast datasets.

- **Recommendation:** For in-service verification under variable conditions, the authors' view is that applying correction and normalisation is preferable to relying on a very large number of tests or extensive filtering to manage environmental variability. This preserves more usable data and avoids unnecessary testing and data rejection which reduces WAPS uptime and realised savings. Where "controlled" conditions are assumed, verify this against recorded measurements and apply corrections where deviations are material. In particular: (i) baseline-drift correction (OFF-ON-OFF): represent the OFF condition during each ON window by time-weighted interpolation between OFF periods either side of the ON period, rather than relying on a single OFF snapshot; (ii) wind-measurement correction: apply a consistent anemometer selection and correction workflow (least-disturbed sensor by direction, interference/height corrections, and, where available, calibration against an upstream reference such as LiDAR) so apparent wind inputs are comparable between ON and OFF and suitable for comparison to predictions based on hindcast wind data.

4.4. Uncertainty Treatment

The ITTC procedure explicitly acknowledges that uncertainty is not yet well established and advises against contractual use of sea-trial results, positioning transparency and standardisation ahead of quantified confidence. The DNV practice places uncertainty quantification at the centre of the methodology, with formal confidence intervals and minimum data requirements linked to defined use

cases. The ANEMOI verification methodology manages uncertainty implicitly through data filtering, paired test design, and model calibration quality, but does not prescribe a formal probabilistic uncertainty framework as a primary output. As a result, DNV practice is currently best aligned with regulatory or third-party verification contexts that require explicit confidence bounds, while ANEMOI verification is optimised for performance understanding and prediction.

- **Recommendation:** Where verification outcomes are intended for third-party review, regulatory purposes, or contractual decisions, explicit and transparent uncertainty quantification is recommended (e.g., measurement and sampling uncertainty, and where applicable environmental correction/model uncertainty and statistical confidence).

4.5. Prediction and Extrapolation

The ITTC methodology treats prediction as a supporting element to the sea-trial, using ITTC 7.5-02-03-01.9, *ITTC (2024b)*, to define the nominated prediction at the trial conditions and then to extend results to route or representative weather savings via a power saving matrix and voyage simulation once agreement is demonstrated. Sea-trial results are used to validate consistency with those predictions, and where agreement is not satisfactory, the prediction provider is expected to update the complete matrix (not only the tested conditions), while any updates to the prediction are treated as an external modelling step rather than prescribed within the trial procedure. DNV practice treats modelling as optional and secondary, mainly for comparison or operational insight rather than as part of the verification deliverable. By contrast, the ANEMOI methodology links measurement and prediction in a closed loop, using ON/OFF tests to calibrate vessel-specific coefficients before applying the validated model to long-term route simulations, enabling extrapolation beyond the tested envelope but increasing reliance on model structure, corrections, and calibration quality.

- **Recommendation:** Where performance verification outcomes are intended to inform long-term, voyage-specific, or commercial decisions, a model-based approach is recommended as it enables extrapolation beyond the tested conditions, thereby reducing the volume of testing required and maximising realised savings. The methodology should require a clearly documented and auditable, step-by-step prediction workflow that defines how sea-trial or in-service evidence are used to calibrate and validate prediction models, with explicit requirements for full-scale calibration, validation and acceptance criteria, and transparent reporting of model assumptions, scope, and limits of extrapolation.

5. Conclusions

This paper has compared three established approaches for verifying the performance of WAPS: the ANEMOI in-service ON/OFF verification and model-calibration methodology, *Anemoi (2025)*, the ITTC guidelines for sea trials and supporting performance prediction (ITTC 7.5-04-01-02 and ITTC 7.5-02-03-01.9), *ITTC (2024a, b)*, and the DNV practice for in-service ON/OFF testing (DNV-RP-0686), *DNV (2025)*. While all three rely on ON/OFF comparison logic, they differ in objective, scope, treatment of environmental variability, and how uncertainty and extrapolation are handled.

The ITTC guidelines are aimed at controlled, short-term verification of performance predictions at the trial conditions, with separate guidance outlining standards-level expectations for building performance prediction models across different fidelity levels (without prescribing implementation details). The DNV practice provides a statistically robust framework for long-term in-service assessment. The ANEMOI methodology integrates medium-term in-service ON/OFF verification with an explicit vessel-specific calibration step, to support voyage- and route-level past and future fuel savings estimation.

Overall, the comparative analysis shows that the three methodologies are complementary rather than competing. ITTC guidelines provide a structured and controlled verification baseline, DNV practice provides a statistically rigorous framework for long-term in-service assessment, and ANEMOI

verification methodology demonstrates how in-service measurements can be combined with modelling to deliver actionable fuel-saving predictions.

Future standardisation efforts would benefit from combining controlled verification principles, explicit uncertainty treatment, and validated performance prediction within a unified framework. Such convergence would improve comparability, strengthen confidence in reported savings, and support wider adoption of WAPS.

Acknowledgement

We would like to thank Rob Tustin (LR), Sofia Werner (RISE), Johanna Tranell, Eivind Ruth, and Heikki Hansen (DNV) for their careful review of the manuscript and for providing valuable feedback.

References

ANEMOI (2025) *Performance Verification of Wind-Assisted Ship Propulsion Systems by On-Off Testing*, Anemoi, London, https://anemoimarine.com/wp-content/uploads/2025/07/Anemoi-LR-Performance-Verification-Paper.pdf

DNV (2025) *Recommended Practice DNV-RP-0686 - Performance of wind assisted propulsion systems*, DNV, standards.dnv.com/explorer/document/6D3EBF2A07294448927ECA668691FC1D/2

IMO (2020) *Fourth Greenhouse Gas Study 2020*, Int. Mar. Org., London, www.imo.org/en/ourwork/environment/pages/fourth-imo-greenhouse-gas-study-2020.aspx

ITTC (2024a), *Recommended Procedures and Guidelines 7.5-04-01-02 – Sea trials for assessing the power saving from wind assisted propulsion*, ittc.info/media/11974/75-04-01-02.pdf

ITTC (2024b), *Recommended Procedures and Guidelines 7.5-02-03-01.9 - Predicting the Power Saving of Wind Powered Ships*, ittc.info/media/11806/75-02-03-019.pdf

LR (2019), *Maersk Pelican Rotor Sail Sea Trial Report (1706-0009 - D2.7), Lloyd's Register*

LR (2024) *Energy Efficiency Retrofit Report: Applying Wind-Assisted Propulsion to Ships*, https://maritime.lr.org/WAPS-report

UNCTAD (2024) *2024 Review of maritime transport*, https://unctad.org/publication/review-maritime-transport-2024

VOUTILAINEN, J.; PAAKKARI, V. (2025), *Continuous Performance Measurement of Wind Propulsion*, 10th HullPIC Conf., Mülheim, pp.141-149, http://data.hullpic.info/HullPIC2025_Mulheim.pdf

Appendix 1: Detailed Comparison of ANEMOI Verification, ITTC Guidelines & DNV-RP-0686 Methodologies

Table II: Objective and performance metric(s)

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Objective and intended use	Verify in-service WAPS performance and calibrate vessel and WAPS model using ON/OFF tests; use calibrated model (FSAM) for voyage fuel-saving prediction.	Short-term sea trial to verify power saving due to WAPS at limited conditions; used to validate a performance prediction matrix.	Framework for performance evaluation/verification of WAPS based on documentation and measured data; supports sea trial and in-service assessment with uncertainty estimation.	All use ON/OFF comparison logic; ITTC is short-term verification, ANEMOI and DNV are in-service/long-term oriented.
Performance Metric	Net change in forward force (and non-dimensional coefficient); net fuel saving from calibrated FSAM.	Power difference (ΔP) at sea-trial conditions; optional normalisation to reference condition; optional voyage average power saving.	Procedure I: power saving (speed change <1 knot). Procedure II: energy per distance (speed change >1 knot). Performance reported as confidence interval; fuel saving optional/not focus.	ANEMOI is force-based; ITTC is power-based; DNV is power or energy-per-distance depending on operating mode.

Table III: Test conditions

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Test location and boundary conditions	Normal trading routes; accept variable environment but require quasi-steady conditions within each ON/OFF test.	Normal-service voyage between port calls or dedicated trial area, with defined boundary-condition requirements (deep water, avoid land effects, manage current influence, etc.).	Typically, in-service routes; may include sea trials; long-term evaluation may use planned/random ON/OFF tests over representative operational pattern.	ITTC minimises variability via site selection; ANEMOI and DNV manage variability via filtering and/or large sample sizes.
Wind conditions	Apparent wind speed ≥10 kn (5 m/s, ideally 8 m/s); cover apparent wind angle from ahead to astern, both sides; include some negative-thrust angles if feasible.	TWS should allow a measurable effect (≥0.3 kn speed change) or be at least 8 m/s; cover at least 5 wind angles spaced by more than 10°, covering thrust peak and zero-thrust regions; if performance is asymmetric, test five angles per side.	Coverage agreed per use case; performance results should be supported by broad polar coverage, including zero/negative benefit conditions in long-term evaluation.	ANEMOI and ITTC require sufficient wind to produce measurable signal; ANEMOI gives explicit AWS guidance; ITTC gives true-wind guidance; DNV emphasises coverage and stationarity rather than minima.
- Propulsion/ steering control	Keep main engine RPM and ship heading constant during ON/OFF; measure speed and power response.	Maintain steady course with minimal rudder change as per the test plan; propulsion can be controlled by constant power, shaft speed or vessel speed.	Minimise vessel speed change between ON and OFF; maintain constant shaft RPM within each period. Propeller shaft speed during ON/OFF can be: (1) held constant if speed change <1 kn, (2) adjusted to keep total energy input (shaft power + WAPS operating power) constant, or (3) adjusted to maintain equal vessel speed. Maintain constant P/D for CPP; maintain a steady vessel course throughout each set.	DNV formalises control logic to minimise speed-correction uncertainty; ANEMOI is agnostic to propulsion control strategy; ITTC allows multiple control modes but is less prescriptive on speed matching. All require steering control.

Table IV: Measurement framework

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Key measurements	Shaft power/torque/RPM; SOG and STW; AWS/AWA (selected anemometer); rotor RPM and PTI; weather data for air density; time synchronised.	GNSS track/COG/SOG; STW; shaft power/RPM; propeller pitch (CPP); heading; AWS/AWA; WAPS PTI and setting; time (UTC); waves/sea state as available.	AWS/AWA; vessel speed (SOG/STW and/or virtual speed); shaft power/torque/RPM; WAPS power (rotor/suction/control) and operational mode event data; plus, optional quantities (humidity, fouling, pitch ratio, etc.).	All require wind, propulsion and speed measurements; DNV and ITTC explicitly require WAPS auxiliary power; ANEMOI includes it in net fuel saving.
Wind measurement definition	At least two anemometers (forward/aft); select most upwind anemometer by AWA range; LiDAR may remove need for correction.	Recommend LiDAR or pre-trial calibration/correction of anemometer; multiple anemometers and/or CFD to identify least disturbed	At least one anemometer required; wind LiDAR regarded more accurate and recommended (especially if multiple WAPS are	All recognise anemometer disturbance as a key risk; ANEMOI is most prescriptive on correction workflow; ITTC and DNV

		location.	installed); anemometers should be unobstructed; lidar can calibrate anemometers/uncertainty.	strongly recommend LiDAR.

Table V: Test procedure

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Test sequence	Baseline #1 (OFF) → Test (ON) → Baseline #2 (OFF); repeated cycles (often 1 hr ON / 1 hr OFF).	Paired runs with WAPS ON/OFF at similar conditions; repeat across wind angles/conditions.	ON/OFF set: approach → measurement (10-15 min) → switch buffer → approach → measurement; can chain sets and reuse measurement periods to increase samples.	ANEMOI explicitly uses baseline-test-baseline to interpolate expected conditions; ITTC uses run pairs; DNV uses structured sets with defined buffers and approach time.
Approach / settling time	Settling time required for the net forward thrust from all tests to approach its asymptotic value. Depends on vessel type and loading condition.	Approach time required for the running averages of shaft rpm, shaft torque and vessel speed to reach stationary conditions; time between runs should be minimised.	Approach time required for shaft speed, torque, and vessel speed to reach stationary conditions; minimum approach time given as function of displacement; typically ~30 min for a large cargo ship	ANEMOI and DNV explicitly separate 'settling/approach' from measurement; ITTC assumes stability within the run and via approach between runs.
Measurement / sampling period	Sampling after settling until next mode change or max period; enforce minimum sampling duration for each Baseline/Test (typically 10-15 minutes).	Run duration ≥15 min for each run (ON or OFF).	Measurement period 10-15 min; not desirable <10 min; equal length for ON and OFF within a set; switch buffer 2-5 min.	ITTC prescribes longer minimum per run; DNV prescribes 10-15 min with gust-process rationale; ANEMOI uses longer cycles but samples only after settling.
Number of tests	No fixed minimum; run many tests across AWA/AWS and loading conditions to reduce scatter and support calibration. Evaluation may cover ~3 months	Minimum program can be done within one day with about five wind conditions.	Depends on use case; long-term evaluation may cover ≥1 year; polar diagram used to show coverage and test density.	ANEMOI and DNV rely on large sample sizes for statistical robustness; ITTC relies on controlled conditions and limited but distributed trials.

Table VI: Data validation, normalisation and correction

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Filtering / stationarity criteria	Reject tests if: (1) current dominates SOG, (2) wind angle varies, (3) engine rpm varies >2, (4) course varies >5°	Reject runs if the running averages of shaft speed, torque, and vessel speed are not stable	Applies Chauvenet's criterion to remove outliers; reject ON/OFF tests if shaft speed, course over ground, heading, wind speed and angle, or water depth exceed defined stationary thresholds	ANEMOI and DNV provides explicit thresholds; ITTC define stability principles but leave thresholds more to practitioner judgement/criteria tables.
Derived quantities	Compute air density; correct AWS/AWA for interference and height; select relevant anemometer; compute mean of primary parameters per period; compute 'Expected' baseline conditions by time-weighted interpolation between Baseline #1 and #2.	Evaluate acquired data; compute mean and standard deviation of primary parameters per run; compile run-level dataset for post-processing steps.	Document data processing; compute virtual speed from propeller torque/RPM if used; check steady-state periods; evaluate data consistency (time-lag, KQ curve).	All rely on derived quantities; ANEMOI emphasises 'expected' interpolation to mitigate linearly varying conditions; DNV emphasises virtual speed and consistency checks.
Current and speed metric handling	Correct SOG for current using SOG-STW difference from Baselines; assumes approximately linear current variation over test.	No analytical current correction; select trial location/time to minimise currents; use GNSS-SOG when current is negligible/approximately constant; otherwise use STW where current varies and the speed log is verified reliable.	No current correction: assumes environmental conditions including current) are equal in ON and OFF segments; virtual speed is calculated from shaft signals to reduce speed uncertainty.	ITTC avoids current by test design; ANEMOI corrects within each test; DNV manages current sensitivity primarily via speed choice (STW/SOG) and virtual speed rather than explicit current correction.
Speed-power curve use (baseline mapping vs speed	Estimate resistance change with speed (typically V^2; exponent adjustable); include propeller efficiency via open-water KT, KQ.	Use speed-power curve/direct power method per ITTC speed and power (P) trial guidance; convert measured (V, P) to ΔP at	For Procedure I: Convert ΔV to equivalent ΔP using a representative speed-power curve; the curve should be based on WAPS ON but	ANEMOI uses a generic resistance-speed law unless better data; ITTC uses the no-WAPS curve primarily as a baseline mapping for

correction)		matched conditions; if speed-power curve is unavailable for the trial speed and loading, include additional speed variation tests	may be based on WAPS OFF if the speed change is < 1 knot and the relative ΔP-ΔV behaviour is expected similar. Establish the curve from sea trials/model tests/CFD or supporting ON-ON / OFF-OFF tests using ≥3 shaft speeds within ±1 knot of the target speed. Procedure II: No speed correction.	ΔP at trial/reference conditions; DNV Procedure I uses a curve primarily for speed correction ($\Delta V \rightarrow \Delta P$) and treats curve representativeness as a major bias source.
Aerodynamic corrections and OFF state drag	Wind corrected to upstream using hindcast data (e.g., ERA5); tests with large wind change excluded. Vessel aero drag (with/without WAPS) and WAPS OFF state drag included in modelling/prediction.	Correct superstructure air resistance when wind differs between run pair; subtract idle WAPS/foundation drag if not retracted/tilted.	No superstructure correction; assumes environmental conditions are equal in ON and OFF segments; aerodynamic resistance from WAPS in OFF period should be removed as per ISO 15016:2015; hindcast data may only be used for validation.	ITTC includes explicit aerodynamic correction terms; ANEMOI includes aerodynamics primarily in the calibrated model; DNV assumes constant environmental conditions and focuses on statistical robustness rather than explicit superstructure correction.
Wind height correction / reference height	Scale corrected wind to WAPS mid-height using 1/9 power law.	Report results corrected to 10 m above water using 1/9 power law; apply heel correction if heel>5°	Wind should be representative at WAPS location; no single exponent mandated; when comparing to prediction, a gust factor for performance may be applied.	ANEMOI explicitly scales to WAPS mid-height; ITTC standardises reporting to 10 m; DNV prioritises representativeness and uncertainty rather than a fixed exponent.

Table VII: Performance analysis

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Performance calculation	Use propeller open-water chart: derive KQ/KT from shaft torque and RPM; compute propeller thrust and effective force; derive net change of forward thrust from ON/OFF including induced drag effects.	Compute ΔP between WAPS ON/OFF at sea-trial conditions as the sum of direct shaft-power reduction and equivalent power from speed increase.	Procedure I: compute instantaneous power saving (corrected for speed change). Procedure II: compute energy-per-meter saving. Both procedures: results aggregated and reported with confidence interval.	WAPS benefit is obtained from differences between ON and OFF periods and should include WAPS operating power; ANEMOI explicitly derives thrust and calibrates model coefficients; ITTC focuses on power difference verification at trial conditions and optional normalisation; DNV focuses on statistical performance estimation over many samples and formal presentation of uncertainty.
Model calibration and validation	Predict net change in force per test using FSAM; convert to force coefficient vs AWA; tune hull/rudder/WAPS coefficients to align predicted and measured curves; apply calibrated model to voyage fuel-saving prediction.	Compute ΔP between WAPS ON/OFF at sea-trial conditions as the sum of direct shaft-power reduction and equivalent power from speed increase.	Primary output is measurement-based performance with uncertainty rather than tuning of predictions by comparison with measurements.	ANEMOI includes an explicit tuning loop; ITTC requires model update if mismatch but does not prescribe tuning method; DNV is prediction-agnostic and centred on verified measured performance.
Long-term prediction and route savings	Net fuel-saving prediction uses baseline vessel performance plus WAPS forces, auxiliary power and four DOF effects; integrate across time steps in voyage simulation to obtain voyage net fuel saving.	If prediction matrix is available and validated, compute average saving potential for a route using the prediction procedure (ITTC 7.5-02-03-01.9 or MEPC.1/Circ.896); include measured WAPS power consumption in average power saving calculation.	Long-term framework emphasises continuous measurements and many ON/OFF tests to build confidence; results can be used to reflect impact on indices such as Carbon Intensity Index (CII) and support operational monitoring and reporting.	Long-term savings require extrapolation beyond the tested envelope using a prediction model. ITTC treats voyage simulation as optional post-step; ANEMOI provides an end-to-end fuel saving computation; DNV focuses on in-service performance evaluation with statistical confidence and verification.

Table VIII: Uncertainty quantification.

Step	ANEMOI	ITTC 7.5-04-01-02	DNV-RP-0686	Comments
Uncertainty estimation and propagation	Not formally prescribed; mitigated via filtering, multiple tests and calibration fit quality (scatter vs AWA).	Uncertainty not well established; recommends not using sea trials for contractual guarantees; report mean/SD and traces for transparency.	Formal uncertainty evaluation (Sec 6) based on valid ON/OFF tests; confidence intervals (e.g., Student's t); recommend ON-ON/OFF-OFF tests to quantify baseline variability.	DNV is the only methodology with a defined statistical uncertainty framework; ANEMOI and ITTC acknowledge uncertainty but are less formal.

Appendix 2: Alternative Approaches to WAPS Performance Assessment

Direct thrust measurement on the WAPS unit

Direct thrust measurements can provide valuable insight into the aerodynamic behaviour of the WAPS, and the forces generated by the device itself. However, this approach is not yet mature for robust full-scale savings verification and does not capture the coupled hydrodynamic response of the vessel, changes in propulsive efficiency, or operational adjustments such as rudder usage and speed control, which ultimately determine the realised fuel and emissions savings, *LR (2019)*, *Voutilainen and Paakkari (2025)*.

ISO 19030-based analysis, comparing vessel performance before and after WAPS installation using long-term operational data

While this approach benefits from large datasets, it requires extended monitoring periods to obtain statistically meaningful results. In its modification for WAPS, it requires higher frequency data to be able to capture the changes in wind conditions and performance. While the approach can capture hull fouling degradation through time, special care is needed to treat hull cleaning, propeller polishing and dry-docking events. In addition, the strict filtering and steady-state requirements of ISO 19030 result in significant data rejection under variable wind conditions, limiting its practicality for isolating WAPS effects, *LR (2019)*.

Data-Driven Decision-Making for Effective Hull Fouling Control: A Case Study for Very Large Crude Carriers

Pavlos Zouridis, Neda Maritime Agency, Piraeus/Greece, pavlos.zouridis@nedamaritime.gr
Barry Kidd, AkzoNobel, International Paint, Newcastle/UK, barry.kidd@akzonobel.com

Abstract

Neda Maritime used a data-driven approach to select and manage fouling-control coatings on three VLCCs, working with AkzoNobel's International® marine coatings business. Operating globally, with about 50% of time in high biofouling risk waters, the vessels used a combined biocide free silicone foul release and silyl methacrylate antifouling system that delivered consistent performance. Hull roughness remained below fleet averages, hull efficiency losses stayed under 2% with no hull cleanings required, and routine propeller polishing maintained losses to near zero. The results highlight how performance analytics and coating expertise enable efficient lifecycle management for large tankers.

1. Introduction

Managing maritime fleets has become increasingly complex as environmental expectations and operational uncertainty continue to evolve. While profitability once dominated strategic decision-making, emissions reduction and biosecurity compliance now play a central role in shaping fleet-management practices. Regulatory frameworks introduced by the International Maritime Organization (IMO) and the European Union reflect a shift toward more stringent environmental governance, linking operational efficiency with financial exposure and placing greater emphasis on maintaining clean, well-performing hulls.

In parallel, several regions have adopted local biofouling requirements that exceed IMO guidance, requiring documented biofouling management and, in some cases, imposing inspections, cleaning orders, delays, or penalties when hull condition cannot be demonstrated. These developments create additional pressure on operators and coating manufacturers, particularly in relation to coating selection, antifouling performance, niche-area specification, and long-term fouling-control planning. Such decisions increasingly rely on vessel-specific data rather than generic assumptions.

Neda Maritime Agency Co. Ltd., one of Greece's oldest independent shipping companies, manages a modern fleet of large tankers and bulk carriers trading worldwide. In navigating the evolving regulatory landscape, the company evaluates investment decisions carefully to balance compliance, operational efficiency, and financial performance. Among the many engineering and operational choices available, selecting an appropriate fouling-control coating and strategy is particularly important, as hull condition directly affects fuel consumption, emissions, and lifecycle cost.

Fouling control coating selection is influenced by vessel speed, activity levels and trading routes, especially in tropical regions where fouling pressure is high. Coatings must remain effective for the full in-service period, be compatible with existing systems, and be applied to hulls whose condition reflects years of wear, repairs, and accumulated roughness. Availability of preferred products, regional chemical restrictions, and the specialised expertise required for silicone-based systems further shape dry-dock planning and cost. These factors make a data-driven approach essential for selecting and maintaining coatings that support both operational and environmental objectives.

Operational uncertainty adds further complexity. Traditional long-haul VLCC trading patterns have given way to somewhat irregular routing, variable idle periods, and shifting charter requirements, all of which influence fouling risk and coating behaviour. Seasonal conditions, prolonged anchorage, slow-speed operation, and exposure to ice can accelerate fouling or damage coatings, making performance difficult to compare across vessels. These realities increase the importance of structured monitoring procedures and consistent evaluation methods.

To address these challenges, Neda Maritime has developed a Biofouling Management Plan (BFMP) aligned with IMO guidelines, providing ship-specific procedures for monitoring, documenting, and managing biofouling. Central to this approach is the ability to reliably quantify and distinguish hydrodynamic efficiency losses, whether hull or propeller related, so that fouling control decisions can be made proactively and based on evidence. This capability is supported by integrated vessel-data systems that collect and transmit all relevant sensor outputs for shoreside analysis.

Against this backdrop, the present case study examines the collaborative approach between Neda Maritime and AkzoNobel's International® marine coatings business to review vessel performance and evaluate the effectiveness of its coatings applied to three Very Large Crude Carriers (VLCCs). The analysis draws on Neda's integrated performance-monitoring framework, combining vessel activity data, environmental exposure, and coating characteristics to quantify hydrodynamic losses and assess coating behaviour under diverse operational profiles. The following sections outline the methodology used to process and interpret the data, describe the operational context of the three vessels, and present the key findings that informed Neda Maritime's fouling-control strategy.

2. Case Study Overview

This case study focuses on three VLCCs with similar design specifications, Table I.

Table I: Subject VLCC Vessel Details

Name	Age (Years)	Design Speed (kn)	Deadweight (MT)	Length (m)	Breadth (m)	Draught (m)
VLCC A	14	15.2	319,330	333	60	22.5
VLCC B	9	15.3	299,323	333	60	21.6
VLCC C	6	15.5	299,999	333	60	21.6

During the coating selection process, all suppliers provided coatings scheme recommendations in accordance with the vessel operational summary provided in Table II.

Table II: Subject Vessel High Level Operational Profile

Typical operating speeds (range):	12.0-14.0 kn
Percentage of activity	75 % Sailing / 25 % Idle
Trading area:	Worldwide trading Frequent operation in high biofouling risk regions No operation in ice
Duration between dry-docking	5 Years

Building on the strong performance record of International® fouling control technologies, a scheme was applied comprising an ultra-performance Intersmooth® silyl methacrylate antifouling coating on the boot top and flat bottom sections, complemented by an ultra-performance Intersleek®, biocide free silicone foul-release coating on the vertical sides. This arrangement applies the super-smooth foul-release coating to the lower vertical sides of the hull, an area of high hydrodynamic significance, thereby improving flow characteristics and maximising the coating's fouling-release capability. This specification supports the objective of achieving optimal hull performance and aligns with a proactive fouling-management strategy, which aims to maintain fouling levels within the low to medium slime range. This scheme arrangement was established over several years and has been proven to deliver ~3% fuel savings versus a scheme comprising a typical Self Polishing Copolymer based antifouling coating.

This work examines the subject vessel operations, out-docking and in-service hull and coating condition, the vessel performance analysis discipline and the resulting performance insights.

3. Reviewing Vessel Operations

Understanding vessel operations is essential for selecting effective fouling control coatings, ensuring long-term performance. Coatings like biocide free silicone fouling release coatings are designed to prevent or release marine biofouling, and their success depends on aligning the vessel's operational profile, such as type, idle periods, speed, and activity with the coating's capabilities.

Intertrac®, developed by International®, supports this process by combining AIS data with proprietary biofouling risk models. It visualises operational metrics including speed, time at sea, trade routes, and environmental factors like seawater temperature and fouling risk.

By analysing changes in these metrics, users can better understand coating performance, identify trends from hull inspections, and link biofouling risk to vessel speed deviations. In short, mismatched coatings can lead to increased hull fouling and reduced vessel efficiency.

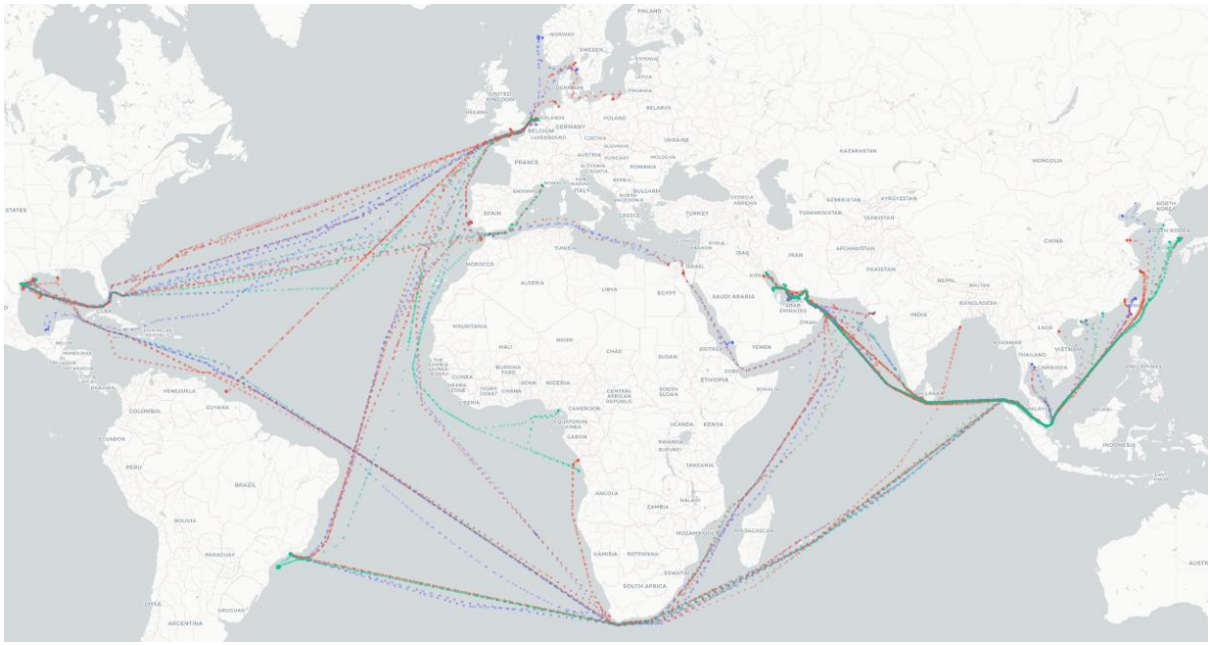


Fig.1: Intertrac® vessel operations visualised for all three VLCC vessels, where the blue, red and green markers represent VLCC A, VLCC B and VLCC C operations, respectively.

Fig.1 provides a visual overlay of the trading pattern for all three VLCC. Although each vessel shows periodic variations linked to specific voyages, all three operate regularly across multiple regions and can therefore be regarded as conducting worldwide trading.

Table III provides further high-level operational insights highlighting a high degree of consistency in the vessel operational dynamics such as speed, activity level, sea miles per month and biofouling risk. Given the limited variation in their operational profiles, and the similarity in their design characteristics, it is reasonable to evaluate the performance trends of these vessels collectively as a single group.

Table III: Intertrac® vessel operational metrics

	Months In-Service (since last drydock)	Most Common Speed (kn)	Average Sea Miles per Month	% Activity > 3 kn	% Time in High to Very High Biofouling Risk Regions
VLCC A	35	12	6,504	76	58
VLCC B	46	12	6,370	76	49
VLCC C	27	12	6,544	78	48

A thorough understanding of vessel operations is a critical foundation before exploring hull inspections, fouling control coating performance, and vessel performance metrics. While these high-level metrics provide a useful insight into the vessel operations, the International® hull performance team typically conducts more detailed analyses that fall outside the scope of this paper.

Building on this operational overview, the next section will evaluate the physical condition of the hull and coating performance through inspections and Average Hull Roughness (AHR) measurements.

4. Hull Condition and Coating Performance

In line with standard practice among many vessel operators, the quality of the coating application during dry-docking was assessed. In collaboration with International®, AHR measurements were undertaken and assessed. In the case of the subject vessels of this study, the vessel out-docking AHR measurements highlight that, the drag penalty of vessels due to underlying hull and coating roughness was lower than the International® database average for similar vessel types and conditions. This insight provides reassurance that the quality of the application was very good for all three vessels.

In this case study, underwater dive inspection material was available for all three vessels, allowing a review of hull condition and coating performance at regular intervals. Dive inspections can offer important visibility into hull and propeller condition, but they present several inherent limitations. The clarity and relevance of inspection material can be affected by water quality, lighting, and diver stability, while the absence of consistent industry standards often results in subjective assessments of biofouling. These factors, combined with partial hull coverage and variable reporting practices, mean that inspection imagery may not accurately represent the condition of the entire underwater surface. As a result, it can be difficult to compare inspections carried out at different times or across a fleet.

These challenges highlight the need for expert interpretation when aligning inspection observations with vessel performance information. Without appropriate context, inspection material can easily lead to incorrect conclusions regarding coating performance, hull resistance, or the cause of change in vessel performance indicators. The International® hull performance team support this process through established procedures, specialist and extensive experience in biofouling characterisation to check, verify, and reinterpret dive inspection material to ensure the correct action is taken. This involvement reduces subjectivity, ensures that observations are evaluated in accordance with recognised standards, and improves reliability when determining the true condition of the coatings and the hull.

Combining structured inspection assessments with performance trend analysis, provides a dependable foundation for decision making. This integrated approach helps identify whether performance variations are truly related to the hull and propeller or are influenced by external operational or data quality factors. It also prevents unnecessary interventions, such as unneeded hull cleaning, and supports operators in maintaining consistent performance and compliance objectives.

Inspection material for all three vessels shows that the coatings applied continued to perform to a high standard throughout the in-service period, with only very light to medium levels of microfouling or slime growth across both the vertical sides and flat bottom. No significant macrofouling organisms, heavy slime layers, or areas of coating breakdown were identified, and coating damage was limited to small, localised regions associated with docking and berthing activity. Furthermore, periodic increases in vessel speed provided additional hydrodynamic shear that supported the foul-release behaviour of the coatings, helping to maintain consistently low fouling levels throughout the operational period.

5. Neda Maritime Vessel Performance Analysis Discipline

Neda Maritime applies its in-house performance analysis framework, Propulsion Diagnosis, to quantify efficiency degradations across machinery, hull, and propeller systems. The model produces discrete efficiency-loss percentages relative to benchmark values, enabling a structured interpretation of vessel performance over time.

Real-time data collection and transmission are handled through Neda Maritime in-house developed Data Acquisition system, designed for online transfer, filtering, verification, and archiving of operational data. The dataset incorporates inputs from bridge navigation equipment, engine control room systems for main and auxiliary engines, steering systems, auxiliary boilers, fuel consumption records, and cargo control room instrumentation covering tank conditions and loading states. Fleet-wide data are gathered via high-bandwidth connections to support continuous monitoring of machinery operation, vessel efficiency, and location. The process follows the principles of ISO 19030-2, supplemented by additional treatments developed through operational experience.

Navigation, combustion, and general sailing information are processed using correlation and autocorrelation techniques to filter and evaluate the influence of weather, currents, swell, and fouling. Results are cross-checked against outputs from the International® model, providing an external and independent reference point. Diver inspections are also used to confirm the physical condition of the hull and propeller, offering an additional layer of verification for the interpreted trends. Taken together, these comparisons support the accuracy of the filtered dataset and the reliability of the resulting performance assessments.

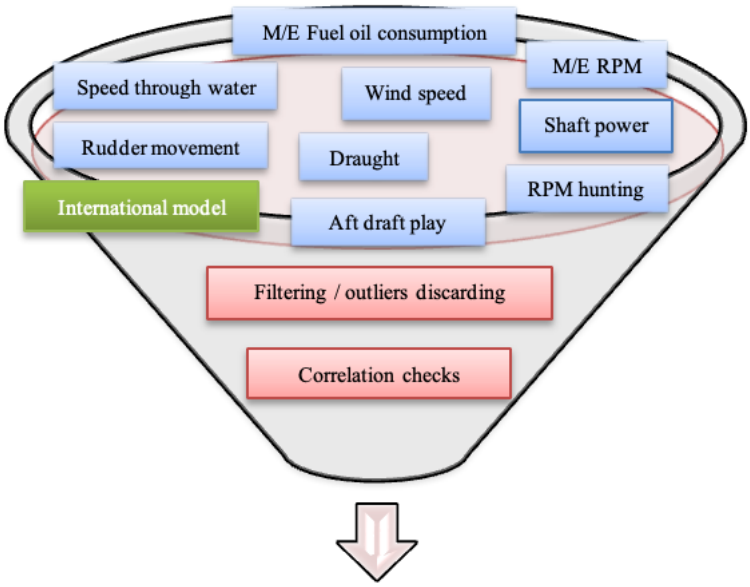


Fig.2: Propulsion Diagnosis model overview

The Propulsion Diagnosis model, Fig.2, integrates data from multiple sources, including shop tests, sea trials, and benchmarking periods following dry-docking or major retrofits. Eq.(1) represents the core formulation of the model, which incorporates the Specific Fuel Oil Consumption (SFOC)-power relationship, wake coefficient function (W_f), hull friction coefficient, vessel characteristic lengths, propeller load (C_α), and propeller jet flow. The resulting figure is compared with real-time measurements to identify distinct hull, propeller, and machinery performance losses.

$$Propulsion\ Diagnosis = \frac{\text{foc} \times C_{NCV} \times (1 - \%s) \times C_{\text{SFOC}}}{W_f \times N^2 \times V_s \times C_\alpha} \quad (1)$$

foc: Fuel oil consumption rate [Kg/day]
C_{NCV}: Net calorific value coefficient
C_{SFOC}: Specific fuel oil consumption coefficient
V: Vessel speed
W_f: Wake coefficient function
N: Main engine revolution rate
s: Slip
C_α: Propeller load coefficient

The wake coefficient function (W_f) represents the quantity of seawater flow upstream of the propeller and influences propeller loading. Under typical hydrodynamic conditions (Re ~10^9), W_f depends on ship speed (V_s), friction velocity (u*), and hull geometry, including shape, roughness, parallel body characteristics, hull lines, and draught. The initial W_f also reflects improvements from any energy-saving devices installed near the propeller to restore rotational inflow to more uniform streamlines.

Propeller load (α), together with the charter-party speed order, determines engine power and propeller revolutions. The load accounts for non-uniform inflow in the propeller's vicinity, the degree of cavitation, and the broader influence of hull resistance.

Propulsion Diagnosis is used to calculate the expected (benchmarked) fuel oil consumption. Deviations from the measured value represent the total efficiency losses (LT). Fig.3 provides a visualisation of these total losses.

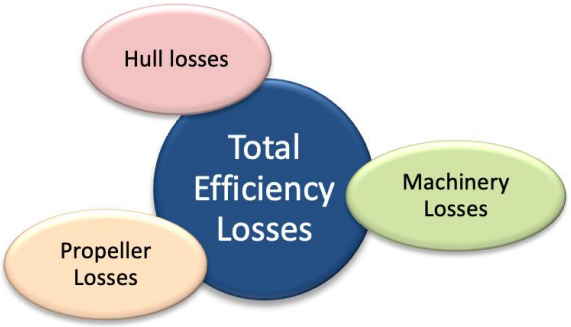


Fig.3: Vessel total efficiency loss visualisation

Machinery efficiency losses (LMACH) are derived by comparing the ISO-corrected shop-test SFOC with the measured SFOC. The influence of the main engine's SFOC on performance measurements is evaluated using detailed shop-test data, which have proven to be a consistent basis for generating the SFOC–power curve.

Hydrodynamic efficiency losses (LHYDRO), which include both hull and propeller effects, are calculated from the correlation between propeller load deviation (α) and the differential between total and machinery losses [LT-LMACH]. Deviation from the benchmarked propeller jet coefficient is used to estimate propeller efficiency losses (LPROP), after which the remaining component is attributed to hull efficiency losses (LHULL). The wake model then translates these hull losses into an equivalent average roughness value, although this final step remains an area of ongoing refinement and should be interpreted with appropriate caution.

Propulsion Diagnosis is continuously improved as additional operational experience, inspection data, analytical insights become available, and its refinement benefits from ongoing collaboration with external experts, ensuring that the methodology evolves gradually and remains grounded in practical evidence rather than representing a fixed or definitive approach.

With the vessel performance data processing and diagnostic approach now defined, the next section examines the resulting performance insights in the context of management practices and other relevant data sources.

6. Vessel Performance Insights

All vessels in this study are reviewed using Neda Maritime's Propulsion Diagnosis framework, which provides the data required to visualise hull efficiency loss and propeller efficiency loss, the two primary indicators examined in this paper. Events such as dry-docking, hull cleaning, and propeller polishing are annotated within the performance visualisations to support interpretation of the efficiency-loss

trends. Propeller polishing is proactively scheduled every six months or sooner, regardless of the observed loss levels. Since the vessels in this study operate with uncoated propellers, polishing remains the only fouling control mechanism available to remove calcium deposits and microfouling from the blades.

Propeller-related degradation can typically be identified quickly and addressed through targeted maintenance. Hull-related degradation, however, requires longer timeframes and larger datasets to interpret reliably. This distinction underscores the importance of separating hydrodynamic degradation sources, specifically hull versus propeller, when reviewing performance data.

Following the discipline set out in ISO 19030, the first 12 months after dry-docking are designated as the reference period, during which the vessel's speed–power relationships are established. After this reference period, the vessel enters its evaluation period, and hull and propeller efficiency losses are derived. These indicators can be reviewed as simple in-service averages or explored in more detail by examining how they evolve over time. Where useful, moving-average trends are also applied to the period preceding drydock to support interpretation of the performance reset.

Neda Maritime's policy of regular, proactive propeller polishing ensures that propeller related efficiency losses, although subject to periodic oscillations, remain consistently low. Polishing events are typically followed by an immediate reduction in propeller losses, reflecting the direct impact of the intervention on restoring thrust performance. As this is a preventive strategy, there are occasions where the loss trend remains largely unchanged, consistent with the fact that polishing is sometimes performed before measurable degradation occurs.

For hull-related losses, any sustained degradation above 4–5% for more than one month is treated as a clear indicator that action is required to check data integrity and hull condition. These thresholds account for inherent data uncertainty arising from filtering processes, operational variability, and limitations in acquisition systems. For example, during super-slow-steaming conditions, the vessel operates outside the design envelope for optimal hull and propeller inflow, which can manifest as a short-term fluctuation in hull and propeller loss. Newbuilding specifications typically target maximum efficiency around the Normal Continuous Rating of the main engine and the corresponding service speed, which is generally above 65–70% of maximum continuous rating, making very low power operation less favourable from a thermo-hydrodynamic perspective.

As highlighted earlier, hull condition is monitored through periodic diver inspections, which the International® hull performance team checks and verifies, and the resulting observations are used to cross-reference the vessel performance trends. In those cases where both the measured and observed conditions support the interpretation of the hull-related loss trends, a hull cleaning would normally be initiated focussing on suitable vendors proposed by the International® hull performance team.

Prioritising cleaning services to use qualified vendors helps reduce risk and ensures that any cleaning conducted supports coating longevity and maintaining fouling control performance. Access to these specialist resources, however, can be limited across VLCC trading patterns, which offer infrequent calls to major ports such as Singapore, Rotterdam, or Fujairah. Even when suitable ports are available, securing preferred diving companies can remain difficult due to availability and scheduling constraints. To limit dependence on physical cleaning, the vessels in this study apply a proactive foul release strategy in-service, periodically increasing speed to around 15 kn to aid the release of early-stage biofouling from the fouling control coatings. This manoeuvre takes advantage of the ultra-smooth, low biofouling adhesion characteristics of the Intersleek® coating applied to the lower vertical side section of the vessel, enabling early-stage biofouling to detach more easily during periods of increased vessel speed. When performed promptly, once fouling risk is identified, this operational approach effectively restricts slime fouling accumulation, protects coating integrity by removing the need for hull cleaning, and helps maintain hull losses within acceptable limits. For VLCCs equipped with silicone-based systems, this provides a practical balance between operational performance and long-term coating protection.

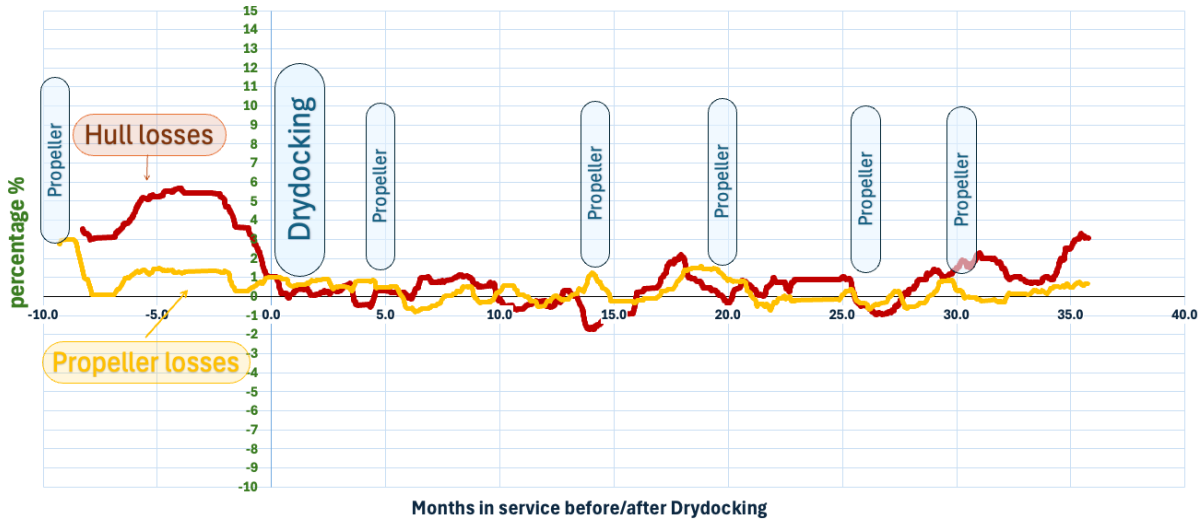


Fig.4: VLCC A hull and propeller efficiency losses 10-day moving average trend

Fig.4 shows that VLCC A undergoes a clear performance reset immediately after dry-docking. Hull efficiency losses drop sharply at the docking mark, reflecting the combined effect of surface preparation quality, the application of a fresh coating system, and the resulting reduction in roughness and hydrodynamic resistance. This establishes a new reference condition for the subsequent operational cycle.

Over the following 30+ months, hull losses remain consistently low. Minor periodic fluctuations appear, likely linked to operational conditions, idle periods, or short-term biofouling dynamics, but the overall trend remains stable. Throughout the post-docking period, average hull losses stay mostly within the 0–1% range, indicating that the fouling control coating scheme and maintenance standard effectively preserved hull condition.

A gradual increase in hull losses is observed in the final months of the evaluation period. Although moderate in magnitude, it represents a deviation from the otherwise stable trend. This development is under review, and an International® verified dive inspection will be sought at the first operational opportunity to aid diagnosis and take corrective action if needed.

Propeller efficiency losses remain tightly controlled throughout the cycle. Although the moving average trend illustrates periodic oscillation, these are short-lived and consistently corrected by polishing events. Each intervention corresponds to a reduction in propeller losses, illustrating the expected improvement in thrust efficiency.

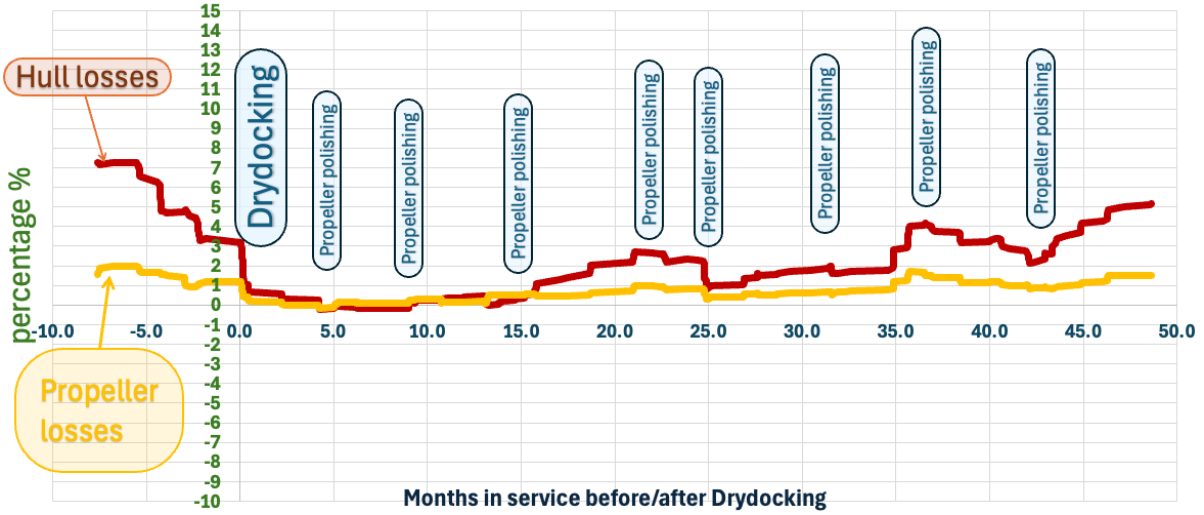


Fig.5: VLCC B hull and propeller efficiency losses 10-day moving average trend

Fig.5 presents the performance trends for VLCC B, showing a clear improvement in hydrodynamic condition immediately after drydocking. The reduction in hull and propeller losses reflects the effectiveness of the surface preparation standard and the application of a new silicone-based coating scheme. During the first 12-month reference period, hull losses remained close to zero.

Beyond the reference period, hull losses follow a gradual upward trajectory but remain well within the 4-5% acceptance level. The previously described speed increase strategy has been applied on several occasions, to aid the release of observed slime. A representative example occurred between months 12 and 14: 30% light to medium slime coverage was observed on the lower vertical side sections by month 12, which reduced to around 10% light slime by month 14 following a speed increase interval after departure from port. Around month 25, the effect of varying power orders, to their extremities (super slow steaming to design speed) can be seen in the form of a 1.5% improvement in hull efficiency losses. Propeller losses evolve differently. Multiple polishing events are visible throughout the operational cycle, each producing a drop in propeller-related degradation.

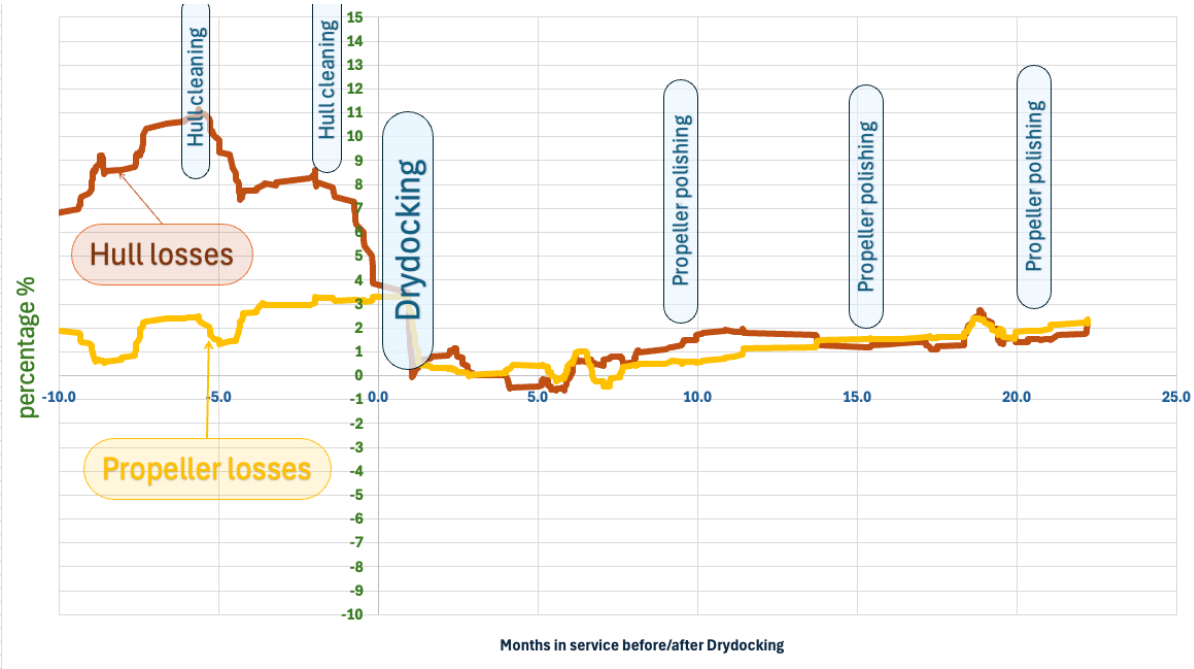


Fig.6: VLCC C hull and propeller efficiency losses 10-day moving average trend

Fig.6 illustrates the performance trends for VLCC C. Following drydock, both hull and propeller losses dropped sharply, marking the beginning of the 12-month reference period. Hull efficiency has remained mostly stable throughout this period, with losses confined to a narrow 1–2% margin, well within the acceptance threshold of 4-5%.

Propeller efficiency losses show periodic reductions following polishing events in April 2024, November 2024, September 2025, and December 2025. These recurring drops, once again highlight the impact regular propeller maintenance in sustaining propulsion efficiency and limiting mechanical degradation.

7. Conclusion

This case study demonstrates the value of combining vessel-specific performance data with coating-technology insight to support informed hull-management decisions across the VLCC lifecycle. By analysing efficiency-loss trends and correlating them with operational events, Neda Maritime, supported by International® hull performance team, has been able to reliably distinguish between different sources of hydrodynamic degradation and shape maintenance strategies that preserve operational efficiency.

Across the three VLCCs, average in-service hull losses consistently remained between 0-2%. These values are well below the internal intervention threshold of 4-5% and indicate no sustained hydrodynamic deterioration that would require cleaning action. All three vessels experienced a clear performance reset following drydock, with hull losses dropping to 0-1%. Out-docking AHR values were lower than the International® fleet database average, confirming excellent coating application quality. Despite significant time spent in high biofouling-risk regions, underwater inspections revealed only light slime accumulation and no macrofouling. This result validates the durability and consistent performance of the Intersleek® biocide free silicone foul-release and the Intersmooth® silyl methacrylate antifouling coatings.

Propeller performance was also well maintained, with polishing every 5 to 7 months consistently resetting losses to near zero, effectively minimising thrust degradation and reducing avoidable fuel penalties. Operational measures, including strategic speed increases, effectively controlled slime accumulation and helped stabilised hull losses by utilising the foul-release properties of the Intersleek® coating, eliminating the need for hull cleaning. In one representative case, observed slime coverage reduced from approximately 30% to 10% over a two-month period following timely speed increases.

These results highlight the importance of a structured and data-driven monitoring framework. Regular review of hull and propeller efficiency trends, supported by robust filtering, verification methods, and validated underwater inspections, provides reliable early warning of performance drift and helps avoid unnecessary or reactive interventions. Selecting fouling-control coatings suited to operational patterns is equally important, particularly for silicone systems that rely on consistent hydrodynamic shear.

Finally, continued collaboration between operators, coating specialists, and analysts enhances data interpretation, improves modelling accuracy and ensures that operational practices evolve with real-world insight. Together, these elements support a lifecycle approach to hull performance management that optimises efficiency, extends coating longevity and supports environmental responsibility.

Acknowledgment

The authors gratefully acknowledge the support of the Neda Maritime Agency Management for allowing the reproduction of the fleet data in the preparation of this paper and the technical analysis support and coating performance insights provided by AkzoNobel's International® marine coatings hull performance team.

References

DELIGIANNIS, P.; ZOURIDIS, P. (2018), *Global Hull Performance: propulsion power optimization and operational diagnosis tool,* 3rd HullPIC, Redworth, pp.62-75

IMO (2011), *2011 Guidelines for the control and management of ships' biofouling to minimise the transfer of invasive aquatic species*, Res. MEPC.207(62), Int. Mar. Org., London

IMO (2023), *2023 Guidelines for the control and management of ships' biofouling to minimise the transfer of invasive aquatic species*, Res. MEPC.378(80), Int. Mar. Org., London

ISO (2015), *Ships and marine technology – Measurements of changes in hull and propeller performance*, ISO 19030-2, Int. Standard Ord., Geneva

JONAS, P.; MAZUR, O.; URUBA, V. (2009), *Determination of Skin Friction on a Rough Surface*, Colloquium Fluid Dynamics, Prague, pp.1-8

JOURNEE, J.; MEIJERS, J. (1980), *Ship routeing for optimum performance,* Delft University of Technology

LOGAN, K.P. (2011), *Using a Ship's Propeller for Hull Condition Monitoring*, ASNE Intelligent Ships Symp.IX, Philadelphia, pp.1-20

NN (2009), *Compendium Marine Engineering,* Seehafen Verlag

SCHLICHTING, H. (1979), *Boundary-Layer Theory,* McGraw-Hill

TENNEKES, H.; LUMLEY, J. (1972), *A First Course in Turbulence,* MIT, Cambridge

Performance Monitoring based Antifouling Selection Service CMP-MAP

Hirohisa Mieno, Chugoku Marine Paints, Hiroshima/Japan, hirohisa_mieno@cmp.co.jp
Sebastian Stoldt, Chugoku Paints (Germany), Hamburg/Germany, s.stoldt@cmpeurope.eu
Massimo Zanone, Chugoku-Boat Italy, Genova/Italy, m.zanone@chugoku-boat.it

Abstract

This paper describes the information technology service (CMP-MAP) for the selection of antifouling paint based on performance monitoring and simulation. "FIR theory" visualizes the effects of low friction antifouling and surface treatment such as full blasting by Friction Increase Ratio calculated from roughness measured by 3D hull roughness analyzer. "Performance analysis" visualizes hull performance using various performance indicators defined in ISO 19030. "Operational profile analysis" helps to find more appropriate antifouling specifications using operational profile calculated by AIS or GPS data. Fuel oil consumption simulation is conducted by "PIR simulator" using power increase ratio parameter which considered adverse effect due to aging (hull roughness) and fouling.

1. Introduction

Global warming due to both CO_2 emissions and energy conservation has been a long-term international concern. Both of these problems must be overcome in the near future for all industries will be grow continuously for the future. The International Maritime Organization has decided to revise Annex VI of the MARPOL treaty in July 2011. According to the revision, both the estimation of the energy efficiency design index (EEDI) in the design stage and its verification during sea trials must be completed in a new building stage by shipbuilders and ship classification societies. Since then, new regulations such as EEXI (Energy Efficiency Existing Ship Index) for existing vessels similar to EEDI and the annual CO_2 emission-based indicator CII (Carbon Intensity Indicator) have been also come into effect from 2023 by revision of same regulation. As a result, the shipping industry faces a sustainable-growing urgency to reduce CO_2 emissions more over. Friction resistance that develops between the seawater and the surface of the ship's hull is one of the important components of the ship's resistance. "Fouling" and "Hull roughness" which are accumulated on ship hull cause adverse effects on ship performance. Antifouling coatings fulfil a fundamental role in mitigating the increase in resistance caused by hull roughness and biofouling., it is need find more practical ways to use of the technologies such as combination with information technologies.

2. CMP-MAP

CMP - Monitoring & Analysis Program (CMP-MAP) developed based on Chugoku Marine Paints (CMP) years of experience. The methods visualise the hull performance through the triple "CMP-MAP" approach. "FIR theory" visualises the effects of low friction antifouling and surface treatment such as full blasting by Friction Increase Ratio calculated from roughness measured by 3D hull roughness analyser. "Performance analysis" visualises hull performance using performance indicator defined in ISO 19030. "Operational profile analysis" helps to find more appropriate antifouling specification using operational profile calculated by AIS or GPS data. The methods continuously improve hull performance through our original PDCA cycle called "Hull-PDCA", Fig.1. CMP will provide the program as solution provider, also using the IoS-OP data platform.

2.1. FIR theory

"Hull roughness" is increase by repeated recoating, physical damage, and cumulative repairs over time. Therefore, evaluating the effects of hull roughness on frictional resistance is vital for effective performance management and taking proactive measures. Traditionally ship hull roughness was measured by the BSRA (British Ship Research Association) hull roughness gauge. The gauge measured the AHR, defined as the average maximum peak-to-trough height over a 50 mm sampling length in the

whole hull. To meet the recent demands for more accurate estimation of friction resistance for improve ship efficiency, it is necessary to evaluate the relationship between hydrodynamic phenomena and roughness textures more precisely. CMP conducted friction resistance test as following Fig.2 (Long flat plate) and Fig.3 (Double cylinder test) to evaluate the wider speed range of skin friction resistance.

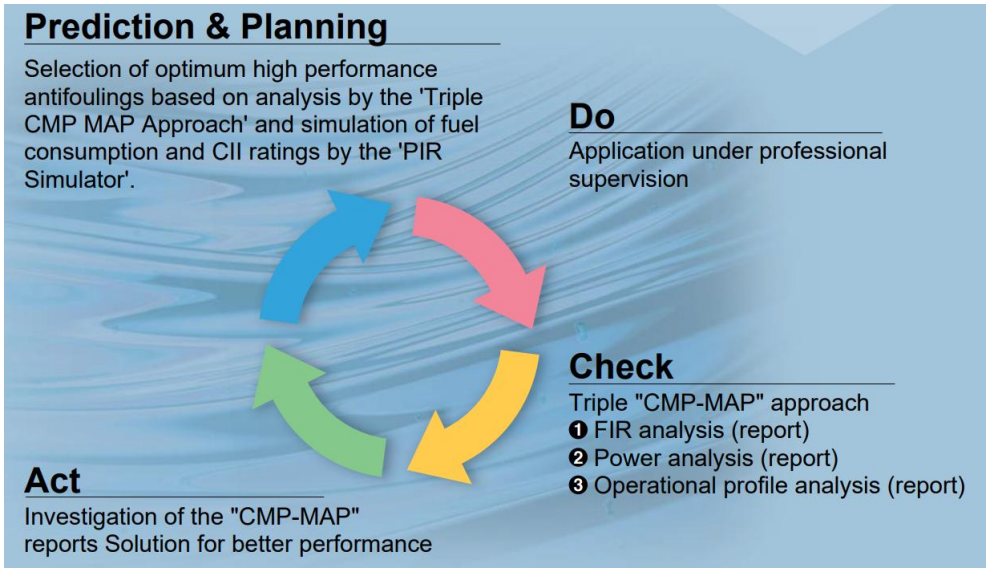


Fig.1: Concept of Hull PDCA

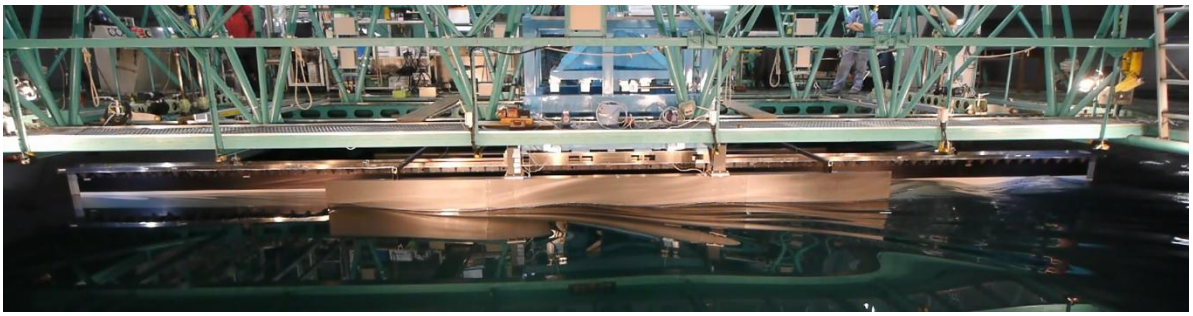


Fig.2: Long flat plate test


Fig.3: Double rotating cylinder test

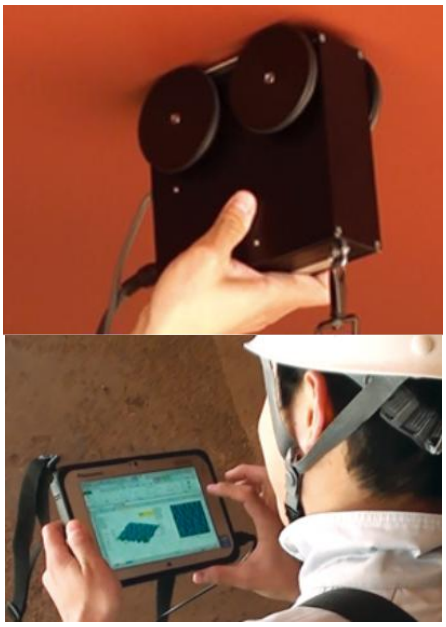


Fig.4: 3D hull roughness analyser

Recognizing the critical need to evaluate surface texture parameters such as wavelength, a laser 3D hull roughness analyser was developed to replace BSRA roughness analyser, Fig.4. Roughness wavelength influence is critical for assessing frictional resistance. Research reveals that the Friction Increase Ratio (FIR) rises with shorter wavelengths but decreases with longer wavelengths, even at an identical roughness amplitude. Based on this initial finding, a more in-depth hydrodynamic investigation was conducted. The results revealed that the Friction Increase Ratio (FIR) is proportional to the frontal projected area per unit area that protrudes from the viscous sublayer, Figs.5 and 6. Further evaluations revealed a transition-like zone between the low-speed regime (flat plate tests) and the high-speed regime (cylindrical tests). Across this zone, a significant shift occurs in the dimensionless distance y^+, which is used as the criterion for calculating viscous sublayer thickness. As the results based on most recent FIR theory ver. 3.0, the calculated results for a 200 m vessel (with an average roughness height Rc = 66-micron metre and RSm = 2740 μm at a reference speed of 12.5 kn are shown in Fig.7. The 10-17 kn range appears to constitute a transition region. Within this zone, the FIR exhibits a unique reversal, shifting from an initial increase to a temporary decrease before rising again-a phenomenon specifically observed and verified in the cylindrical test regime, *Mieno (2020)*. Given the high complexity of these hydrodynamic computations, the FIR calculator is provided as a cloud-based software solution, https://www.CMP-MAP.com.

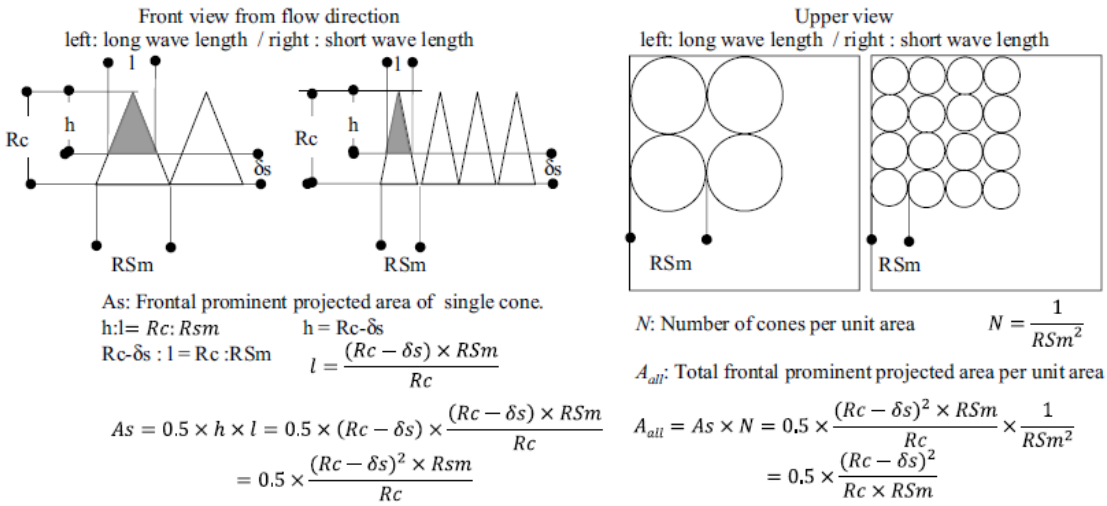


Fig.5: Calculation method of frontal projected area per unit area from viscous sub layer δs based on conical approximation

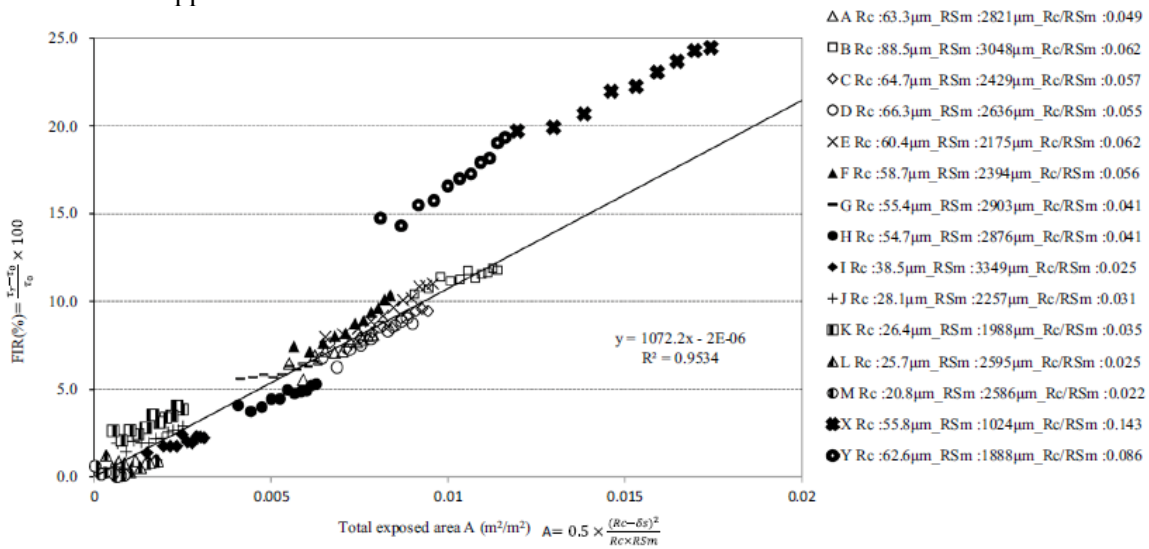


Fig.6: Relation between FIR and frontal projected are per unit area from viscous sub layer

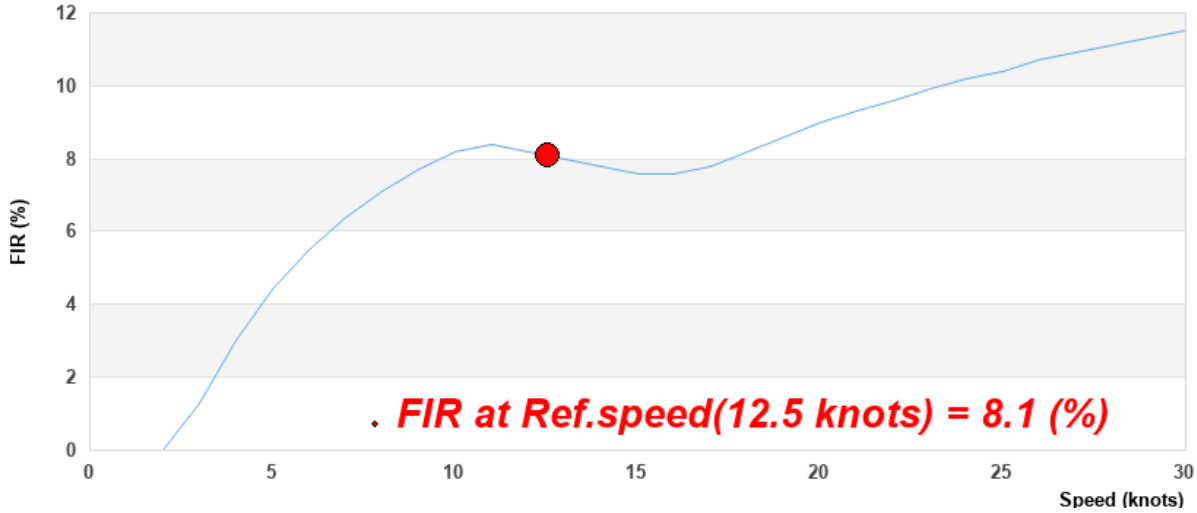


Fig.7: Calculation result by FIR theory ver 3.0 (FIR calculator)

2.2. PIR simulator

"Prediction and Planning" is another foundational component of the CMP-MAP solution of “Hull PDCA cycle”. As established, both FIR (representing initial roughness parameters) and speed loss (representing fouling and aging parameters) are equally critical indicators of vessel performance. Accordingly, we have developed the "PIR Simulator," a comprehensive system that integrates both FIR and speed loss variables. By converting both FIR and Average Speed Loss into a unified 'PIR' parameter, this simulator allows for the seamless integration of initial roughness and subsequent fouling data, thereby facilitating more precise performance forecasting. An example of this analysis is illustrated in Fig.8-11. The integration of the PIR Simulator with real-time monitoring specifically through the Triple CMP-MAP approach establishes a continuous feedback loop known as "Hull PDCA." By providing data-driven insights into fuel consumption trends, Return on Investment (ROI) (factoring in both coating (include surface treatment) and fuel costs), and carbon footprints (including CII ratings), this simulator facilitates holistic decision-making that balances environmental sustainability with financial performance. By continuously verifying simulation results against empirical monitoring data, we can achieve consistent enhancement of vessel performance throughout its lifecycle. Given the high complexity of these simulations involving various parameters, the PIR simulator is provided as a cloud-based software solution, accessible via dedicated website: https://www.CMP-MAP.com.

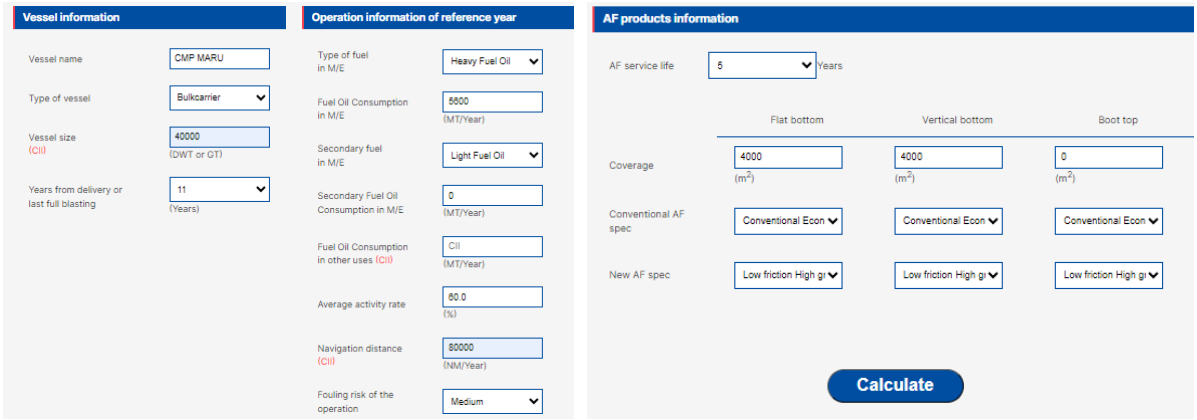


Fig.8: Information for PIR simulation (Left Vessel / Right AF specification)

	Initial factor			Fouling factor	
	FIR (%)	Form factor	PIR (%)	Average speed loss (%)	Average PIR (%)
(c) Conventional AF - Spot blasting	23.85	0.8	19.08	5.0	15.0
(d) New AF - Spot blasting	16.95	0.8	13.56	0.9	2.7
(e) Conventional AF - Full blasting	15.0	0.8	12	5.0	15.0
(f) New AF - Full blasting	1.2	0.8	0.96	0.9	2.7

Fig.9: Factor for PIR simulation

	1st year (MT/year)	2nd year (MT/year)	3rd year (MT/year)	4th year (MT/year)	5th year (MT/year)
Conventional AF - Spot blasting	5,782	6,118	6,454	6,790	7,126
New AF - Spot blasting	5,322	5,379	5,436	5,493	5,550
Conventional AF - Full blasting	5,373	5,685	5,997	6,309	6,621
New AF - Full blasting	4,612	4,662	4,711	4,761	4,810

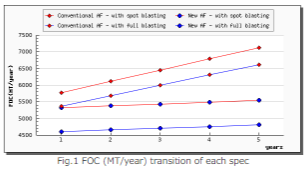


Fig.1 FOC (MT/year) transition of each spec

Fig.10: Annual fuel oil consumption simulation

	1st year	2nd year	3rd year	4th year	5th year
Conventional AF - Spot blasting	B	C	C	D	E
New AF - Spot blasting	A	B	B	B	C
Conventional AF - Full blasting	A	B	C	C	D
New AF - Full blasting	A	A	A	A	A

Total performance

	FOC (MT/Syears)	FOC reduction (MT/Syears)	CO2 reduction (MT/Syears)
Conventional AF - Spot blasting	32,270		
New AF - Spot blasting	27,180	5,090	15,850
Conventional AF - Full blasting	29,985	2,285	7,115
New AF - Full blasting	23,556	8,714	27,135

Fig.11: Simulation of CII and FOC reduction

2.3. Operational profile analysis

As already explained in PIR simulator section "Prediction and Planning" is foundational component of the CMP-MAP solution of “Hull PDCA cycle”. The operational profile (vessel’s operating condition) is also vital factor in predicting biofouling and, consequently, in designing optimal antifouling specifications. CMP has developed proprietary operational profile analysis software and established an extensive database that allows for the visualization of diverse cross-sectional profiles of vessel operation throughout the antifouling coatings service life. This data-driven approach enables the selection of more appropriate painting specifications including coating type and dry film thickness (DFT) tailored to the specific requirements of each individual vessel. Specifically, this software quantifies and visualizes critical parameters that directly influence biofouling risk, such as ship speed distributions, idle-time ratios (stay periods), and exposure to seawater temperatures across various geographical regions. By synthesizing these data points, "Operational Profile Analysis" can estimate the cumulative environmental fouling stress on the hull. This methodology facilitates the precise determination of the necessary antifouling performance, ensuring that the coating system remains effective against the specific fouling pressures encountered on a vessel's particular trade route, Figs.12-16.

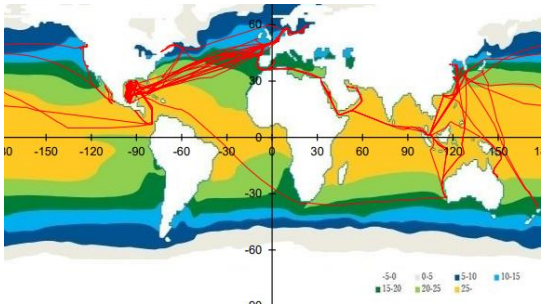


Fig.12: Ship course

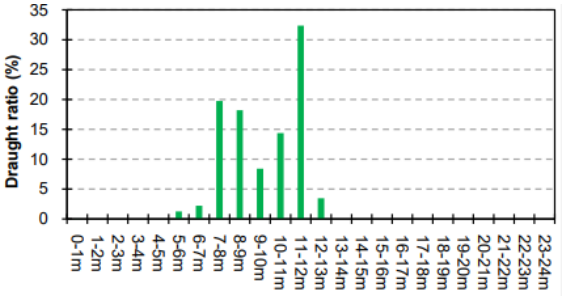


Fig.13: Draught histogram

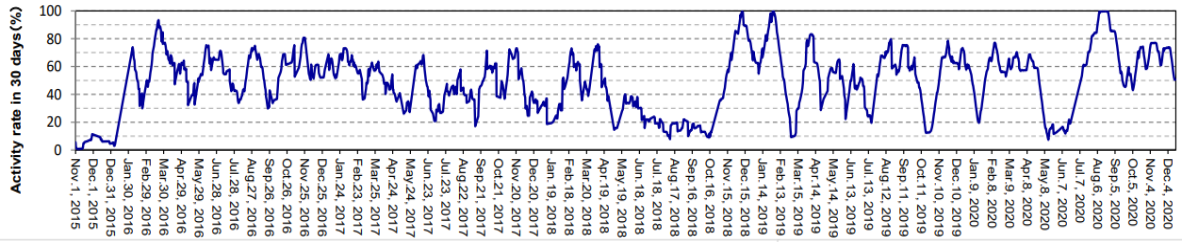


Fig.14: 30 days operating rate

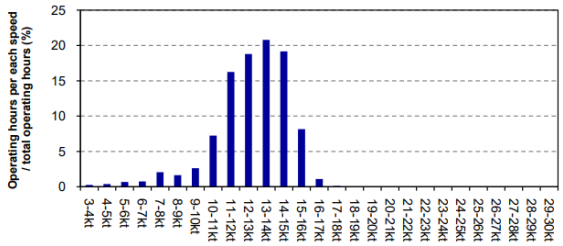


Fig.15: Speed histogram

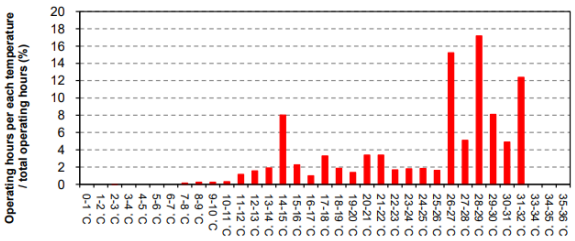


Fig.16 Sea water temperature histogram

2.4. Performance analysis

"Check" is also another foundational component of the CMP-MAP solution of “Hull PDCA cycle”. Hull performance analysis is conducted in accordance with the ISO19030 standard. This is an example of performance analysis based on ISO19030. 8,900TEU containership completed a five-year operational period, Fig.17. Notably, an economical grade of coating had been applied specifically to the boot-top area of the hull. While biological fouling was observed on the boot-top area, the vertical bottom which had been coated with high-performance antifouling paint remained free of fouling. The results of the hull performance analysis are shown in Fig.19. Speed loss V_d (%) from the estimated speed V_e on fitting line as Fig.18 is calculated. Trend of V_d is shown in Fig.19. Throughout the five-year operational period, the increase of V_d remained remarkably constrained, demonstrating the sustained efficacy of the applied coating system. Average speed loss (K_{HP} by in service performance) is -0.5%. Leveraging these empirical results to demonstrate the efficacy of high-performance antifouling coatings, as well as to establish effective countermeasures against biofouling such us hull cleaning provides significant practical and strategic utility based on the data. Furthermore, these performance data and analysis results are seamlessly shared through the IoS-OP (Internet of Ships Open Platform), a universal ship data distribution infrastructure. As a solution provider, CMP leverages this platform to deliver comprehensive analysis reports to ship owners and managers. In a notable case, CMP conducted an advanced analysis applying ISO 19030 to data shared by a shipowner via IoS-OP. Both the resulting analysis data and the underlying analytical process were officially verified by ClassNK as a third party, which issued a Statement of Fact (SOF). These reports provide transparent, data-driven insights that facilitate objective decision-making. By cross-referencing these results with the “Operational Profile Analysis” identifying precisely when and why speed loss occurred and formulating corresponding countermeasures such as the design of subsequent antifouling paint specifications becomes a critical element of the Check-Act phase within the “Hull PDCA cycle”. Ultimately, this insight leads seamlessly

into the next 'Prediction and Planning' stage, ensuring a continuous evolution of vessel efficiency throughout its lifecycle.

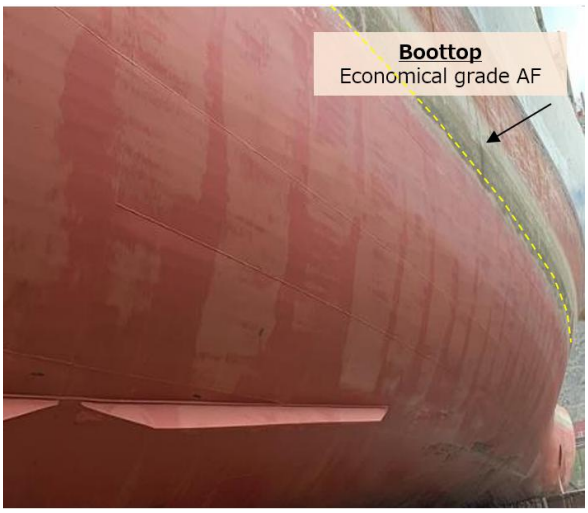


Fig. 17 Dry dock result of 8,900TEU containership

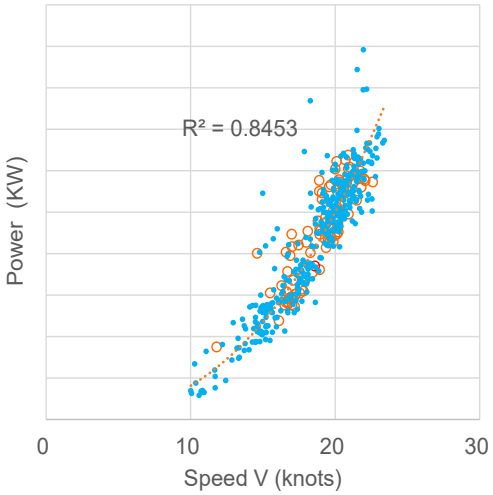


Fig.18: Speed-power curve and fitting

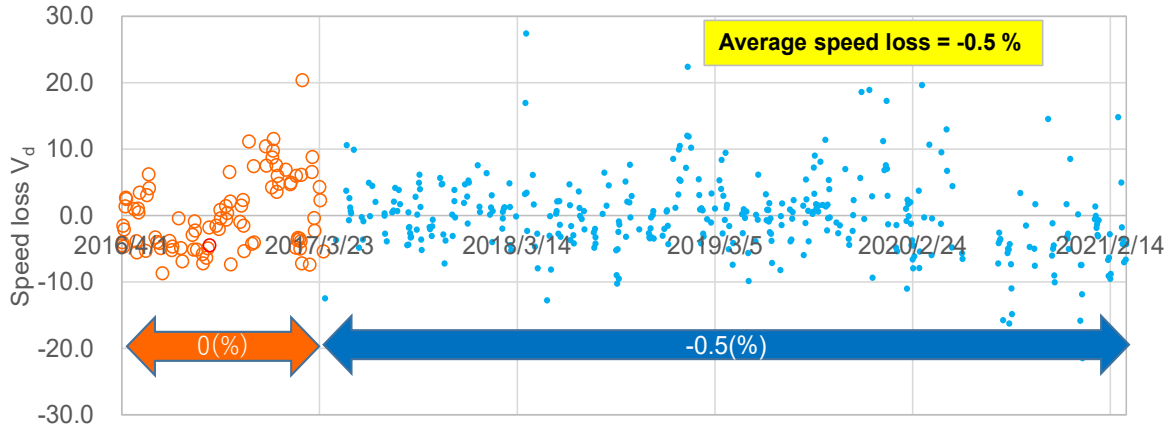


Fig.19: Average speed loss analysis result (in service performance) of the containership

3. Conclusions and Future Prospects

This report identifies ‘CMP-MAP’ as a critical data-driven solution based on the ‘Hull PDCA’ cycle. Within this framework, CMP-MAP integrates the PIR Simulator, an advanced friction resistance estimation tool (FIR theory), ISO 19030-compliant monitoring, and operational profile analysis. Notably, these core components of the CMP-MAP solution - specifically the combination of FIR theory and the PIR Simulator, alongside the integration of performance and operational profile analysis - have been granted the Innovation Endorsement (Product & Solution) certification by ClassNK. As international environmental regulations, such as EEDI, EEXI, and CII, become increasingly stringent, these integrated tools enable the global shipping industry to achieve a critical balance between environmental sustainability and financial performance. This solution represents a significant advancement in maritime technology, ensuring long-term operational excellence through the synergy of digital analysis and high-performance antifouling coatings.

Acknowledgement

Part of the CMP-MAP technology was developed under the 'Next-generation environmental technology development support program' of the Japanese Ministry of Land, Infrastructure, Transport and Tourism (MLIT), and as joint research with ClassNK, both titled 'Method to estimate the changing ratio of in-

situ friction resistance using the data of reduced surface roughness and roughness parameters of hull coating.'

Furthermore, we would like to express sincere gratitude to the Tokyo University of Science (TUS), Tokyo University of Agriculture and Technology (TUAT), and the National Maritime Research Institute (NMRI) for their fundamental research collaboration over many years. We also acknowledges the invaluable support of the global CMP network offices, whose continuous efforts in data collection and field verification were essential to demonstrating the practical efficacy of the “Hull PDCA cycle” and CMP-MAP.

References

MIENO, H. (2020), *Development of a method for estimating frictional resistance increase due to roughness parameters by ship hull coating surface,* Doctoral Thesis, Kobe University

Impact of Data Frequency and Data Block Length on Ship Performance Evaluation in Onboard Monitoring

Naoto Sogihara, National Maritime Research Institute, Tokyo/Japan, sogihara@m.mpat.go.jp
Yoshihiko Sugimoto, Mitsui O.S.K. Lines, Tokyo/Japan, yoshihiko.sugimoto@molgroup.com
Hiroyuki Sano, Mitsui O.S.K. Lines, Tokyo/Japan, hiroyuki.sano@molgroup.com

Abstract

This paper presents findings on two topics. First, in order to investigate the impact of data sampling frequency on performance evaluation in ship performance monitoring, we conducted performance evaluations using both log and sensor data and compared the results. Second, we examined how sensor data block length affects performance evaluation. Block length is set to calculate means and standard deviations from continuously high-frequency measurements. This examination was conducted on merchant ships and training vessels. We believe that their work will contribute to future technical revisions of the existing ISO standard, ISO 19030.

1. Introduction

Advances in measurement and transmission technologies over the past few decades have enabled operators on land to monitor the performance in actual seas of vessels in real time while they are underway. Furthermore, the evaluation of ship performance in actual seas using sensor data transmitted to shore is being vigorously pursued worldwide. Conducting this evaluation requires the construction of large-scale monitoring systems. Consequently, only large-scale companies, such as major shipping lines, can realistically implement evaluations using onboard monitoring data. Companies lacking such monitoring systems must, in practice, rely on ab-log data recorded manually onboard to assess ship performance in actual seas.

The difference between ab-log data and sensor data is outlined below. Ab-log data refers to data recorded daily in the logbook by crew onboard the vessel. Specifically, crew record the values displayed on indicators shown in the bridge into the logbook. Consequently, the data frequency for ab-log data is 24 hours. On the other hand, sensor data has the advantage of being measured automatically and continuously. On the other hand, sensor data offers the advantage of being measured automatically and continuously. This means sensor data is measured at a higher frequency than ab-log data. Sensor data is well-suited for capturing changes in ship performance accompanying shifts in sea conditions. While sensor data is recorded at high frequencies, such as once per second, it is common practice to calculate mean over a specific period from the time history of sensor data for each parameter to handle the ship's steady-state conditions and use these for performance evaluation.

Both ab-log data and sensor data fall under the category of onboard monitoring data. While sensor data is well-suited for evaluating ship performance in actual seas, its use requires utmost care. This is because it necessitates determining the “specific period” for calculating the mean mentioned above. Regarding this point, ISO 19030, which describes methods for evaluating hull and propeller performance changes, recommends setting this specific period (referred to in this paper as the “block data length”) to 10 minutes. However, no technical verification has been reported for the 10-minute block data length.

This paper first examined the applicability of ab-log data and sensor data in evaluating ship performance in actual seas. Using a medium-range tanker in operation as a case study, we conducted performance evaluations using ab-log data and sensor data respectively and compared the results. Then, we investigated the impact of block data length on evaluating a vessel's ship performance in actual seas. This investigation covered seven vessels: the aforementioned medium-range tanker, one training ship, and five merchant ships. We set multiple block data lengths, calculated means for each block length, and evaluated ship performance in actual seas. Based on these investigations, an appropriate data block length for evaluating ship performance in actual seas was discussed.

2. Evaluation of Ship Performance Using Onboard Monitoring Data

2.1. Methodology

This paper treats both ab-log data and sensor data as onboard monitoring data. Since the methodology for evaluating ship performance in actual seas using onboard monitoring data was presented at a past HullPIC conference, *Sogihara et al. (2020)*, we briefly outline the evaluation methodology here.

The evaluation methodology consists of three phases, Fig.1. The first phase is "Measurement." In this phase, it is crucial not only to specify the measurement items but also to determine the block data length, which is the primary focus of this study. Data extraction to ensure the ship's time-steady is performed based on the mean obtained at the determined block data length. This aims to exclude data measured during acceleration or deceleration of the ship, or data measured during maneuvering.

Phase 2 is the "Analysis" phase. For the data extracted in Phase 1, this phase corrects the ship speed for displacement and corrects the engine speed and power for sea states. Filtering for apparent slip ratio is applied to the corrected data to extract data ensuring the accuracy of the ship speed through water.

Phase 3 is the "Evaluation" phase. Using the data extracted via apparent slip ratio filtering in Phase 2, the ship's performance is evaluated. This performance evaluation is based on the Resistance Criteria Method (RCM) which is developed, *Sakurada et al. (2020)* and validated. *Sogihara et al. (2021)* A key feature of the RCM is data extraction based on resistance increase rates. Conventional methods, including ISO 19030, extracted data required for evaluating performance in calm seas using thresholds related to wind speed. A disadvantage of the conventional method is its inability to account for the fact that the impact of sea conditions on performance depends on the ship's size. To overcome this, the RCM incorporates a relative value, the resistance increase rate required for each ship, into the evaluation. In addition to evaluation using the RCM, evaluation utilizing tank test results is also permitted.

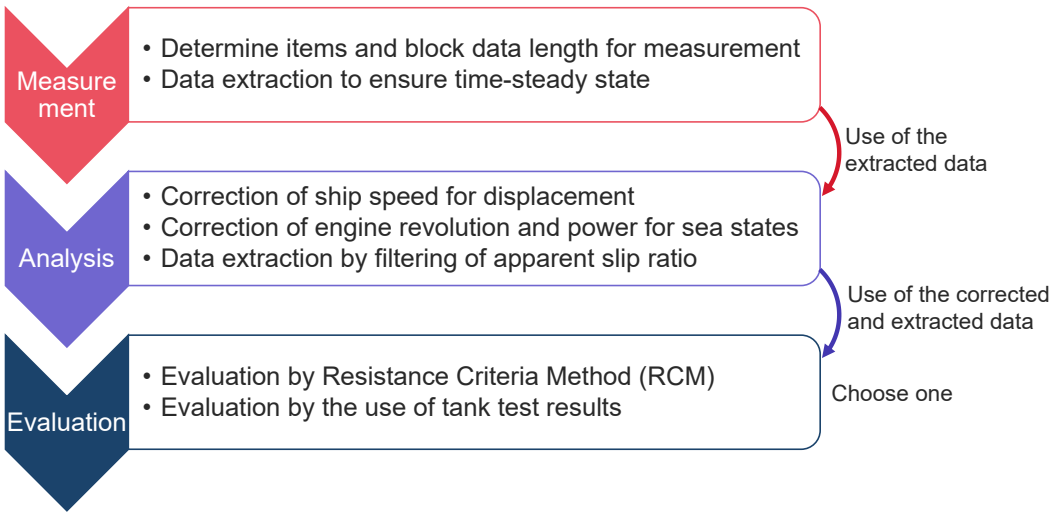


Fig.1: Outlines of the methodology for evaluating ship performance using onboard monitoring data

2.2. Resistance Criteria Method (RCM)

The RCM is outlined in Fig.2. In order to evaluate ship performance in calm seas, the RCM introduces a resistance increase rate δR defined by Eq.(1) which is calculated in the process of 'Correction of engine revolution and power for sea states' in Fig.1.

$$\delta R = \frac{\Delta R}{R_{ms} - \Delta R} \tag{1}$$

ΔR denotes the increase in resistance due to wind and waves and R_{ms} represents the resistance in actual seas which can be obtained in the correction of the engine revolution and power.

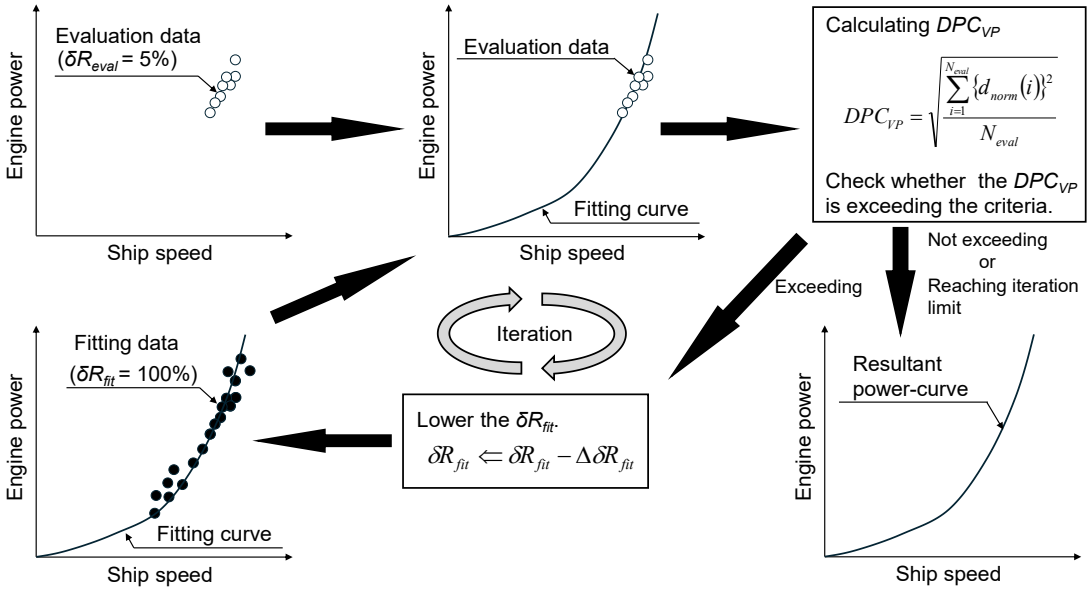


Fig.2: Outline of ship performance evaluation by Resistance Criteria Method

The RCM consists of two types of filtering and they are carried out in parallel. One is filtering to extract data measured in the state where waves and wind are negligible with much smaller δR (δR_{eval} expressed in Fig.2) and the data extracted is called 'evaluation data'. The other is filtering to extract the data with larger δR (δR_{fit} expressed in Fig.2) for drawing the power-curve in a wide range of engine power and the data extracted is called 'fitting data'. The obtained fitting data is applied to Eqs.(2) and (3), which gives the performance power-curve called 'fitting curve'.

$$N_{id} = d_{nv} \cdot V_S \tag{2}$$

$$P_{Bid} = a_n \cdot {N_{id}}^{b_n} \tag{3}$$

V_S, N_{id}, and P_{Bid} represent ship speed, the corrected engine revolution, and the corrected engine power, respectively, where a_n, b_n, and d_{nv} are for the fittings. Whether the fitting data is appropriate as the power-curve is judged by data scatter of the evaluation data around the fitting curve. The scatter is expressed as DPC_{VP} which is defined by Eq.(4) using the number of the evaluation data (N_{eval}) and the nominal distance ($d_{norm}(i)$) between the evaluation data and the fitting curve in the relationship between ship speed and the corrected engine power.

$$DPC_{VP} = \sqrt{\frac{\sum_{i=1}^{N_{eval}} \{d_{norm}(i)\}^2}{N_{eval}}} \tag{4}$$

The criteria for DPC_{VP} is given 2.0, *Sakurada et al. (2021)*. If the DPC_{VP} is less than the criteria, the fitting curve is accepted as the result of the evaluation, which is regarded as the best evaluation. If the DPC_{VP} exceeds the criteria, less δR_{fit} is given and the filtering for the fitting data and the drawing of the fitting curve is conducted again, which are iterated until the DPC_{VP} falls below the criteria. If the DPC_{VP} does not fall below the criteria even after the iteration, the power-curve with the smallest δR_{fit} is regarded as the result of the evaluation with a notation of less reliability than the best evaluation.

2.3. Source of Wave Data

Generally, ships are equipped with an anemometer, enabling the collection of wind data encountered by the vessel. On the other hand, ships are rarely equipped with instruments for observing waves. Ab-log data may include sea state data based on visual observations by crews, but its utility is low due to the absence of data on wave period, among other factors.

Wave radar is also manufactured, but it is not widely used for onboard monitoring due to its high cost and measurement accuracy being affected by weather conditions. Wave forecast data, which is simulation results based on numerical prediction models, has become widely used in recent years. Various institutions and organizations provide wave forecast data. The dataset (ERA5) provided free of charge by ECIMF is well-known as one of wave forecast data and has many applications. This study utilized wave hindcast data provided by the Japan Weather Association, *Sato and Matsuura (2019)*.

3. Ab-log Data versus Sensor Data

3.1. Description on items corrected as ab-log data and sensor data

We investigated the impact on ship performance evaluation using both the ab-log data and the sensor data collected over a period of three and a half years. This investigation was conducted on a medium-range tanker. The items collected as ab-log data and sensor data are listed in Fig.3.

Items (Sampling interval)	Ab-log data (1 day)	Sensor data (1 hour)
Speed over ground	X	X
Speed through water		X
Heading angle		X
Course		X
Rudder angle		X
Wind speed and direction	Beaufort scale	X
Wave height, period, and direction	Visual observation	Forecast
Engine revolution	X	X
Engine power	Calculated from FOC	X
Draft and displacement	X	X

Fig.3: Items corrected as ab-log data and sensor data

Items collected as the ab-log data do not include speed through water, bow angle, course, or rudder angle. While the use of speed through water is essential for this performance evaluation, speed over ground is used instead of the speed through water in this study.

Regarding wind, relative wind speed and direction are collected. Further, Beaufort scale number, which is defined by the WMO based on true wind speed, are recorded in the logbook. For waves, while swell height and direction are recorded, the period is unknown, which means that the added resistance due to swell cannot be estimated. Therefore, this study treats only wind waves and estimates the significant wave height (H) of wind waves according to the Beaufort scale shown in Fig.4.

The mean wave period (T) for wind waves is estimated using Eq.(5), which is valid under the assumption of fully developed waves, *Price and Bishop (1974)*. The direction of wind waves is assumed to be equal to the true wind direction relative to the bow.

$$T = 3.86\sqrt{H} \tag{5}$$

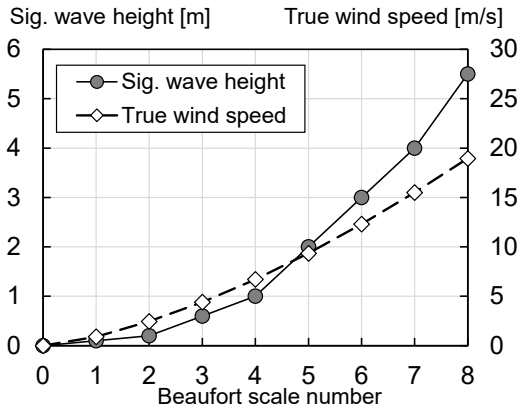


Fig.4: Relationship between significant wave height, true wind speed and Beaufort scale number

The displacement of the ship in operation is calculated based on its draft. The displacement calculation utilizes the draft measured at departure. The calculated displacement is used for both performance evaluation using the ab-log data and performance evaluation using sensor data.

3.2. Cases Study on Performance Evaluation

This study is subject to a medium range (MR) tanker whose length is 175 meter. A system for collecting the sensor data is installed on the MR tanker and calculates the mean for every one hour. This study uses the ab-log data and the sensor data collected for about three years, which provides 925 datasets for the ab-log data and 16,856 datasets for the sensor data for the case study. The comparison between the ab-log data and the sensor data for a voyage on ship speed over ground, engine power, wind speed, and significant wave height is shown in Fig.5. Here, V_{des} and MCR represents the designated ship speed and maximum continuous rate.

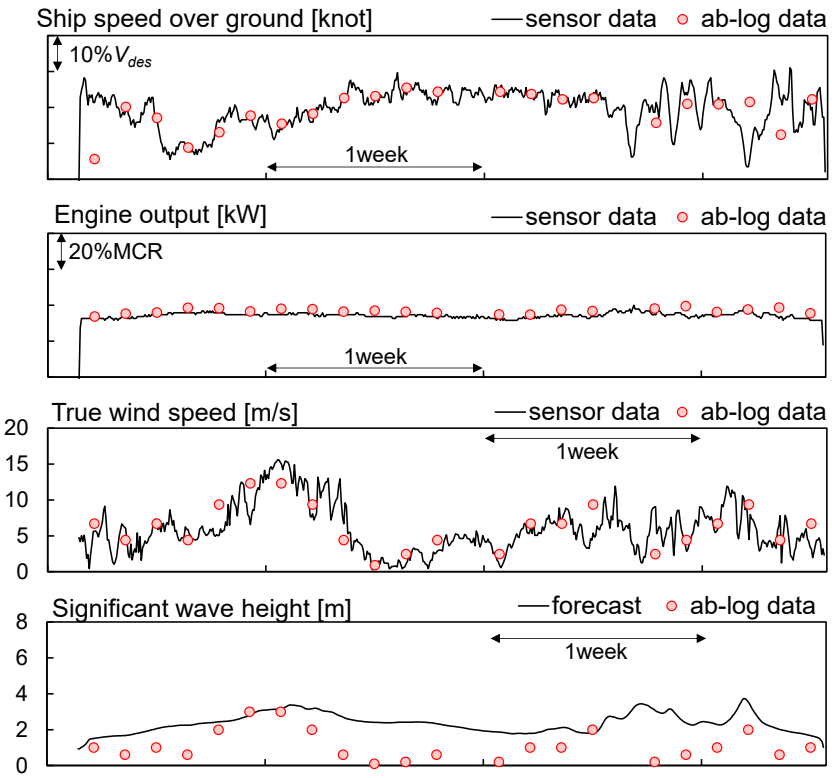


Fig.5: Comparison between ab-log data and sensor data in one voyage

The performance of the MR tanker in calm seas is evaluated using the ab-log data and the sensor data in accordance with the methodology described in chapter 2. The estimation of the parameters required for the correction of the engine revolution and power such as propeller open characteristics, self-propulsion factors, and lateral projected area above waterline are performed according to the simplified method. The estimation of added resistance due to waves and that due to wind is carried out using the estimated hull form data based on the principal particulars. These estimations are conducted utilizing the OCTARVIA Web Applications, *Sogihara et al. (2024)*.

The evaluated power-curve in calm seas is shown in Fig.6. The figure shows that the scatter of the evaluation data around the obtained power-curve based on the ab-log data is larger than that based on the sensor data. DPC_{VP} by the ab-log data yields to 3.8% while that by the sensor data yields to 2.3%, which mean that the use of the ab-log data for evaluating the ship performance can result in the loss of the reliability of the evaluation.

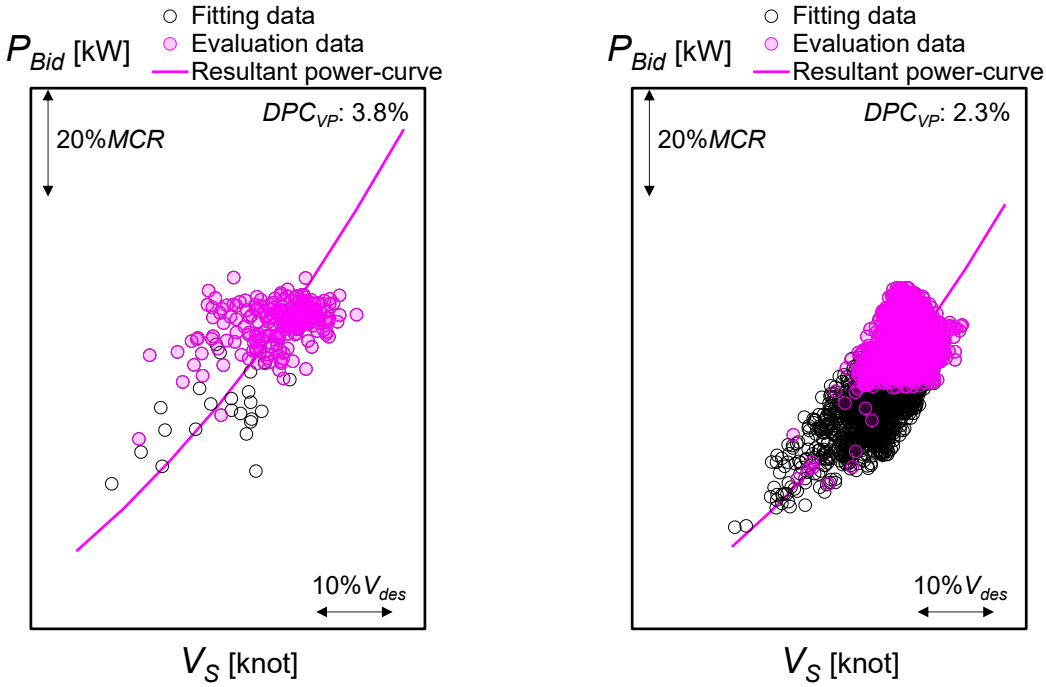


Fig.6: Evaluated power-curve in calm seas using ab-log data and sensor data
(left: ab-log data, right: sensor data)

The reason for the larger scatter by using the ab-log data can include:

1) Use of the speed over ground: In terms of hydrodynamics the flow field around the ship is usually expressed as relative speed to the ship, which requires the use of the speed through water in evaluating the ship performance. In other words, using the speed over ground cannot eliminate the effect of current, which cannot be error source deteriorating the accuracy for the evaluation.

2) Use of the engine power calculated from fuel oil consumption (FOC): The property such as density and viscosity of the fuel oil provided for the ship affects the relationship between FOC and the engine power. Unless the fuel property is considered appropriately in the calculation of FOC, the engine power cannot be calculated with accuracy.

3) Sea state representation by Beaufort scale number: The significant wave height and wave period is estimated based on the true wind speed under the relationship shown in Fig.4. The wave height and wave period are varied irrespective of the true wind speed, which can fail to ensure their accuracy if the estimated wave height and period are used.

4) Use of the ab-log data: The sampling interval for the ab-log data is one day while that for the sensor data is one hour. Longer sampling interval cannot capture the change of sea state and concurrently the change of ship performance such as ship speed and FOC. Further, the use of the ab-log data cannot be clarified whether the ship is in unsteady-condition such as acceleration or steering, which can contain the data measured in unsteady-condition in the evaluation.

The comparison between the evaluated power-curve based on the ab-log data and that based on the sensor data is demonstrated in Fig.7, representing that the red and black line correspond to the lines in Fig.6. Here, *diff. P* is defined by Eq.(6), where $P_{Bid|ab\text{-}log}$ and $P_{Bid|sensor}$ are the evaluated engine power using the ab-log data and the sensor data, respectively.

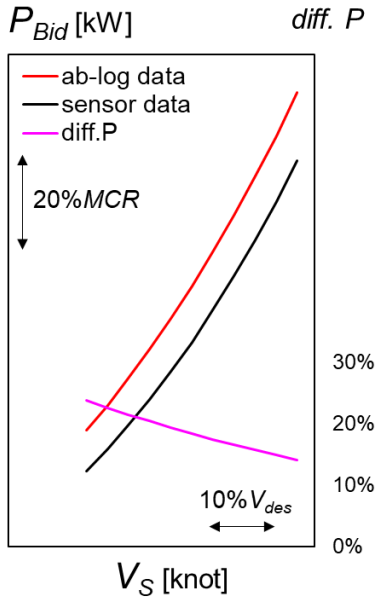


Fig.7: Comparison on the evaluated power-curve

$$diff.P = \frac{P_{B_{id}}|_{ab-log}}{P_{B_{id}}|_{sensor}} - 1 \tag{6}$$

Fig.7 indicates that the evaluated engine power using the ab-log data is 20% larger than that using the sensor data. Considering that the data scatter in the case the sensor data is used is smaller, the ab-log data should not be used and the use of the sensor data is recommended for evaluating ship performance in calm seas.

4. Investigation of Block Data Length

4.1. What is Block Data Length?

In the evaluation of ship performance using sensor data, the raw data for each sensor are measured with high frequency such as 1 Hz. Fig.8 shows an example of raw data of wind speed measured by an anemometer. In terms of data handling, it is recommended to deal with the mean in a certain period, not to use the raw data as it is. This study defines such period as 'block data length (BDL)'.

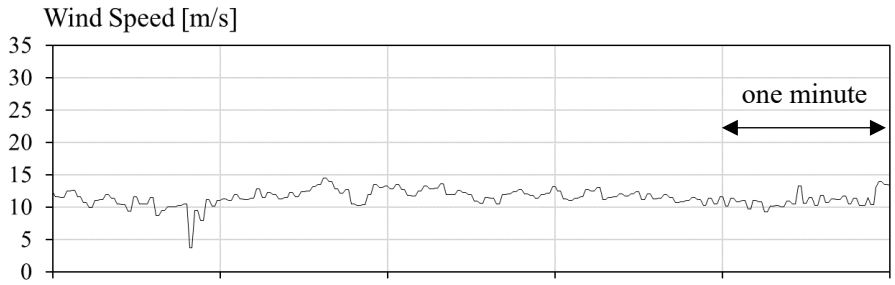


Fig.8: Example of raw data

4.2. Description in ISO 19030

ISO 19030 (2016) stipulates the procedure for measuring changes in hull and propeller using onboard monitoring data. This standards describes the BDL as follows.

> "Furthermore, the data set shall be split in non-overlapping, consecutive blocks of 10 min and for every block of 10 min, the mean, the standard deviation, the maximum, and the minimum value for every parameter shall be computed.
>
> NOTE It is noted that 10 min data blocks can be useful for analysis."

ISO 19030 recommends the use of 10 minutes as the BDL, which are not validated by the researches or publications. Furthermore, whether another BDL is applicable or not in the evaluation is not mentioned in ISO 19030. Taking into consideration that the DBL can differ up to the monitoring system, it is very important to investigate the impact of the BDL to the ship performance evaluation, which is the motivation of chapter 4.

4.3 Investigation on Impact of Block Data Length to Ship Performance Evaluation

This study uses the datasets of the sensor data collected on the following ships:

- ✓ Oceangoing container ship 1 (CS-1)
- ✓ Oceangoing container ship 2 (CS-2, sister ship of oceangoing container ship A)
- ✓ Oceangoing container ship 3 (CS-3, sister ship of oceangoing container ship A)
- ✓ Oceangoing pure car carrier (PCC)
- ✓ Very large ore carrier (VLOC)
- ✓ Medium range tanker (MRT, same ship as used in chapter 3)

For the ships above, the datasets compiled under the various DBL were used in the investigation. The DBL was given 5, 10, 20, 30, 60, 90, and 120 minutes. The example of the mean with the DBL given 10, 60, and 120 minutes is illustrated in Fig.9. Using the datasets of the six ships under seven DBL mentioned above, the ship performance in calm seas is evaluated in accordance with the methodology described in chapter 2.

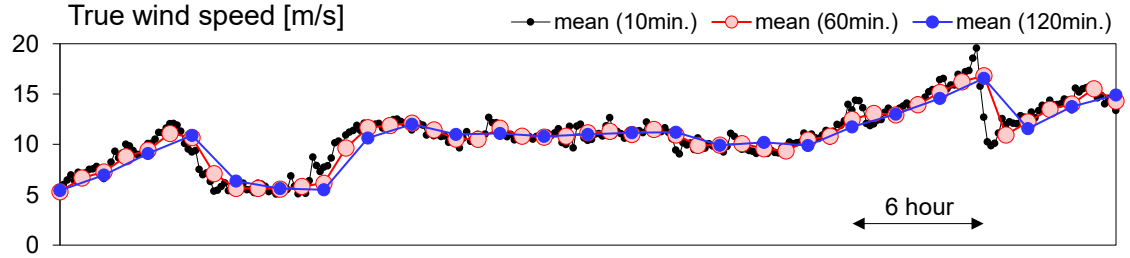


Fig.9: The mean with the DBL given 10, 60, and 120 minutes

As the results of the evaluation, seven power-curves are obtained. The impact of the DBL on the evaluation of the ship performance is quantified by the coefficient of variation CV_P for each ship which is defined by Eqs.(7) to (9).

$$CV_P = \frac{\sigma_P}{\mu_P} \tag{7}$$

$$\mu_P = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{8}$$

$$\sigma_P = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mu_P - P_i)^2} \tag{9}$$

P_i means the engine power in calm seas at the designated ship speed based on by the datasets of each DBL. n means the number of the DBL, which is given $n = 7$.

Fig.10 shows that CV_P is less than 1.0% for ships other than MRT while CV_P of MRT is remarkably larger than that of the other ships. This means that, excluding MRT, the BDL does not have a significant impact on the evaluation of the ship performance in calm seas. Fig.11 indicates the obtained power-curve by seven DBL and shows that seven power-curves almost overlaps for CS-3 and PCC. Fig.11 also shows that seven power-curve for MRT shows a different slope depending on the DBL.

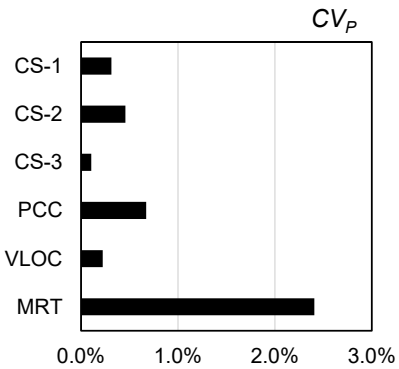


Fig.10: Coefficient for variation of the engine power at designated ship speed

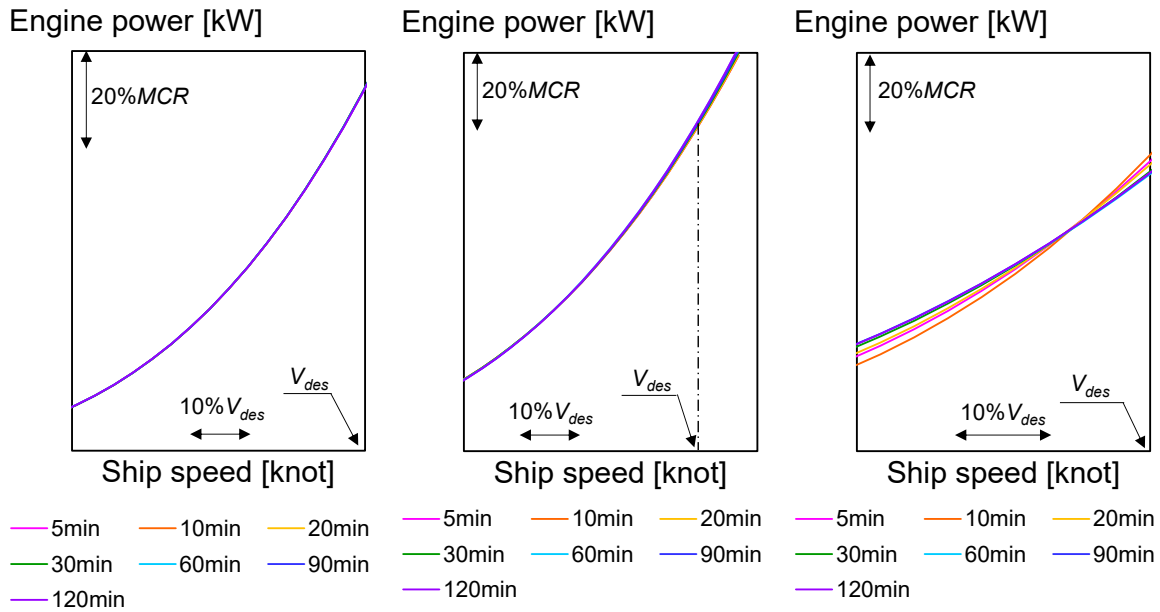


Fig.11: Power-curve for seven DBL (left: CS-3, center: PCC, right MRT)

The datasets of MRT are examined to reveal the reason why the resultant power-curve of MRT shows a different trend from the other ships. Focusing the time history of engine power under the DBL 5 minutes and 60 minutes results in Fig.12. The figure shows that the time history of the engine power under the DBL 5 minutes contains noises whilst such noises does not appear under the DBL 60 minutes.

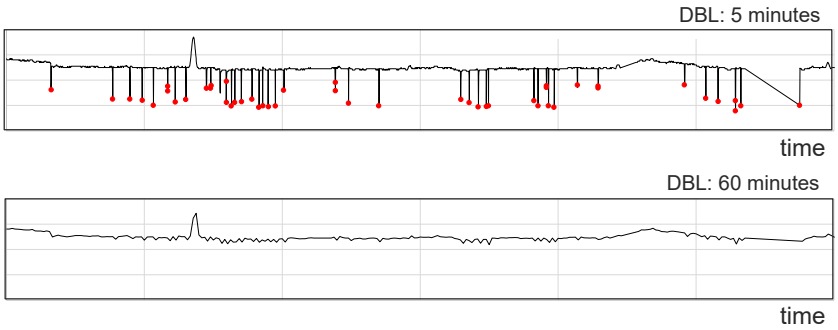


Fig.12: Time history of engine power (upper: DBL 5 minutes, lower: DBL 60 minutes)

Fig.12 implies that the noises can affect the accuracy for evaluating ship performance. We consider that the filtering by the standard deviation during the DBL can be applied to eliminate the noises from the time history. Specifically, we propose to introduce a criterion of 1 kW with respect to the standard deviation during the DBL for eliminating the noises. Introducing the criteria above and evaluating the ship performance using the datasets for seven DBL results in Fig.13, showing that the criterion of 1 kW is effective for ensuring the reliability of the evaluation of the ship performance.

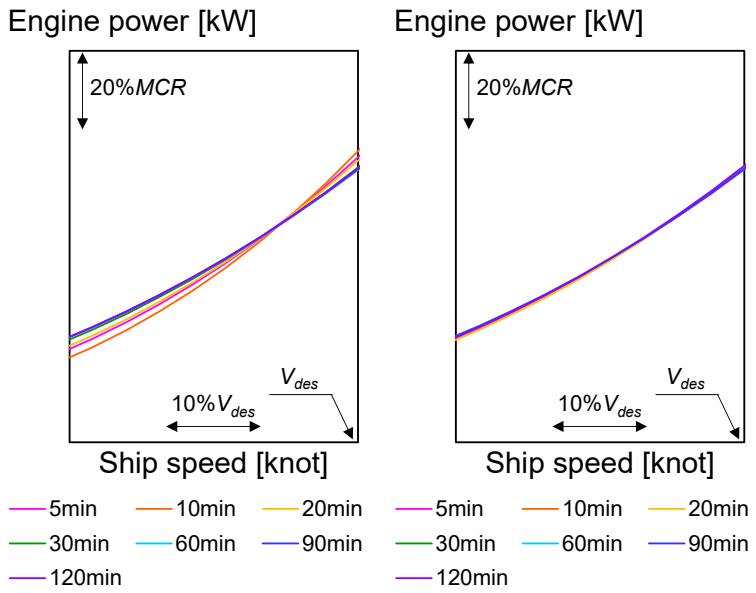


Fig.12: Power-curve of MRT
(left: without criterion of 1 kW, right: with criterion of 1 kW)

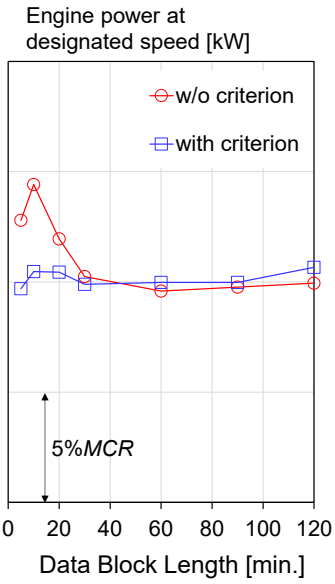


Fig.13: Impact of DBL on the evaluated power

Figs.11 and 12 suggest that the DBL can be freely given as far as it is between 5 minutes and 120 minutes. Nevertheless, we believe that the time history of the sensor data under the longer DBL, specifically 90 minutes and 120 minutes, cannot follow the change in waves and wind in actual seas. Furthermore, Fig.13 demonstrates the criterion of 1 kW is required for the reliable evaluation when the DBL is less than 20 minutes. Accordingly, we recommend that the DBL determined for the use of the ship performance evaluation should be between 10 minutes and 60 minutes, considering the recommendation by ISO 19030.

5. Conclusions

This paper investigated two key aspects of ship performance evaluation using onboard monitoring data: the impact of data sampling frequency and the effect of block data length (BDL) on evaluation accuracy. Regarding the data sampling frequency, the comparison between ab-log data (24-hour sampling interval) and sensor data (1-hour sampling interval) on a medium-range tanker revealed significant differences in evaluation reliability. The performance evaluation using ab-log data resulted in DPC_{VP} of 3.8%, while sensor data achieved 2.3%, indicating that sensor data provides significantly better accuracy. Furthermore, the evaluated engine power using ab-log data was approximately 20% higher than that obtained from sensor data. These discrepancies are attributed to several factors: the use of speed over ground instead of speed through water, inaccuracies in engine power calculation from fuel oil consumption without proper fuel property consideration, limitations of Beaufort scale-based sea state representation, and inability to capture changes in sea state and ship performance due to longer sampling intervals. Based on these findings, the use of sensor data is strongly recommended over ab-log data for ship performance evaluation in actual seas.

Concerning the impact of block data length, the investigation was conducted on six vessels (three container ships, one pure car carrier, one very large ore carrier, and one medium-range tanker) using seven different BDL settings: 5, 10, 20, 30, 60, 90, and 120 minutes. The results showed that for most ships, the coefficient of variation CV_P was less than 1.0%, indicating that BDL does not significantly affect performance evaluation. However, the medium-range tanker (MRT) exhibited a remarkably larger CV_P, with obtained power-curves showing different slopes depending on the BDL. Detailed analysis revealed that shorter BDL (e.g., 5 minutes) introduced noise in the engine power time history, while longer BDL (e.g., 60 minutes) effectively eliminated such noise.

To address the noise issue observed in the MRT data, this study proposes introducing a filtering criterion based on standard deviation during the BDL. Specifically, a criterion of 1 kW for the standard deviation of engine power was proposed and tested. The application of this 1 kW criterion proved effective in eliminating noise from the time history and ensuring the reliability of ship performance evaluation, resulting in consistent power-curves across different BDL settings. Based on these results and taking into account the recommendation by ISO 19030, we recommend that the BDL determined for the use of ship performance evaluation should be between 10 minutes and 60 minutes.

Acknowledgement

This study was conducted by the Initiative on Evaluation of Ship Performance in Actual Seas, the "OCTARVIA Project," a Japan Maritime Cluster Collaborative Research project and its phase 2 project (OCTARVIA2). We are grateful to all the parties concerned who discussed the content of this study in the working groups.

References

ISO 19030 (2016), *Ship and marine technology -Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

PRICE, W.G.; BISHOP, R.E.D. (1974), *Probabilistic Theory of Ship Dynamics*, Chapman and Hall, London, pp.157-163

SAKURADA, A.; SOGIHARA, N.; TSUJIMOTO M. (2020), *Development of a Filtering Method for the Evaluation of Performance in Calm Sea Based on Onboard Monitoring Data*, J. Japan Society of Naval Architects and Ocean Engineering 31

SAKURADA, A.; SOGIHARA, N.; TSUJIMOTO, M.; SATO, H. (2021), *Development of resistance criteria method for ship performance evaluation by onboard monitoring data*, Conf. Full Scale Ship Performance, pp.19-26

SATO, Y.; MATSUURA, K. (2019)*, Forecasting and hindcasting*, The Naval Architect, November, pp.31-34

SOGIHARA, N.; SAKURADA, A.; TSUJIMOTO, M.; SATO, H. (2020), *Validation of the Evaluation Method of Ship Performance Using Onboard Monitoring Data -Development of Resistance Criteria Method with Apparent Slip Ratio and Quality Management*-, 5th HullPIC Conf., pp.82-95

SOGIHARA, N.; SAKURADA, A.; TSUJIMOTO, M. (2021), *Validation of Filtering Method for Evaluating Ship Performance in Calm Sea Using Onboard Monitoring Data*, J. Japan Society of Naval Architects and Ocean Engineering 33, pp.25-33

SOGIHARA, N., KURODA, M., AKAMATSU, T., YANAGIDA, T., SUGIMOTO, Y., and ITO, R. (2024), *Simulation of Fuel Oil Consumption of Ships Based on Big Data Analysis by OCTARVIA Web Applications*, 9th Hull Performance & Insight Conference, pp.45-55

Introduction of the New ISO Initiative for the Evaluation of Fuel Consumption and Propulsion Performance in Actual Seas (ISO 25817)

Tetsuro Ashida, Mitsui O.S.K. Lines, Tokyo/Japan, tetsuro.ashida@molgroup.com
Mariko Kuroda, National Maritime Research Institute, Tokyo/Japan, kuroda@m.mpat.go.jp

Abstract

This study introduces the new ISO initiative (ISO 25817), which aims to develop the world's first fair and transparent method for quantitatively evaluating the fuel consumption and propulsion performance in actual seas, including winds and waves effect, using operational data collected during voyages. While ISO 19030 evaluates relative performance deterioration, ISO 25817 enables absolute performance evaluation in actual seas. The method consists of performance prediction and validation using 14 essential onboard parameters, allowing the consistency between theoretical design models and the ship's actual monitored performance. This enables consistent and quantitative evaluation and supports decision-making for design and operational optimization.

1. Introduction

ISO 25817 is a new initiative agreed upon at the ISO subcommittee meeting held in Baltimore in June 2025 to advance the development of the standard, *ISO (2025)*. This initiative aims to support the design and operation of ships that reduce greenhouse gas (GHG) emissions, using fair and transparent evaluation methods. It is expected to significantly contribute to reducing loads on the environment by the shipping sector. The working group was established to develop a standard method for calculating fuel consumption and ship performance in actual seas using monitoring data and ship design data, and to establish a standard method for calculating a ship's lifecycle fuel consumption.

This paper will outline the draft ISO 25817, the initiative's plan, and presents practical examples demonstrating how the effects of fouling and aging, derived from onboard monitoring data analysis, can be used to evaluate fuel consumption over a ship's lifecycle and how the results can be applied to improve operational efficiency.

2. Outline of ISO 25817

ISO 25817 consists of two parts:

- ISO 25817-1 Ships and marine technology- Evaluation of fuel consumption and propulsion performance in actual seas- Part 1: Method for evaluating fuel consumption and propulsion performance
- ISO 25817-2 Ships and marine technology- Evaluation of fuel consumption and propulsion performance in actual seas- Part 2: Index for life cycle fuel consumption

2.1. Outline of ISO 25817-1

Part 1 provides methods for estimating ship propulsion performance (speed, power output, fuel consumption) in actual seas, as well as methods for verifying the estimation results using monitoring data of a ship operating in service.

For estimating ship performance in actual seas, this method takes ship design data with reference to energy-saving technology installation conditions as inputs. The effects of winds and waves on a ship in actual seas is estimated through calculations and tank tests. It then calculates the power curve—the relationship between speed, power, and RPM—and predicts ship speed, power, and fuel consumption based on engine operating conditions.

In verifying the predicted ship performance results in actual seas, a method utilizing monitoring data of a ship operating in service is adopted. Regarding ship monitoring, while the rapid spread of the system raises expectations for accurately evaluating a ship's true performance through such monitoring, it is a fact that the sequence of procedures − measurement, analysis, and evaluation − is based on each company's methodology. This hinders the objective evaluation of ship performance in actual seas. ISO25817 presents an approach for objectively evaluating and verifying actual ship performance through ship monitoring. This includes determining what measurement items are necessary and the required accuracy for each, followed by data filtering techniques before analysis and analysis methods concerning disturbances.

Even if the same measurement items are acquired with the same accuracy, the reliability and quality of analysis results vary depending on the encountered weather conditions and operational conditions. Therefore, a grade classification approach to indicate the reliability and quality of results is also proposed in ISO25817.

Fig.1 shows a conceptual diagram illustrating the content of Part 1. Using the performance estimation method and verification method enables comparison of power curves under designated weather conditions. For example, results such as those shown in Fig.2 can be obtained.

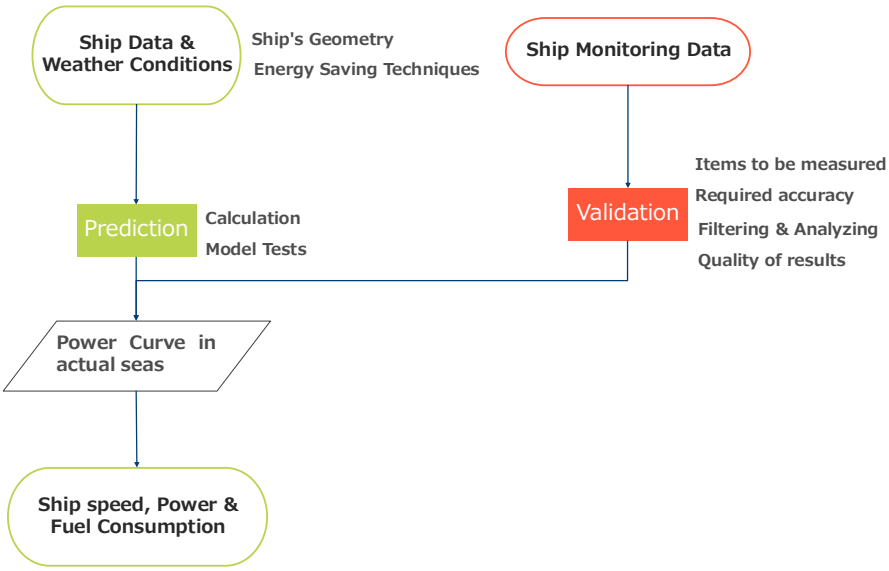


Fig.1: Summary of Part1 of the proposal

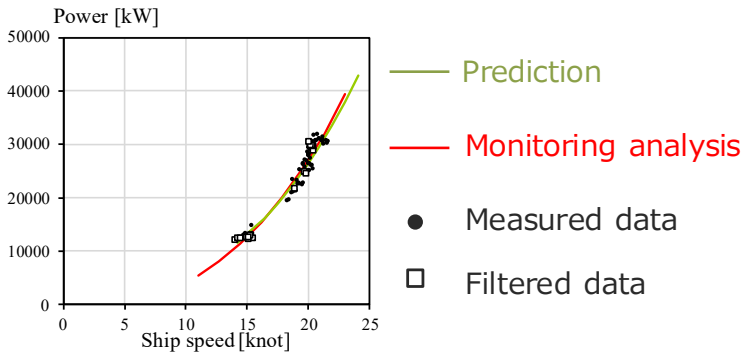


Fig.2: Example of a result of Part 1, *Sogihara (2021)*

2.2. Outline of ISO 25817-2

Using a performance estimation method for actual seas, fuel consumption under designated weather conditions can be calculated. Furthermore, to address the issue that evaluation conditions vary

depending on operational patterns and make comparisons difficult, Part 2 adopts a standard operational model and presents a method for calculating a ship lifecycle index representing fuel consumption and GHG emissions.

The standard operational model used for calculating a ship lifecycle index is shown in Table I. To objectively evaluate ships in actual seas, standard evaluation conditions must be established. However, the environmental conditions affecting ships in actual seas vary depending on ship specifications, such as type and size, as well as the operational status of each ship (route, season, loading condition, etc.) and broader economic conditions, making it impractical to define a single set of conditions. The standard operational model therefore provides a framework that systematically organizes evaluation criteria to uniformly define performance conditions in actual seas across different operational patterns.

Table I:Standard operational model

Item	Remarks
Weather conditions	Weather conditions corresponding to the assumed route must be specified. ISO 25817 provides weather conditions for the following representative routes based on seasonal statistical data. – North Pacific, – West Pacific, – Asia-Europe via Suez, – Asia-Europe via Cape, North Atlantic, – Asia-Middle East – World-wide,
Loading condition	Information for the outward and homeward voyages must be specified.
Designated ship speed	Information for the outward and homeward voyages must be specified.
Rate of operating days per year	The required input must be specified to calculate the total fuel consumption during operation.
Effects of fouling and aging	Input data on biofouling and aging degradation conditions affecting hull resistance, propeller efficiency, and fuel consumption rate must be specified for each calculation step.
Evaluation period	The required input must be specified to calculate the total fuel consumption during operation.

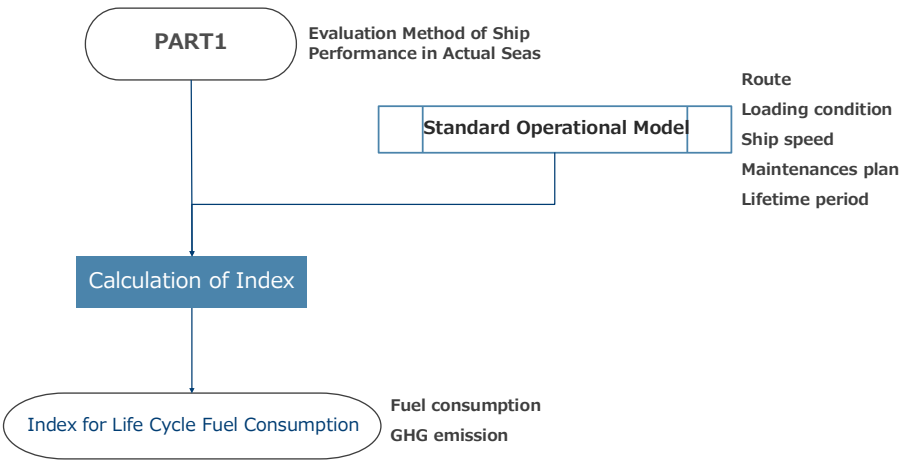


Fig.3 Summary of Part2 of the proposal

The introduction of alternative fuels and energy-saving technologies to reduce greenhouse gas emissions has progressed. Consequently, evaluating fuel efficiency in actual seas has become important not only for short-term assessments, such as a single voyage or one year of operation, but also from a lifecycle perspective. Accordingly, the standard operational model enables long-term, comprehensive

evaluation of fuel efficiency under actual operational conditions by considering not only encountered weather conditions, such as wind and waves, but also the effects of aging and fouling deteriorations. Fig.3 shows a conceptual diagram illustrating the scope of Part 2.

3. Evaluation of Life Cycle Fuel Consumption

To evaluate lifecycle fuel efficiency, models for fouling and aging deteriorations are utilized. Although these effects can sometimes be estimated from factors such as paint quality, they remain parameters that are difficult to define quantitatively. One practical approach is to use the rate of increase in required power derived from the analysis of onboard monitoring data of ships in service. ISO 19030-2, Annex K, also provides a method for calculating this power increase rate as part of the PPI calculation, which can be utilized for this purpose. In addition, performance changes are mainly influenced by factors specific to the hull and propeller. By modeling these elements separately, it becomes possible to independently evaluate measures such as hull coating improvements and underwater propeller cleaning. Therefore, ISO 25817 adopts the following four parameters to represent fouling and aging deteriorations:

- Increase in hull resistance due to fouling
- Increase in hull resistance due to aging
- Deterioration in propeller efficiency due to fouling
- Deterioration in fuel efficiency due to aging

These separated parameters allow the individual treatment of hull aging, hull fouling, propeller fouling, and main engine aging. However, since the output increase rate obtained using this model does not correspond uniquely to that obtained using methods such as ISO 19030-2 Annex K, *ISO (2016)*, conversion work is required. The sequence of steps and how to interpret the results is going to be explained using a sample.

3.1. Input for the Evaluation of Life Cycle Fuel Consumption

To evaluate fuel consumption over a ship's lifecycle, the following information is required in addition to the effects of aging and fouling:

- Weather conditions
- Rate of operation
- Loading conditions
- Ship speed
- Baseline: Speed-Power curve in calm seas for the clean condition
- Evaluation period (number of years)

These factors should be considered in combination. For example, if the ship is operating fully loaded, the encountered weather conditions, ship speed, rate of operation, and baselines are required for the loading condition.

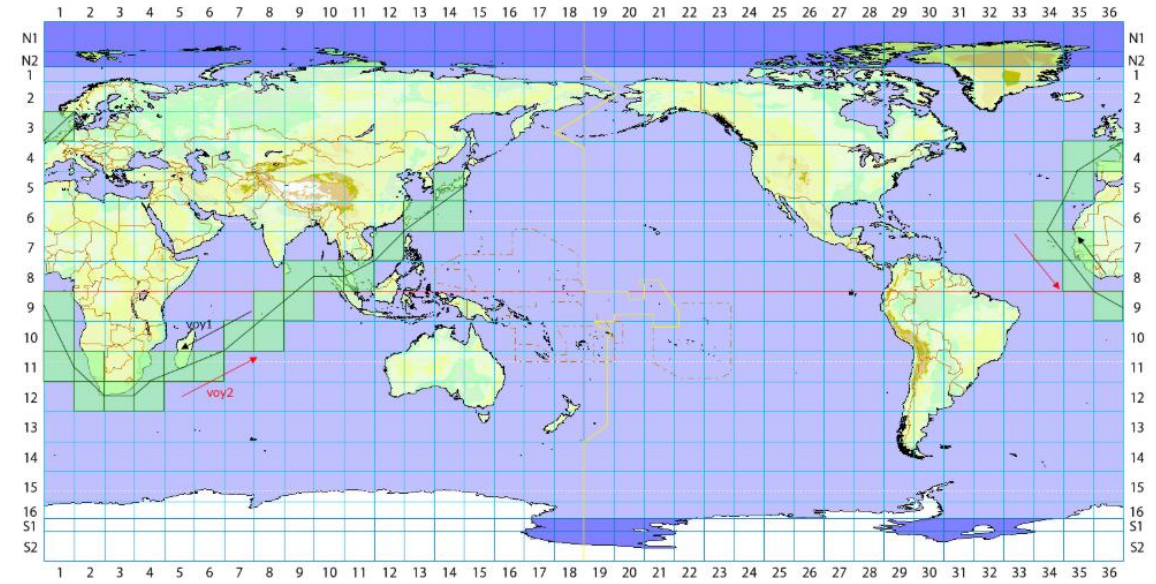
In this evaluation, the input data shown in Table III were used for a pure car carrier, Table II. Here, the EC in the frequency distribution in Fig.4 refers to the weather conditions represented by the combination of mean wind speed V_w, significant wave height $H_{1/3}$, and mean wave period T_{01}, Table IV.

Table II: Principal parameters of an assumed pure car carrier, *Sasaki et al. (2009)*

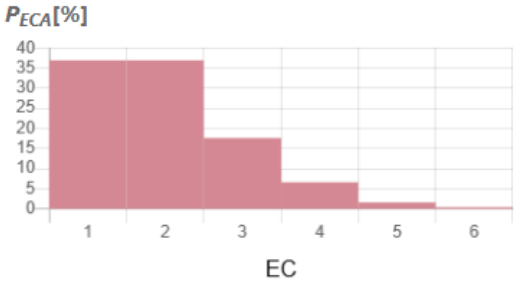
Length between perpendiculars: L_{pp}	190.00 m
Breadth: B	32.26 m
Maximum continuous rating of engine: MCR	15,920 kW
Engine revolution at MCR: N_{MCR}	112.4 rpm
Deadweight: DWT	16,500 t

Table III: Input for the Life Cycle Fuel Efficiency Evaluation

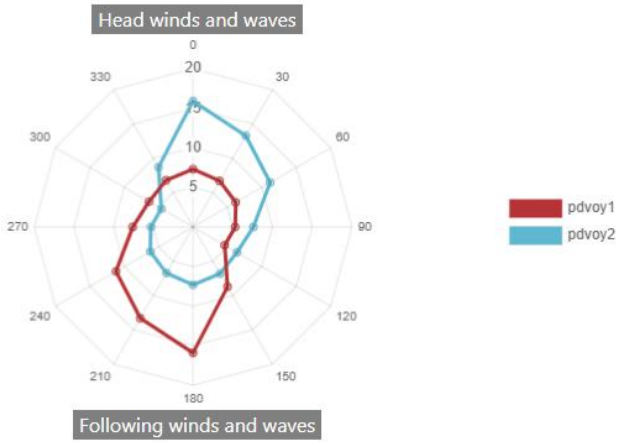
Weather conditions	Assuming an Asia–Europe route via the Cape; Weather conditions are applied based on statistical data, *Kuroda et al.* (*2022*)	As shown in Fig.4
Rate of operation	Assumption	80%
Loading condition	Assumption	Design full load condition
Ship speed	Assumption	18 kn
Base line	Tank test results	As shown in Fig.5
Evaluation period	Assumption	15 years



(a) Assumed ocean areas



(b) Frequency distributions for EC



(c) Frequency distributions for directions

Fig.4: Weather conditions

Table IV: Evaluation conditions (EC)

EC	V_w	$H_{1/3}$	T_{01}
1	4.4 m/s	1.25 m	4.3 s
2	6.9 m/s	2 m	5.5 s
3	9.8 m/s	3 m	6.7 s
4	12.6 m/s	4 m	7.7 s
5	15.7 m/s	5.5 m	9.1 s
6	19.0 m/s	7 m	10.2 s

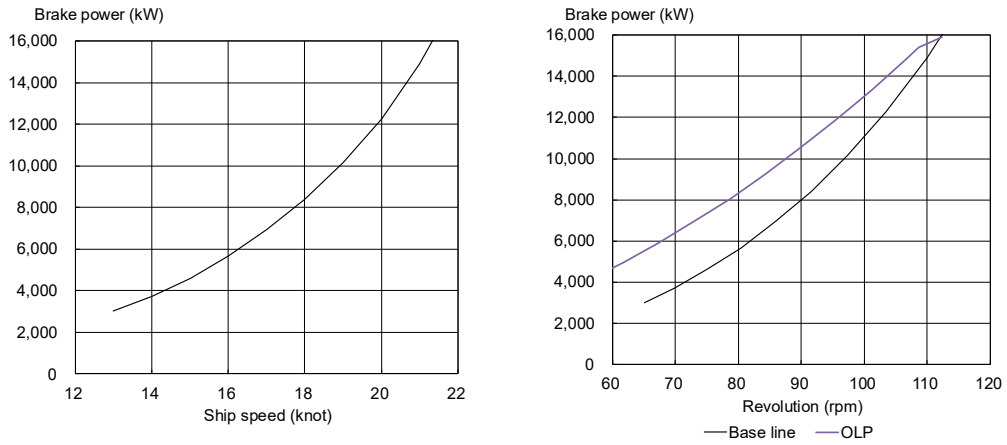


Fig.5: Base line for the pure car carrier

3.2. Incorporation of Effects from Fouling and Aging Deteriorations

The results of analyzing fouling and aging deteriorations using the RCM method, *Sakurada et al. (2021),* were utilized. *Sakurada et al. (2026)* determines the rate of power increase or the rate of the decrease of ship speed between events (dry docking, propeller cleaning). The analysis results for the power increase rate (at constant speed) during the target ship's 5-8 years of service are shown in Fig.6.

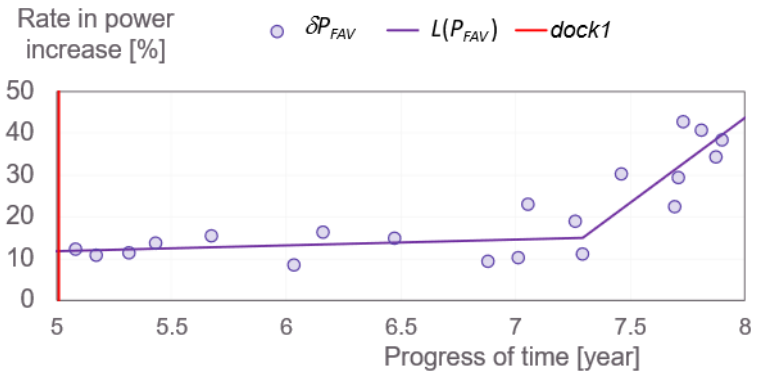


Fig.6: Example of analysis results for rate in power increase using onboard monitoring data, *Sakurada et al. (2026)*

In evaluating lifecycle fuel consumption, it is desirable to separate the output increase rate into effects attributable to the hull and effects attributable to the propeller. The required power in clean conditions in calm seas is expressed using resistance, ship speed and propulsion efficiency as shown in Eq. (1).

$$P_0 = \frac{R_0 V}{\eta_0} \tag{1}$$

P_0 is the brake power for the clean condition, R_0 is the hull resistance for the clean condition, V is the ship speed, η_0 is the propulsion efficiency for the clean condition.

Fouling effects originate from both the hull and the propeller. In such cases, both hull resistance and propeller efficiency can influence the power in Eq. (1). However, separating the power increase obtained from monitoring data analysis into these two components requires a comparison of power before and after either propeller cleaning or hull cleaning independently, which cannot be reliably ensured using monitoring data during normal service operation. Therefore, in this study, all fouling effects are assumed to be attributable to hull fouling. Under this assumption, when expressing power P as a time series with elapsed time denoted by t, it can be represented by Eq. (2).

$$P(t) = \frac{R(t)V}{\eta(t)} \tag{2}$$

$$R(t) = R_0\big(1 + r(t)\big) \tag{3}$$

$$\eta(t) = \eta_0 \tag{4}$$

In Fig.6, the approximation line for the period from 5 to 7.3 years is drawn at 1.40%/year, and the approximation line for the period from 7.3 to 8 years is drawn at 43.57%/year. Furthermore, according to the referenced paper, the power increase rate immediately after dry docking in the 8th year of operation was 11%. Then, the output increase rate due to aging degradation was set at 1.38%/year. The impact of fouling was calculated by subtracting 1.38 from each value, resulting in a fouling impact of 0.02%/year from 5 to 7.3 years and 42.19%/year from 7.3 to 8 years. Expressing this as the resistance increase rate yields Fig.7.

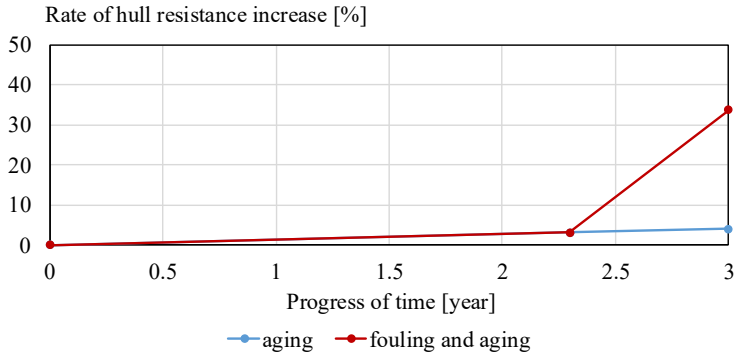


Fig.7: Setting of rate in resistance increase

Based on the settings in Fig.7, the recovery interval is extended to every 3 years over the 15-year life cycle assessment period to evaluate fuel consumption. Furthermore, since the results in Fig.7 show a sharp performance deterioration after 2.3 years, evaluations are also conducted for recovery intervals of every 2.0 years and every 2.5 years to investigate their effects. Fig.8 shows the resistance increase rates for each case.

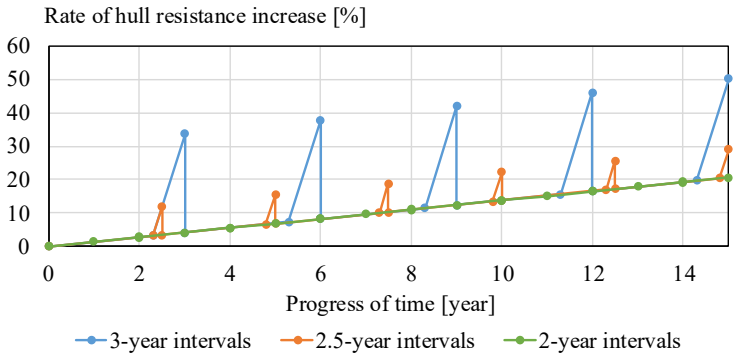


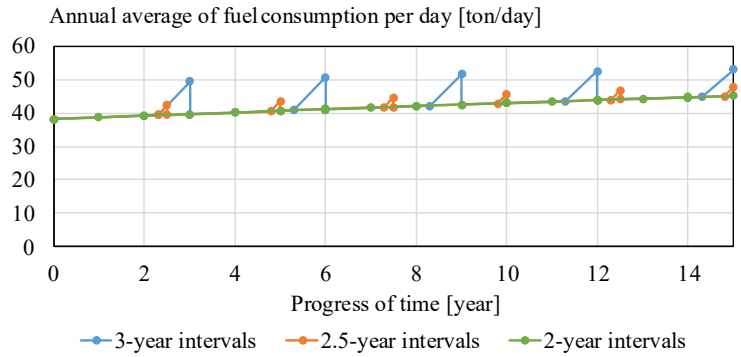
Fig.8: Setting of rate in resistance increase for the life cycle fuel consumption

3.3. Results of the Life Cycle Fuel Consumption

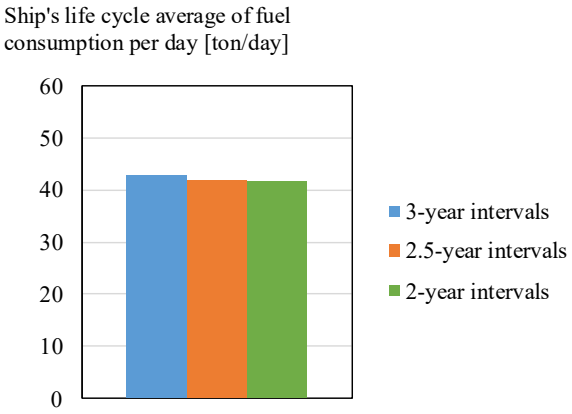
The life cycle fuel consumption was calculated based on the input conditions described in the preceding section. This assessment includes the effects of weather conditions (winds and waves) and the impacts of fouling and aging as factors representing actual seas.

The results are shown in Fig.9. From this, it is calculated that fuel consumption differs by approximately 2.5% (equivalent to 5,800 tons of fuel consumption over the lifecycle) between maintenance intervals of 3.0 years and 2.0 years.

By utilizing such quantitative evaluation results, it is possible to determine the appropriate interval for repainting during dry docking. This enables practical reductions in fuel consumption and GHG emissions while considering economic factors.



(a) Annual average of fuel consumption per day



(b) Ship's life cycle average of fuel consumption per day

Fig.9: Results of life cycle fuel consumption

4. Conclusions

This paper outlines the new initiative ISO 25817 concerning the evaluation of fuel consumption and propulsion performance in actual seas, which began last year, and presents concrete examples of evaluations based on lifecycle fuel consumption as specified therein. Using this methodology enables performance evaluation in actual seas covering a ship's lifecycle, considering not only ship design but also operational profile. Since the evaluation is based on quantitative fuel consumption calculations, the derived values can contribute to realistic fuel consumption reduction and GHG reduction initiatives.

References

ISO (2016), *ISO 19030: Ship and marine technology -Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

ISO (2025), Resolutions ISO TC 8/SC 2 2025 Plenary, Int. Standard Org., Geneva

KURODA, M.; SOGIMOTO, Y. (2021), *Development of the Evaluation Method for Life Cycle Ship Performance*, J. Japan Society of Naval Architects and Ocean Engineering 34, pp.19-27

KURODA, M.; SOGIMOTO, Y. (2022), *Evaluation of ship performance in terms of shipping route and weather condition*, Ocean Engineering 254

SOGIHARA, N. (2021), *Evaluation of Ship Performance in Actual Seas Using Onboard Monitoring Data -Application of Resistance Criteria Method to Designated Weather Condition-*, Proc. Japan Society of Naval Architects and Ocean Engineering 32

SAKURADA, A.; KURODA, M.; TSUJIMOTO, M.; SANO, H.; ASHIDA, T. (2026), Analysis *of Fouling and Aging Effect Using Monitoring Data Obtained from Pure Car Carrier*, 45th Int. Conf. Ocean, Offshore and Arctic Engineering (OMAE)

SAKURADA, A.; SOGIHARA, N.; TSUJIMOTO, M.; SATO, H. (2021), *Development of resistance criteria method for ship performance evaluation by onboard monitoring data*, Conf. Full Scale Ship Performance, pp.19-26

Development of Hull Fouling over the Docking Period – A Comparison of Two Fleets

Falko Fritz, Albis Marine Performance, Hamburg/Germany, falko.fritz@albis-mp.com
Lukas Schmidt, Albis Marine Performance, Hamburg/Germany, lukas.schmidt@albis-mp.com

Abstract

This paper discusses the results of hull condition monitoring over the full docking period of multiple vessels in two fleets. By comparing the trends on a broader scale, the effectiveness of different approaches in hull condition management may be evaluated. Relevant influences in the data generation and drivers in the hull cleaning strategy are highlighted, before discussing improvements that were already implemented during the evaluated period or have been developed to increase the confidence in the results in the future.

1. Introduction

Hull condition monitoring is a process that affects a vessel and its performance over its entire lifetime. New methods to schedule cleanings, optimize the preservation of the hull coating, more frequent inspections with underwater drones, etc., don't show their benefits immediately. It takes years to identify their full potential. As the supplier of the vessel performance manager V-PER and its hull condition evaluations, Albis has the privilege of having access to the performance data of multiple customer fleets, with up to 13 years of continuous high-frequency sensor data from the same vessels. Monitoring numerous vessels operated by different clients over the complete docking intervals allows for a bird's eye view on the development of hull conditions and what factors influence the outcome.

Two customer fleets with 20-30 vessels each, called A and B for confidentiality reasons, were chosen for the discussion in this paper. In both cases, the vessels were monitored over a full docking period of five years or more. Both fleets operate globally and include vessels between 11 000 and 55 000 DWT.

2. Development of Hull Fouling over the Docking Period

Each vessel is equipped with fuel meters that were continuously maintained to record accurate M/E fuel consumption data. In addition to the fuel consumption, NMEA data including GPS position, ship speed over ground (SOG), ship speed through water (STW), heading, wind speed and direction, etc. were logged in high frequency.

Evaluations to filter and normalize the vessels' M/E fuel consumption to a defined reference condition and track this result over time, very similar to the methods defined in ISO 19030:2016 Part 1, *ISO (2016)*, were carried out on a monthly basis since January 2015. All results were checked by experienced members of the Albis reporting team.

Baseline consumptions were agreed in coordination with the clients. These are typically based on the recorded data after a drydock and don't take any results from the original shipyard trials into account, so that the baselines reflect the actual capabilities of the vessels in real day-to-day operation. Since 2022, AI models have been added to the empirical vessel model to increase the rate of evaluable datasets and improve the certainty of results. The general approach of normalizing and identifying fuel consumption trends has been consistent throughout the entire time, though.

Fig.1 and Fig.2 show box plots of all monthly M/E overconsumption results of the two fleets in percent, ordered by the time since leaving drydock in months. Comparing the two plots, fleet B has higher overconsumptions and a larger data scatter than fleet A, even though both fleets were equipped with the same monitoring system and the clients received similar information for their decision-making. The reasons for the differences are discussed in chapter 4.

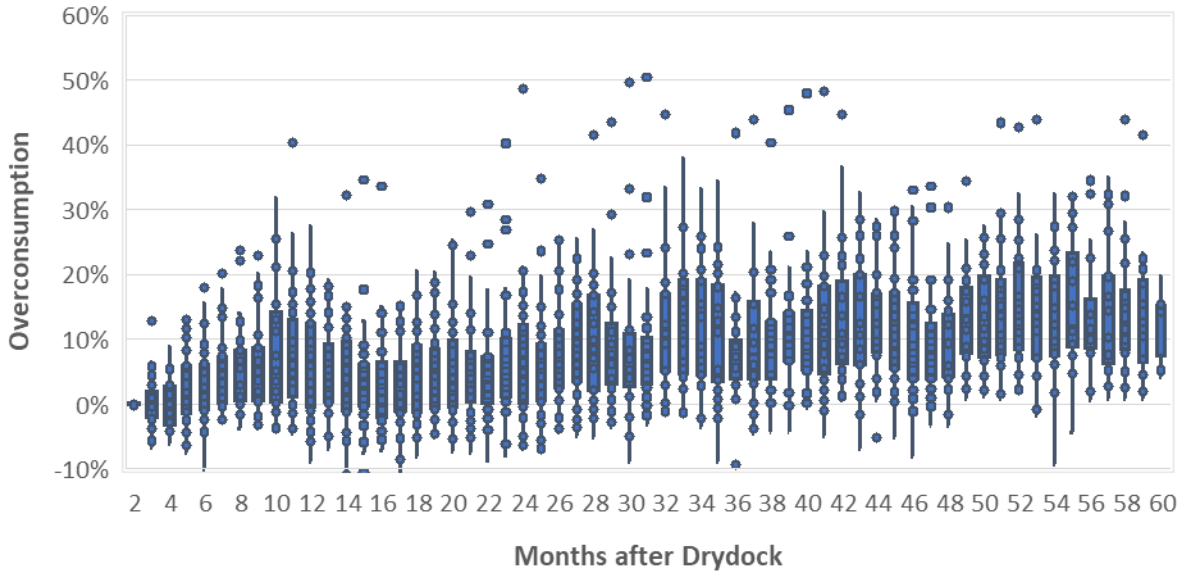


Fig.1: Evaluated overconsumption due to hull fouling over 60 months, fleet A

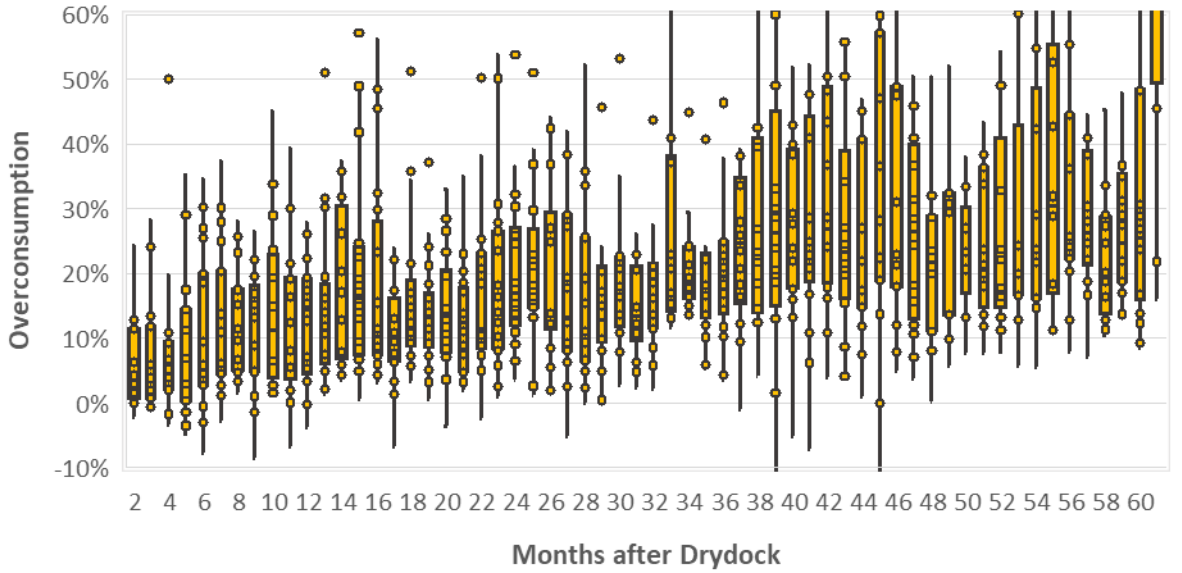


Fig.2: Evaluated overconsumption due to hull fouling over 60 months, fleet B

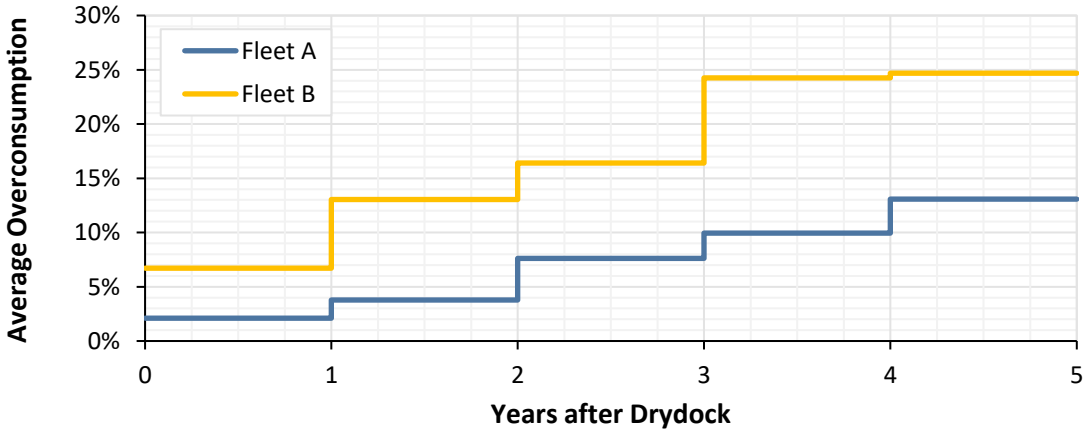


Fig.3: Average overconsumption of fleets A and B

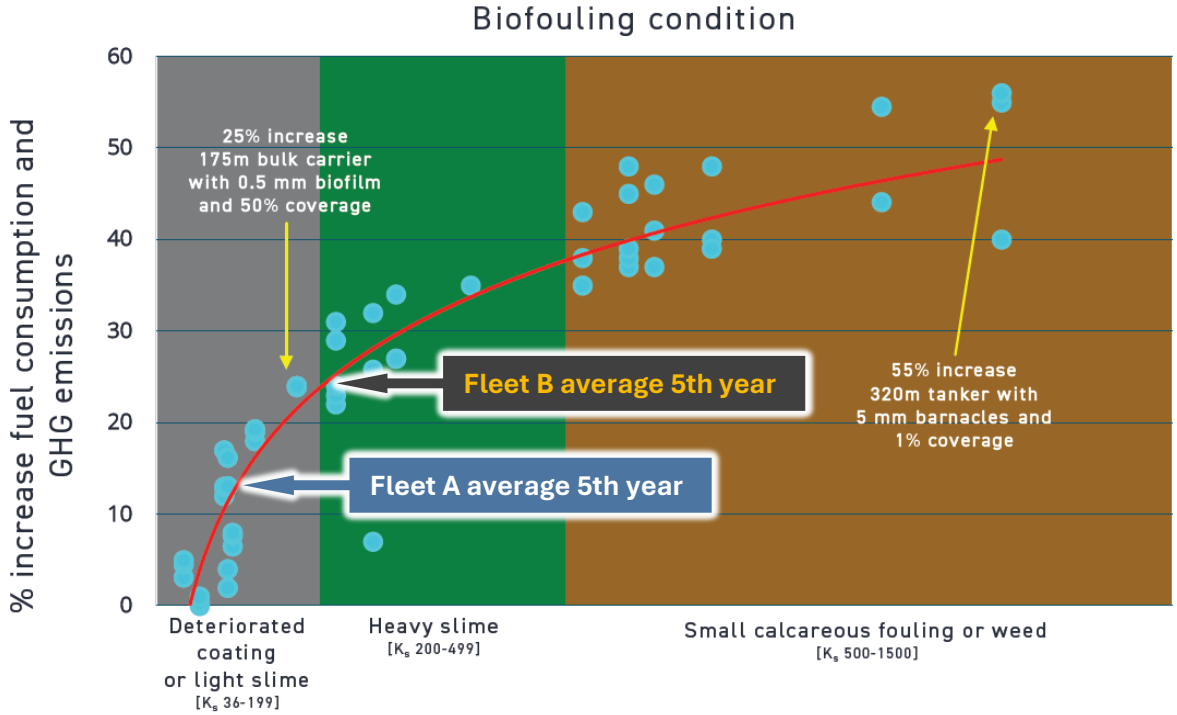


Fig.4: Fleet average overconsumption in year 5, compared to IMO GloFouling study results

Looking at the average fleet overconsumption per year after drydock, fleet A only reaches an average of 13% in the final year before a new coating is applied, while fleet B hits 24-25% already in year 4, as summarized in Fig.3. However, with regards to the overconsumption figures published in the IMO GloFouling study, *IMO (2022),* shown in Fig.4, both 13% and 25% are still at the lower end of the curve of consumption increase due to biofouling. Managing the vessels in such a way that they don't exceed the limit of deteriorated coating before entering drydock, should be seen as a very good result in general.

3. Financial Impacts

The added fuel costs caused by biofouling are substantial. Assuming that an exemplary fleet

- consists of 25 vessels,
- each of them consumes 25 tons of fuel oil per day at sea for propulsion,
- the ships spend 250 days at sea per year, and
- that each ton of fuel costs USD 600,

each percentage point of overconsumption across the fleet costs USD 0.4 million per year.

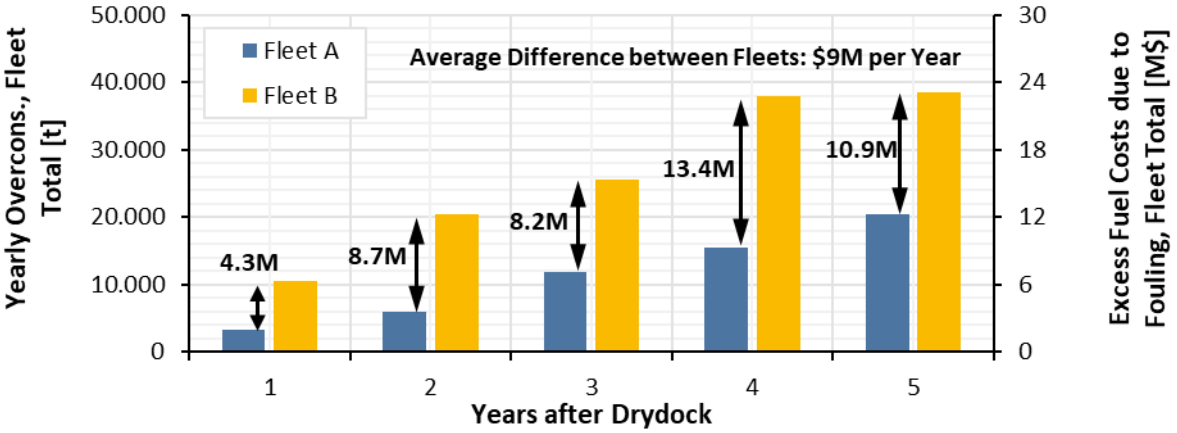


Fig.5: Financial impact of overconsumption, exemplary results for a fleet of 25 ships, consuming 25 tons per day for 250 days per year at a fuel price of USD 600 per ton

Fig.5 illustrates the financial impact of the overconsumptions calculated for fleets A and B, applied to this exemplary fleet of 25 ships. The outcome is a lower total fuel cost of USD 9 million per year on average, if the hull cleaning strategies of fleet A are followed, rather than those of fleet B.

4. Causes of Increased Overconsumption and Data Scatter

Since both fleets A and B are equipped with the same monitoring system, the reasons why fleet B has a higher overconsumption and significantly higher added fuel costs due to hull fouling must be found elsewhere. Judging by the available information and extensive communication with both clients, the crucial difference that could be identified is the market situation in which the fleets operate. Fleet A is active on trade routes where fuel costs seemingly have a significantly higher impact on the competitiveness, while the freight owners chartering the transport work with fleet B accept added costs more readily. As a consequence, the owners of fleet A have a stronger focus on keeping the hulls in an optimal condition than the operators of fleet B.

However, while the market situation may indicate why the vessels of fleet B get fewer hull cleanings on average, it would not explain why the monthly overconsumption results illustrated in Fig.2 show a significantly higher data scatter compared to fleet A, Fig.1. According to the evaluation, fleet B also had a higher rate of vessels where the ship speed log was defective or inconsistent, so that numerous calculations were based on SOG. The importance of a functioning ship speed log for hull condition monitoring was already discussed in *Fritz (2023)*.

More data scatter leads to lower confidence in the results, which is then easier to dismiss in the decision-making process, possibly leading to fewer cleanings and less attention paid to hull condition in general. This increases the risk that the ship speed log won't be well maintained either. This is further influenced by the market situation, so there is also an indirect connection between the trade routes and the data scatter.

5. Employing Metocean Data to Model STW

Since ship speed logs continue to be somewhat unreliable in practice, more research was conducted to evaluate alternatives. A commonly used method to infer STW is by using SOG and correcting for currents by using data from a global ocean model. While this approach is simple and straight-forward to implement due to the availability of many global ocean current data products, it is in most cases much less exact than using a direct measurement of STW. Ocean near-surface currents are the result of a variety of driving forces, some of which are physically complex. They all add up with different degrees of influence to create a three-dimensional field of currents. The ocean surface is far more turbulent and variable than models can capture. Wind and waves create complex circulation patterns and turbulent mixing in the upper ocean that change current patterns over distances of hundreds of meters and time scales of hours. Recent research shows these small-scale processes can account for up to one-third of the actual current variations experienced by ships, *Dong (2024)*, but they occur at scales too small for operational ocean models to resolve or predict accurately.

Among the main drivers of near-surface currents are wind stress, wave induced stokes drift, density differences caused by temperature and salinity, tidal forces and coastal geometry and bathymetry. Wind forcing at the surface is incorporated quite well by most global ocean current products, due to the use of bias-corrected ECMWF reanalysis wind fields combined with satellite scatterometry observations, *EU (2024)*.

However, variation of currents by wind and waves is not constrained to the surface in two dimensions but extends downwards with significant gradients in strength and direction, even within the uppermost few meters in which ship hulls move. These gradients depend on differing factors, like wave induced turbulence or temperature stratification, which are themselves subject to large variations on short timescales, and can be quite hard to model according to *Röhrs (2023)*.

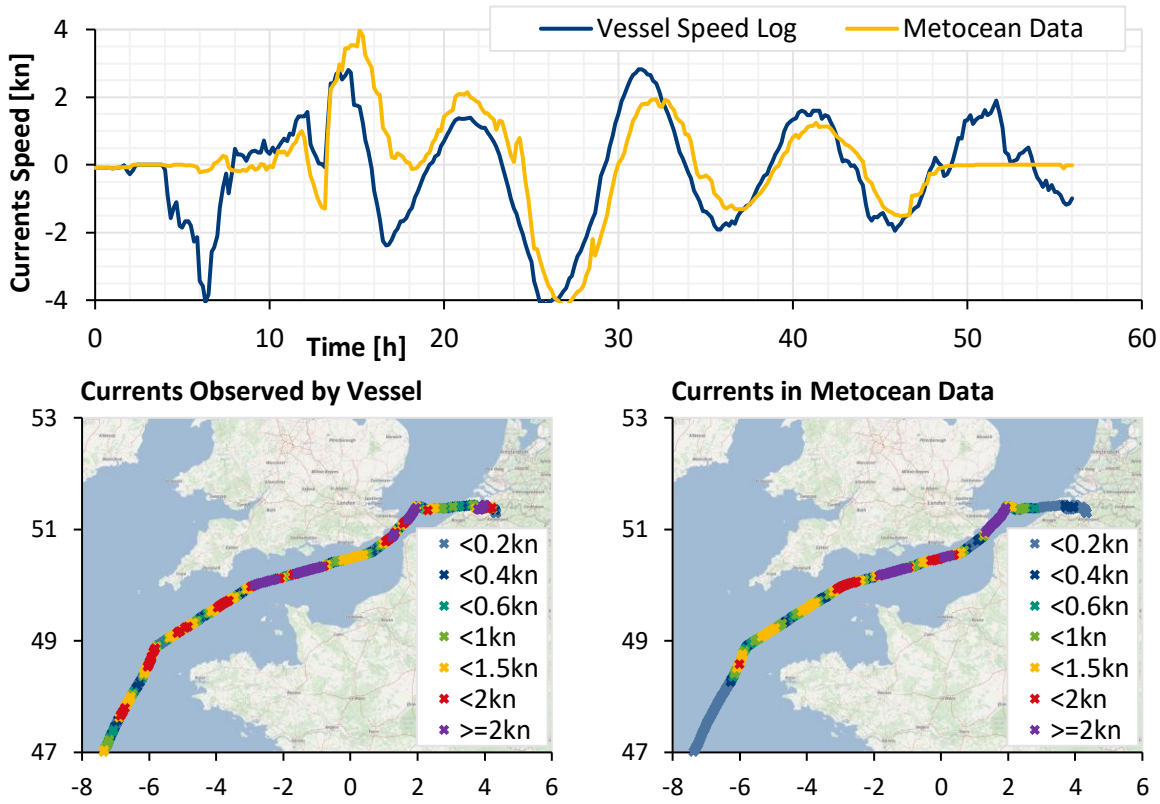


Fig.6: Currents in the English Channel, indicated by metocean data

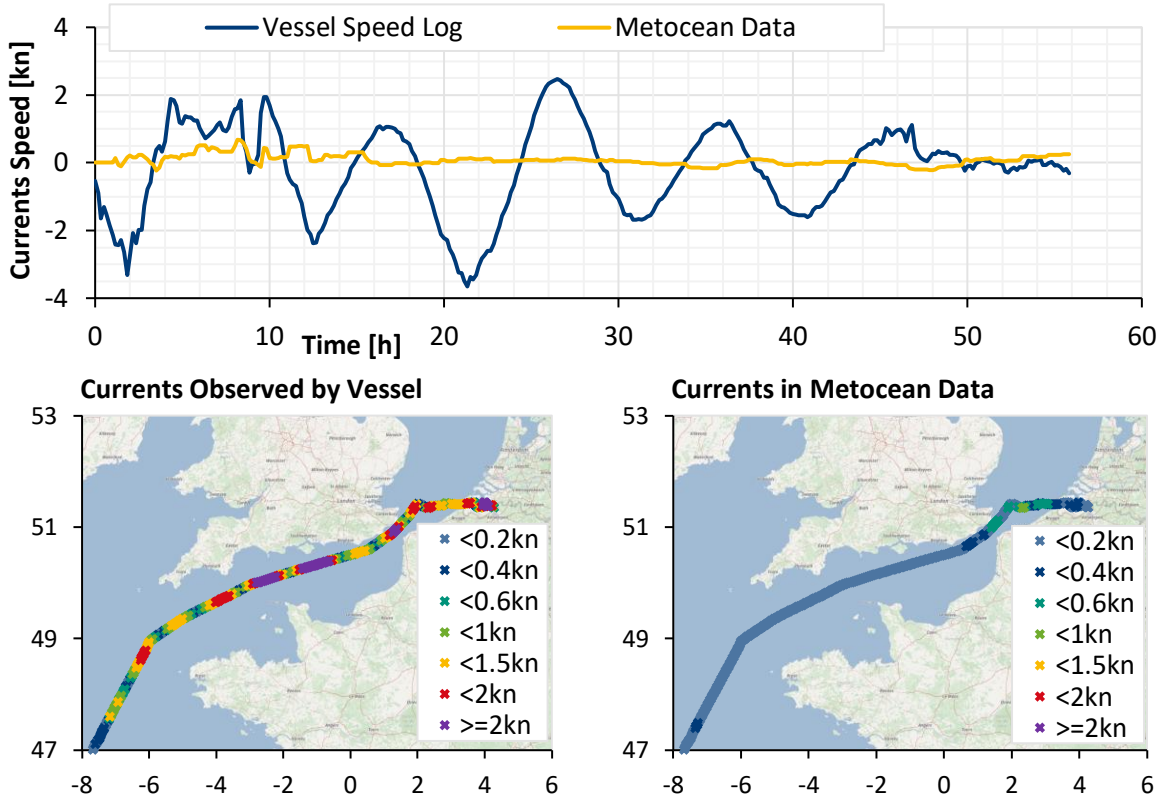


Fig.7: Currents in the English Channel, this time not indicated by metocean data

Tidal currents on the other hand are not included in many common data products. They are a key component in many sea areas, especially in the vicinity of coasts, and can explain a large part of

deviations between models and measurements, as discussed in *Aijaz (2023)*. Other typical weaknesses of current models on near the surface on scales relevant for ships include wave induced stokes drift and Langmuir circulation, capturing complex near-surface currents from thermohaline turbulence patterns and limited model resolutions in the horizontal and vertical. Speed logs on the other hand directly measure what matters: the actual flow of water past the hull.

The issue of tidal currents was checked with regards to the metocean data used in the V-PER system, with mixed results. E.g., in some cases where a vessel navigated the English Channel, the currents were identified very accurately, as shown in Fig.6. In other instances, the data received from the weather and metocean data provider did not resolve the tidal currents at all, Fig.7.

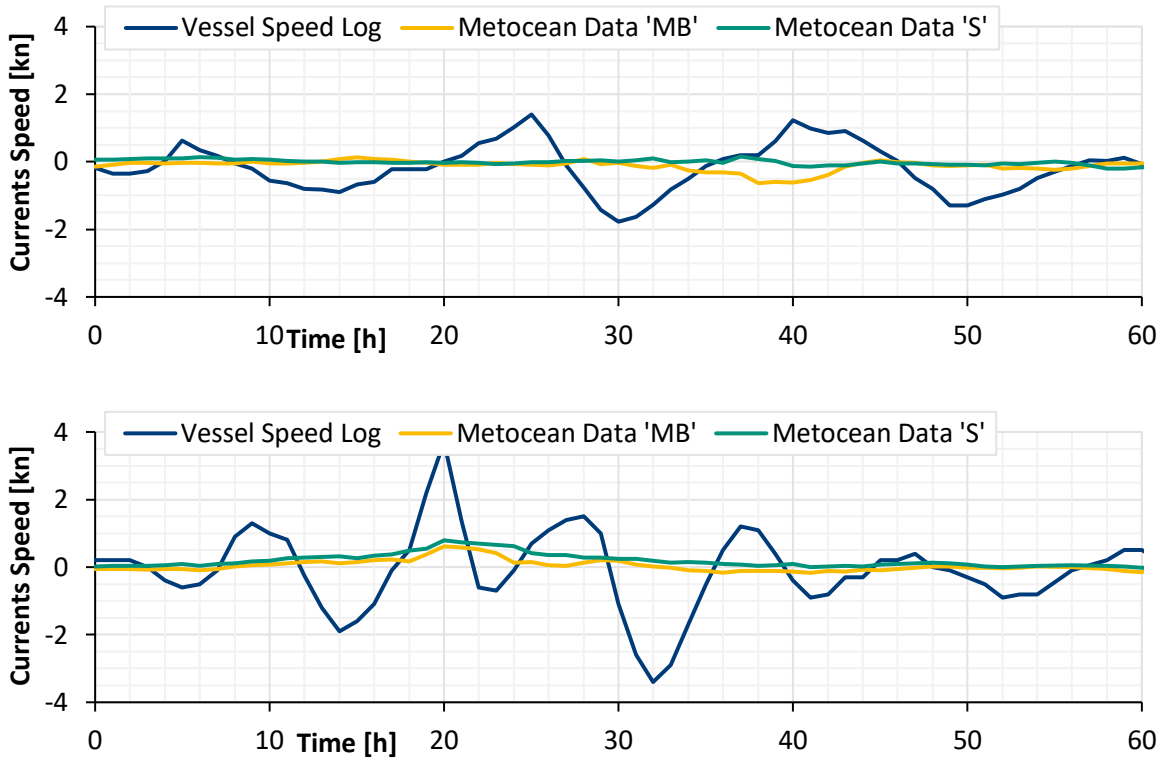


Fig.8: Similar currents in the Real Vessel Data Challenge dataset, not indicated by metocean data

It was then examined if this inconsistency is specific for the data provider used in V-PER or also present in competing meteorological and oceanographic models. Thankfully, two separate surface current data series were made available in the Real Vessel Data Challenge datasets, *Tsarsitalidis (2025)*, calculated by the undisclosed provider "MB" and "S". Judging by the data structure and statistical dependencies of wind speed and wave height, these are not the same provider as used in V-PER. SOG and STW patterns similar to the ones recorded in the English Channel could be found in the Real Vessel Data Challenge dataset, presumably also caused by tidal currents. Neither the metocean data from provider "MB" nor "S" modelled the surface currents as they were observed by the vessel, Fig.8. Taking both the literature and the examined data into account, there is good reason to still use the STW measured on board as the main parameter for ship speed, as defined in *ISO (2016)*.

6. Established Improvements

Since the beginning of the monthly hull condition evaluations in January 2015, two main changes have been implemented to improve the modelling accuracy and inspire more confidence in the results:

- Metocean data were added to the available data scope. While these are not sufficient to fully substitute a ship speed log as discussed in chapter 4, they improve the reliability of STW

correction functions. When the ship speed log is implausible or not available, it is still preferred to use a simulated STW than SOG alone.

- With AI and machine learning, the amount of valid data in the hull condition evaluation could be more than doubled. The empirical vessel model was limited to good weather conditions to calculate normalized fuel consumption rates. With AI, a much wider range of weather conditions can be included in the assessment. Improved outlier detection methods are an added bonus that helps to reduce scatter.

The results in chapter 2 must therefore be seen as the outcome of a monitoring system that wasn't yet developed to the full capabilities it has today. Major advances were only realized in recent years and the confidence in the monthly results is much higher now than it was 5-10 years ago.

7. Conclusion

A high-quality vessel performance monitoring system proves essential for understanding and managing the long-term development of hull fouling. The comparison of two customer fleets over full docking cycles highlights how operational practices, market conditions, and data quality jointly influence both fuel efficiency and the confidence in performance evaluations. While fleet A's more proactive hull cleaning strategy resulted in significantly lower overconsumption and avoided excess fuel costs as much as possible, the results of fleet B underscore the importance of reliable sensors. Advances in recent years, like the integration of metocean data and AI-enhanced modelling, have already improved the robustness of the monthly results. Continued refinement of on-board and external data, raising the awareness of the crucial role of the ship speed log and novel modelling techniques will further strengthen decision-making in the years ahead and help to meet the constantly increasing demand for ships with better fuel economy.

References

AIJAZ, S.; BRASSINGTON, G.B.; DIVAKARAN, P.; REGNIER, C.; DREVILLON, M.; MAKSYM-CZUK, J.; PETERSON, K.A. (2023), *Verification and intercomparison of global ocean Eulerian near-surface currents*, Ocean Modelling 186, https://doi.org/10.1016/j.ocemod.2023.102241

DONG, J.; FOX-KEMPER, B.; WENEGRAT, J.O. (2024), *Submesoscales are a significant turbulence source in global ocean surface boundary layer*, Nat Commun 15, 9566, https://doi.org/10.1038/s41467-024-53959-y

EU (2024), *Wind and Stress from Scatterometer and Model*, Copernicus Marine Service, https://doi.org/10.48670/moi-00181

FRITZ, F. (2023), *Data Recording on Board – Are We Getting it Wrong from the Start?*, 8th HullPIC Conf., Pontignano, pp.62-69, http://data.hullpic.info/HullPIC2023_Pontignano.pdf

ISO 19030-1:2016 (2016), *Ships and marine technology — Measurement of changes in hull and propeller performance — Part 1: General principles,* Int. Standard org., Geneva

IMO (2022), *Analysing the Impact of Marine Biofouling on the Energy Efficiency of Ships and the GHG Abatement Potential of Biofouling Management Measures*, GloFouling, Int. Mar. Org., London, https://www.glofouling.imo.org/_files/ugd/34a7be_afd9d183df9a4526bd088007436c1079.pdf

RÖHRS, J.; SUTHERLAND, G.; JEANS, G.; BEDINGTON, M.; SPERREVIK, A.K.; DAGESTAD, K.F.; LACASCE, J.H. (2023), *Surface currents in operational oceanography: Key applications, mechanisms, and methods*, J. Operational Oceanography 16(1), pp.60-88, https://doi.org/10.1080/1755876X.2021.1903221

TSARSITALIDIS, V.; PONKRATOV, D.; TSOULAKOS, N. (2025), *A First-of-a-Kind Workshop on Modeling, Insights and Actionable Outcome*, 10th HullPIC Conf., Mülheim, pp.117-120, http://data.hullpic.info/HullPIC2025_Mulheim.pdf

Enabling Improved Performance Analysis through Distributed Ocean Sensing

Pieter Smit, Sofar Ocean, San Francisco/USA, pieter.smit@sofarocean.com

Abstract

Accurate vessel performance assessment requires high-quality vessel and environmental data. Vessel performance analysis has long been constrained by limited environmental data, typically derived from sparse daily reports. As the maritime industry adopts data-driven operations, high-frequency vessel measurements have improved, but environmental observations lag behind. We describe efforts to close this gap using Sofar's intelligent ocean network of over 500 sensors. By combining these ocean observations with advanced vessel monitoring, we enable more accurate, data-driven assessments of vessel performance.

1. Introduction

Reliable vessel performance assessment depends on accurate knowledge of prevailing environmental conditions, particularly waves, winds, and surface currents at the vessel location. In practice, however, these conditions are often poorly constrained. Onboard weather instrumentation may be unavailable or sub-optimally installed (e.g., anemometers positioned in the aerodynamic shadow of the superstructure). Visual sea-state observations are inherently imprecise and cannot reliably distinguish between wind sea and swell in complex mixed conditions. Operational reporting frequently relies on noon reports, which provide a single daily snapshot and are often biased toward the worst encountered conditions rather than representative averages.

As a result, numerical weather prediction models are typically used to estimate metocean conditions operationally. Although forecast skill has improved substantially over recent decades (Ritchie, 2024), model accuracy remains fundamentally constrained by observational input through data assimilation (DA). Indeed, DA is widely recognized as contributing as much to operational forecast skill as model physics itself, *Kalnay (2002), Buizza et al. (2005).* A persistent limitation is the scarcity of open-ocean observations, in particular spectral wave measurements, in the marine boundary layer. Coastal buoy networks provide localized coverage and are insufficient to constrain global wave models. Satellite altimetry offers broader reach but typically only samples bulk wave parameters, suffers from saturation in extreme sea states, and is temporally sparse.

To address this gap, Sofar has developed a globally distributed network of spectral wave buoys and an operational data assimilation framework designed specifically to improve open-ocean wave state estimation. Drifting buoys are among the most impactful observation types in operational forecasting systems when evaluated per observation, *NASA (2026)*, making distributed networks combined with data-assimilation a highly efficient way to improve ocean state estimation. The resulting observation-informed analyses feed directly into physics-informed vessel performance models and operational decision-support tools.

This paper focuses on three elements. First, we describe the architecture and global deployment of the distributed spectral buoy network and its operational integration. Second, we present the wave data assimilation framework used to incorporate buoy and satellite observations into a global wave model, and quantify the resulting improvements in wave-state accuracy. Third, we evaluate how reductions in environmental error translate into improved vessel performance estimation through a simplified but representative performance modeling exercise. By connecting improvements in environmental estimation directly to downstream performance analytics, we aim to demonstrate that closing the open-ocean observation gap is a needed step toward robust, data-driven maritime optimization.

2. Sofar's Global Observation Network

Starting in December 2018, Sofar developed a dense, globally distributed network of drifting ocean sensors measuring waves, wind, drift, barometric pressure, and sea surface temperature, *Houghton et al. (2021), Dorsey et al. (2023)*. The observational platform, the Sofar Spotter, is a compact directional wave buoy that samples three-dimensional motion at 2.5 Hz and performs onboard processing to estimate directional wave spectra, *Raghukumar et al. (2019)*. The Spotter provides hourly updates of the frequency variance-density spectrum and the four lowest directional moments. This spectral capability is a key differentiator, enabling advanced assimilation and resistance modeling beyond bulk wave height corrections.

The system consists of a solar-powered surface float connected via the Iridium satellite network to a cloud-based backend. The backend provides API endpoints for seamless integration into operational models and real-time applications. Approximately 500 buoys are deployed globally, Fig.1, providing long-dwell, real-time observations across major shipping lanes and storm-development regions.

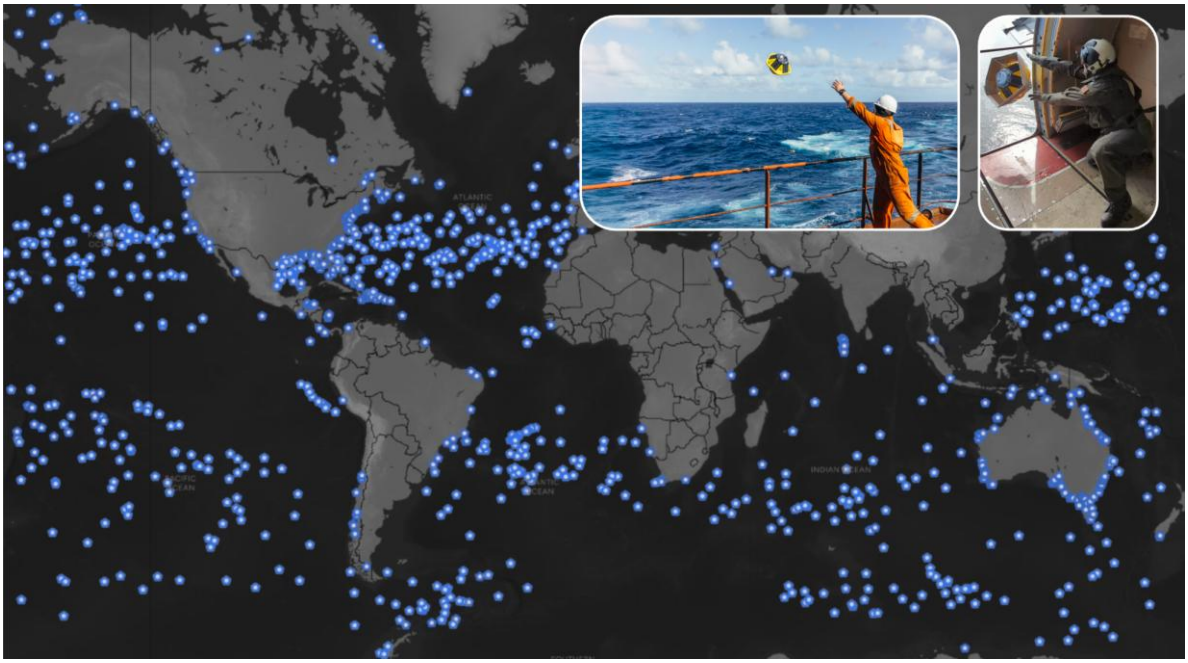


Fig.1: Sofar network of ~500 drifting buoys deployed globally. Every symbol indicates an individual satellite-connected buoy measuring wave spectra, directional moments, sea surface temperature, drift, barometric pressure and inferred wind.

In addition to vessel-based deployments, innovative aerial deployment strategies have been demonstrated ahead of tropical cyclones, *Dorsey et al. (2023)*, enabling targeted observations where conditions are most dynamically relevant. Where required, sensors can be co-deployed with vessels during speed trials to ensure spatial co-location and maximize data reliability.

The primary objective of the network is to provide accurate, real-time environmental intelligence for maritime operations. Its global distribution and spectral fidelity enable not only improved forecasting but also materially improved performance analysis.

3. Sofar's Wave Data Assimilation System

The core differentiator of Sofar's forecasting system is our focus on accurate wave predictions. And in turn, what differentiates our forecasting system is our ability to use wave observations to improve both our model, and estimates of the initial state. The basic idea is as follows. To estimate wave conditions at a certain time, estimates of wave conditions at a previous instance (say 6 hours ago) and estimates of

forcing winds over that period. We can then use a wave model to model wave evolution over the period to arrive at wave conditions at the desired point in time. Evidently, for this to work we need good estimates of previous wave conditions (i.e. the initial conditions), and a good model as both contribute to the accuracy of the model prediction.

Observations can contribute to improved wave estimates in two ways:

- Observations can be used to calibrate model parameters. Wave models contain many parametrizations that need to be calibrated to optimize model accuracy.
- Observations can be used to improve the initial conditions. Improving initial conditions will improve the accuracy of model predictions at the next instance in time.

Because wave models evolve governing equations dynamically, assimilation updates propagate in space and time. For example, correcting swell near the Aleutian Islands can improve downstream swell predictions near Hawaii, *Smit et al. (2019)*. This propagation is particularly valuable for shipping routes far from direct observation locations.

4.1. Model setup and calibration

Sofar produces a best-in-class wave forecast based on a cloud based implementation of NOAA's WAVEWATCH3 model, *Tolman et al. (2019)*, that is implemented over the global ocean at 0.25° horizontal resolution and forced by near-surface winds from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System, *Houghton et al. (2022)*. Model physics (non-linear interactions, wave generation, dissipation etc.) use ST4 source terms, *Ardhuin et al. (2010)*. The system is calibrated primarily by tuning the strength of the air-sea interaction through the Miles parameters (often referred to as `β-max' in the literature). Calibration of model parameters is performed through scalable algorithms for Bayesian optimization, *Daulton et al. (2022)*. Importantly, optimization targets performance-relevant conditions rather than global mean statistics, aligning model tuning with maritime use cases.

4.2. Data Assimilation - why this matters for vessel performance estimates

The data-assimilation strategy is built upon a previously established optimal interpolation framework described by *Smit et al. (2021)*. The basic idea is as follows. Sparse observations of a wave-property (say wave height) are compared with model estimates to obtain "departures" (differences between model and data at those grid-points). These departures are estimates of local model error. Through a parametrization of spatial error-covariances we may correlate errors at the observed site to nearby points. Say, a buoy observes 1 m higher waves than predicted in a large extra-tropical storm. While we would not anticipate the model error to be exactly 1 m at a vessel located 20 nm away from the buoy, we do anticipate that the error is similar, and may choose to increase the forecast by a fraction of 1 m depending on the anticipated correlation. This procedure may be formalized mathematically (for details, see *Smit (2021)* and references therein).

Such direct updates applied to wave heights alone are the de-facto standard for wave data assimilation of satellite observations of wave heights by operational centers. However, while these methods may scale the total energy in the waves (through wave height), they cannot correct for errors in distribution of energy in the energy spectrum between, say, sea and swell. While this method was found to produce improvements in both model nowcasts and forecasts, substantially more information is available from the Spotter buoys beyond significant wave height, specifically the variance density spectrum and the four Fourier coefficients. To fully utilize these observational data to update the initial state of the operational forecast model, we augmented the optimal interpolation framework to update the full two-dimensional model spectra. For details, see *Houghton et al. (2023)*.

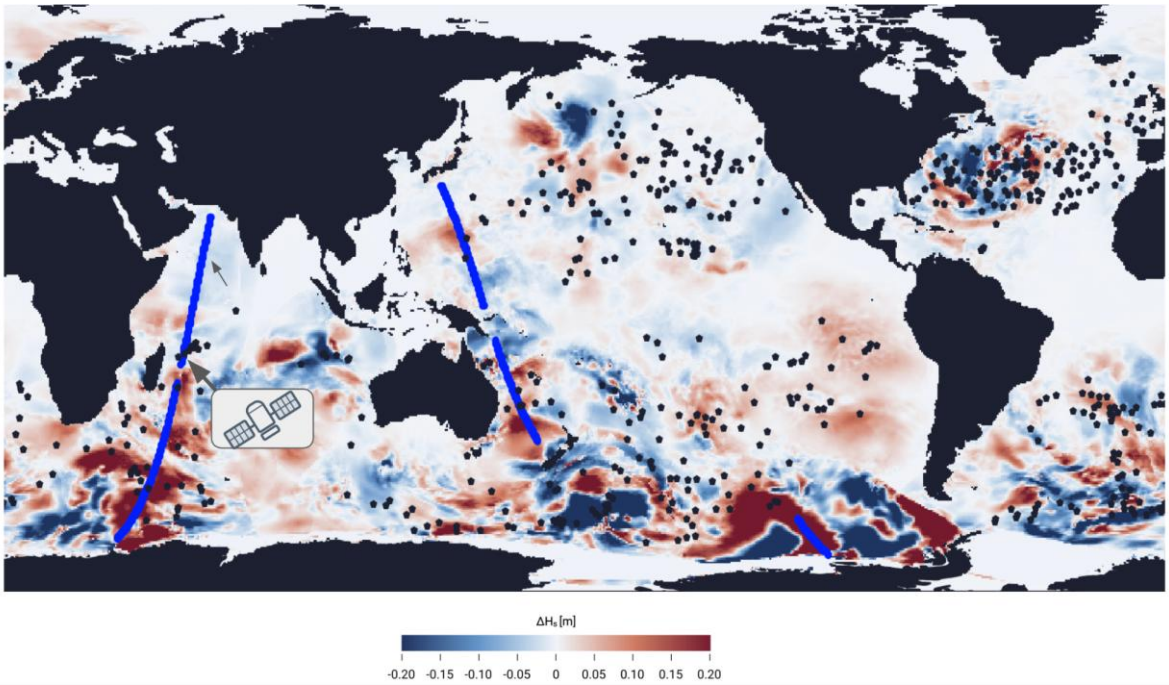


Fig.2: Snapshot of difference between estimates of wave height with and without data assimilation of Spotters and satellite data. Combining drifters and satellites ensures that through data assimilation we can improve estimates over large sections of the open ocean. Note that differences occur throughout the world's oceans.

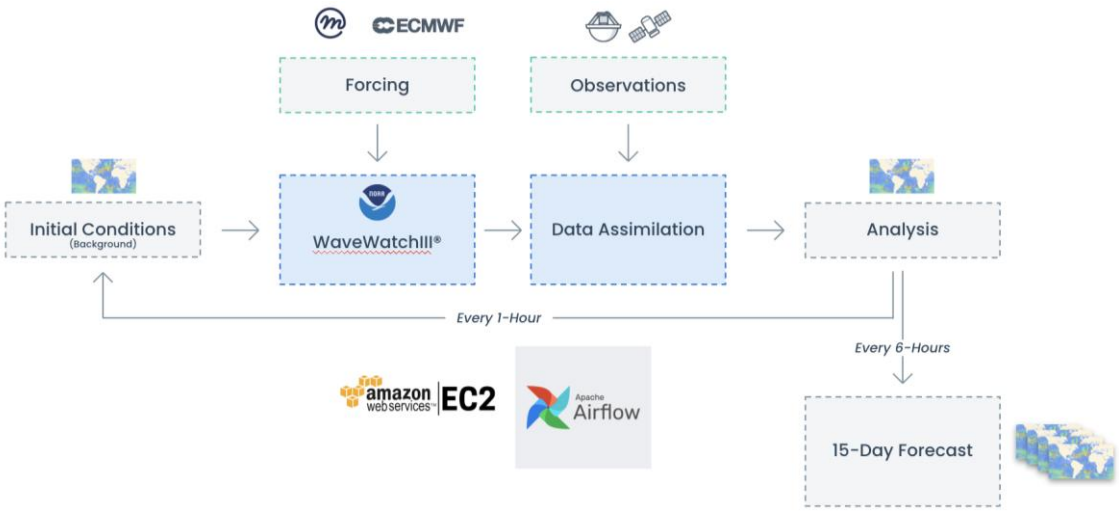


Fig.3: Sofar's operational data assimilation and forecast infrastructure. Every six hours we perform 6 hourly data assimilation cycles. After the 6th cycle a 15 day forecast is performed.

This spectral data assimilation is unique to our integrated observation-forecast system, and is only possible because of spectral data that is delivered and reported by the buoys.

Why does this matter? The primary advantage of this system is that we are able to provide best estimates not only of total energy - but also of how energy is distributed in the wave field. As a consequence, estimates of wave direction, sea/swell distribution, and wave periods are greatly improved. In turn, this admits the use of more advanced wave resistance models. For example, wave resistance in the SNNM method, *Liu and Papanikolaou (2020)*, depends on spectral information.

While the system primarily relies on data from the buoy network, we also make use of all available data from satellite observations, including data from Jason-3; Sentinel 6A; Saral and SWOT. At any given

time, the combined system will correct estimates of wave conditions in a large fraction of the world's ocean, Fig.2. It is this property of data assimilation, i.e. it effectively uses the model to extrapolate the observation in space and time, that makes it so impactful. As a consequence, observation-informed estimates of wave conditions will differ from those based on models alone, and in turn may impact estimates of vessel performance.

4.3. Operational Implementation

Sofar's operational models run on a cloud based infrastructure (AWS, Airflow scheduler). Every six hours the system uses the last 6 hours of observations to produce hourly analysis (or best estimate) wave fields, Fig.3. These wave fields are stored and subsequently used to evaluate and train Sofar's Vessel Performance models. After this, a 15-day operational forecast is initialized from the latest available analysis. These forecasts are subsequently used to help provide context to captains, and to generate daily optimized guidance for vessel route geometry and speed. The operational model can produce full spectral output (frequency/direction) at any desired location, and a large historical forecast archive is available for training purposes. Data from the system is available through API access.

4.4. Data assimilation impact

Forecast and model accuracy are monitored daily, and we have extensively evaluated the improvement the system provides as we developed and improved the system, and we refer to published work for overall details, *Smit et al. (2021), Houghton et al. (2022,2023)*. Here we will specifically consider the difference between a year-long model run (including daily 5-day forecasts) with (labelled WW3_DA) and without data assimilation (labelled WW3_noDA). This specifically highlights the impact of the observations.

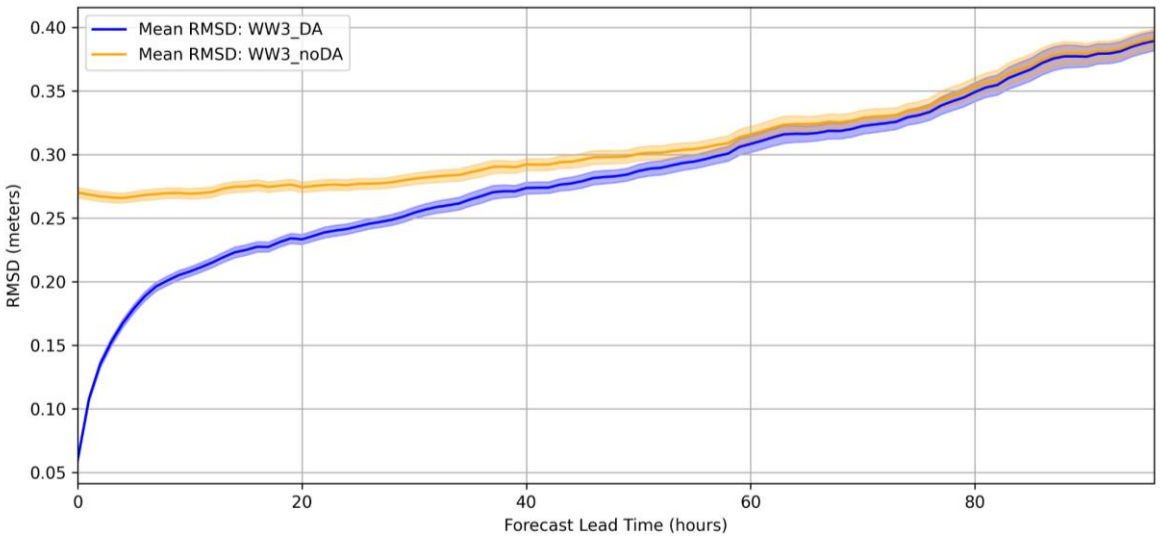


Fig.4: Waveheight root-mean-square difference as function of forecast lead time. Comparison between a model with DA and the baseline (no-DA) model. Difference is estimated with respect to observations from the global Spotter grid. Bands indicate uncertainty in error estimates.

We observe the following. Compared with observations the data-informed estimates perform substantially better over the first 24 hours. While at short lead times (analysis time) we are effectively just looking directly at the buoy observations (which naturally have low difference when evaluated with themselves) it is important to remember that these differences are spatially projected - and we thus anticipate similar error reductions near the buoys, and once the model evolves in time, down stream of the buoys. As the forecast increases, the model relaxes back towards errors in the forcing winds, which is particularly visible in the wave heights, whereas (through swell) the periods retain improvements. That said - for application to vessel vessel performance, the improvements in short lag-time are primarily of interest, and here we see significant reductions in errors (>50% reductions). Further, we stress that

while a reduction from 25 cm to 5 cm error may not seem significant (i.e. a 25cm error may seem acceptable) we stress these are average global errors. The mean error is dominated by good weather, where errors are small because waves are small. But in energetic systems corrections of 0.5m or more to wave height are not uncommon in the system.

5. Discussion

5.1. Impact on performance

Assuming errors of 0.5 m in wave height are not uncommon, the question that remains is: does it matter? To make this concrete we will use a simplified version of Sofar's VPM to estimate the impact of such errors. Specifically, consider fuel consumption estimates of a Capesize bulk carrier. We will estimate fuel consumption assuming that at any speed propulsive and resistive forces balance; and estimate forces based on literature expressions for propulsion, *Oosterveld (1975)*, residual resistance, *Holtrop and Mennen (1978)*, skin friction, *ITTC (2021)*, and added wave resistance, *Liu and Papanikolaou (2020)*. Solving the force balance for a given speed (here 11 kn) will determine propellor rpm, which in turn informs propellor torque, engine power and, through an engine efficiency model (here based on a simple ~50% thermodynamic efficiency and a standard estimate for energy density of fuel oil), fuel consumption. We will assume waves from the bow, parametrize the spectrum based on *Pierson and Moskowitz (1964)*, with 30° directional distribution and consider different wave heights with periods determined by deep water equilibrium estimates, e.g. *Holthuijsen (2010)*, and consider the impact of wave height on fuel consumption for a vessel at 11 kn, Fig.5.

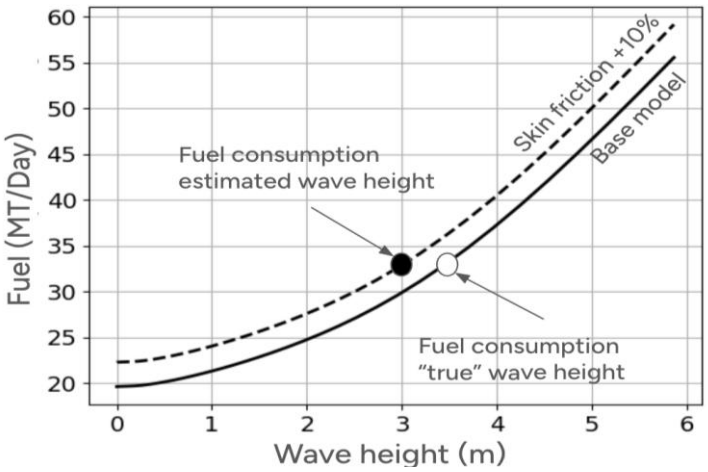


Fig.5: Estimates of fuel consumption for a Capesize vessel for different wave heights based on a base estimate of the vessel, and an estimate where we increased the influence of skin friction by 10%. An error of 0.5 m in wave height can be misinterpreted as corresponding to a 10% difference in hull resistance.

If we (say), assume that the vessel reports 33 MT/day consumption for 3 m waves, we may attribute that to an incorrect estimate of skin friction. However, if the true waves were 3.5 m, the error vanishes. The mismatch was merely due to incorrect weather conditions. However, misjudging hull performance will have consequences, as performance estimates inform operational decisions such as whether hull cleaning is needed, economic speed of a vessel, and e.g. charter party agreements. Evidently, this example is academic, as in general decisions will be based on multiple data points. Here we merely wish to illustrate that error reductions in wave estimates can and will meaningfully reduce noise in data, which in turn can lead to better decision making.

5.2. Accurate weather data will matter more in the future

Arguably, weather accuracy has historically been of less importance because of issues with noon reports. Manual data entry issues, unavailability of direct observations (e.g. shaft power meter), and the low

reporting frequency (hourly data) all contribute to high noise that may overwhelm other issues. This is, however, rapidly changing as high frequency data and automated logging becomes more widely adopted in industry. In this context, accurate weather information will become crucial. A single weather prediction over the day is insufficient if speed, power and fuel consumption are monitored at much higher frequencies. More so because it is primarily weather that will induce large changes in any of these signals; under normal operations a vessel will maintain near constant speed and consumption unless weather conditions change. Consequently, accurately capturing how and when this change occurs is crucial to explain changes in observed signals. Hence, as the industry shifts towards high-frequency data, and data-driven methods to help explain trends and estimate hull/vessel performance, we are convinced that to unlock the potential of these methods, accurate weather data is crucial.

6. Conclusion

Accurate environmental forcing is a foundational requirement for reliable vessel performance analytics. While onboard monitoring systems have advanced rapidly, open-ocean environmental observations have remained comparatively sparse.

We presented Sofar's globally distributed spectral buoy network and a corresponding operational data assimilation system that integrates buoy and satellite observations into a global wave model. Assimilating distributed spectral observations reduces short lead-time wave errors by more than 50% relative to a non-assimilated baseline. Because assimilation propagates information dynamically, improvements extend across large ocean regions.

We demonstrated that wave-height errors on the order of 0.5 m - common in energetic systems - can translate into apparent hull resistance differences on the order of 10% in simplified performance models. In operational contexts, this level of environmental uncertainty can materially distort performance interpretation and optimization decisions.

As maritime operations increasingly rely on high-frequency data and digital optimization frameworks, reducing environmental uncertainty becomes not merely desirable but necessary. Distributed spectral ocean sensing combined with advanced data assimilation provides a scalable path toward more reliable, data-driven vessel performance optimization.

Acknowledgements

It takes a village to develop a system. The work presented here could not have happened without contributions from Sofar colleagues past and present (too many to name individually), funding from the office of naval research (ONR), and feedback from and collaboration with our users.

References

ARDHUIN, F.; ROGERS, E.; BABANIN, A.V.; FILIPOT, J.F.; MAGNE, R.; ROLAND, A.; WESTHUYSEN, A.v.d.; QUEFFEULOU, P.; LEFEVRE, J.M.; AOUF, L.; COLLARD, F. (2010), *Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation*, J. Physical Oceanography 40(9), pp.1917-1941

BUIZZA, R.; HOUTEKAMER, P.L.; PELLERIN, G.; TOTH, Z.; ZHU, Y.; WEI, M. (2005), *A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems*, Monthly Weather Review 133(5), pp.1076-1097

DAULTON, S.; ERIKSSON, D.; BALANDAT, M.; BAKSHY, E. (2022), *Multi-Objective Bayesian Optimization over High-Dimensional Search Spaces,* 38th Conf. Uncertainty in Artificial Intelligence (UAI 2022), https://arxiv.org/pdf/2109.10964.pdf

DORSAY, C.; EGAN, G.; HOUGHTON, I.; HEGERMILLER, C.; SMIT, P.B. (2023), *Proxy observations of surface wind from a globally distributed network of wave buoys*, J. Atmospheric and Oceanic Technology 40(12), pp.1591-1603

DORSAY, C.; HOUGHTON, I.; DAVIS, J.; THOMSON, J.; SMIT, P.; STACKPOLE, E. (2023), *Aerial deployment of spotter wave buoys during hurricane Ian*, IEEE OCEANS Conf., pp.1-5

HOLTHUIJSEN, L.H. (2010), *Waves in oceanic and coastal waters*, Cambridge University Press

HOLTROP, J.; MENNEN, G.G.J. (1978), *A statistical power prediction method*, Int. Shipbuilding Progress 25(290), pp.253-256

HOUGHTON, I.A.; HEGERMILLER., C.; TEICHEIRA., C.; SMIT, P.B. (2022), *Operational Assimilation of Spectral Wave Data from the Sofar Spotter Network*, Geophysical Research Letters

HOUGHTON, I.A.; PENNY, S.G.; HEGERMILLER, C.; CESARETTI, M.; TEICHEIRA, C.; SMIT, P.B. (2023), *Ensemble-based data assimilation of significant wave height from Sofar Spotters and satellite altimeters with a global operational wave model*, Ocean Modelling 183

HOUGHTON, I.A.; SMIT, P.B.; CLARK, D.; DUNNING, C.; FISHER, A.; NIDZIEKO, N.J.; CHAMBERLAIN, P.; JANSSEN, T.T. (2021), *Performance statistics of a real-time Pacific Ocean weather sensor network*, J. Atmospheric and Oceanic Technology 38(5), pp.1047-1058

ITTC (2021), *Recommended Procedures and Guidelines, 7.504-01-01.1, Preparation, Conduct and Analysis of Speed/ Power Trials*, International Towing Tank Conference

KALNAY, E. (2002), *Atmospheric Modeling, Data Assimilation and Predictabilit*y, Cambridge University Press

LIU, S.; PAPANIKOLAOU, A. (2020), *Regression analysis of experimental data for added resistance in waves of arbitrary heading and development of a semi-empirical formula*, Ocean Engineering 206, 107357

OOSTERVELD, M.W.C.; VAN OOSSANEN, P. (1975), *Further computer-analyzed data of the Wageningen B-screw series*, Int. Shipbuilding Progress 22(251), pp.251-262

PIERSON, W.J.; MOSKOWITZ, L. (1964), *A proposed spectral form for fully developed wind seas based on the similarity theory of SA Kitaigorodskii*, J. Geophysical Research 69(24), pp.5181-5190

RAGHUKUMAR, K.; CHANG, G.; SPADA, F.; JONES, C.; JANSSEN, T.; GANS, A. (2019), *Performance characteristics of "Spotter," a newly developed real-time wave measurement buoy*, J. Atmospheric and Oceanic Technology 36(6), pp.1127-1141

RITCHIE, H. (2024), *Weather forecasts have become much more accurate; we now need to make them available to everyone*, Our world in data

SMIT, P.B., HOUGHTON, I.A.; JORDANOVA, K.; PORTWOOD, T.; SHAPIRO, E.; CLARK, D.; SOSA, M.; JANSSEN, T.T. (2021), *Assimilation of significant wave height from distributed ocean wave sensors*., Ocean Modelling 159, 101738

TOLMAN, H.L.; ABDOLALI, A.; ACCENSI, M.; ALVES, J.H.; ARDHUIN, F.; BABANIN, A.; BARBARIOL, F., BENETAZZO, A.; BIDLOT, J.; BOOIJ, N.; BOUTIN, G.; BUNNEY, C.; CAMPBELL, T.; CHALIKOV, D.; CHAWLA, A.; CHENG, S.; COLLINS, C.O.; FILIPOT, J.F.; FLAMPOURIS, S.; FOREMAN, M.; GINIS, I.; HARA, T.; HESSER, T.; JANSSEN, P.A.E.M.; LECKLER, F.; LI, J.G.; LIND, K.; LIU, Q.X.; MEIXNER, J.; MENTASCHI, L.; ORZECH, M.; PADILLA-HERNANDEZ, R.;

PERRIE, W.; RAWAT, A.; REICHL, B.; ROGERS, E.; ROLAND, A.; SÉVIGNY, C.; SHEN, H.; SAULTER, A.; SIKIRIC, M.D.; SZYSZKA, M.; SMITH, J.M.; TOULANY, B.; TRACY, B.; YOUNG, I.R.; ZHAO, X. ZIEGER, S.; LIANG, Z. (2019), *User manual and system documentation of WAVEWATCH III R version 4.07*, Environmental Modeling Center, Marine Modeling and Analysis Branch, Maryland

Proactive Hull Management with Biocide-Free Hard Foul-Release Coatings: Lessons from an LPG Carrier after One Year in Operation

Ryan Ingham, GIT Coatings, Halifax, Canada, ryan.ingham@gitcoatings.com
Michael Jacot, GIT Coatings, Halifax, Canada, michael.jacot@gitcoatings.com
Philippos Sfiris, GIT Coatings, Athens, Greece, philippos.sfiris@gitcoatings.com

Abstract

This paper presents a one-year operational case of Proactive Hull Management (PHM). The case focuses on an LPG carrier using a biocide-free hard foul-release coating with onboard robotic grooming. A ship-specific plan was developed through biofouling-risk modelling and port-feasibility assessment. Over twelve months, hull condition and performance were monitored via inspection footage and speed-power analysis. Despite delayed or partial cleanings, the hull remained free of hard fouling and the power trend stayed relatively stable: 6% fuel savings out of dock, <2% added power over the year, and 3-5% fuel savings regained after each grooming event. The study highlights operational learnings related to crew involvement, port practices and the importance of advisory support, and outlines the framework that enabled successful implementation. Findings indicate that PHM, when supported by the right coating, grooming technology and advisory model, offers a scalable, sustainable pathway to meet emerging biofouling management requirements.

1. Introduction

1.1. Three schools of thought in keeping the hull clean

Today, most ship operators have one of three broad approaches to control the growth of biofouling on the hulls of their ships, Fig.1.

School of thought	Coating type (indicative)	Ship owner/operator profile	Benefits	Risks
Reactive cleaning when needed	*SPC (Self-polishing antifoulings)*	• Cost-driven owners • Clean only after performance loss or visible fouling • Trade and application limitations	• Proven tech • Moderate cost	• Removing hard fouling may be too late • Damages the paint and increases roughness • Paint thickness loss • Biocide release • Risk of microplastic pollution
'Apply & forget'	*Silicone*	• Seeking optimized fuel savings with minimal maintenance	• Low friction • Fuel savings	• Sensitive to damage • Reactive cleaning can cause damage, especially after idle periods in aggressive waters • Risk of microplastic pollution
Proactive cleaning	*Biocide-free, 'groomable' coatings designed for regular slime-level cleaning*	• Innovative forward-thinking • Use of onboard robots or ROVs in port • Total Cost of Ownership mindset • Managing biofouling as a routine	• Keeps hull smooth and clean • Reduced risk of water pollution and invasive species transfer	• Robotic tech still maturing • Port restrictions • Crew involvement • Planning required

Fig.1: Three schools of thought for hull management

1) Reactive cleaning on self-polishing biocidal coatings
Conventional SPC antifouling coatings are applied at drydock and the hull is left largely unattended until performance loss is detected or visible macrofouling is observed. In practice, macrofouling is usually confirmed only after divers are sent to inspect the hull, following a fuel or speed penalty. Cleaning at that stage is often aggressive and time-pressured, increasing hull roughness, removing film thickness and releasing additional biocides and microplastic particles into the water, *Soon et al. (2024)*. Hull cleanings of this nature shorten coating life and erode the coating's original fuel-saving benefit.

2) "Apply & forget" silicone foul-release systems
Silicone coatings deliver a smooth low friction surface and good fuel performance as long as the coating remains intact. These coatings are attractive for owners who monitor performance and seek optimised fuel savings with minimal maintenance. However, they are relatively soft and not designed for frequent mechanical cleaning. After long idle periods in warm waters, operators may still need to clean. If the cleaning is too harsh, the coating can be damaged locally, shortening life and creating uncertainty about when cleaning helps and when it harms coating performance.

3) Proactive cleaning with groomable, biocide-free coatings
A third, emerging school of thought uses biocide-free coatings specifically designed for regular slime-level cleaning. Microfouling (FR20) is removed before it develops into more resistant fouling, and hull care is treated as a routine operational task rather than an exception. This approach typically combines "groomable" coatings with onboard robots or in-port ROV services and requires planning, crew involvement, and a mindset that considers the total cost of ownership of the asset.

1.2. Evolving regulations and standards lead to proactive biocide-free solutions

Regulatory and standardisation trends are pushing the industry away from unplanned, reactive cleaning of biocidal coatings and toward proactive, verifiable biofouling management with reduced discharge of toxic chemicals and microplastics to the marine environment.

On the biofouling-management side:

- Revised IMO Biofouling Management Guidelines (MEPC.378(80)) introduce Biofouling Management Plans and Record Books, a standardized Fouling Rating scale (FR 0-5), and a stronger emphasis on underwater hull inspections and in-water cleanings, *IMO (2023), Scianni et al. (2023).*
- MEPC.1/Circ.918 provides guidance on in-water cleaning, defining coating compatibility, capture and environmental safeguards, *Tamburri et al. (2022), IMO (2025)*.
- MEPC 83 has agreed to develop a legally binding framework on biofouling management, with adoption targeted before 2030, https://maritime.lr.org/MEPC-83-Summary-Report
- Port and regional authorities (e.g. New Zealand, California, Brazil, Norway, Australia) are already translating IMO and ISO guidance into enforceable rules, often requiring clean-hull entry and restricting cleaning of biocidal coatings to captured operations, *Scianni et al. (2023).*

On the coating-chemistry side:

- Authorities are tightening control on the use and concentration of antifouling biocides, with several active substances under re-evaluation or phase-out, *Paz-Villarraga et al. (2021), Weber et al. (2023), Lagerström et al. (2026)*.
- The EU Microplastics Restriction under REACH explicitly covers coatings, focusing on the release of synthetic polymer particles over service life, *Tamburri et al. (2022), De-la-Torre et al. (2023)*.
- Mechanical wear and in-water cleaning of polymer-based foul-release and antifouling coatings can generate microplastic particles that enter the marine environment, and their persistence and behaviour are increasingly under scientific and regulatory scrutiny, *Bork et al. (2025), Soon, et al. (2024), Tamburri et al. (2022)*.

In this context, there is growing interest in hull-management models that avoid biocide release, minimize microplastic emissions from coating wear and provide traceable records of inspections and cleanings to satisfy emerging IMO regulations, ISO standards and port requirements. The need for frequent inspection and proactive cleanings via condition monitoring was first addressed in last year's

HullPIC conference proceedings by Jotun's Hull Skating Solutions, *Levantis and Johansen (2025)*. The goal of this study is to further advance the understanding of the potential benefits of this concept, while highlighting the elements that differentiate its application across coating technologies and associated implementation framework.

The following sections document a full-year case where such a model was implemented on a commercially trading LPG carrier: a biocide-free graphene-based hard foul-release coating combined with proactive robotic grooming and an advisory-led methodology provided by GIT Coatings.

2. Proactive Hull Management (PHM) Methodology

GIT Coatings' PHM methodology is structured as a 10-step process that links data-driven planning, grooming feasibility, technology selection and in-service support into a single end-to-end hull management system. The main goal of this methodology is to provide vessel owners, managers, and operators with the proper framework and tools to maintain an optimal hull condition over the drydock interval. This methodology is introduced prior to the vessel's drydock so that the appropriate stakeholders can be informed on what is expected of them once the vessel launches and resumes its commercial trade.

1) Biofouling risk assessment: An overview of the vessel's historical trading pattern over the previous 12 months in service is analyzed to determine the biofouling exposure and the likelihood of fouling development over time.

2) Recommended grooming plan: A custom-tailored grooming plan is developed based on an in-house fouling prediction model that uses the biofouling risk profile derived from the vessel's historical trading pattern. In this model, biofouling is represented as a single level $B(t)$ that changes over time based on where the vessel's trading region and activity level. AIS data provides position and speed at time t; the position and time are used to look up local ocean conditions $E(t)$ (for example sea water temperature, salinity, and biofouling productivity proxies). In practice, $G(E(t), v(t))$ increases under environmentally favourable conditions and decreases toward zero during sustained transit speeds. Model parameters are chosen using literature guidance and then calibrated using real vessel histories by setting $B(t)$=0 at documented full hull cleanings and tuning the model so its predicted levels align with subsequent underwater inspection observations. The result is a predictive tool that can be run forward using a planned route when available or using recent historical trading patterns when the future route is unknown.

$$B(t + \Delta t) = max\left(0, B(t) + G\big(E(t), v(t)\big)\right), \qquad \Delta t = 1\ hour$$

$B(t)$: biofouling level at time t
$E(t)$: local ocean conditions at time t
$v(t)$: vessel speed /activity status at time t
$G(E(t), v(t))$: fouling growth added during time step t, computed from ocean conditions $E(t)$ and vessel speed/activity $v(t)$
$max(0,)$: prevents the fouling level from going below zero

The model estimates when the hull reaches a grooming threshold corresponding to Fouling Rating FR20 (light slime) on both the Vertical sides (VS) and Flat bottom (FB) areas independently.

The FR20 grooming threshold was chosen because:

- It is the point where drag penalties begin to be significant in fuel terms.
- Fouling is still soft and can be removed with one gentle pass, limiting cleaning time and coating wear.
- It aligns with the recommendation in the IMO 2023 Biofouling Guidelines

The grooming plan recommends the frequency (in months) for the vertical side and flat bottom hull areas to be groomed to maintain cleanliness at or below the FR 1 / NSTM FR 20 level.

3) Grooming model: Based on the vessel's criteria (vessel size, speed, time in port, etc.) the appropriate grooming model is recommended by GIT's advisory team to support the vessel's ability to follow the recommended grooming plan. The selection is made from three main types of grooming models:
    - Model A: onboard robotic grooming operated by the vessel's crew
    - Model B: in-port grooming by ROV or diving services company
    - Model C: in-transit grooming of the vertical sides and in-port grooming of the flat bottom and niche areas

4) Feasibility assessment of grooming plan: Once the grooming plan is developed along with the method for which the recommended grooming events will be carried out, a feasibility assessment is conducted to determine if the vessel would have been able to execute each grooming event throughout its previous 12 months in service. Port regulations and time in port are evaluated, and the vessel receives a feasibility score that aims to give the vessel's owner confidence that the vessel's trade allows for PHM.

5) Selection of grooming technology: The vessel owner can choose from a list of available grooming technologies that fit the selected grooming model. These technologies have been verified as compatible with GIT's coating by a class society. This class society has reviewed compatibility testing provided by GIT and the technology provider to ensure that there is minimal impact on the coating's surface properties after repeated grooming cycles with the selected cleaning equipment.

6) Total cost of ownership and payback analysis: The PHM approach with GIT's high-performance antifouling coating XGIT-Fuel is compared against alternative methods of antifouling control (SPC, High Tier SPC, Silicone Foul Release, etc.) in a simulation of added resistance/power over the duration of the vessel's drydock cycle. The total cost of ownership is forecasted using estimates for coating cost, drydock cost, fuel cost and cleaning costs to determine the predicted return on investment and payback period.

7) Documentation for port authorities: Ports often require detailed information on the coating technology and grooming technology prior to issuing approval for a grooming event to occur in their local waters. The GIT advisory team works with the cleaning provider and the vessel's captain to ensure the port receives all documentation required for approval.

8) Fouling warnings: The vessel's biofouling risk is continuously monitored over the drydock interval and fouling warnings are issued to the vessel if the predicted biofouling risk surpasses that of FR20 and or if the risk does not align with the original grooming plan. These fouling warnings signify the need for an underwater inspection at the next available port or anchorage to understand the hull's condition and then a decision on the need to perform a hull grooming event can be made based on inspection findings.

9) Analysis of inspection and performance data: The GIT advisory team works closely with the cleaning provider to ensure all underwater footage is stored and analyzed accordingly. The team will provide timely feedback on the coating condition and if there is a need to update the recommended grooming plan. Performance data is also reviewed periodically to check if the current hull condition matches the expected hull performance of the vessel over time.

10) In-service hull grooming advisory: This service aims to assist the vessel in making informed decisions on where and when to perform a hull grooming event in the case of port restrictions,

excess idle time, or ROV breakdown. The end goal of this service is to ensure that the vessel can employ PHM even if unforeseen complications arise throughout the drydock interval.

In summary, the PHM methodology defines how proactive hull management should work in principle for any vessel. The next sections show how it was applied in practice to a specific LPG carrier.

3. PHM Implementation on an LPG Carrier – Pre docking analysis

For the purposes of this paper, the case study focuses on selected elements of the 10-step PHM methodology: pre-docking biofouling risk assessment, the ship-specific grooming plan and feasibility assessment, and the in-service execution and performance results. Before the vessel left drydock, a ship-specific PHM plan was developed based on its historical trading pattern and fouling-risk profile. The vessel is a 120m LPG carrier and was trading mainly in the Mediterranean Sea and Black Sea prior to drydock coating application.

3.1 Assessment of biofouling risk based on vessel's historical trading pattern

Analysis of the previous 12 months showed roughly 40% activity (time at sea) and approximately 225 idle days, many of them in warm water ports and anchorages. Using internally developed tools, the trade data (speed profile, idle periods, water temperature and salinity) were processed into a biofouling risk assessment, Fig.2. The vessel's pattern places it firmly in the high to extreme exposure range, with frequent stops in warm, nutrient-rich waters and long idle periods in high-risk fouling zones. Without early maintenance, fouling would be expected to grow quickly on such a profile, making the vessel an ideal and demanding candidate for proactive hull management.

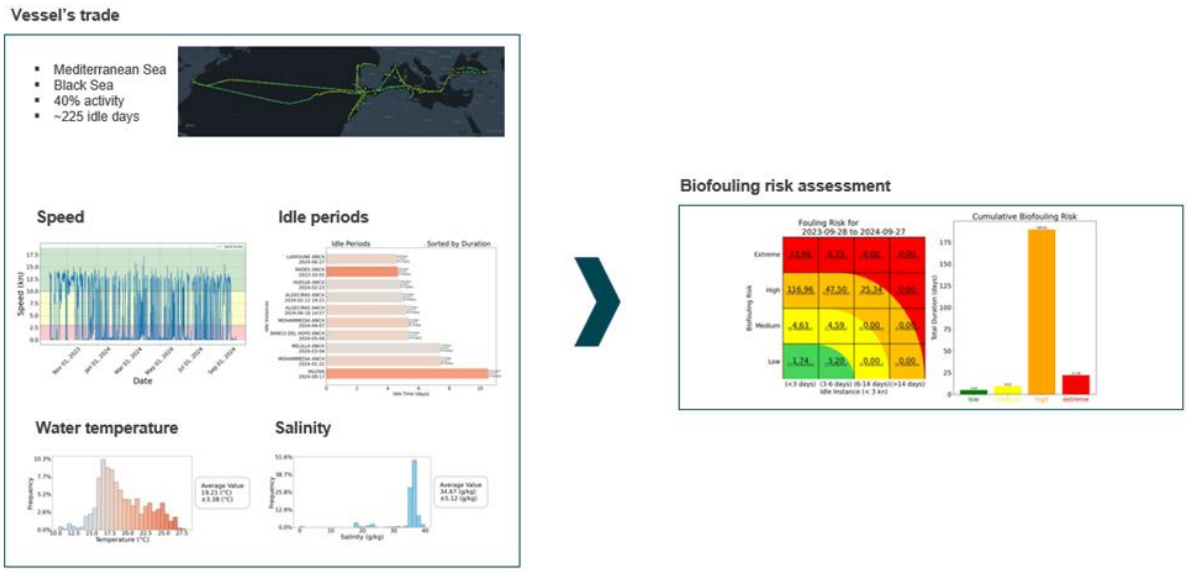


Fig.2: Vessel's historical trade and biofouling risk assessment

3.2. Ship-specific grooming plan based on biofouling growth modelling

Once the biofouling risk was quantified, a biofouling growth model was used to simulate how quickly microfouling would build up on the LPG carrier under its expected trade, Fig.3.

From the biofouling growth model, a ship-specific grooming plan was defined for the LPG carrier based on its historical trading pattern:

- Vertical sides
  - Inspection every ~30 days.
  - Grooming approximately every ~90 days.

- Flat bottom
    - Inspection every ~60 days.
    - Grooming approximately every ~180 days.

These intervals formed the backbone of the PHM plan and were used to identify when, and in which trading regions the grooming events would need to be scheduled.

Fig.3: Vessel simulation of biofouling growth over time and recommended grooming plan

3.3. Grooming feasibility assessment (port analysis)

To make the ship-specific grooming plan operational, a feasibility assessment was carried out to identify where grooming could be performed with the onboard robot within the planned inspection and grooming intervals. The study focused on three questions for each port and anchorage on the vessel's typical routes:

- Is in-water grooming/cleaning permitted under local port regulations and terminal rules?
- Is the average time in port sufficient to complete a grooming cycle with the onboard robot (≈6-8 hours per side (based on vessel size) plus deployment and inspection time)?
- Are there nearby ports or anchorages that could serve as alternative grooming locations within the same trading leg?

Using the vessel's historical AIS data and a database of port-stay durations and in-water cleaning restrictions, each relevant call the vessel made was classified as either:

- Feasible: time and local conditions allow a planned robot deployment within the grooming interval.
- Not feasible: either too little time or in-water work is not permitted for the onboard system.

The outcome was a port-feasibility map linking the FR20-based grooming intervals (Section 3.2) to ports and anchorage locations where the plan could be executed with the robot onboard. For the legs in the Mediterranean Sea and Black Sea, at least one "anchor port" fell within each three-month window, meaning that a grooming opportunity was available before the hull was expected to reach the FR20 threshold.

In parallel, the same map recorded the presence of approved in-port cleaning providers as a contingency plan, Fig.4. Where feasible ports also hosted these service providers, they were marked as backup options in case the onboard robot became temporarily unavailable or if local rules required a provider to perform the work.

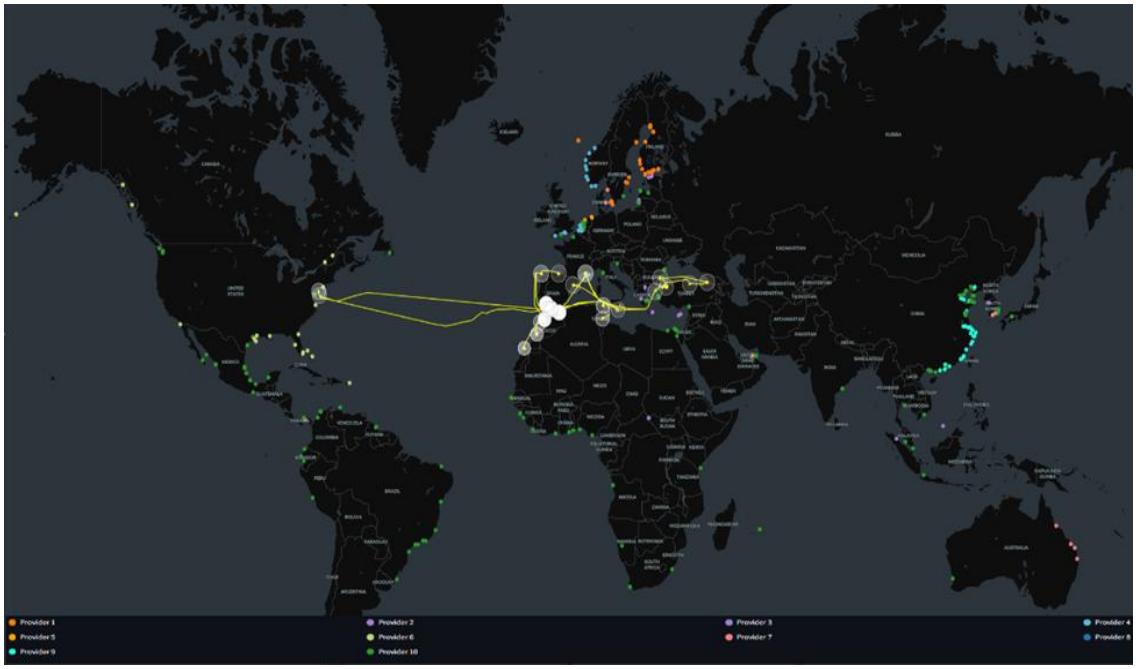


Fig.4: Map of global ports for which hull grooming is feasible by various providers

3.4. Selection of grooming method

Different grooming methods were evaluated to help the owner decide how grooming events should be performed in practice:

- Model A: Onboard robot operated by crew.
  - Best suited for vessels up to around 150 m LOA, with predictable idle periods where the crew can deploy and recover the robot effectively.
- Model B: In-port ROV service.
  - Using local approved service providers that bring their own equipment to the vessel in port. Suitable where onboard robots are restricted, where LOA is large, or where crew resources are limited.
- Model C: Grooming in transit.
  - Tools operated during ocean legs, suited to high-speed, long-haul trades where in-port cleaning windows are scarce.

For this LPG carrier, Model A was chosen as the primary method:

- Vessel size allows full-hull coverage in reasonable time.
- The owner was prepared to train a small deck team to operate the robot.
- Idle periods in anchorages and ports create natural windows for grooming.

At the same time, Model B was retained as a structured back-up method:

- To treat niche areas (bilge keels, sea chests, complex geometries) that may be less accessible to the robot.
- To provide contingency if the onboard robot is temporarily unavailable or if local rules restrict its use in specific ports.

Model C was not considered to be a primary option due to the vessel's typical service speeds and the need to proactively clean the flat bottom every 6 months.

3.5. Proposed solution: Biocide-free graphene-based coating and grooming robot onboard

Given the high fouling risk and the owner's decarbonization ambitions, the proposed solution was to apply GIT's biocide-free graphene-based hard foul-release coating (XGIT-Fuel) paired with an onboard hull-grooming robot.

The onboard grooming system is a semi-autonomous magnetic crawler with soft brushes. It can map the hull and record fouling coverage, perform inspection passes and remove slime-level fouling in a single light pass. For this vessel size, experience shows that once mapping is completed, a full grooming cycle requires roughly 6-8 hours per side of vertical hull, plus shorter periods for deployment, post-inspection and reporting.

Together, the coating and robot form a simple but disruptive combination: a durable, low-friction surface engineered for frequent light cleaning, and a tool that allows crew to remove early-stage slime before it can become more persistent fouling.

After the first year in service, the feasibility assessment was compared with the vessel's actual trade and grooming history. All three grooming events, as well as the inspections that informed scheduling decisions, were carried out in ports and anchorages that had been pre-classified as feasible locations in the study. This alignment indicates that the pre-docking feasibility assessment correctly identified where grooming could be performed within the planned intervals, with the approved cleaning partners available if needed as contingency.

Having established the plan, technology and feasibility before leaving drydock, the next step is to see how PHM performed once the LPG carrier returned to commercial service.

4. PHM Implementation on an LPG Carrier – Execution & monitoring

4.1 In-service monitoring & advisory support

Once in service, the PHM plan was managed as a live process rather than a static grooming schedule. GIT's advisory team supported the operator through:

- Inspection footage reviews after each robot deployment to ensure there is no signs of macrofouling
- Fouling warnings triggered once predicted growth levels exceeded FR20 (typically due to long idle periods in high-risk waters).
- Monthly performance reports which combined performance insights from high-quality noon reports with inspection and grooming logs.
- Recommendations on when and where to groom next, based on the grooming plan, feasibility map and future port calls.
- Documentation material for port approvals, including XGIT-Fuel's enhanced antifouling type approval from a class society and the vessel's biofouling management plan.

Regular three-way communication between the owner, the robot supplier and GIT's team allowed the plan to be adjusted when planned grooming windows were missed or shortened, robot technical issues delayed a cleaning, port regulations or operational constraints changed at short notice.

4.2 Operational timeline - cleaning events

During the first 12 months after drydock, the vessel completed:

- 7 underwater inspections.
- 3 grooming events using the onboard robot, targeting full-hull coverage where time allowed.

The operational timeline, Fig.5, shows the vessel's actual inspection/grooming schedule over the first 12 months.

- The robot was launched immediately after drydock, and the vessel's crew were trained on how to deploy and recover it.
- The first hull grooming event brought the hull back to a clean condition after initial development of a slime layer, confirming that the coating tolerated microfouling removal without visible damage from the type of cleaning equipment used.
- Subsequent grooming events were affected by typical commercial realities including tight port schedules, time-sharing with cargo operations, and occasional robot reliability issues. This led to partial cleanings (e.g. one vertical side only) and delayed cleanings relative to the intended FR20 thresholds.

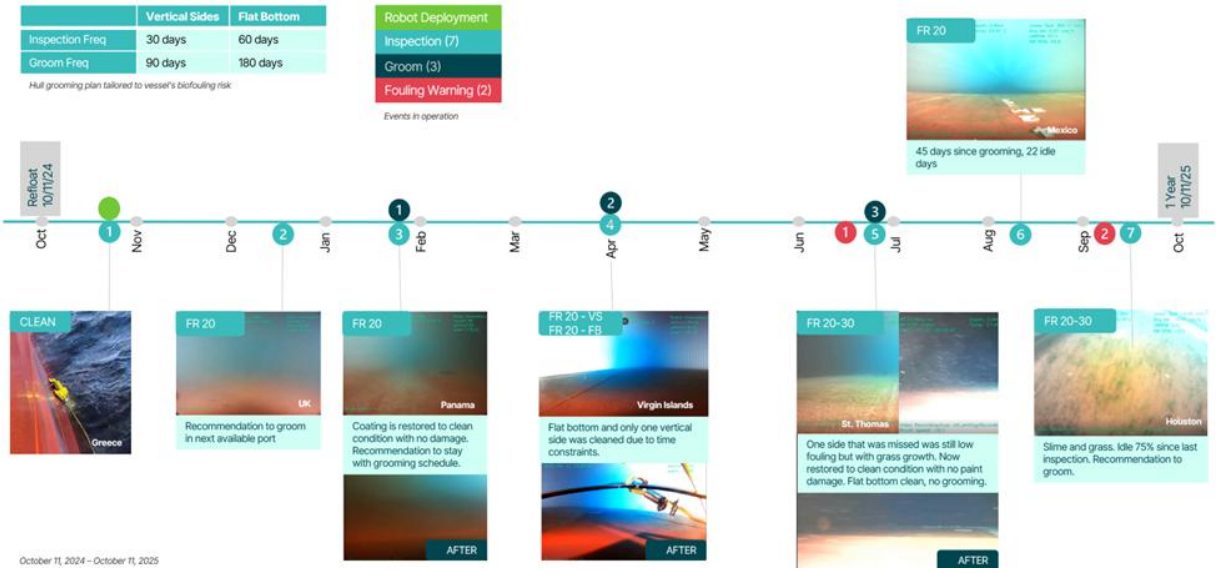


Fig.5: Operational timeline of events for the LPG Carrier's first year in-service

Despite these challenges, GIT's advisory team continued to issue fouling warnings and updated grooming recommendations based on the actual trade. Each completed grooming successfully restored the treated areas to a clean condition.

Video footage from inspections and grooming events provided a consistent picture of the hull condition over the year:

- The hull remained at FR ≤ 20 (slime only) for most of the period.
- When grooming was delayed, local grass growth appeared on one side and in some niche areas but was removed at the next cleaning without apparent damage to the coating.
- No hard fouling was recorded on the underwater hull.
- No coating detachment, blistering or visible roughening was observed after multiple grooming events.

These observations support the conclusion that the graphene-based coating maintained its smoothness and integrity under repeated grooming events at the slime-stage. In addition to this visual evidence of hull condition, quantitative analysis of speed-power performance was carried out to understand how PHM affected fuel efficiency over the first year in service.

4.3. Hull performance evaluation

The vessel's hull performance was evaluated based on high quality noon report data that was collected by the vessel owner. Once received, the data was cleaned and the shaft power was normalized to an

average sailing speed of 12 knots in good weather conditions (BF wind force rating ≤4). Using this normalized dataset, a baseline hull performance was established for laden condition to represent a clean hull. The changes in hull performance were then quantified before and after each inspection and hull grooming event as denoted by the vertical-coloured lines in Fig.6.

This performance assessment aims to quantify the discrete impact of hull grooming events under representative commercial operating conditions. While ISO 19030 provides a standardized framework for long-term hull and propeller performance monitoring, the objective of this approach was to isolate and evaluate the short-term performance changes attributable to individual grooming interventions over a smaller time frame. It should be noted that the uncertainty associated with this approach is expected to be higher than that of the ISO 19030-2 default method. The primary driver is measurement frequency, as noon report data provides one aggregated data point per day, whereas ISO 19030-2 typically relies on high-frequency measurements at intervals of ~15 s. Reduced measurement frequency increases statistical variability and limits the ability to average out short-term fluctuations. For context, ISO 19030-3-3 estimates a performance indicator uncertainty of approximately 4.8% at a 2 sigma, 95% confidence interval when comparing two 3-month in-service periods. The uncertainty associated with this event-centered approach using noon reports is therefore expected to be of similar order due to the cluster size and environmental variability of the vessel's daily operations.

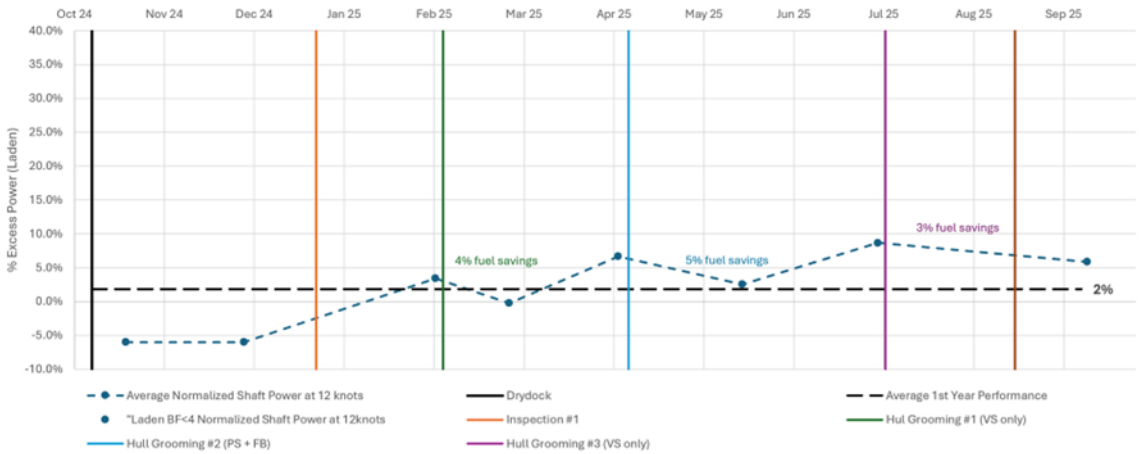


Fig.6: % Excess Power over time trend for the LPG Carrier in Laden condition.

Key findings:

- A 6% power gain out of dock compared to the vessel's previous high-tier SPC antifouling, under comparable trading conditions. The gain is attributed mainly to lower roughness and frictional resistance which has been previously tested and validated in a fully turbulent flow channel experiment, *Ravenna et al. (2022).*
- Only ~2% added power over the first year, despite ≈180 idle days in extreme fouling exposure. The power trend remained relatively stable.
- Average of ~4% fuel saving regained after each grooming (3-5% range), measured as the reduction in excess power immediately after grooming vs pre-grooming. This indicates the value of timely removal of slime-level fouling.

To put these results in context, the data were compared with two simplified theoretical curves, Fig.7:

1. An ideal proactive grooming model, assuming full-hull, on-time grooming at FR20 with no operational constraints.
2. A typical high-tier SPC antifouling baseline, with no in-water cleaning in year one but its hull roughness increases by ~200 microns per year due to dry film thickness loss and early-stage microfouling development over time.

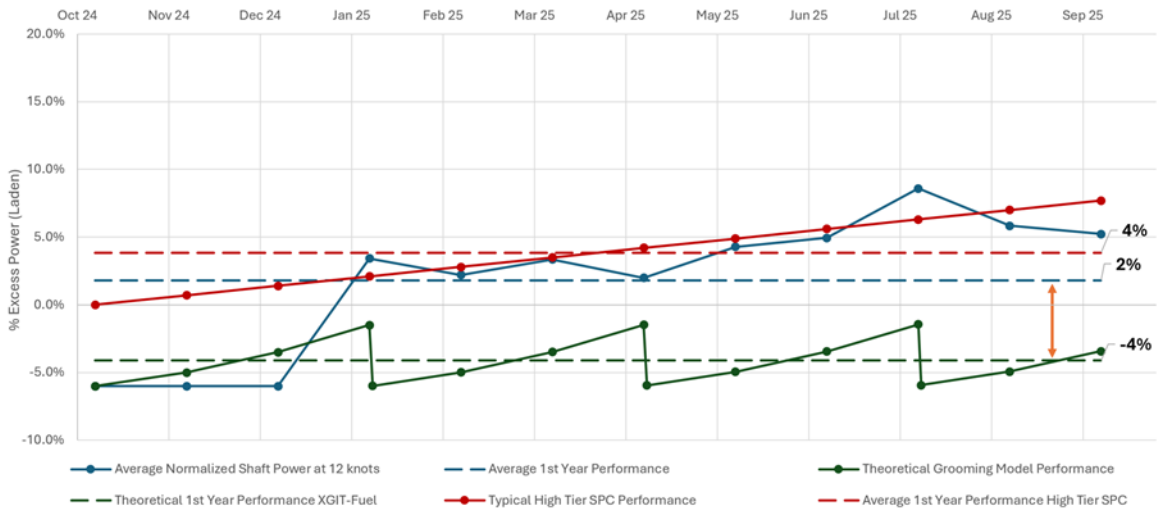


Fig.7: % Excess Power over time trend for the LPG Carrier (theoretical vs. practical scenarios)

The actual performance of the coating-plus-grooming bundle sits between these two curves but closer to the ideal proactive model:

- Even with partial and delayed cleanings, the vessel delivered around 2% better annual performance than the high-tier SPC baseline in the same trade and fouling exposure.
- The modelling indicates up to an additional 6% in annual efficiency potential is attainable if the performed grooming events can move closer to the ideal schedule through better planning, wider port coverage, consistent full-hull grooming and remote operation with skilled personnel.

5. PHM Implementation on an LPG Carrier – Key learnings and challenges

After one year of operation, several clear lessons emerge from this case study:

1) It can be done. A biocide-free graphene-based hard foul-release coating combined with proactive grooming and advisory support kept an LPG carrier's hull clean in day-to-day commercial trade, under high fouling pressure and with real-world operational constraints. No hard fouling developed, coating integrity was preserved, and hull performance stayed close to the "as-docked" levels.

2) It can perform better than the conventional alternative. Even with non-ideal execution, this approach outperformed a theoretical high-tier SPC baseline, and there is still measurable upside if the PHM execution is optimised.

3) It requires a managed program, not just a coating. Success depended on the combination of:

   - Trade-based biofouling risk assessment and FR20 grooming thresholds.
   - A port-feasibility assessment that proved realistic when checked against actual trade and provider availability.
   - In-service monitoring and monthly reviews that kept everyone aligned and allowed the plan to be adjusted as reality evolved.
   - Coating chemistry and robotics were necessary but not sufficient; the advisory framework was essential.

4) People and collaboration matter. As crew confidence with the robot increased, coverage and consistency improved. Close collaboration between the owner, robot supplier and GIT's advi-

sory team enabled quick resolution of early technical issues and supported decision-making on when and where to groom.

5) There is still room to improve. The first year highlighted clear improvement areas:
    - The importance of trade route identification - Spot charter trade is unpredictable and does not allow for future planning of grooming events since the future port calls are typically not known in advance. Whereas time-chartered vessels have scheduled routes and are better suited with respect to the applicability of PHM.
    - Some ports do not allow hull grooming to be done using the hull grooming robot while in inner anchorage areas even if the vessel explains that the hull coating is biocide-free and the previous month's inspection report shows that hull is almost clean and only slime removal is needed.
    - There is a need to educate port states and other port control authorities on the PHM concept and that it entails regularly schedule grooming that targets early stage microfouling on a biocide-free coating that is designed to be cleaned. There also needs to be clear messaging that hull grooming robots carried on board are not to be confused with underwater free-floating drones, since the grooming robot should always remain attached to the hull.
    - Broader port coverage and approved service partners to complement onboard robots and cover niche areas.
    - Consistent full hull grooming is achieved for each event, whenever time and conditions permit.
    - Remote robot operation to reduce crew workload and standardise cleaning quality.
    - Further refinement of the grooming feasibility model, leveraging additional trade with good sailing days and long-term performance data.
    - Future work will benefit from increased measurement frequency through direct sensor integration, enabling improved noise reduction, tighter confidence intervals, and formal uncertainty propagation aligned with ISO standard methodologies.

Overall, the first-year operation of the LPG carrier demonstrated that a proactive, biocide-free hull-management model is no longer a theoretical concept. With the right planning, feasibility assessment and in-service support, it can work and already delivers better performance and environmental outcomes than some conventional approaches.

Acknowledgement

We wish to thank the reviewers of this paper for their thoughtful feedback and constructive suggestions, which helped improve its quality and content. We also gratefully acknowledge the vessel owner for providing the candidate vessel and implementing GIT's proactive hull management methodology, as well as the robotic grooming supplier for supplying the ROV onboard, supporting crew training, and providing technical assistance during operations.

References

BORK, K.; NIELSEN, M.; DAM-JOHANSEN, K. (2025), *Microplastics released from antifouling coatings: Comparison of coating types under realistic wear*, Int. Antifouling Conf., pp.112-119

IMO (2023), *Guidelines for the control and management of ships' biofouling to minimize the transfer of invasive aquatic species*, Int. Mar. Org., London, https://wwwcdn.imo.org/localresources/en/KnowledgeCentre/IndexofIMOResolutions/MEPCDocuments/MEPC.378(80).pdf

IMO (2025), *Guidance on in-water cleaning of ships' biofouling*, Int. Mar. Org., London, https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Biofouling%20pages/MEPC.1-Circ.918%20-%20Guidance%20On%20In-Water%20Cleaning%20Of%20Ships%27%20

Biofouling%20(Secretariat).pdf

LAGERSTRÖM, M.; GRANHAG, L.; YTREBERG, E. (2026), *Impact of antifouling paint regulation on copper and zinc emissions from leisure boats*, Marine Pollution Bulletin 191

LEVANTIS, M.; JOHANSEN, M.S. (2025), *Proactive Cleaning: Leveraging Data for a Successful Beginning*, HullPIC Conf. Mülheim, pp.184-194, http://data.hullpic.info/HullPIC2025_Mulheim.pdf

PAZ-VILLARRAGA, C.; MUÑOZ-BRAVO, J.; ESPINOSA, A. (2021), *Biocides in antifouling paint formulations currently registered for use*, Env. Science and Pollution Research 28, pp.50825-50842

RAVENNA, R.; INGHAM, R.; SONG, S.; JOHNSTON, C.; TEZDOGAN, T.; ATLAR, M.; DEMIREL, Y.K. (2022), *Predicting the effect of hull roughness on ship resistance using a fully turbulent flow channel*, J. Marine Science and Engineering 10(12)

SCIANNI, C.; GEORGIADES, E.; MIHAYLOVA, R.; TAMBURRI, M.N. (2023), *Balancing the consequences of in-water cleaning of biofouling to improve ship efficiency and reduce biosecurity risk*, Frontiers in Marine Science 10

SOON, Z.Y.; TAMBURRI, M.N.; KIM, T.; KIM, M. (2024), *Estimating total microplastic loads to the marine environment as a result of ship biofouling in-water cleaning*, Frontiers in Marine Science 11

TAMBURRI, M.N.; GEORGIADES, E.; SCIANNI, C.; MOSER, C.S. (2022), *Understanding the potential release of microplastics from antifouling coatings*, Frontiers in Marine Science 9

WEBER, S.; KINNER, P.; ROSSHART, C.; ROSENHAHN, A. (2023), *Marine biofouling and the role of biocidal coatings: Regulation driven innovation*, Biofouling 39, pp.739-754

Strategic Hull & Propeller Performance Degradation Management

Richard Marioth, Idealship, Itzehoe/Germany, rm@idealship.de
Sylvain Julien, Berge Bulk, Singapore, Sylvain.Julien@bergebulk.com
Thomas Callahan, Independent Consultant, San Diego/USA thcallahan.consulting@gmail.com

Abstract

This paper discusses key strategies in the context of fleet performance management. It first presents options that allow shipping companies to analyze the development of hull and propeller performance over time at a fleet level. Such as aggregated views enabling fleet managers to identify vessels with the most severe degradation and track the long-term effectiveness of hull and propeller performance management initiatives. Corrective actions can then be planned based on further assessments and in depth vessel performance studies. Based on these overview methods, the paper then explores fleet performance forecasting options that support data driven decision-making for energy efficiency investments and budgeting the overall company fuel and emission reduction strategy.

1. Introduction

1.1. Importance of the topic

Each vessel has its own technical characteristics and operating history. The technical vessel managers in charge have traditionally focused on optimizing individual ships and sister class series. Fleet-level views on fuel performance and efficiency were often of secondary importance, particularly during periods of low fuel prices. This environment has fundamentally changed. Regulatory frameworks such as FuelEU Maritime and the EU Emissions Trading System (EU ETS) have transformed greenhouse gas emissions into a direct and increasingly material cost component while allowing pooling mechanism that extends optimization strategy beyond individual vessels.

Regulatory requirements are expected to strengthen further, with potential global alignment under the IMO Net Zero Framework (NZF). At the same time, financial institutions and charterers are increasingly incorporating emissions performance into lending terms and commercial decisions. Vessel efficiency is therefore no longer solely a fuel optimization issue, but it has become a determinant of asset competitiveness and long-term commercial viability.

In this context, inefficient vessels risk becoming structurally disadvantaged as carbon-related costs rise and performance thresholds tighten. Fleet-level performance transparency is therefore essential. Consistent aggregation of vessel data enables operators to quantify degradation trends, evaluate coating and retrofit effectiveness, forecast regulatory exposure, and support informed capital allocation decisions.

This paper outlines how vessel performance can be systematically aggregated and presented at fleet level to support strategic decision-making processes and explores typical options for performance forecasting.

1.2. Regulatory compliance vs. actual propulsion performance

Regulatory performance indicators such as the Carbon Intensity Indicator (CII) or the Energy Efficiency Existing Ship Index (EEXI) serve important compliance purposes. However, these metrics are not designed to isolate or diagnose the current propulsion performance of a vessel in its current state.

The CII is based on highly aggregated fuel consumption and transport work data, without consideration given for effects like weather conditions, operational speed changes, trading pattern and idle periods; the CII is in many ways the reflection of a trade efficiency rather than a description of the vessel

performance. On the other hand, the EEXI is based on a ship's technical specification at construction verified through a sea trial; this measure focuses on a single operation point at the time when the vessel was built, rather than the diverse operational profile a ship experiences in reality. A good CII or EEXI value does not necessarily mean a good fuel performance of a vessel, *Marioth (2025)*, and therefore, regulatory metrics should not be used to:

- Assess antifouling paint effectiveness
- Quantify degradation trends
- Evaluate retrofit impact
- Reliably compare sister vessels
- Judge the fuel efficiency of vessels at chartering processes

For strategic performance management and capital allocation decisions, a separate layer of performance analytics is essential to isolate propulsion efficiency from operational variability.

1.3. Sample vessels within the case study

To illustrate the concepts introduced in this paper, a case study based on 13 bulk carrier vessels with three years of anonymized operational noon data is used within this study. A brief overview of the fleet composition is provided in Table I.

Table I: Overview of Sample Bulk Carrier Fleet

Group name	No. of vessels	DWT	Year Built	Last DD
Series A	4	165 kt	2016-2018	2022-2023
Series B	4	180 kt	2011-2012	2020-2023
Series C	2	185 kt	2006-2007	2022-2023
Series D	3	170 kt	2010-2014	2024-2025

2. Prerequisites for an overview of vessel performance on a fleet level

To monitor performance at fleet level, shipping companies require structured processes that review the performance of each vessel. Overviews of the fleet performance are based on the results of these vessel level assessments, which may be conducted fully automated or with human in the loop concepts.

2.1 Data and methodology foundations

Reliable performance analysis at fleet level requires consistent and methodically sound evaluations at vessel level. As with any performance review over time, the central question is how to achieve sufficient accuracy for sound decision making. Structured processes for sensor accuracy assessment and data quality control are therefore essential. Before meaningful fleet level insights can be derived, the following prerequisites must be in place:

- Structured handling of sensor/reporting inaccuracies
- Defined data filtering criteria (e.g., weather limits, draft bands, speed ranges)
- Established data quality control processes
- Consistent baseline models across comparable vessel groups
- A standardized methodology for vessel performance evaluation

The objective is to evaluate all vessels on a comparable basis. This can be achieved either through robust concepts for determining evaluation power and evaluation speed or by limiting comparisons to the minimum common measurement equipment installed across the fleet. In practice, most approaches compare observed speed or current corrected observed speed with the reported fuel consumption, as these parameters are generally available across the fleet. The choice of filtering criteria and the

availability of sufficient data within the selected ranges strongly influences the robustness of the results. Once these prerequisites are fulfilled, aggregated fleet indicators can be computed.

2.2. Estimation of the current hull & propeller performance on vessel-level

In order to track hull performance developments of a vessel over time, performance indicators are required which are independent of the weather condition, draught and other environmental conditions. For instance, ISO 19030, *ISO (2016),* uses the Percentage Speed Loss for this purpose. But there are multiple options. An overview of such indicators can be found in *Marioth and Julien (2023)*. In the following the Excess Fuel Consumption Percentage Indicator is used. This value subtracts the estimated fuel consumption related to environmental effects from the observed propulsion fuel consumption and compares it with fuel consumption of a reference model at the same operation conditions:

$$EFC\% = \frac{FC_{Observed} - FC_{WeatherEstimate}}{FC_{Model}(v,T)}$$

EFC%: Excess Fuel Consumption Percentage for propulsion of an observation
$FC_{observed}$: Fuel Consumption observed at a manual or automated data point
$FC_{WeatherEstimate}$: Estimated impacted of weather related aspects
$FC_{Model(v,T)}$: Fuel consumption of reference model at a given draft and speed

Shipping companies that have introduced vessel performance analysis processes can estimate the Excess Fuel Consumption Percentage using defined baseline models and weather correction models. The most recent estimate of the hull performance indicator represents the current propulsion performance status of the vessel. This value can be used, for example, to compute through the baseline model fuel consumption curves or to support short-term operational planning in regard to biofouling management.

In practice and in order to reduce the influence of short-term fluctuations, the current propulsion performance of a vessel is determined via a data trend or a rolling average over a defined recent evaluation period. The achievable accuracy depends on the data availability, data quality and operating conditions. Particular challenges arise when vessels are idle for extended periods. During such phases, reliable performance data under representative sailing conditions is not available, while biofouling can still develop. As a result, the actual propulsion condition may deteriorate without being immediately reflected in measured performance data. The current performance estimate must therefore always be interpreted in the context of operational activity and data coverage.

2.3. Sister class or individual baselines?

Shipping companies that use vessel performance analytics to track performance over time require propulsion performance reference models. It has become common practice to use the Sea Trial recording and build such a reference or baseline model using this information. Sea Trials accuracies have certain limits, *Ponkratov and Strujik (2023)*, but they can still be regarded, as representing the achievable optimum performance of a vessel. Using appropriate estimation approaches, such as those presented in *Marioth and Raj (2021),* feasible base line models can be derived.

A remaining question concerns sister vessel series, where multiple sea trial documents are typically available. At hull and propeller performance related tasks, a sister vessel would have the same propulsion setup. When the propulsion setup changes by the installation of a retrofit that has a strong impact on the performance characteristics, like a bulbous bow modification, the models shall be updated and adjusted for this and not treated as sisters. But for vessels with identical propulsion setups: Should a single baseline model be applied to the entire series, or should each vessel be assigned an individual model using its own Sea Trial? When the focus is strictly on the hull performance development of an individual vessel over time, the use of vessel-specific sea trial data can be considered the most accurate approach. Sister vessels can exhibit notable performance differences *Werner and Gustafsson (2020)*.

However, a number of arguments support the use of sister class baseline models:

- Maintaining and updating a single baseline model per sister class significantly reduces modelling effort and improves transparency and reproducibility across the fleet.
- Excess fuel consumption percentages are comparable between sister vessels and therefore allow a direct comparison of hull condition and performance status.
- Sea trial quality can be heterogeneous across a sister series. Using a sister-class baseline enables the application of the most reliable sea trial data to all vessels in the series.
- In some cases, additional performance-related data, such as trim optimisation results from CFD or model tests, are available. Under these circumstances, baseline models are typically developed to incorporate this supplementary information, making vessel-specific sea trial baselines less relevant.
- A larger combined data basis across sister vessels allows further refinement in terms of hull performance indicator related draft dependency and stabilisation of the baseline models.

Most vessel performance analytics companies now use sister class models to adjust for draught and speed variations due to the arguments above. Using sister class reference models means that speed vs. power curves have the same slope for all vessels of a sister class series. The models may still be calibrated to a certain condition, e.g. out of dock, so that the absolute values are different. It depends on the use case, such as tracking overall fleet performance in overview charts.

3. Fleet performance overviews

Building upon the previously defined performance indicators and baseline concepts, this section presents practical methodologies used in the industry to track hull and propeller degradation over time. These approaches aim to translate technical performance metrics into structured fleet-level insights that support coating evaluation and maintenance planning.

By aggregating vessel-level indicators into a consistent overview, performance deviations become transparent across the fleet. This enables the identification of outliers and underperforming vessels, allowing targeted corrective actions and more effective allocation of maintenance resources.

3.1. Fleet performance overview referenced to dry docking

The time since dry docking has typically a very high impact on the hull & propeller performance development of a vessel. This effect is particularly pronounced for vessel types with a high fouling risk, such as tankers and bulk carriers, which may remain idle for extended periods in tropical ports.

Once the current performance indicators of each vessel have been determined, they can be plotted against the time since the last dry docking. An example of this plot is shown in Fig.1 for the test group. Each data point represents the current performance state of an individual vessel as compared to its baseline, while the trendline illustrates the overall development across the fleet.

The main advantage of this visualization is that it provides an immediate overview of the biofouling status of the entire fleet. The slope of the fleet performance trendline can be interpreted as an indicator of the average degradation rate and may serve as a reference for fleet-level hull performance development. A flatter slope may indicate a more effective antifouling strategy. However, interpretation requires caution as the observed slope reflects the combined influence of antifouling coating characteristics, surface preparation quality, cleaning frequencies, time since last cleaning, operational profile and environmental exposure.

In addition, the view may give a rather absolute impression, but both estimations of current vessel performance and underlying baseline models are subject to uncertainty. These uncertainties are not represented in the visualization and should therefore be considered when drawing conclusions.

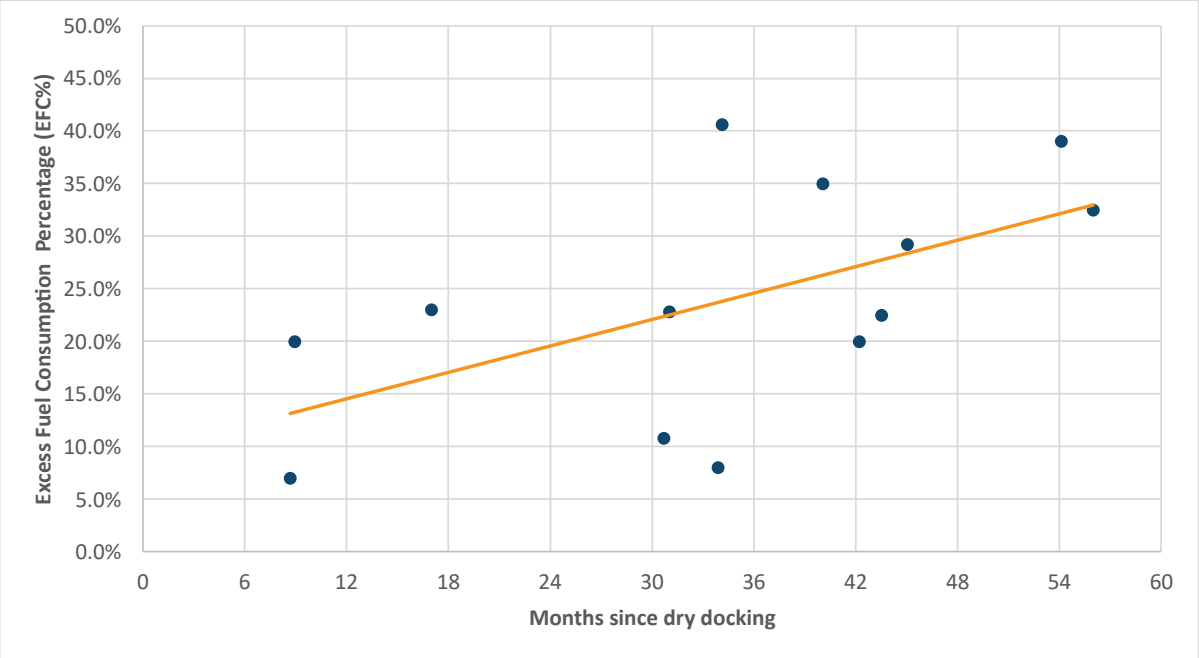


Fig. 1: Fleet performance overview of sample group

3.2. Fleet level overview by out of dock reference and degradation trend

Coating manufacturers may offer paint performance guarantees. These provide a contractual framework that links the expected in-service performance of an antifouling coating to a defined operational profile and application conditions over a specified dry-docking interval. In essence, they formalise what level of hull condition and performance degradation can be expected, and under which circumstances the coating supplier may be held responsible.

Paint guarantees specify defined performance limits that must not be exceeded within a contractual period following dry docking. These limits are commonly formulated as maximum allowable speed loss at constant power or maximum power increase at constant speed. The reference condition is typically the post dry docking performance, so a baseline model as described in chapter 2.3 calibrated to the performance data after the docking.

Considering these principles paint performance development at fleet level can be analysed via two complementary views. The first view focuses on a defined point in time after dry docking. Within paint guarantees, this reference is typically twelve months in service. For each vessel, the relevant performance indicator at this defined age since docking is determined and compared across the fleet. This approach enables a structured comparison under similar exposure durations. By fixing the time since application, short-term operational fluctuations are reduced and differences between coating systems become more visible.

When comparing the performance between the vessels, it must be considered that the average performance indicators do not represent the coating effect alone. Provided that the baseline model remained the same and was not modified due to a major retrofit installation, the measured indicator reflects the combined influence of the antifouling system, installed smaller energy saving devices (PBCF or similar), changes of structural and surface condition since delivery, as well as differences in coating specification and application quality. The comparison therefore captures the integrated outcome of all technical measures implemented during docking and the accumulated degradation state of the vessel over a dry-docking cycle.

Fig.2 presents a bar chart of the results based on this approach. Each bar reflects the performance indicator of a vessel at the selected reference age. In addition, the chart includes the corresponding

value from the previous dry docking or, where applicable, from the initial baseline after delivery. So, this view allows a direct comparison between consecutive docking intervals and across the fleets. It provides transparency on recent docking performance, long-term degradation trends, and the effectiveness of retrofit measures.

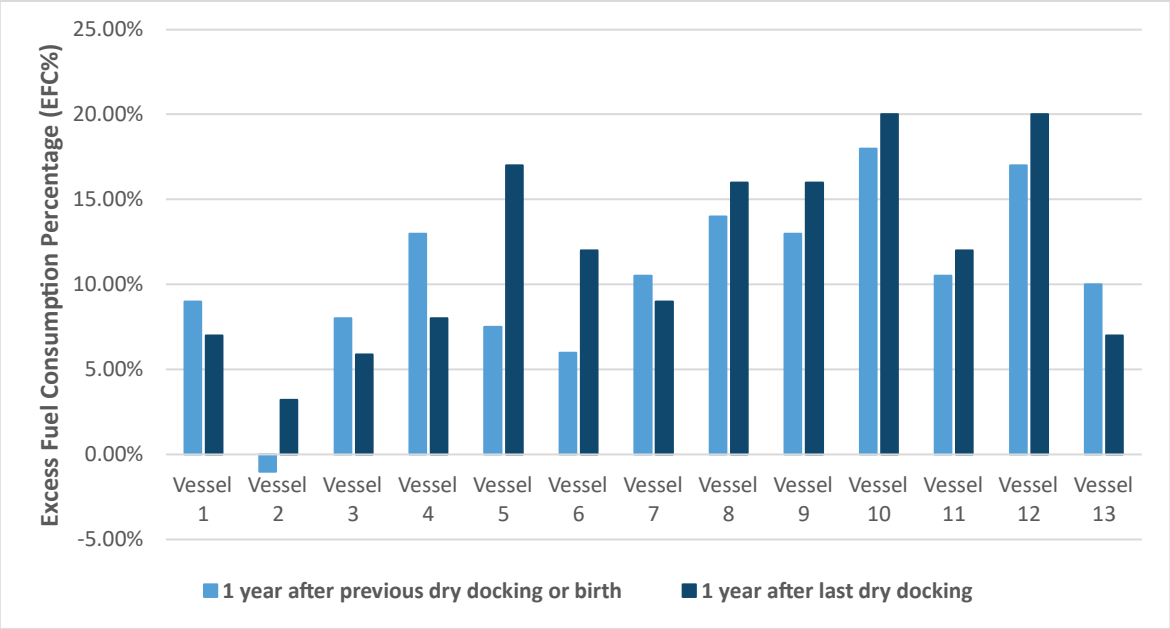


Fig.2: Average Excess Fuel Consumption Percentage of sample group, 1 year after docking

The second view in Fig.3 addresses the performance development over time using baselines that are calibrated with data after the dry docking. In this second view the current performance indicator is plotted against the time elapsed since dry docking and compared with the average degradation trend of the entire fleet. The average paint degradation of the fleet can be understood as the average fouling curve.

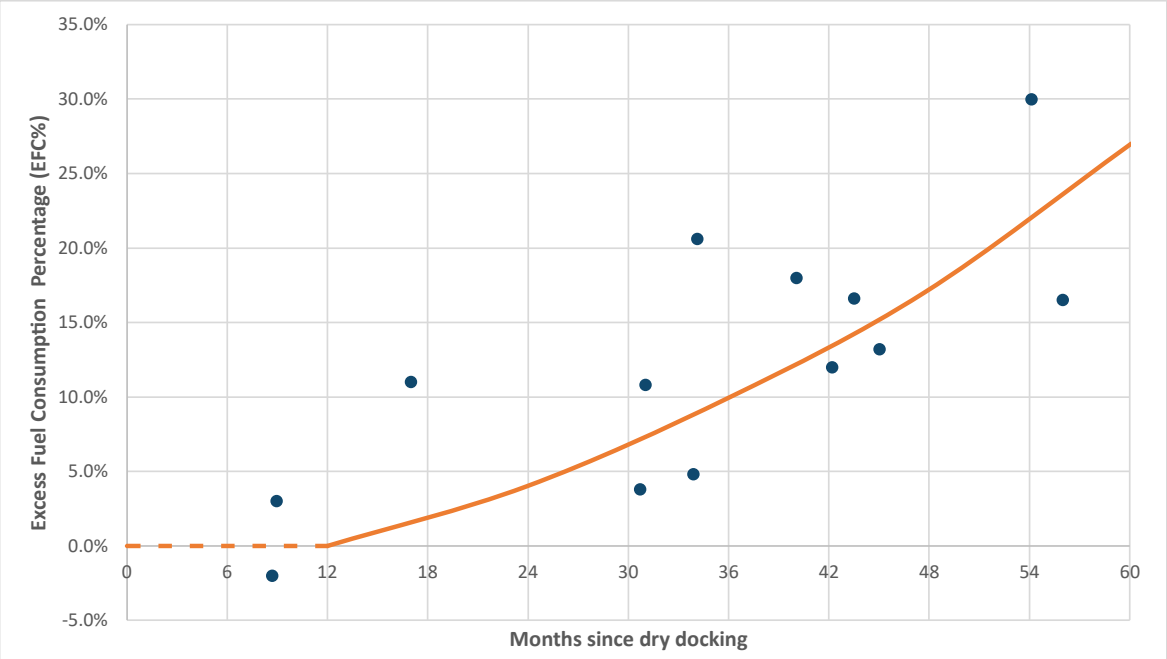


Fig.3: Excess Fuel Percentage development using out-of-dock performance as computational reference

This time series representation enables an assessment of the degradation rate and supports the identification of vessels that deviate from the expected average. When fleets do have significant sizes and performance evaluations are working well, then one can create separate views to assess different paint types and create separate expected paint degradation models.

The degradation effect is captured by this method, Fig.3, more accurately than in the approach described in chapter 3.1, Fig.1, as the baselines are calibrated to the post docking performance. This methodology helps to distinguish the effects of dry docking and retrofits from antifouling management. However, fouling pressure, idling time, and hull cleaning events differ from vessel to vessel, so this approach, while useful for more detailed reviews or for developing paint-performance models, still has its limitation. Even so, estimating average fouling rates remains important for tracking fleet performance management efforts and related monitoring concepts, as outlined next.

4. Tracking the fleet performance development over time

4.1. Estimating the reference conditions

By combining the described illustration and processes, shipping companies can monitor the overall performance development of their fleets over time. The central challenge in this context is the definition of an appropriate reference condition. This reference must reflect the expected performance level of each vessel at a given point in time, considering coating degradation, aging effects, and operational exposure. A physically consistent representation of the reference fuel consumption accounts for these multiple degradation mechanisms acting simultaneously. In simplified form, this can be expressed as:

$$Fc_{Ref} = Fc_{Optimum} \cdot (1 + f_{wT} \cdot t_{SinceBirth} + f_P \cdot t_{SinceDD} + f_{FP} \cdot (f_{FRP} \cdot t_{SinceDD} + f_{CE} \cdot t_{SinceHC}))$$

FC_{Ref}: Reference fuel consumption used to determine excess consumption
$FC_{Optimum}$: Fuel consumption of validated baseline model (chapter 2.3)
f_{wT}: function of resistance effects due to structural aging and wear and tear
$t_{SinceBirth}$: Age of the vessel
f_p: Expected paint performance changes over time regardless of fouling exposure
$t_{SinceDD}$: Time elapsed since the last dry-docking
f_{FP}: Fouling pressure at the vessel, due to operation and regional characteristics
f_{RPP}: Fouling resilience capability of the applied paint
f_{CE}: Expected effectiveness of cleanings
$t_{SinceHC}$: Time elapsed since the last hull cleaning

This formulation separates structural aging from fouling-related degradation. The first two terms capture gradual and largely irreversible effects such as hull roughness increase due to coating aging, corrosion, and minor structural deformations. The second term represents the time-dependent accumulation of biofouling, which is influenced by trading patterns, water temperature, idle periods, and the effectiveness of the applied antifouling system. Maintenance events reset or partially reset this component. However, some parameters are interdependent. Frequent hull cleanings for instance impact the paint performance, in particular when the hull cleaning method does not sufficiently consider characteristics of the applied paint system. This can reduce the fouling resilience of the coating and thereby influence subsequent degradation rates.

Although this formulation reflects the underlying physics well, its practical implementation requires reliable quantification of fouling pressure, coating-specific performance parameters and cleaning effectiveness. Some of these inputs are no easily quantifiable; For this reason, many operators apply a simplified approximation that captures the dominant effect, which is the time elapsed since the last dry docking. Vessels are usually dry docked at approximately five-year intervals, typically including a full renewal of the underwater hull coating. Assuming that coating degradation represents the primary driver of performance deterioration within a docking cycle, the reference fuel consumption can be approximated using a linear trend based on time since dry docking:

$$Fc_{Ref} = Fc_{SeaTrial} \cdot (1 + m_{Paint} \cdot t_{SinceDD})$$

m_{Paint}: Slop of the linear trendline
$t_{SinceDD}$: Months since last dry-docking

The slope m_{Paint} can be derived empirically from fleet-level performance development, for example by evaluating the aggregate trend of excess fuel consumption percentage versus time since docking, so basically the slope of the entire fleet over time, Fig.1. This approach effectively condenses multiple degradation mechanisms into a single observable parameter. Note that this linear approximation remains a strong simplification and where larger datasets or segmented sub-fleets are available, non-linear or segmented degradation models may provide improved accuracy of the reference condition.

4.2. Overview graphs and results

When one has determined the reference fuel consumption by an approach described above and the current fuel consumption of the vessel excluding weather impacts, then one can compute the total fleet excess fuel consumption in the following way:

$$EFC\%_{Fleet} = \frac{\sum_{i=1}^{n} FC_{Calm_i}(EFC\%_i)}{\sum_{i=1}^{n} FC_{Ref_i}}$$

$EFC\%_{Fleet}$: Average Excess Fuel Consumption over defined reference condition (i.e. speed and draft)
FC_{Calm}: Current calm-water fuel consumption of each vessel as function of its EFC% at reference condition
FC_{Ref}: Expected reference fuelconsumption of each vessel (see 4.1)

This value can be computed over time to track and monitor the effectiveness of the hull & propeller performance maintenance efforts. A sample of what one can derive by this over time is shown in Fig.4. Keeping the overall accuracy of the calculations in mind one can use these calculations to estimate achieved fuel savings by hull performance management initiatives.

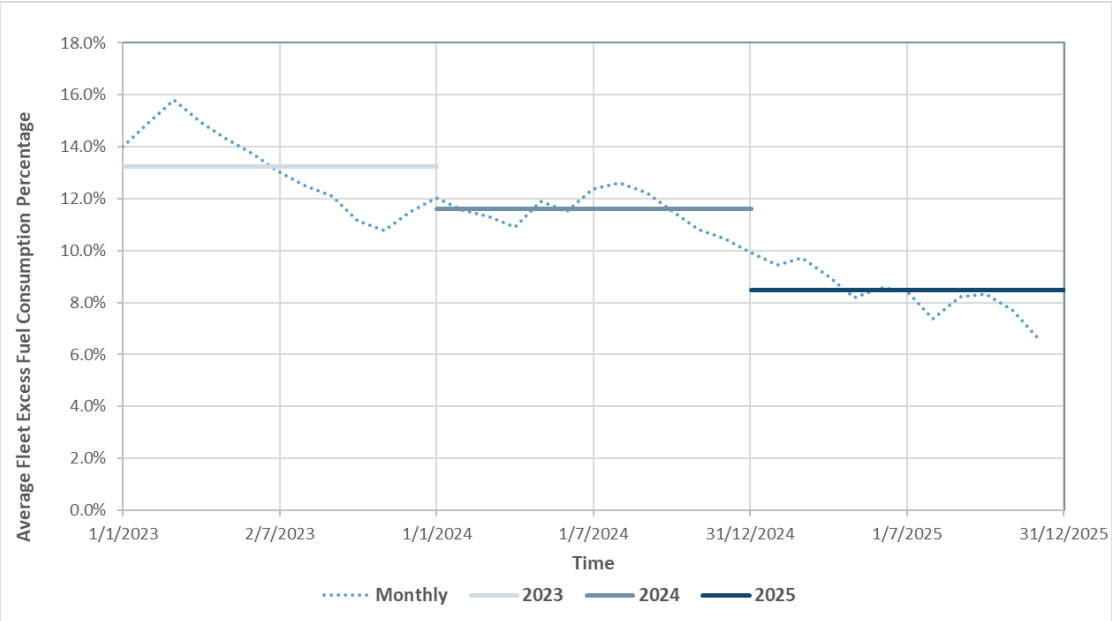


Fig.4: Average Excess Fuel Percentage development of managed fleet over time

Tracking performance of the past and demonstrating success is an important aspect of optimization processes. Another important aspect is to use this knowledge for forecasting and future predictions. Options for this are explored in the next section.

5. Forecasting

5.1. Forecasting principles

Performance forecasting in fleet management must be embedded in a clearly defined decision context. The operational or strategic question to be answered determines the relevant forecast horizon, the tolerable uncertainty, and the level of modelling detail required. Without this alignment, projections may lack practical relevance, either by oversimplifying complex degradation effects or by introducing unnecessary analytical complexity.

Within the scope of this paper, the objective is to forecast the future development of a hull and propeller performance indicator. The forecasted performance indicator can then be utilized to determine a speed vs. draft vs. propulsion fuel consumption table. Such tables can be used as an input of future fleet operation scenario computations.

In practice, two distinct time horizons can be differentiated. Short-term projections rely on recent performance trends to support operational decisions such as cleaning scheduling and bunker planning. Long-term projections estimate expected degradation based on current condition, maintenance history, and representative fleet behavior, thereby providing a basis for budgeting and strategic planning. In the following more details and possible solutions are described for both tasks.

5.2. Short-term forecasts through linear models

In the short-term horizon, vessel performance development can be approximated by extending the linear trend observed in recent performance data. This approach assumes that the current degradation pattern continues over a limited time interval and therefore relies strongly on high quality and sufficiently dense operational data. An example of this method is shown in Fig.5.

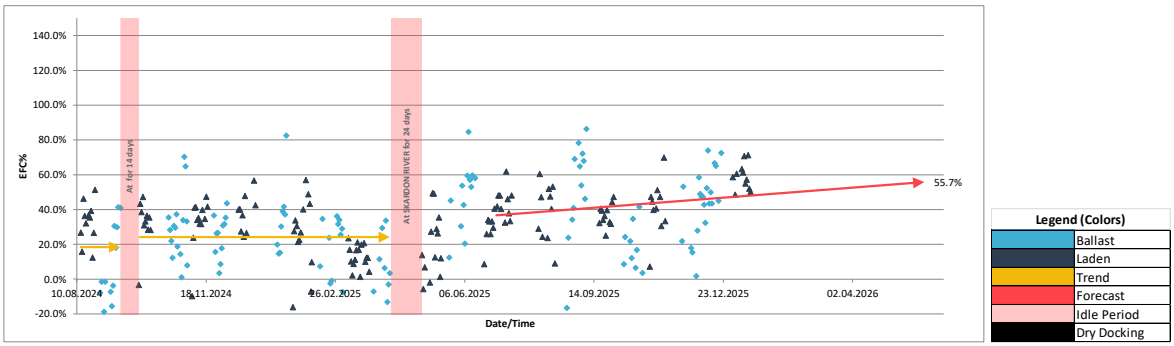


Fig.5: Hull Performance over time – Forecasting capability via a linear trendline in red

The linear forecast model assumes that the recently observed rate of performance change continues over the selected forecasting periods. In reality, hull and propeller performance are influenced by interacting factors such as biofouling dynamics, data quality and model accuracy. The resulting performance development may therefore deviate from a purely linear pattern.

Particular attention must be given to the slope of the trendline. Steep gradients do not necessarily indicate accelerated physical deterioration; they may also reflect measurement inaccuracies, inconsistent filtering, or limited data density. In addition, biofouling can develop rapidly under favorable conditions, leading to nonlinear effects that are not captured by such a linear forecast model. Consequently, linear extrapolation is suitable primarily for short-term application. In practice, a projection horizon of up to approximately three months can provide reasonable directional guidance, while uncertainty increases significantly beyond this range.

Within these limits, the method can support operational decision-making, for example in bunker planning or in assessing the urgency of hull cleaning. On a fleet level, aggregated short-term forecasts can provide an indication of the expected overall performance position in the coming months, while acknowledging the underlying uncertainties. An example of how the forecast can be used in the fleet performance overview chart, which was described earlier, is shown in Fig.6.

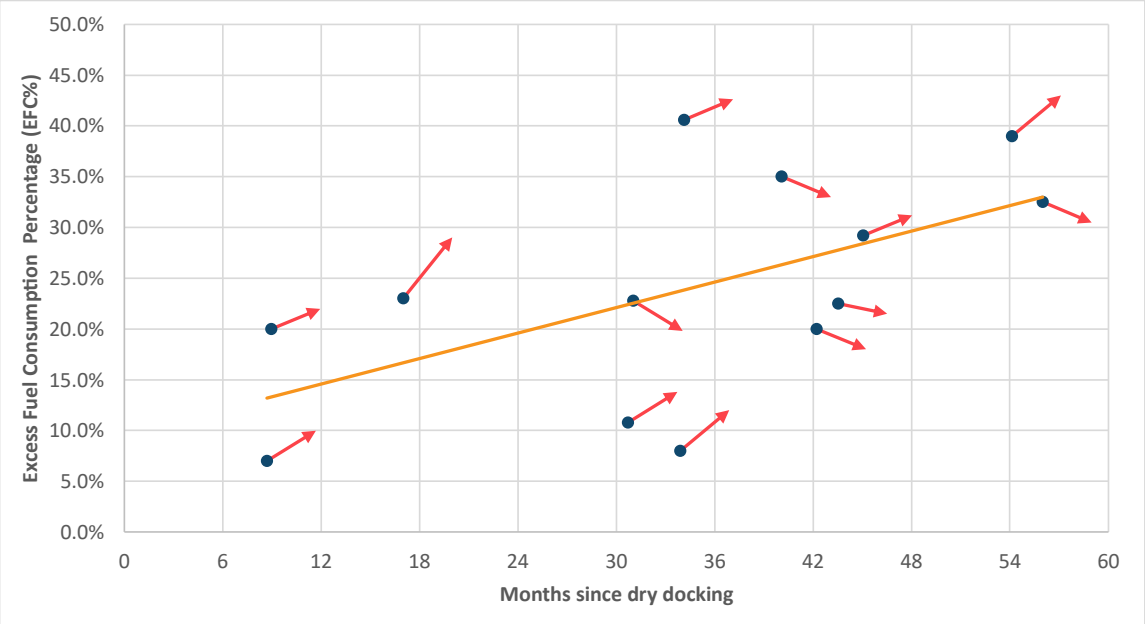


Fig.6: Fleet performance overview of sample group & Forecasting via linear trendlines in red

5.3. Long-term forecasting (strategic horizon)

Long-term projections serve a different purpose. They do not aim to predict the exact vessel level performance several years into the future, as current operational data provides limited predictive power at that horizon. Instead, long-term forecasting focuses on estimating expected fleet-level degradation behavior under defined operational and maintenance assumptions.

Rather than extending recent short-term regression slopes indefinitely, the projection gradually converges toward an expected degradation path derived from statistically stable fleet-average behavior. Where sufficient segmentation is possible, fleet-average fouling curves should be developed separately for comparable subgroups, such as vessels with similar coating types, surface preparation standards, trading patterns, or vessel classes. This reduces non-uniformity and improves forecast robustness.

The future Excess Fuel Consumption Percentage can be derived by assuming that the vessel performance will follow certain fouling curve characteristics. The forecasted expected fouling curve of each vessel provides a structured reference that helps position and calibrate each individual vessel projection. Fig.7 shows how such a projection could be made for a single vessel of the case study group. The long-term projection (red) anchors at the current state but follows the fleet fouling curve going forward, gradually converging toward the fleet average curve as the forecast horizon extends. This reflects the structural expectation that individual anomalies normalize over time.

When a dry docking has occurred within the forecasting horizon period of an individual vessel, then the docking effect would need to be estimated and the vessel-level projection would follow a new forecasted fouling curve, which can be either coating specific or based on the fleet average fouling curve. Similarly, effects on propulsion performance related to retrofits can be added as a decrease of the Excess Fuel Consumption Percentage in the future, provided that they change primarily the propulsion fuel curve magnitude and not its slope.

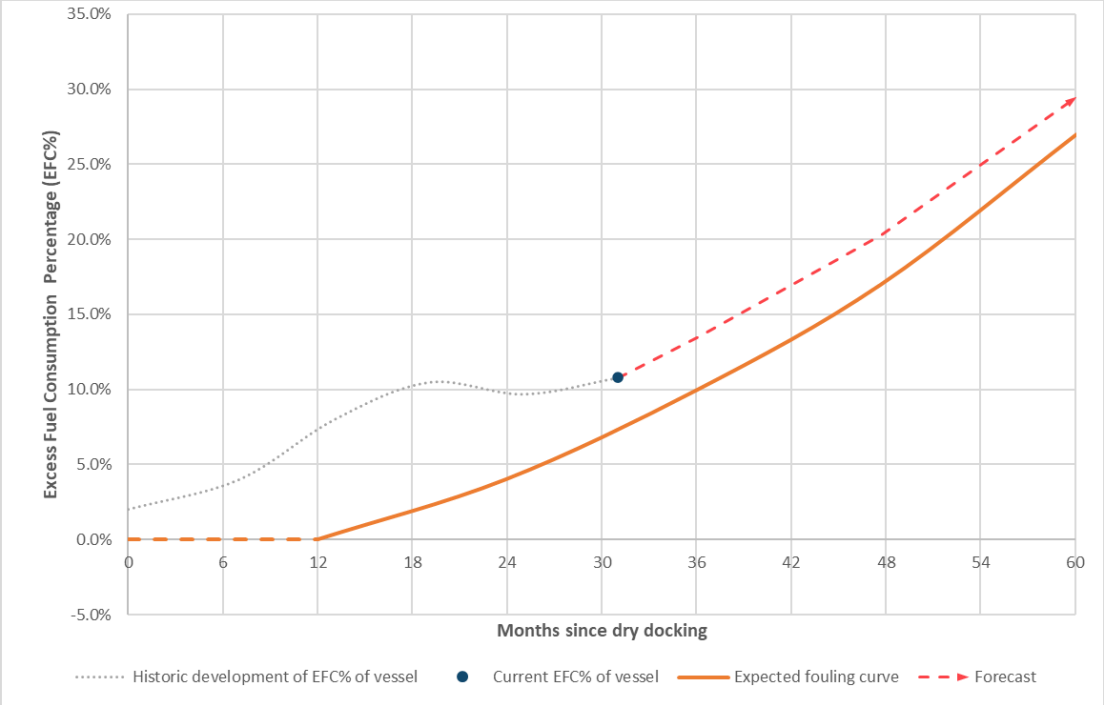


Fig.7: Expected development of the Excess Fuel Consumption % for long-term forecasting

These projections provide essential input for strategic planning, including estimation of future fleet fuel consumption, regulatory compliance exposure, and the evaluation of maintenance and retrofit strategies. While individual vessel forecasts remain uncertain over long horizons, fleet-level projections enable robust scenario analysis and support informed operational and investment decisions.

6. Summary

This paper has presented practical methodologies for analyzing and forecasting hull and propeller performance at fleet-level. The aggregation of individual vessel performance developments into structured fleet-level indicators provides a quantitative foundation for strategic performance management. While the focus of this work lies on data-driven assessment and forecasting, vessel-level strategic decisions should additionally consider trading pattern and exposure profile, as these may influence coating durability and performance dynamics.

By combining consistent vessel-level hull and propeller performance assessment with structured aggregation methods, shipping companies can establish transparent fleet performance overviews. These allow continuous monitoring of overall degradation trends, systematic comparison of vessel condition across the fleet, and quantitative evaluation of maintenance measures such as hull cleaning, coating renewal, and technical retrofits.

Using these analytical foundations, forecasting approaches were introduced to support both operational and strategic decision-making. Short-term projections provide guidance for operational measures such as cleaning scheduling and bunker planning, while long-term projections enable estimation of future fleet fuel consumption and expected degradation behaviour under defined assumptions. By comparing projected and realized fleet performance shipping companies can assess whether implemented strategies are effectively shifting the fleet-wide degradation trajectory downward.

When applied consistently, hull performance indicator based fleet overview and forecasting methods provide a structured quantitative basis for performance management. They support regulatory compliance planning and enhance the robustness of fuel consumption and emissions forecasts at strategic fleet level.

References

ISO (2016), *ISO 19030 - Ship and marine technology – Measurement of changes in hull and propeller performance*, Int. Stand. Org., Geneva

MARIOTH, R.; JULIEN S. (2023), *Savings as Expected? - Guidance Towards a Successful Benefit Tracking Process of Propulsion and Hull Improvements*, HullPIC Conf., Pontignano, pp.44–54

MARIOTH, R.; RAJ P. (2021*), Improving true lies – Creating sophisticated baselines out of woefully little*, HullPIC Conf., Pontignano, pp.161–171

MARIOTH, R. (2025), *Assessing and Rating the Performance when Chartering Unknown Bulk Carrier Vessels*, HullPIC Conf., Mulheim, pp.150–162

PONKRATOV, D.; STRUIJK, G.D. (2023), *Quantifying the Uncertainty of High-Fidelity Speed/ Power Trials*, HullPIC Conf., Pontignano, pp.131–138

WERNER, S.; GUSTAFFSON L. (2020), *Uncertainty of Speed Trials*, HullPIC Conf., Hamburg, pp.28–36

Strategies for Full-Scale Measurement of Fuel Savings from Wind-Assisted Propulsion

Sofia Werner, RISE SSPA Maritime Centre, Sweden, sofia.werner@ri.se

Abstract

Accurately measuring fuel savings from wind-assisted propulsion presents a technical and operational challenge. Nevertheless, such measurements are essential for assessing commercial viability, supporting investment decisions. Moreover, it can be used for verifying regulatory metrics such as the reward factor in FuelEU Maritime. Future IMO regulations, including the anticipated Greenhouse Gas Fuel Standard (GFS), are also expected to rely on full-scale data for compliance verification. This paper presents and evaluates strategies for full-scale measurement of wind-assisted propulsion performance. Approaches include short-term and long-term ON-OFF testing campaigns, as well as direct force measurements on wind propulsion devices. These methodologies are demonstrated on a vessel currently in operation, providing practical insights into their implementation and reliability. The findings aim to support both industry stakeholders and regulatory bodies in the development of robust assessment frameworks for wind propulsion technologies.

1. Introduction

Interest in wind assisted propulsion has grown as the shipping sector seeks practical methods to reduce fuel consumption and limit greenhouse gas emissions. By early 2026, there are nearly 100 commercial vessels with wind-assisted propulsion in operation. As the industry matures, the demand for dependable full scale performance assessments has increased. Ship owners, charterers and technology developers need verified fuel savings to support investment decisions and to evaluate commercial potential. Regulatory frameworks also require transparent and consistent methods that can be applied across different vessel types. Future IMO regulation, including the expected Greenhouse Gas Fuel Standard (GFS), is anticipated to rely on real-life data for compliance verification. Such methods are also required for cost saving sharing and for financial schemes that use pay as you save models.

Reliable measurement of fuel savings from wind propulsion remains a significant challenge. Conventional performance monitoring methods may not always be well suited for this purpose. Retrofit installations are usually carried out together with other maintenance activities, which means there is no period where the ship operates under identical conditions before and after the installation. For new builds, a reference period does not exist at all. In addition, the effect of wind propulsion depends strongly on both wind speed and wind direction, and most vessels rarely encounter matching weather conditions that allow direct comparison with earlier voyages. The requirement for fair comparison using conventional performance monitoring or fuel logging can provide indications of savings in *some* cases, but the approach does not deliver reliable results for all ships. For regulatory use, methods are needed that work consistently across different vessel types and operating patterns. This paper therefore proposes alternative approaches that are better suited to isolating the specific contribution of wind propulsion technology.

The paper presents methods that can be used both for verification of wind propulsion savings (Section 3) and for long term logging of the savings during regular operation (Section 4). The work builds on the study presented at last year's HullPIC conference, *Werner (2025),* and extends it with further developments and practical implementation experience. The previous study outlined general approaches for evaluating wind propulsion performance. The present work advances this by demonstrating the methods on real operational data from a vessel currently in service. Two approached are discussed: ON-OFF testing campaigns and direct force measurements on the wind propulsion device. The paper introduces an acceptance criterion and provides a discussion of uncertainty and its influence on the assessment.

The aim is to support the development of robust and practical procedures for full scale performance evaluation. The results are intended to assist both industry stakeholders and regulatory bodies as wind assistance becomes an integrated part of commercial shipping.

2. Case description – vessel and full-scale data

The methods presented in this paper are demonstrated using data from a full-scale sea trial campaign carried out on a large full block vessel fitted with Norsepower rotor sails. For proprietary reasons, no further information about the vessel can be disclosed.

The campaign consists of ON-OFF runs, i.e. short consecutive runs with the wind propulsion devices turned on and off, collected over a period of five months in regular commercial operation. The trials cover a service speed close to eleven knots and include several draught conditions. The runs were around 20 minutes long and with sufficient time between to reach stable condition. Data was recorded through the ship's standard logging system. The signals include wind measurements, ship speed, shaft power, shaft speed and global navigation satellite system position.

The Norsepower rotor sails were also instrumented with force measurement in each device, providing both longitudinal and transversal force components. These measurements are based on surface pressure readings obtained from pressure transducers mounted on the rotor sails. See *Voutilainen and Paakkari (2025)* for further details.

The ON-OFF tests were conducted by the crew following detailed instructions by RISE. The high-frequency data was transmitted to RISE via electronic files extracted from the ship's performance monitoring system, and from the rotor force collection system.

3. Method for verifying wind propulsion power saving

This work applies a methodology where estimation of fuel saving from wind propulsion is divided into two phases. The first phase is the determination of a performance model that expresses power saving for a matrix of wind speed, wind angles and ship speeds. The second phase combines this power saving matrix with wind statistics for a given route. The process to "verify wind propulsion performance" refers to verification of the first phase, the power saving matrix, through dedicated sea trials. Once verified, the power saving matrix can be used to extrapolate the saving to any operational route using suitable wind statistics. This approach was first proposed by *Werner (2021)* and was adopted as recommended procedure by ITTC in 2024. Similar methods are also adopted by other organisations, see for example *DNV* (*2025)* and *Anemoi (2025).*

Sea trials for confirming wind propulsion savings is, just as conventional trials, consisting of a series of short runs. The key difference is that the outcome is not the absolute value of the speed power curve, but the reduction in required power due to the wind propulsion system. The effect is determined by comparing the measured speed and power of paired runs with and without the wind propulsion system engaged under similar wind conditions, commonly known as ON-OFF tests. The resulting speed difference is then translated into a power difference using the speed power curve. The measured signals correspond to those used in a conventional speed power trial, including ship speed, power, wind speed and wind direction. Further procedural details can be found in the ITTC recommended procedure for sea trial of wind assisted ships, *ITTC (2024:2)*.

<u>Bias Uncertainties</u>

After publication of the ITTC procedures for sea trial of wind assisted ships in 2024, additional studies have examined the uncertainties associated with wind propulsion sea trials. For bias uncertainties, the dominating source is, not surprisingly, the measured wind, as noted by *Werner (2026).* This is mainly caused by airflow disturbance from the ship above water rather than limitations of the instrument itself. Such disturbances influence both wind speed and wind angle relative to free

stream conditions. CFD studies conducted at RISE indicate that disturbances at the anemometer location can lead to variations between zero and fifteen percent in measured wind speed, depending on the direction of the incoming wind, and deviations of about plus or minus five degrees in wind angle. These findings agree with published literature, for example *Ponkratov and Struijk (2025).*

The consequence of bias uncertainty in measured wind for the final calculated saving is substantial. Accuracy can be improved by applying corrections to account for the disturbance around the hull. Three approaches are available: far field wind measurement using Lidar, the use of CFD, and finally comparison with meteorological hind-cast data. The case study presented in this paper uses a combination of CFD and hind-cast data.

Precision uncertainty

Precision uncertainty can be assessed in various ways. In scientific publications it can for instance be estimated by determining the precision of each measured signal and then applying a Monte Carlo simulation. In practice, for full scale experiments, this approach often underestimates the total precision uncertainty. For such cases, then "Type B uncertainty assessment", *ITTC (2008),* is more useful. It is based on "previous experience, calibration records or established best practice". A hands-on example of this approach was described by *Ruth et.al (2024).*

In this work we illustrate the precision using yet another variant of Type B uncertainty estimate. In an ideal experimental setup, measurements would be repeated under identical conditions, and the uncertainty could be derived directly from the statistical spread of the repeated point measurements. In wind propulsion sea trials, the environmental conditions vary from run to run. Since identical wind conditions cannot be reproduced, the conventional method of estimating precision uncertainty from repeated measurements is not directly applicable. One way to address this is to derive a "power-saving coefficient":

$$C_{\Delta P} = \frac{\Delta P}{\frac{1}{2}\rho_a \cdot A \cdot AWS^2 \cdot V} \tag{1}$$

ΔP is the power saving due to the wind propulsion, A is wind propulsion system's projected area, V is ships speed, AWS is the apparent wind speed.

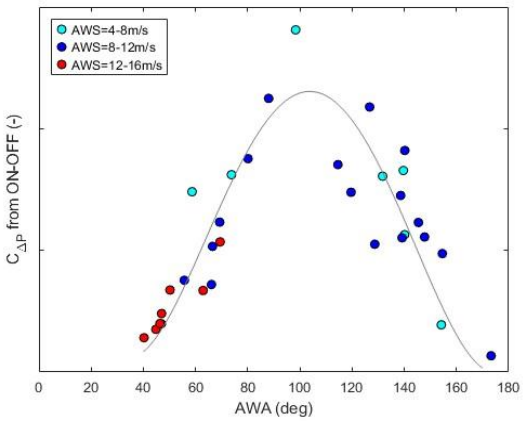


Fig.1: Power saving coefficients derived from ON-OFF tests

Fig.1 shows the result for the current test case. Results of this kind provide information on the level of precision that can be expected and, in turn, the number of runs that are required to achieve a desired level of precision uncertainty in the averaged result. The ITTC guideline specifies a minimum test program that includes five wind angles. Based on the present results, as well as several other test campaigns conducted at RISE, it is evident that using only five runs would lead to a large overall uncertainty. We therefore recommend that the test program be extended to include a wider range of conditions collected over a longer period of operation.

Verification acceptance criteria

The purpose of the sea trial is to verify and, if needed, to correct the calculated Power Saving Matrix. For this reason, a clear indicator is required to determine when validation can be considered acceptable. Currently, no such indicator exists in any guideline or published method. The following proposal is presented as an initial concept, intended as a starting point for further discussion rather than a fully matured process.

We propose to define a comparison error $\bar{E}$ representing the error of the power saving on a route, by calculating the weighted average error calculated over the tested points *1…k:*

$$\bar{E} = \left(\sum_{k=1}^{N} \Delta P_{pred_k} \cdot w_k - \sum_{k=1}^{N} \Delta P_{st_k} \cdot w_k \right) / \sum_{k=1}^{N} w_k \quad (2)$$

expressed in percentage of $\overline{\Delta P} = \sum_{k=1}^{N} \Delta P_{pred_k} \cdot w_k \,/\, \sum_{k=1}^{N} w_k$

ΔP_{st_k} is the derived ΔP from the sea trial ON-OFF run k
ΔP_{pred_k} is the ΔP from the performance prediction, interpolated to the same condition as ΔP_{st_k}
w_k is the corresponding weight in a weather probability matrix, for example the GWM used for EEDI (*IMO, 2021*).

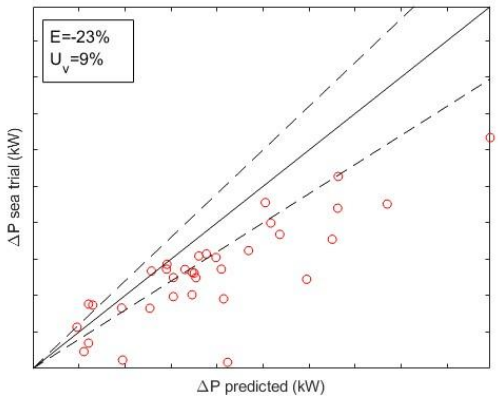
To prevent the average being skewed, distribution of points must be fairly even over the wind matrix.

In analogy with the ITTC procedure regarding CFD validation (*ITTC, 2024:2*), as an indicator of the correctness of the predicted we define the following criteria:

Given the required validation tolerance U_{req} and validation uncertainty U_V :

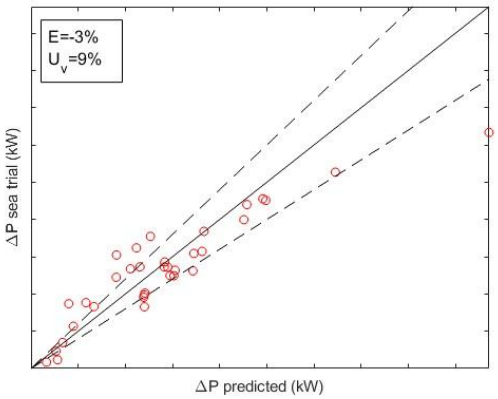
- If $|E| < U_V < U_{req}$, validation is successful at U_V level
- If $U_V < |E| < U_{req}$, validation is successful at $|E|$ level

The estimated uncertainty of the sea trial and the simulations, U_V must hence be estimated for each case. An example of the acceptance criteria on real trial data is shown in Fig.2. The agreement is initially very poor. After more detailed modelling, a significantly improved comparison is obtained, as shown in Fig.2(b). The comparison error, as defined in Eq.(2), is satisfactory.



$|\bar{E}| > U_V$,
Validation is achieved to 23%
Correction of the model can be made

Fig.2(a): ΔP, derived from sea trial and predicted with initial prediction model



$|\bar{E}| < U_V$
Validation is achieved to validation tolerance 9%

Fig.2(b): ΔP, derived from sea trial and predicted with improved prediction model

4. Methods for continuous logging of wind propulsion savings

The previous Section discussed verification of wind propulsion performance. The present section will instead deal with quantification of the saving from wind propulsion during operation. As outlined in the introduction, the purpose here is to identify methods that can quantify the effect of wind propulsion more directly than by tracking accumulated fuel consumption or total delivered power. Two approaches are proposed:

1. "Wind logging"
2. "Force logging"

Both approaches request some degree of computational modelling to complement the continuous measurements, but "Force logging" method requires less modelling than "Wind logging". In addition, both require a limited number of sea trial runs to confirm and adjust the modelling so that the logged data can be interpreted with sufficient accuracy.

4.1. "Wind logging" method

This method was developed at RISE to address the industry's demand for a cost-effective and robust process for saving-splitting, *Werner (2025)*. The same methodology can be used for other purpose that requires an assessment of continuous power saving from wind propulsion. The procedure involves two phases:

1. Establish power saving matrix: verification/calibration of a power saving model
2. Logging of wind during operation and estimate accumulated saving

Phase 1. Establishing the power saving matrix

The first step is to predict the power saving provided by the wind propulsion system for a range of wind conditions, ship speeds and draughts. In this context, saving refers to the difference in required power between the ship operating with wind assistance and the same ship operating without wind assistance, while maintaining identical ship speed.

The predicted power saving matrix is then verified through sea trial runs, as described in Section 3. If the acceptance criteria are not met, the simulation model must be adjusted and the verification repeated until the predicted and measured performance are in sufficient agreement.

Phase 2. Operation, data logging and saving estimation

During operation, the following signal are logged:

- Position and time fore retrieving hind-cast wind *OR* anemometer readings, if reliable
- Ship speed and draught
- Operational state of the wind propulsion system (ON or OFF state)

From the logged positions and time stamps, hind cast weather data can be retrieved from an agreed meteorological source, such as ECMWF reanalysis, for the periods when the wind propulsion system was engaged. Alternatively, wind conditions can be taken from the ship anemometer, provided the measurements have been corrected for hull induced disturbance.

The estimated saving for the operational period is then obtained by combining the verified power saving matrix with the logged wind conditions. This allows a continuous estimate of the effect of the wind propulsion system throughout the voyage.

2.3. "Force logging" method

Several providers of wind propulsion systems now offer equipment that measures the forces generated by their devices. This is a relatively new development area with limited published information on measurement accuracy and long-term reliability. However, the trend is clear, and wider adoption can be expected in the coming years.

A challenge when using force measurements is that the measured wind propulsion thrust must be converted into an equivalent propulsion power saving (ΔP). This conversion is affected by the propulsive efficiency of the ship as well as the added resistance due to drift and increased rudder angle induced by the side forces from the sails. At present, no published guidelines define how this conversion should be carried out. The approach below is suggested by RISE and demonstrated using full scale trial data.

The proposed procedure consists of four phases:

1. Predict the power saving using a Performance Prediction Program. Based on the simulation results, construct a conversion table that relates ship speed, wind propulsion thrust force (Fx), wind propulsion side force (Fy) and the resulting power saving (ΔP).
2. Verify the conversion table through sea trials. Conduct a number of ON-OFF tests following the ITTC procedure and compare the measured force signals with the predicted power saving. Adjust the underlying model if discrepancies remain.
3. Log sail force during operation. Continuously record ship speed, thrust force and side force from the wind propulsion devices. After filtering and averaging over a suitable interval, convert the measured forces to power saving using the verified conversion table.
4. Accumulate the converted power saving to derive the energy saving over time. This yields a continuous estimate of the contribution of the wind propulsion system.

With this procedure, the effects of side force and propulsive efficiency are verified once and then assumed unchanged during service. This does not imply that they are constant, but rather that their dependence on wind conditions remains similar to that established during verification. *Voutilainen and Paakkari* (2025) proposed a different approach, where changes in propulsive efficiency during operation, for example due to hull fouling, are included. Their method does not require speed measurements but does require a shaft thrust meter. This requirement limits its applicability to ships equipped with such instrumentation, but it offers increased confidence for those vessels that can use it.

Here follows a demonstration of the method for the ship case described in Section 2.

Phase 1: Construction of conversion table. RISE in-house Power Prediction Program was used to model the relation between rotor force and power saving. The Fig.3 shows the response surface, i.e. the conversion table, for one ship speed. This illustrates that the surface is regular and suitable for interpolation.

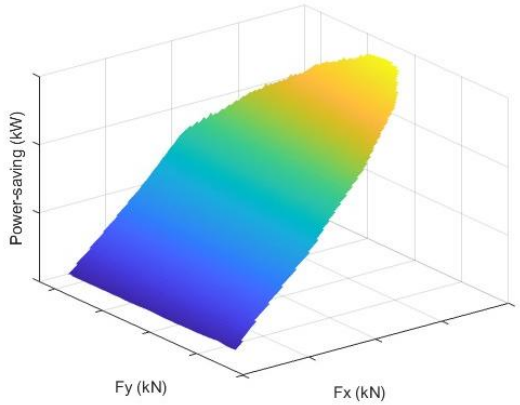


Fig.3: Visualisation of conversion tables for translating wind propulsion thrust force (Fx) and side force (Fy) to power saving (ΔP) for one ship speed.

Phase 2: Verification of conversion table. In addition to the ON OFF sea trial tests described in Section 3, force measurements were also collected by the rotor sail provider Norsepower. Comparison between results from the ON-OF test and the rotor forces measured at the same occasion give a direct means to verify the accuracy of the force to power relationship.

An acceptance criterion can be derived in the same manner as described in Section 3. The comparison error is then computed as:

$$\bar{E} = \left(\sum_{k=1}^{N} \Delta P_{Force_k} \cdot w_k - \sum_{k=1}^{N} \Delta P_{ON-OFF_k} \cdot w_k \right) / \sum_{k=1}^{N} w_k \quad (1)$$

expressed in percentage of $\overline{\Delta P} = \sum_{k=1}^{N} \Delta P_{pred_k} \cdot w_k / \sum_{k=1}^{N} w_k$

ΔP_{ON-OFF_k} is the derived ΔP from the sea trial ON-OFF run k
ΔP_{Force_k} is the ΔP derived from the force measurements, translated to ΔP using the conversion table
w_k is the corresponding weight in a weather probability matrix, for example the GWM used for EEDI (IMO, 2021).

Fig.4 demonstrate this using the real sea trial data, showing a comparison between the power saving derived via force conversion table, and the power saving from the ON-OFF test, measured at the same time. The derived weighted average error $\bar{E}$ shows that the methods agree. Further establishing of the uncertainty U_V, is required to conclude to what tolerance verification is achieved.

Note that the conversion table has been modelled without the use of the force measurements. Therefore, the comparison error can be derived on the whole population. If the conversion is tuned to real measurements, then the test of the model must be done on an independent set of data points.

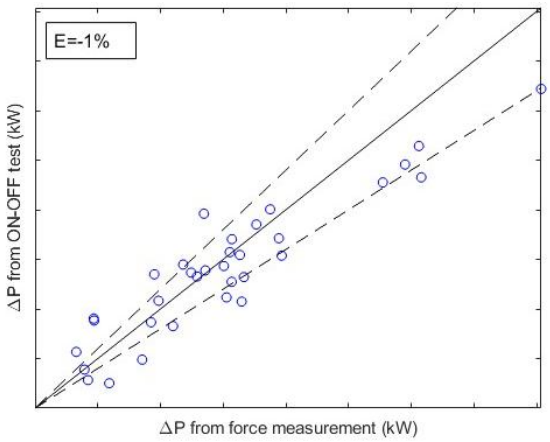


Fig.4: Validation of conversion of wind propulsion force to power saving

Fig.5 compares the power saving coefficient from both methods: from force measurements translated to power saving, and from ON-OFF test. There are scatter and outliers in both methods, but the scatter is larger for the ON-OFF testing. From this we can conclude:

- The force measurements, converted via the conversion model, derives power saving that on average is verified using ON-OFF power saving test
- The force measurement seems to be more precise than the ON-OFF tests (lower precision uncertainty)

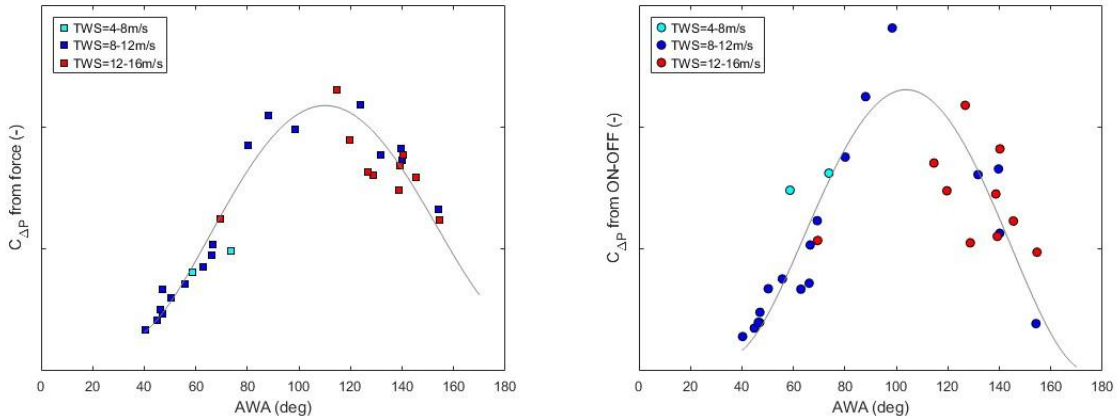


Fig.5: Power saving coefficient derived from force measurements (left) and ON-OFF tests (right)

Phase 3-4: Logging of sail forces and derivation of energy saving. Fig.6 illustrates the power saving and the accumulated energy saving obtained with the two methods. In general, the force logging approach provides a more stable estimate of the performance. The wind logging approach is more sensitive to uncertainty in the wind measurements and therefore exhibits greater variability. Even so, wind logging remains a valuable fallback method. If the advanced force logging system becomes unavailable for any reason, the wind-based method can continue to deliver usable estimates of the saving. In addition, wind logging can be applied using hind cast weather data in situations where the ship anemometer is not operational. This ensures that both methods complement each other and provide redundancy in long term monitoring of wind propulsion performance.

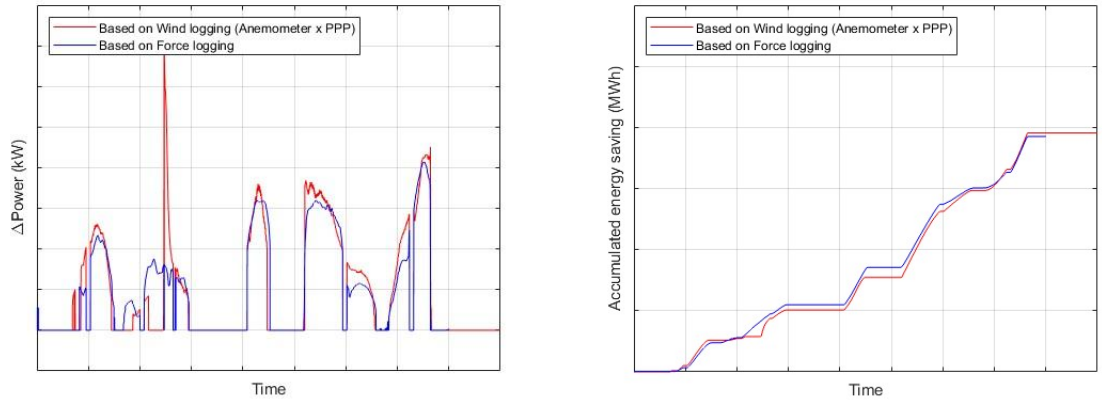


Fig.6: Example of savings monitoring based on Force logging and Wind logging combined with verified model

4. Conclusions

This study has presented practical methods for verifying and monitoring the performance of wind assisted propulsion at full scale. The work demonstrates that a two-phase strategy, where a power saving matrix is first established through modelling and then verified through sea trials, provides a reliable structure for evaluating wind propulsion performance. Once verified, the matrix can be combined with route specific wind statistics to estimate fuel saving during normal operation.

A clear acceptance criterion has been proposed to support transparent verification. The weighted average comparison error offers a straightforward indicator of agreement between model predictions and measured performance. The case study illustrates how this indicator can guide improvements of the performance model until the comparison falls within an acceptable tolerance.

The results confirm that bias in wind measurements is a major source of uncertainty. Disturbance from the ship superstructure affects both apparent wind speed and apparent wind angle. Calibration

using CFD studies or comparison with hind-cast weather data can reduce this bias substantially and improve the accuracy of the final power saving estimate. The analysis of precision uncertainty further shows that reliable assessment requires a wider set of sea trial conditions than the minimum recommended program. A broader range of wind angles, wind speeds and draught conditions, collected over an extended period, is needed to achieve a satisfactory precision.

Two approaches for continuous monitoring have been demonstrated: wind logging and force logging. The force-based approach generally provides more stable estimates, as it relies on direct measurements of thrust and side force translated through a verified conversion table. Wind logging is more sensitive to uncertainties in the wind measurement but remains a useful fallback and can be applied using hind cast weather when the ship anemometer is unavailable. Together, the methods provide redundancy and support long term monitoring of wind propulsion performance.

A structured process for converting measured forces into propulsion power saving has been proposed and applied to full scale data. This addresses a current gap where no guideline exists. Alternative approaches that consider varying propulsive efficiency can offer further improvements when shaft thrust measurements are available, although this instrumentation is not common on most vessels.

The methods presented in this study are relatively new and will benefit from further testing on additional vessels, propulsion technologies and operational profiles. Wider application will help refine the acceptance criterion, improve understanding of uncertainties and support the development of consistent industry practice. Continued collaboration between technology providers, ship operators and research organisations will be important to mature these methods and ensure they meet future regulatory needs.

Overall, the work contributes to the development of robust and transparent procedures for full scale performance evaluation of wind assisted propulsion, supporting reliable assessments as these technologies become more widely adopted in commercial shipping.

Acknowledgement

This work has been made possible thanks to a number of research grants: Swedish Transport Administration FOI project FAIRWINDS TRV 2024/31480. European Climate, Infrastructure and Environment Executive Agency (CINEA) project 101096673 (Orcelle), Korea Planning & Evaluation Institute of Industrial Technology (KEIT) project “Design of Wind-Assisted Propulsion Systems and Hull for Performance Evaluation and Control Technology Development in Wind-Assisted Propulsion Applications”.

References

ANEMOI MARINE (2025), *Performance verification of wind-assisted ship propulsion systems by On-Off testing*, Anemoi White Paper

DNV (2025), *Recommended Practice DNV-RP-0686 – Performance of wind assisted propulsion systems*, DNV, Hovik

IMO (2021), *MEPC.1-Circ.896, Guidance on treatment of innovative energy efficiency technologies for calculation and verification of the attained EEDI and EEXI,* Int. Mar.Org., London

ITTC (2008), *ITTC 7.5-02-01-01 – Guide to the expression of uncertainty in experimental hydro-dynamics*, ITTC Publication

ITTC (2024) *ITTC 7.5-03-01-01 – Uncertainty analysis in CFD verification and validation methodol-ogy and procedures*, ITTC Publication

ITTC (2024), *ITTC 7.5-04-01-02 – Sea trials for assessing the power saving from wind assisted propulsion',* ITTC Publication

PONKRATOV, D.; STRUIJK, G.D. (2024), *Quantifying the uncertainty of high-fidelity speed/power trials*, 9th HullPIC Conf., Tullamore

RUTH, E.; TRANELL, J.; ROGNEBAKKE, O.; HOLLENBACH, U. (2024) *Long-term Verification of ALS and WAPS Systems,* 9th HullPIC Conf., Tullamore

VOUTILAINEN. J.; PAAKKARI, V., (2025), *Continuous Performance Measurement of Wind Propulsion*, 10th HullPIC Conf., Mulheim

WERNER, S.; NISBET, J.; HÖRTEBORN, A.; NIELSEN, R., (2021), *Sea trial Verification for a Wind Assisted Ship*, RINA Int. Conf. Wind Propulsion, London

WERNER, S.; GERHARDT, F.; (2025), *A Feasible Approach for Cost-Saving Splitting of Wind Propulsion Technology Installations,* 10th HullPIC Conf., Mulheim

WERNER, S., (2026), *Sea Trial Methodologies for Wind Propulsion: A Study of Uncertainty,* RINA Int. Conf. Wind Propulsion, London

Challenges Toward Utilizing Monitoring Data for Fully Wind-Powered Merchant Ships

Yuji Arai, Sumitomo Heavy Industries Marine&Engineering, Yokosuka/Japan, yuji.arai@shi-g.com
Takeshi Aono, Sumitomo Heavy Industries Marine&Engineering, Yokosuka/Japan, takeshi.aono@shi-g.com
Kahori Amano, Sumitomo Heavy Industries Marine&Engineering, Yokosuka/Japan, kahori.amano@shi-g.com
Akihiko Masutani, Sumitomo Heavy Industries Marine&Engineering, Yokosuka/Japan, akihiko.masutani@shi-g.com

Abstract

This paper describes Sumitomo experience with harnessing monitoring data in the development and improvement of wind assisted ship propulsion (WASP) systems. Real-time data obtained from onboard sensors is indispensable for achieving precise sail control and power management. This study examines what types of data should be acquired to effectively control an autopilot system incorporating sail handling. Results indicate that efficient control can be achieved by appropriately combining and utilizing information between sensors, without requiring the addition of significantly specialized sensors beyond those normally installed.

1. Introduction

In line with global initiatives to address climate change, the maritime industry is now required to achieve net-zero CO_2 emissions by 2050. Within this trend, the use of wind as a source of propulsion for merchant ships has been receiving renewed attention. In recent years, ships have begun entering service that are equipped with wind assisted propulsion systems, such as sails or rotor sails, operating in motor-sailing mode in which wind-powered thrust supplements the propulsive power produced by conventional internal-combustion engines.

Against this societal background, Sumitomo Heavy Industries Marine&Engineering has been advancing the development of merchant ships equipped with rotor sails that are capable of fully wind-powered operation, in which the propeller can be stopped and propulsion is provided solely by wind when conditions are favourable. The company has an extensive track record in designing and constructing large tall ships currently operating in Japan, Nippon Maru, Kai-wo Maru, and Miraie (originally Akogare), all of which are capable of fully wind-powered operation, Fig.1. Through these projects, substantial technical expertise in tall ship design has been accumulated by *Yoshimura et al. (1991), Amemiya et al. (1992), Tanabe et al. (1987), Takekawa et al. (2004)*.

The present development integrates this experience with the latest rotor-sail technology to realize a practical sail control system that reduces crew workload by minimizing the need for manual and continuous sail adjustments. In the practical implementation of the sail control system, it should be noted that wind is inherently unsteady. This unsteady wind, combined with variations in waves, tidal currents, and maneuvering conditions, can cause significant fluctuations in thrust, lateral force, and yawing moment. To maneuver toward an arbitrary course while capturing wind power without loss during fully wind-powered sailing or motor-sailing, closed-loop control and operational decision-making based on monitoring data are essential.

The objective of this study is to define the role of monitoring data and the minimum data requirements necessary for sail control applicable to both fully wind-powered and motor-sailing modes. Based on these requirements, the study proposes a system architecture and control design for a sail control system, and validates their effectiveness through model tests.

Fig.1: Nippon Maru (left) and Kaiwo Maru (right) moored at the former Uraga Shipyard

2. Role of Monitoring Data in Fully Wind-Powered Ships

The unsteady nature of wind is one of the major challenges for maintaining stable sailing performance. Wind direction and wind speed fluctuate over short time scales due to geographical, seasonal, and meteorological factors, making accurate prediction difficult and reliant on the experience of navigators. Furthermore, the aerodynamic forces acting on the sails depend not only on the wind conditions but also on ship speed, heading, course over ground, and sail configuration.

In fully wind-powered operation, the sails are the sole source of propulsion, and thus continuous adaptation to frequent wind variations is required for maintaining a desired ship course. Relying solely on crew experience makes practical operation extremely challenging.

Accordingly, to reduce crew workload and allow operation by non-specialist crews, sail control system should rely on monitoring data automatically and periodically acquired from conventional onboard instruments, avoiding any need for specialized sensors or manual inputs.

3. System Architecture

To clarify the overall configuration of a vessel equipped with a wind-propulsion system, the system architecture of a general merchant ship fitted with such a device is summarized below. An illustrative schematic of this architecture is presented in Fig. 2.

Control Equipment
1. Control System Console, etc.

Controlled Objects
2. Rotor Sails
3. Rudder
4. Propeller
5. Main Engine
6. Battery

Sensors
7. GPS (Global Positioning System)
8. Directional Gyro
9. Rate Gyro
10. Anemometer
11. Wave Height Sensor (depending on the ship)

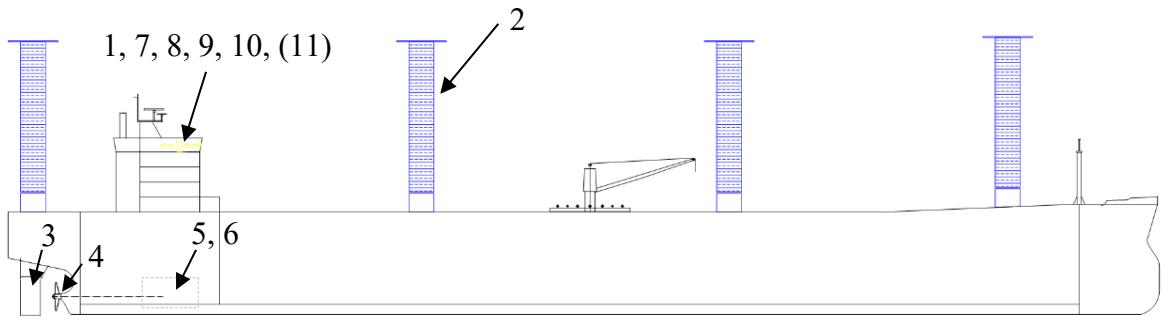


Fig.2: System Architecture of Fully Wind-Powered Ship

4. Data Requirements for Sail Control System

The definitions of the variables used in the control system and the associated coordinate system are presented below and in Fig. 3.

Vs	: Ship Velocity (m/s)
u	: Velocity Component in X-direction (m/s)
v	: Velocity Component in Y-direction (m/s)
COG_T	: Target Course Over Ground (deg.)
COG_M	: Measured Course Over Ground (deg.)
ψ	: Heading Angle (deg.)
r	: Yaw Rate (deg./s)
β	: Drift Angle (deg.)
TWS	: True Wind Velocity (m/s)
TWA	: True Wind Angle (deg.)
AWS	: Apparent Wind Velocity (m/s)
AWA	: Apparent Wind Angle (deg.)
n_RotorX	: Rotating Speed of Rotor Sail (1/s) (X indicates Sail No.)

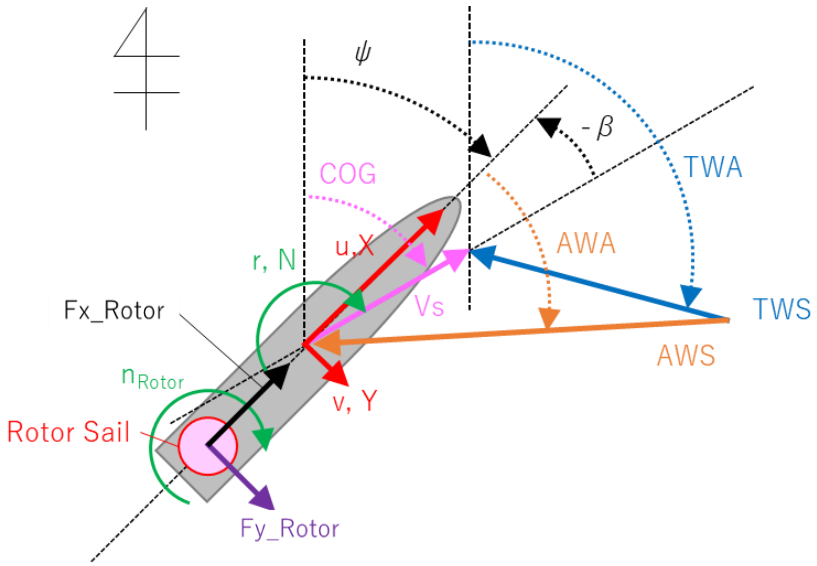


Fig. 3: Coordinate system

The major data used for the control system and their acquisition methods are summarized in Table I.

Table I: Data List

Variables	Acquisition Method
Ψ	Directional Gyro installed on the ship
r	Rate Gyro installed on the ship
Latitude-position	DGPS
Longitude-position	DGPS
COG_T	Input from the system
COG_M	Calculated within the system
Vs	Calculated within the system
u	Calculated within the system
v	Calculated within the system
β	Calculated within the system
AWS	Anemometer installed on the ship
AWA	Anemometer installed on the ship
TWS	Calculated within the system
TWA	Calculated within the system
n_RotorX	Encoder installed in the sail (X indicates Sail No.)

5. Sail Control System Overview

Ship maneuvering control consists of two essential functions: course-keeping and turning (tacking/gybing). The control algorithm must therefore accommodate both behaviors. A block diagram capable of realizing these two functions was developed using only the information obtained from sensors commonly installed on general merchant ships, as listed in Table I, *Arai et al. (2024), Aono et al. (2026)*. An overview of the block diagram is shown in Fig. 4.

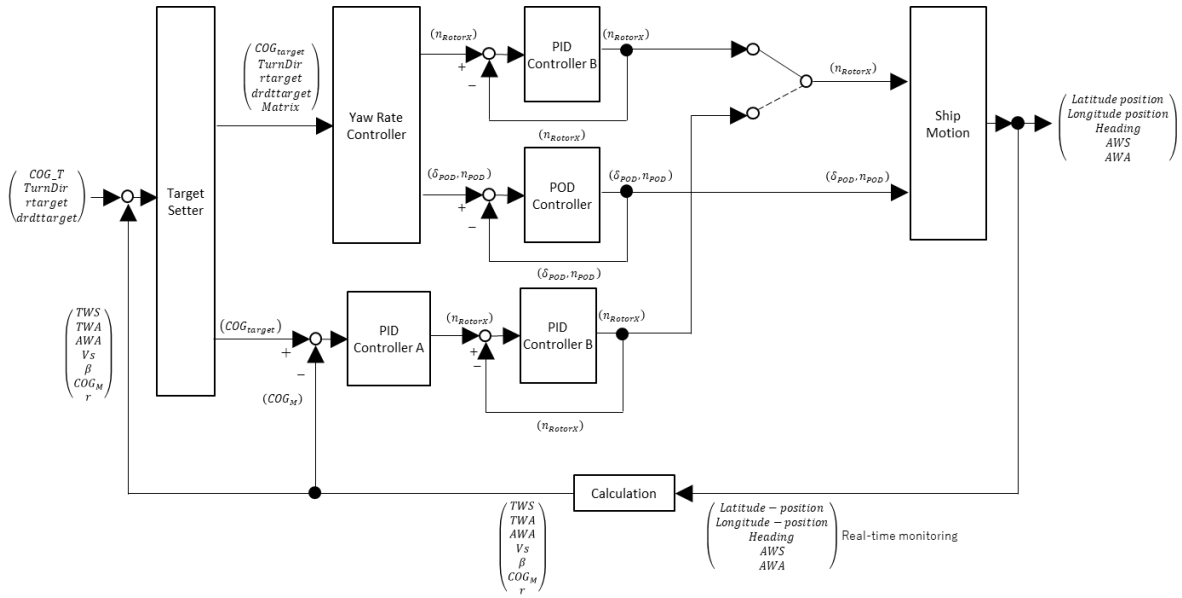


Fig.4: Sail Control Block Diagram

6. Verification Test

6.1. Test Overview

To verify that the sail control can be achieved using only the monitoring data obtained from standard onboard sensors, without adding any special measurement devices, free-running sailing tests were conducted using model ships equipped with rotor sails under both steady and unsteady wind conditions. A free-running test is one in which the model ship is not constrained by a carriage and is allowed to move freely, while its track, attitude, and motions are measured.

The experiments were carried out as part of a joint research program with the National Maritime Research Institute (NMRI) of the National Institute of Maritime, Port and Aviation Technology, using the Actual Sea Model Basin, *Tanizawa et al. (2010)*. Model ships equipped with rotor sail measurement and control system developed by the authors were brought into the facility. The basin is a large indoor tank equipped with wind blowers and wave makers, and was considered optimal for the objectives of this study.

An overview of the measurement and control system used in the experiments is shown in Fig.5. Anticipating future installation on full-scale ships, the instrumentation layout on the model ship was designed such that all data expected to be measured onboard an actual ship would be acquired using sensors mounted on the model. However, ship speed and course over ground (COG), which would normally be obtained from GPS on a full-scale vessel, were instead replaced by non-contact measurements of the model ship's (X, Y) position within the basin. The wind blower was mounted on a carriage that followed the model ship seamlessly while maintaining a constant distance, thereby allowing the model to be exposed to wind from a constant direction.

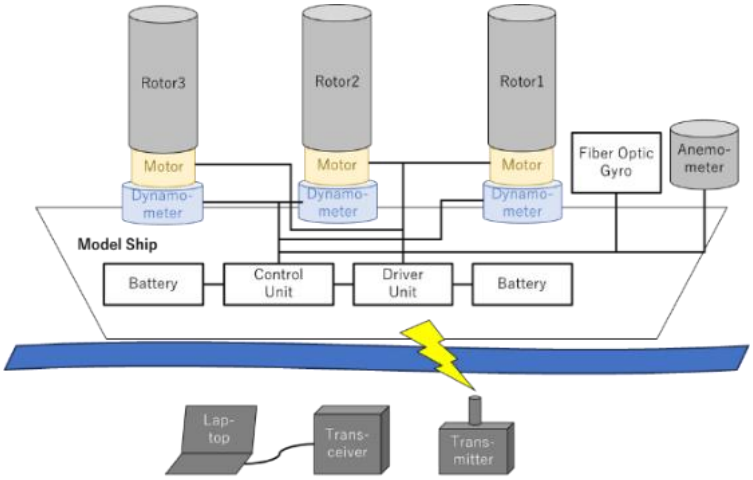


Fig. 5: Overview of the Measurement System for Model Ship Test

The rotor sails and the onboard measurement/control system were powered by batteries installed inside the model and operated remotely from shore via wireless communication. To measure wind conditions essential for control, a single ultrasonic anemometer was installed near the bow at approximately half the rotor sail height. Although potential interaction was anticipated due to the sensor position possibly entering the wake or shadow of the superstructure depending on ship's heading angle and wind direction, ideal wind direction and wind speed measurements are difficult even on full-scale ships. Consequently, the decision was made to evaluate control performance using only the data obtained from this single anemometer.

The measurement and control system consisted of a control unit, which also incorporated data-logging functionality, and driver units that commanded each rotor sail. Real-time control computations were performed onboard the model using the measured data, enabling individual rotational commands to be

issued to each rotor sail. The (X, Y) positional data measured remotely from shore, serving as a substitute for latitude and longitude position from GPS, was transmitted in real time to the onboard system and used together with the onboard sensor data for control calculations. The control outputs and all measurement data were synchronized as time-series records and transmitted in real time to a land-based data-acquisition Laptop.

Two model ships capable of accommodating the measurement and control system were constructed, and experiments were conducted using these models. The model ships were designed as cargo-vessel types intended to be equipped with rotor sails, and their main characteristics are as follows:

(A) Equipped with three rotor sails
(B) Equipped with four rotor sails
(Common to A and B) No propeller installed, and rudder fixed; propulsion and maneuvering forces are generated solely by the rotor sails

The principal particulars of the model ships are listed in Table II, and photographs of the models are shown in Fig6. Although Fig. 6 appears to show the model ship and a carriage above the model ship connected by wires or similar equipment, these equipment units are only used to prevent the model from drifting when measurements are not being taken. During data acquisition, the mechanism is configured that no forces affecting the measurement results are applied to the model ship.

Table II: Model Ship Particulars

(A)	Length	3.720 m	No. of Sails	3
	Breadth	0.645 m	Rotor Height	0.6 m
	Depth	0.515 m	Rotor Dia.	0.1 m
	Draught	0.160 m	Rotating Speed	Max. 50 rps
(B)	Length (m)	3.791 m	No. of Sails	4
	Breadth (m)	0.753 m	Rotor Height	0.6 m
	Depth (m)	0.353 m	Rotor Dia.	0.1 m
	Draught (m)	0.188 m	Rotating Speed	Max. 50 rps

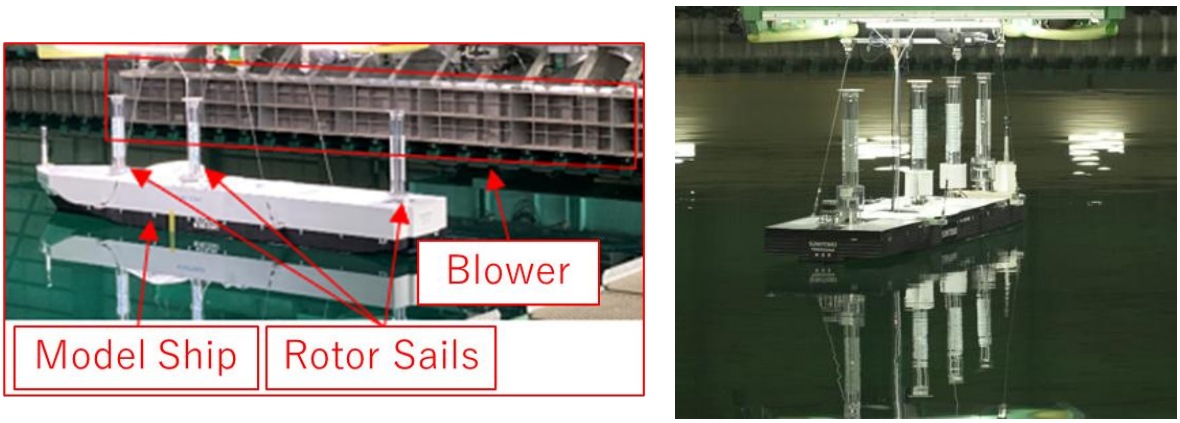


Fig. 6 Model Ship (A) with Rotor Sails (left), Model Ship (B) with Rotor Sails (right)

6.2. Test Conditions and Test Results

6.2.1. Course-keeping Control

Using Model Ship A, the effectiveness of the sail control function was evaluated by comparing tracks and time histories of COG between control ON and OFF conditions. As shown in Figs.7 and 8, the uncontrolled model exhibited continuous starboard turning, whereas under control it stabilized at the target COG = 0° and maintained the scheduled course, confirming that sail control can be achieved using only the assumed sensor information.

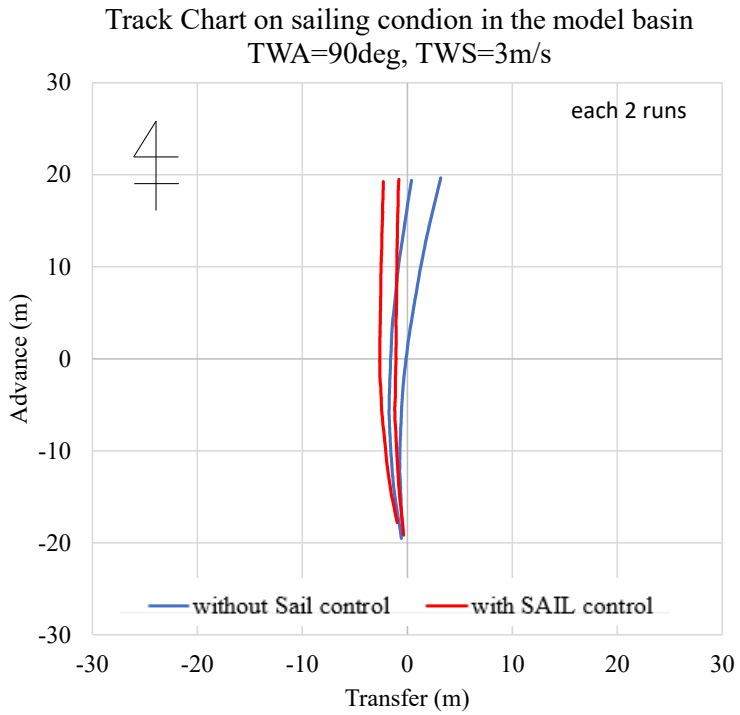


Fig.7: Comparison of Tracks with Sail Control On/Off (A)

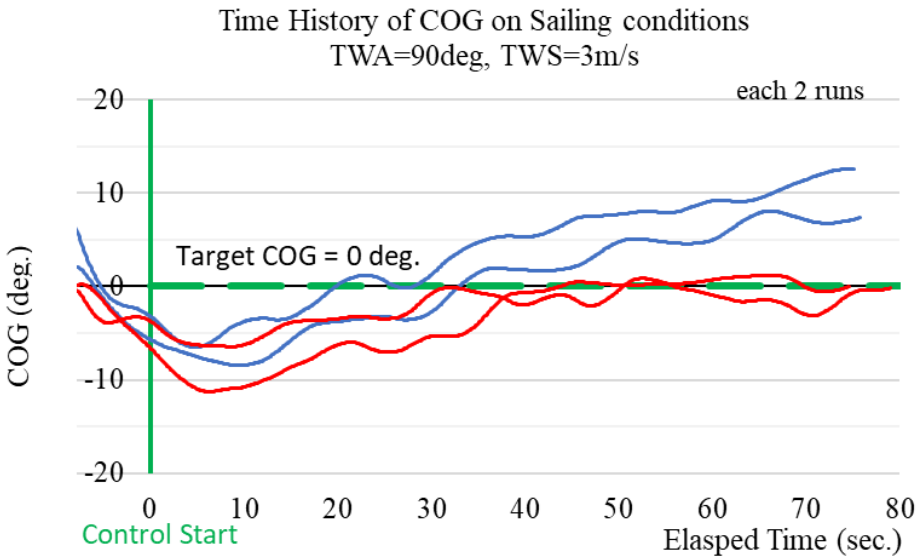


Fig.8: Comparison of COG with Sail Control On/Off (A)

6.2.2. Turning (tacking/gybing) Control

Turning control verification during fully wind-powered operation was conducted using Model Ship B, which is equipped with four rotor sails, since greater maneuvering force is required compared with course-keeping control. Although the basin used for this test is among the largest of its kind worldwide, the turning radius during fully wind-powered sailing becomes very large because both propulsion and turning moment must be generated solely by the rotor sails. As a result, it was not possible to secure sufficient space for the model ship to complete 360° turning circle within the basin.

Therefore, the model tests focused on the region near the wind-axis crossing, where control becomes most challenging due to rapid changes in the turning moment associated with variations in apparent

wind angle. Both tacking and gybing control were successfully achieved in the model tests; however, as it is well known that tacking involves a higher probability of failure compared with gybing - and thus would generally be avoided in commercial operations - the results reported here focus on the gybing control. Fig.9 shows a photograph taken during the gybing control tests and the measurement area used in the basin. The photograph shows the situation immediately after the start of the measurements. Following this, the model ship performed the gybing control maneuver along the track indicated by the red line.

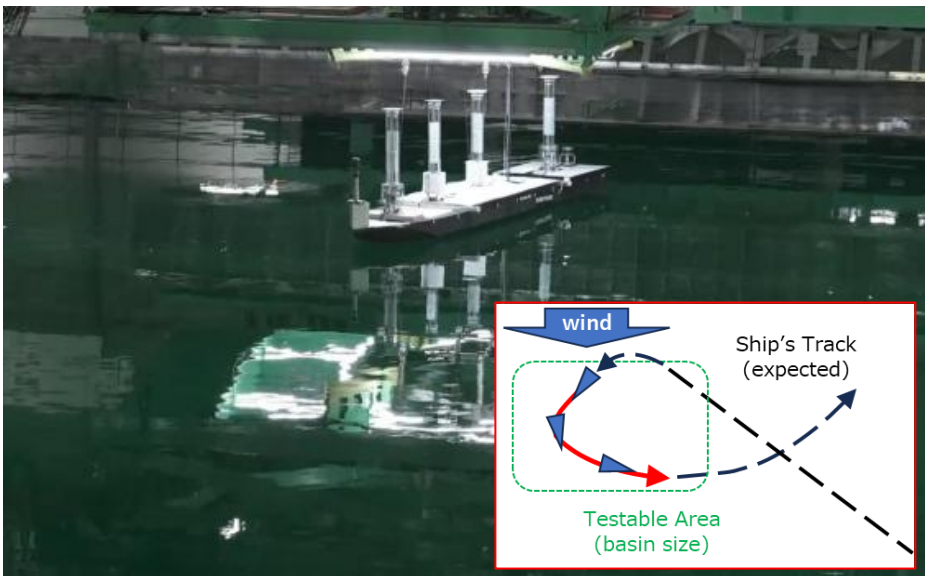


Fig.9: Photograph during Gybing Control Test (B)

The model ship's tracks recorded during the tests are shown in Fig. 10, and the corresponding AWA time histories are presented in Fig.11. These results confirm that turning maneuvers across the wind axis can be achieved by controlling the rotor sails solely based on sensor information.

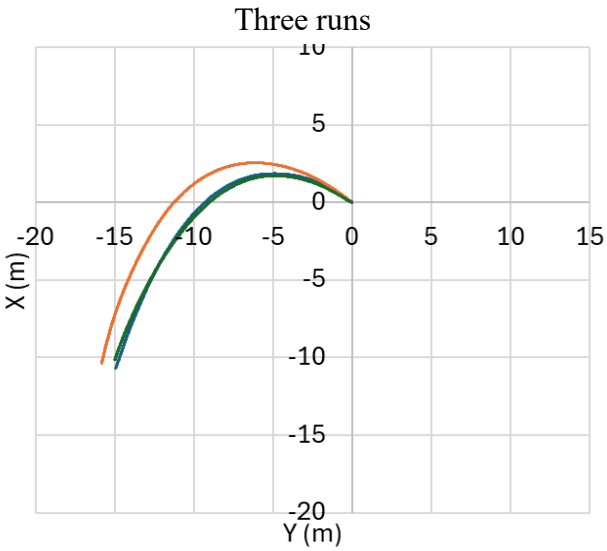


Fig.10 Model Ship Track during Gybing Control Test

Notably, the results also indicate that maneuvering control remains feasible even when the anemometer temporarily experiences interaction with the rotor sails or above-water structures, leading to degraded measurement accuracy. This suggests that, for practical implementation, the robustness of the sail control system can be ensured without resorting to high-end wind-measurement systems - by instead employing measures such as sensor redundancy (multiple anemometers), optimized sensor placement, and refinements to the control algorithm.

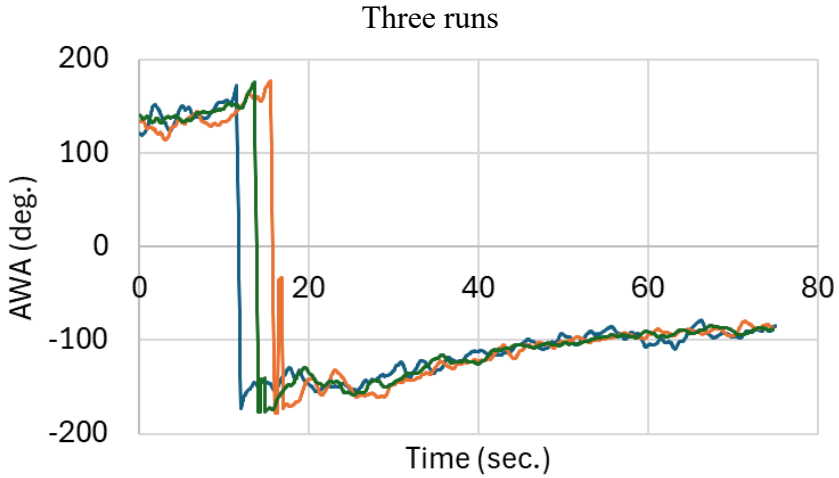


Fig.11 Time Series of AWA during Gybing Control Test

7. Conclusion

To support the global objective of achieving net-zero CO_2 emissions by 2050, Sumitomo Heavy Industries Marine&Engineering has been advancing the development of rotor sail equipped merchant ships and associated sailing technologies that enable fully wind-powered operation, in which the propeller can be stopped under favorable wind conditions.

This study defined the minimum sensor data requirements for practical fully wind-powered control, proposed the corresponding system architecture and control design, and validated their effectiveness through model experiments from the following viewpoints:

- Identification of monitoring data necessary for sail control
- Construction of a unified control algorithm for both course-keeping and turning
- Experimental validation of the control algorithm using monitoring data

The study demonstrates that the proposed sail control system which is built on monitoring data obtained from standard onboard instrumentation can sail control during fully wind-powered operation of rotor sail equipped ships. Consequently, it will enable safe and effective operation even by crews with limited sailing experience or specialized knowledge.

Moreover, the system is applicable in motor-sailing mode in which the propeller and rotor sails are used concurrently. By employing the maneuvering control function of the sail control system in motor-sailing conditions, the use of counter-rudder, an important contributor to added resistance, can be reduced. The associated fuel-saving effect has been estimated at approximately 4% relative to conventional rudder-based maneuvering, *Mizutani et al. (2024)*.

Two principal directions for future work are envisaged:

- Verification of the robustness of the sail control system
- Voyage planning optimized for the proposed control scheme and for fully wind-powered operation

In addition to sail control, other major factors influencing the ship's maneuverability, such as the hydrodynamic characteristics associated with the hull form and the spatial arrangement of the rotor sails, must also be considered. Accordingly, we are currently working on enhancing the robustness of both the control system and the monitoring-data acquisition system to enable their application to a wider variety of ship types. Ultimately, our goal is to advance the system to a level at which it can be used

with confidence even in actual sea conditions, where severe sea conditions and diverse external disturbances are encountered.

Acknowledgement

The free-running tests of the fully wind-powered model ship were conducted as part of a joint research program with the National Maritime Research Institute of the National Institute of Maritime, Port and Aviation Technology. This development was carried out as the Frontier Development Project of the Technology Research Centre of Sumitomo Heavy Industries, Ltd. The authors would like to express their deep appreciation to all individuals and organizations who contributed to this study.

References

AMEMIYA, I.; YOSHIMURA, Y.; et al. (1992), *On the Aerodynamic Characteristics of the Sail Training Ship NIPPON MARU -Aerodynamic Characteristics and Flow Visualization on Each Mast-*, J. Japan Institute of Navigation 87 (in Japanese)

ARAI, Y.; AONO, T.; et al. (2024), *Development of Wind Propulsion System*, Wind Propulsion 2024, London, pp.375-386

AONO, T.; ARAI, Y.; et al. (2026), *Development of a Wind-Powered Ship Control System - The Challenge of Achieving Fully Sailing in Merchant Ships* -, Report on JIME Research Committee for Machinery Engineering Systems No.1 (in Japanese, accepted for publication)

MIZUTANI, N.; ARAI, Y.; et al., (2024), *Navigating a Sustainable Future with Wind-Assisted Ship Technology: NAPA, Norsepower and Sumitomo's Collaborative Journey*, 9th HullPIC Conf., Tullamore, pp.111-128, http://data.hullpic.info/HullPIC2024_Tullamore.pdf

TAKEKAWA, M.; et al. (2004), *Development of an Advanced Sail Assisted Ship*, Conf. Proc. Society of Naval Architects of Japan 4 (in Japanese)

TANABE, Y.; FUKUI, T., et al. (1987), *Analysis of Sailing Performance of Sail Training Ship "Nippon-maru"*, J. Japan Institute of Navigation 94 (in Japanese)

TANIZAWA, K. et al. (2010), *The Actual Sea Model Basin*, Papers of National Maritime Research Institute 10/4 (in Japanese)

TATANO, H.; YOSHIMURA, Y., et al. (1991), *On the Aerodynamic Characteristics and Flow Fields of Square Sails*, J. Japan Institute of Navigation 84 (in Japanese)

UENO, U.; TAKEKAWA, M.; et al. (2004), *Fundamental research for development of an advanced sail-assisted ship*, OCEANS '04 MTS/IEEE Tecno-Ocean '04

YOSHIMURA, Y.; TANABE, Y.; OHSUGI, I.; AMEMIYA, I. (1991), *On the Prediction of the Sailing Performance for a Large Sail Training Ship*, J. Japan Institute of Navigation 84

YOSHIMURA, Y.; et al. (1991), *On the Aerodynamic Characteristics of the Sail Training Ship NIPPON MARU*, J Japan Institute of Navigation 84 (in Japanese)

Digital Fouling Reports – A New Data Source for Hull Performance Management

Solène Guéré, OpenHull, La Sure En Chartreuse/France, solene@openhull.com
Marciel Gaier, OpenHull, Montreal/Canada, marciel@openhull.com
Guastavo Caldas, OpenHull, Montreal/Canada, gustavo@openhull.com

Abstract

This paper describes how hull condition reports can be used as additional data source in hull performance management. Guide 21517 is a newly published AMPP framework. It creates an open, interoperable fouling report that can be generated by underwater service providers and easily read by performance monitoring software. The format provides a standardized template and file type, allowing seamless comparison between hull reports, even from different service providers. By providing more consistent and higher-quality data, it will support the development of machine learning algorithms, ultimately improving the ability to predict fouling conditions.

1. Introduction to Guide 21517

1.1 Rationale

Significant strides have been taken in recent years to enhance the energy efficiency of ship hulls. The generalization of hull optimization algorithms, and the switch from noon reports to high frequency data, have provided a robust framework to fine-tune cleaning schedules.

Still, hull fouling has, until now, remained largely untouched by digitalization efforts. This has left a notable gap in research regarding the harmonization and optimization of the interplay between cleaning technologies and coatings, as well as their alignment with the operational profiles of diverse vessels, *Guéré (2023)*.

Several standards already exist for assessing fouling and generating underwater hull reports, such as those from *BIMCO (2023)* and *IMO (2023)*. However, these reports are still primarily issued as static PDFs or through proprietary applications tied to specific service providers.

To reduce the carbon intensity of the maritime industry, it is essential to deepen our understanding of the relationship between hull cleaning and coating performance. A standardized digital format simplifies the tracking of hull conditions across different ports, service providers, and technologies. Furthermore, it facilitates the comparison of reports over time and across vessels, enabling seamless integration into broader hull performance data analysis.

The goal of the AMPP Guide 21517 is to create an open, interoperable fouling report that can be generated by underwater service providers (divers and ROV operators alike) and easily read by assorted performance monitoring software.

This format is not intended to replicate or replace existing reporting standards, nor does it propose new recommendations for conducting underwater inspections or cleanings. Rather, it provides a standardized template and file type for digitizing key reporting information from traditional hull reports.

The Guide is technology-neutral and can be directly created by diving services as a CSV or XLS format. OpenHull has developed OpenHull Sync, an implementation example: a platform that enables the full integration and historical tracking of digital fouling reports compliant with AMPP Guide 21517. By aggregating this data alongside dry-dock records, users can monitor the evolution of fouling and evaluate coating performance in real time throughout the entire dry-dock cycle.

1.2. Purposes

The Digital Fouling Report (DFR) is designed to fulfill the following strategic objectives:

a) Monitoring Hull Roughness Over Time
The DFR facilitates the seamless integration of inspection data into digital hull management systems, enabling the cross-comparison of reports regardless of the service provider. By ensuring data consistency and integrity, the DFR provides a robust foundation for machine learning algorithms, ultimately enhancing the predictive modeling of biofouling development.

b) Monitoring Coating Condition Over Time
By tracking coating conditions with high-fidelity data, stakeholders can deepen their understanding of long-term coating integrity. This enables a more precise evaluation of cleaning efficiency and helps quantify the potential mechanical impact of various cleaning technologies on coatings.

c) Facilitating Communication between Parties (service providers and port authorities)
The DFR establishes a comprehensive framework for inspections, improving coordination between ship managers and underwater service providers. It allows for the clear definition of standardized inspection protocols and reporting parameters, ensuring a higher overall quality of hull assessments. The DFR provides structured, granular data to support port entry procedures and compliance with increasingly stringent local biofouling regulations.

d) Data Tracking and history:
The DFR enables centralized access to a vessel's entire maintenance history. By consolidating performance and condition data into a single platform, managers can maintain a continuous, auditable record of the hull's lifecycle.

2. AMPP Guide 21517 development

2.1. Working Group composition

Following an industry-wide call for participation, a working group of 56 maritime professionals was established. To ensure the final guide addressed the diverse needs of the sector, specific emphasis was placed on ensuring strategic representation from across the entire value chain:

- Manual Service Providers (Diver-based): Focused on standardizing client template requests and prioritizing the feasibility of manual data entry by personnel.
- Robotic Service Providers: Seeking to standardize reporting formats with a specific focus on the automated generation of DFRs directly from robotic software systems.
- Coating Manufacturers: Ensuring the DFR provides the necessary data to accurately track coating integrity and antifouling effectiveness over time.
- Experts: Leveraging their independent status to provide expert guidance on enhancing report quality and technical rigor.
- Shipowners and Fleet Performance Managers: Advocating for actionable data that serves the dual purpose of Regulatory Compliance (e.g., New Zealand port entry requirements) and Performance Analytics (e.g., hydrodynamic drag estimations).

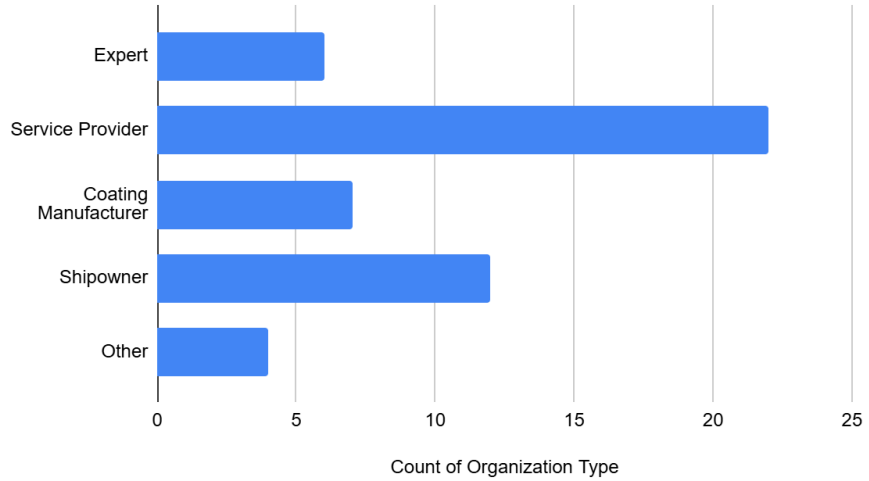


Fig.1: Group members of Guide 21517, sorted by organization type

2.2. Decisions on Hull Zones & Fouling Degrees

Given that existing reporting standards vary regarding hull segmentation and fouling nomenclature, the working group sought to establish an optimal method for quantifying fouling severity and spatial coverage.

2.2.1. Hull images

The working group determined that the DFR would not provide specific guidance on safety, visibility, or image quality. These parameters are already comprehensively addressed in existing inspection standards, such as AMPP SP21487 or AMPP SP21421.

2.2.2. Hull Zones

To ensure the standard is practical for both manual diver inspections and robotic systems, the zoning resolution is scalable based on vessel size. The hull is divided into longitudinal sections of 50 m. Each section consists of:

- Vertical Sides: Two zones per side (Upper Vertical and Lower Vertical).
- Flat Bottom: One central zone.

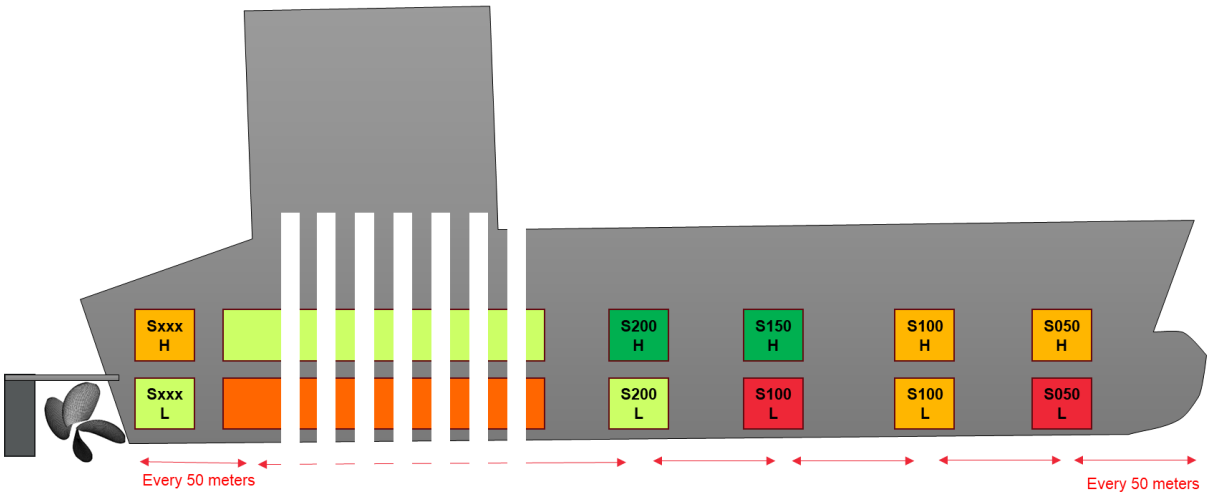


Fig.2: Vertical Sides Zones - the number of section varies depending on the vessel length

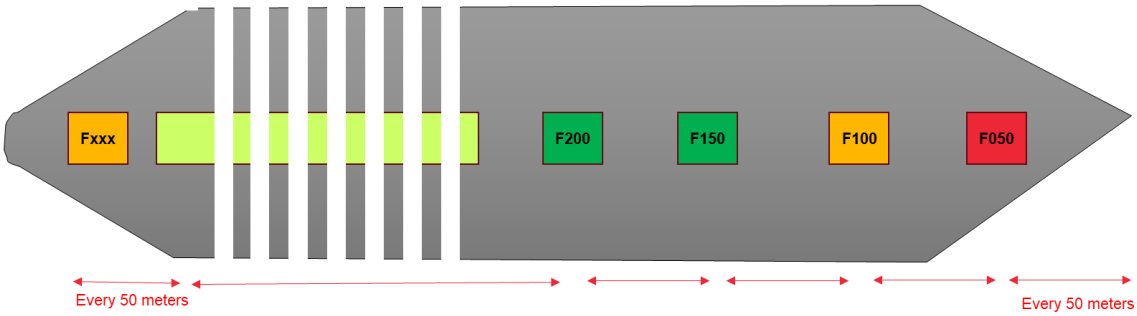


Fig.3: Flat Bottom Zones - the number of section varies depending on the vessel length

2.2.3. Niche Areas

To facilitate regulatory compliance, specifically for New Zealand's Craft Risk Management Standard, *MPI (2023)*, it was decided that the niche area classifications must be exhaustive. Consequently, a comprehensive taxonomy of 93 distinct niche areas was established. For baseline or simplified inspections, a prioritized subset of these niche areas may be utilized, provided they meet the minimum requirements of the relevant port authority.

2.2.4. Fouling Rating

The working group developed a biofouling scale designed to meet the following strategic criteria:

- Compatibility: Building upon existing frameworks, particularly the IMO Biofouling Guidelines.
- Hydrodynamic Precision: Enabling precise hull resistance calculations by differentiating between slime levels and identifying long filamentous species.
- Biosecurity Risk Management: Specifically flagging calcareous species (e.g., barnacles, tube worms) due to the significant biosecurity risks they pose.
- Ease of Use: Ensuring the scale remains accessible to offshore personnel without requiring specialized biological or species-specific knowledge.

Most studies, as aggregated by the *IMO (2023)*, confirm that even slime-level fouling can result in an overconsumption of fuel by as much as 35%, Fig.4.

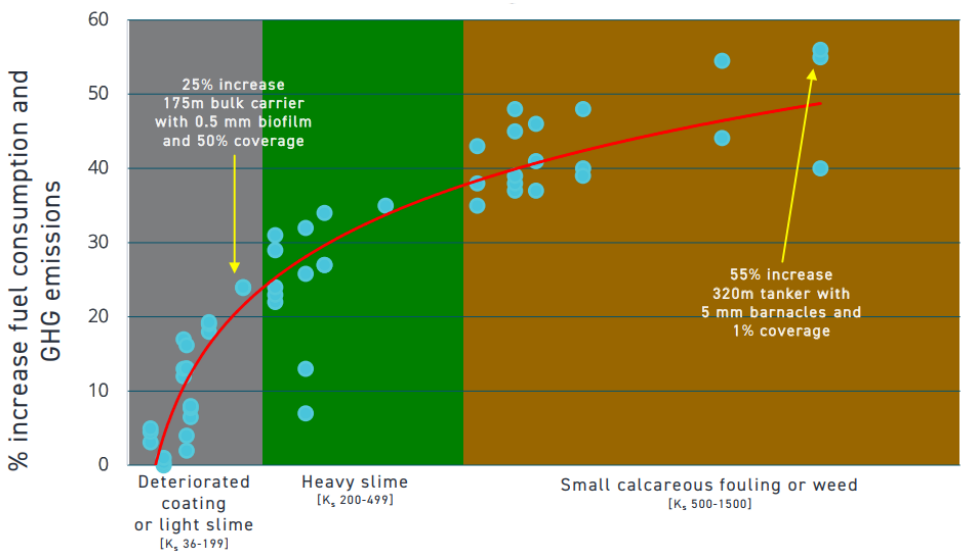


Fig.4: Additional fuel consumption caused by an increasing level of fouling, *IMO (2023)*

Aware of the fuel penalty, most vessels now focus on keeping fouling at slime levels, and avoiding any macrofouling. However, the IMO guidelines does not provide a differentiation between light and medium slime levels, nor do they quantify the spatial extent of coverage - a parameter that most cleaning services already document in their field reports.

On the other hand, the NACE SP21421 standard by *AMPP (2025)*, proposes a high degree of granularity for slime levels but offers limited detail for heavy macrofouling coverage.

The NSTM rating is a comprehensive system that unfortunately lacks a metric for hull coverage, which is essential for accurate hydrodynamic drag estimation.

The working group for Guide 21517 elected to adopt fouling ratings closely aligned with the SP21421 standard. However, an additional metric for coverage exceeding 40% was incorporated to more accurately account for high-density fouling rates and their associated impact on vessel performance.

<table>
<tr><th>Code</th><th>Degree of Biofouling</th><th>Correlation with SP21421</th><th>Correlation with IMO guidelines</th><th>Correlation with NSTM rating</th></tr>
<tr><td>0</td><td>Clean Hull</td><td></td><td>0</td><td>0</td></tr>
<tr><td>L</td><td>Light microfouling (‘thin slime’)</td><td>L</td><td rowspan="2">1</td><td>10</td></tr>
<tr><td>M</td><td>Medium microfouling (‘moderate slime’)</td><td>M</td><td>20</td></tr>
<tr><td>H</td><td>Heavy microfouling (‘thick algal slime’/’emergent beard’)</td><td>H</td><td>2</td><td>30</td></tr>
<tr><td>1</td><td>Up to 1% coverage by macrofouling</td><td>1</td><td rowspan="3">3</td><td>40</td></tr>
<tr><td>5</td><td>Up to 5% coverage by macrofouling</td><td>5</td><td>50</td></tr>
<tr><td>10</td><td>Up to 10% coverage by macrofouling</td><td>10</td><td>60</td></tr>
<tr><td>15</td><td>Up to 15% coverage by macrofouling</td><td>15</td><td rowspan="2">4</td><td>70</td></tr>
<tr><td>40</td><td>Up to 40% coverage by macrofouling</td><td rowspan="2">15+</td><td>80</td></tr>
<tr><td>+</td><td>>40% coverage by macrofouling</td><td>5</td><td>90-100</td></tr>
</table>

Fig.5: Approximative correlation of alternative fouling levels standards with the Guide 21517 scale (on the left)

Guide 21517 adds 3 additional Boolean indicators (checkboxes) to provide a more nuanced understanding of the hull's condition:

- Calcareous species >5%: Specifically flags high-biosecurity-risk organisms.
- Species higher than 1cm in height >5%: Included because long filamentous algae can induce higher hydrodynamic drag than smaller, low-profile calcareous macrofouling.
- Coating condition: : Essential for evaluating the probability of accelerated fouling regrowth.

2.2.5. Fouling Images

The standard permits the inclusion of two images per zone to illustrate both the maximum fouling score and the average fouling score. This dual-image requirement ensures compliance with New Zealand’s entry standards, *MPI (2023),* and facilitates long-term verification of reported fouling ratings.

2.3. Data Output and Schema

In the DFR, each zone is assigned a unique identifier (ID). The output is a structured data package comprising:

- Vessel & Service Metadata (CSV): General data regarding the vessel, inspection tools, and cleaning methods.
- Hull Zone Scores (CSV): Unique IDs mapped to fouling scores for all 50-meter hull sections, both precleaning and post cleaning.
- Niche Area Scores (CSV): Detailed pre- and post-cleaning fouling ratings for all 93 niche area classifications.
- Image Repository: An organized folder of photographs documenting the maximum and average fouling rates for each ID.

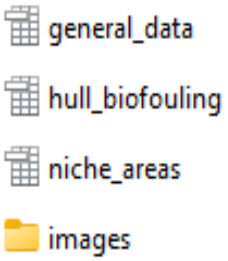


Fig.6: Example of organization of the CSV output

The complete list of parameters and IDs is available for purchase through the AMPP website. Together, these outputs represent the industry's first digital-ready format for hull biofouling. This standardized framework paves the way for advanced insights into cleaning strategies and coating efficiency, serving as a foundational dataset for sophisticated hull resistance calculations and fuel performance modeling.

3. First results and outlook

3.1. Generating Digital Fouling Reports

Shipping companies may now formally request DFRs from their underwater service providers. While adopting this new reporting format may require an initial administrative adjustment during the inspection or cleaning process, it is strategically designed to streamline operational workflows by:

- Standardizing Output: Relying on a single, industry-recognized template to satisfy all ship management requirements.
- Real-Time Reporting: Providing the opportunity to generate reports seamlessly during the active inspection.
- Improving Coordination: Clearly defining niche areas and inspection protocols to eliminate communication gaps between parties.

The more frequently ship managers request this format, the more rapidly the standard will achieve global adoption.

The official Excel template is available for purchase via the AMPP website. Additionally, OpenHull has developed a dedicated application designed to simplify DFR generation.

The OpenHull application can be utilized during active inspections or used to retroactively digitize historical PDF reports. This allows stakeholders to analyze fouling variations across previous dry-dock cycles.

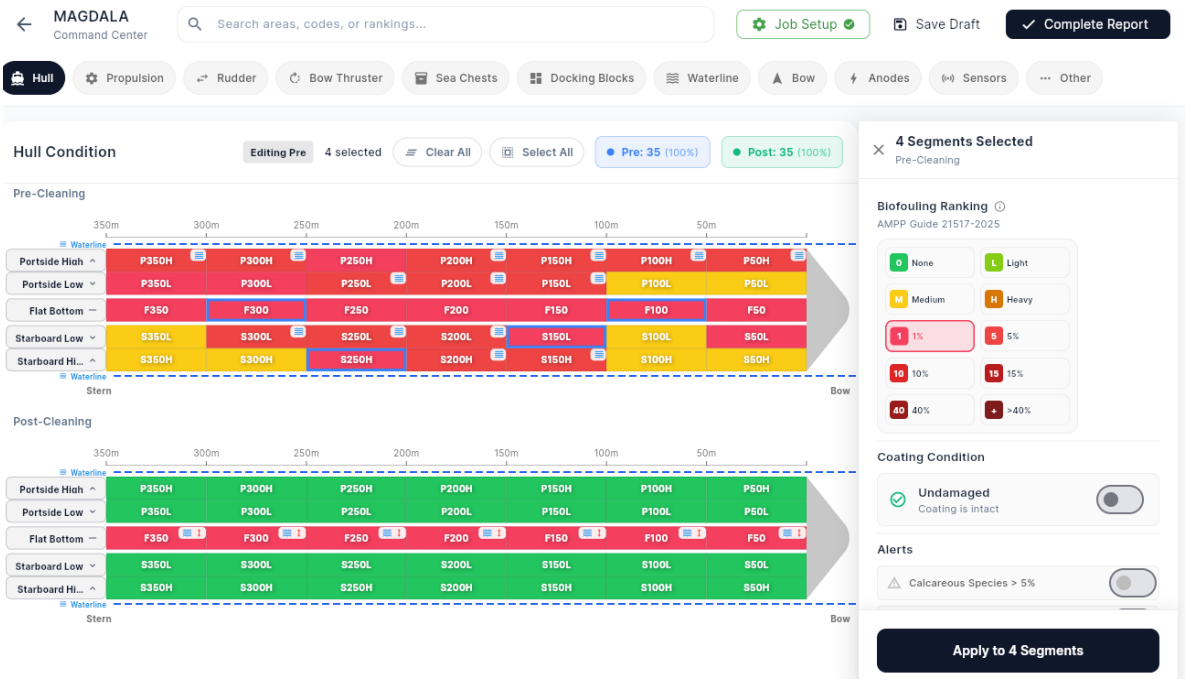


Fig.7: Generation of a Digital Report with OpenHull Sync

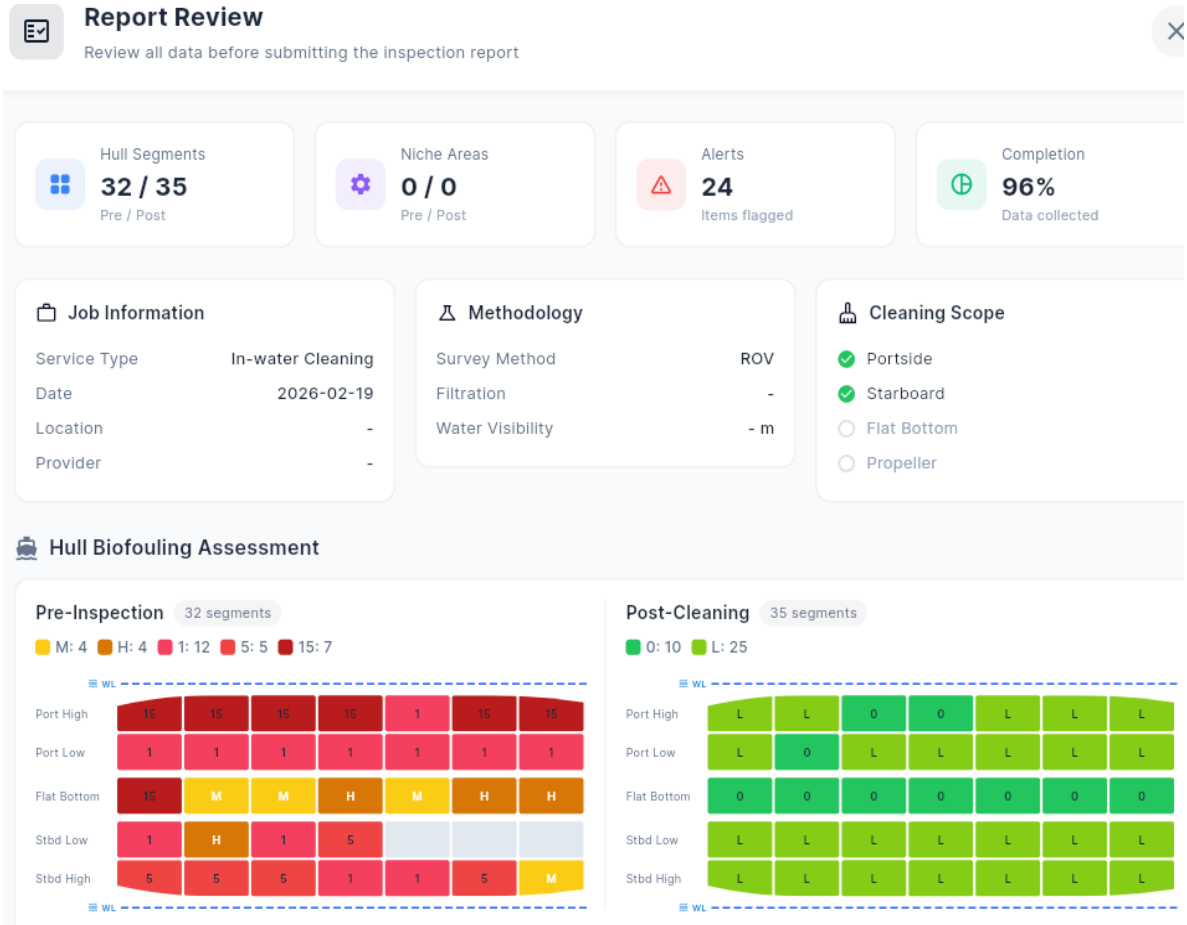


Fig.8: Online DFR viewed on OpenHull Sync

When digitizing legacy PDFs from various providers, the DFR operator must correlate existing hull zones and fouling nomenclatures with the new standardized format. This normalization of legacy data is a powerful tool for unlocking insights from past performance and refining hull resistance calculations.

Regardless of the data source (whether real-time inspection or legacy PDF analysis), the resulting output is hosted in a user-friendly online environment. Data can be exported into standardized Excel spreadsheets or integrated directly into third-party hull performance software via API or data-sharing protocols.

3.2. DFR comparison and value creation - Use case of a Bulk Carrier

We analyzed hull cleaning and inspection data from a bulk carrier throughout a full dry-dock cycle. A DFR, Fig.8, was retroactively generated for each documented hull event.

The DFR analysis provides a granular understanding of coating efficacy over time. By offering a direct overview of fouling severity and distribution, the DFR leads to the estimation of hull roughness, Fig.9.

When compared with traditional speed loss and hull resistance calculations derived from fuel consumption data, Fig.10, the DFR yields similar results through a significantly more straightforward calculation. In this model, hull roughness transitions from a derived output to a direct empirical input.

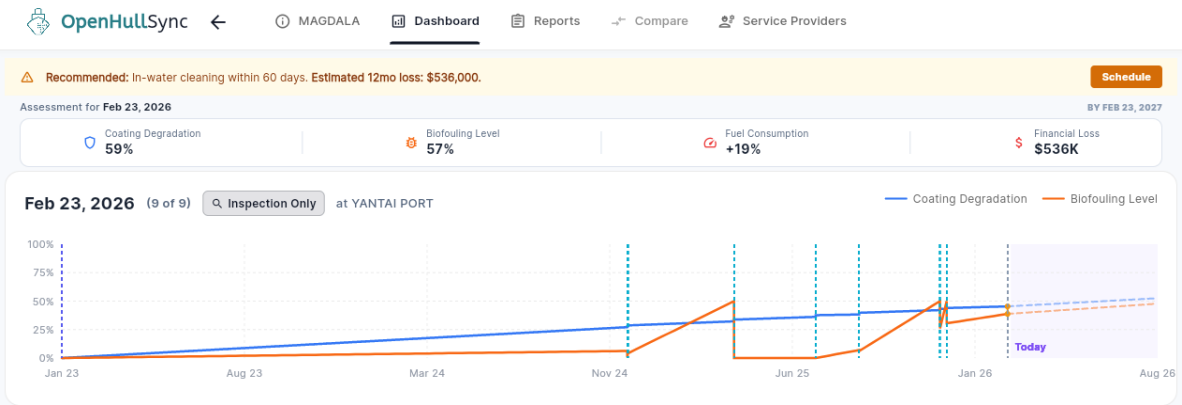


Fig.9: Evaluation of dry docking, fouling and cleanings over time according to the Guide 21517 AMPP digital fouling report

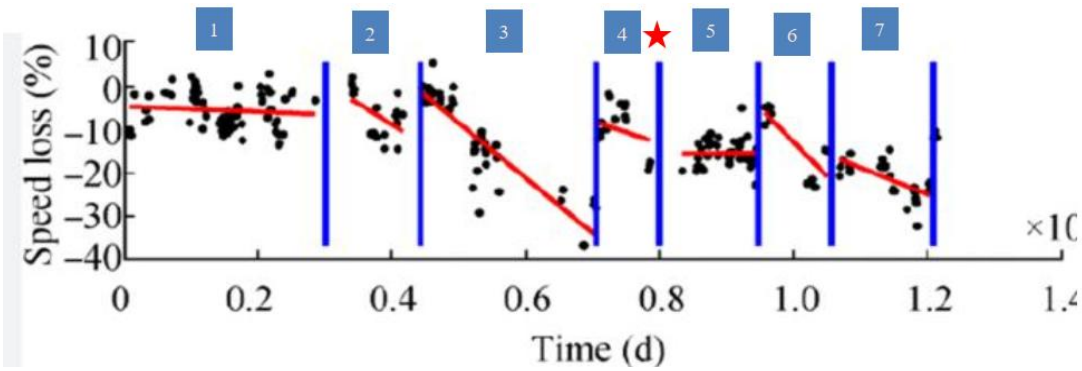


Fig. 10: Corrected Speed Loss over time, with cleaning events in blue, *Oliveira (2017)*

Fig.9 demonstrates that the coating performed optimally for the first 20 months. However, following the November 2024 cleaning event, fouling re-colonization accelerated significantly, necessitating cleanings every 2–3 months. This suggests that the cleaning methodology employed in November 2024 may have been mechanically incompatible with the coating technology, potentially over-polishing the surface.

The DFR provides additional specificity regarding biofouling height and spatial distribution. For example, given that fouling in the bow area has a disproportionate impact on hydrodynamics, this level of detail can allow for a significant refinement of existing hull performance models.

Finally, the DFR can be directly downloaded and shared with port authorities. Including images and the exhaustive list of niche areas, it is fully compliant with the requirements, for example, of the New Zealand MPI, effectively bridging the gap between ship operations and environmental regulation.

3.3. Discussion

The DFR currently presents certain inherent limitations. Depending on the technology employed, fouling degrees are defined by human experts, divers, or AI-driven tools. Given the global diversity of diving services and cleaning technologies, subjective variability between inspectors can decrease the statistical significance of drag estimations.

However, as the adoption of the standard spreads, the resulting library of labeled images and expanding datasets will facilitate the training of more sophisticated Machine Learning models. This will enable an objective assessment of hull conditions and improve the reliability of fouling data. Ultimately, the DFR is positioned to become the foundational dataset for AI-driven coating optimization and predictive maintenance.

The systematic generation of Digital Fouling Reports is currently in its starting stage. While promising, significant work remains to effectively transform granular hull data (such as species height, calcareous presence, and zone-specific density) into robust hydrodynamic drag coefficients. While foundational studies, such as the widely cited work by *Schultz (2007)*, propose resistance factors based on biofouling levels, the resistance models derived from DFR parameters will require gradual, empirical validation and refinement.

To fully unlock their potential, these resistance factors must be correlated with the biofouling pressure (e.g., vessel speed, salinity, water temperature) and specific operational profiles. This holistic approach will enable a deeper understanding of coating efficacy throughout its lifecycle, as well as the mechanical compatibility between specific cleaning technologies and coating chemistries.

To advance this field, future research should prioritize:

- Operational Profiling: Linking DFR fouling scores directly to vessel-specific trade patterns.
- Benchmarking: Developing multi-vessel database benchmarks for fleet-wide performance analysis.
- Degradation Modeling: Quantifying the cumulative impact of various cleaning methodologies on coating integrity.
- CFD Integration: Integrating digital fouling data with Computational Fluid Dynamics (CFD) roughness simulations.
- Biosecurity Metrics: Establishing quantitative risk assessments for regional biosecurity compliance.

4. Conclusion

Hull performance modeling has evolved rapidly over the past decade through the use of high-frequency operational data and advanced resistance algorithms. Yet, until now, one of the most influential variables in hull efficiency, the actual fouling condition, has remained largely fragmented and incompatible with digital performance systems.

AMPP Guide 21517 introduces the first interoperable digital framework for structured hull fouling reporting. By transforming underwater inspection observations into standardized, machine-readable data, the Digital Fouling Report (DFR) enables hull condition to become an input parameter in performance analysis rather than an inferred output derived from fuel consumption trends.

The case study presented demonstrates how structured fouling data can provide clearer insight into coating degradation patterns and cleaning compatibility than fuel-based indicators alone. While further validation is required to fully parameterize fouling-to-drag correlations, the framework establishes the foundation for such work.

As regulatory pressure intensifies (through carbon intensity targets, biosecurity requirements, and environmental scrutiny of in-water cleaning), the industry requires better visibility on hull condition and its performance implications. Standardized digital fouling data offers a practical and scalable path forward.

The maritime industry now possesses the tools to systematically integrate real fouling data into performance management. Broad implementation of digital fouling reports will be a decisive step toward more transparent, data-driven, and environmentally responsible hull management.

We invite all industry stakeholders to collaborate in the rapid and global adoption of the DFR standard.

Acknowledgement

We thank the 56 members of the AMPP Guide 21517 workgroup, and in particular Maria Leon from Biomarine Services for her precious help in writing the standard.

References

AMPP (2025), *NACE SP21421-2025 Pictorial Standard for Underwater Evaluation of Fouling Degree on Ship Hulls,* Association for Material Protection and Performance

BIMCO (2023a), *Independent Testing & Certification of In-water Cleaning Companies*, Baltic and International Maritime Council & International Chamber of Shipping

BIMCO (2023b), *Procedure for Testing and Certification of In-water Cleaning Companies*, Baltic and International Maritime Council & International Chamber of Shipping, p.27

IMO (2023), *2023 Guidelines for the Control and Management of Ships' Biofouling to Minimize the Transfer of Invasive Aquatic Species,* Res. MEPC.378(80), Int. Mar. Org., London, p.62

MPI (2023), *Craft Risk Management Standard: Vessels*, Ministry for Primary Industries, p.19

OLIVEIRA, D. (2017), *The enemy below-adhesion and friction of ship hull fouling*, Master thesis, Chalmers University of Technology, Gothenburg, pp. 56–67

SCHULTZ, M. (2007), *Effects of coating roughness and biofouling on ship resistance and powering,* Biofouling 23, pp.337-341

Vessel Performance Analyses Using Measured Wave, Current and STW Data vs. Using Hindcast Model Data

Gunnar Prytz, Miros, Asker/Norway, gp@miros-group.com
Alfredo Carella, Miros, Asker/Norway, ac@miros-group.com
Vemund Svanes Bertelsen, Miros, Asker/Norway, vsb@miros-group.com
Michael Schmidt, Copenhagen Commercial Platform, Copenhagen/DK, mhs@ccp-platform.com

Abstract

Vessel performance analysis is a critical component in achieving the ambitious emission reduction targets set for the global shipping industry. As regulatory frameworks tighten and sustainability goals become more urgent, understanding and optimizing vessel performance is essential for improving operational efficiency and reducing carbon footprints. This paper explores the impact of using accurately measured wave, current and speed through water (STW) data over model-based hindcast data for assessing vessel performance. We demonstrate that real-world data, collected through advanced monitoring technologies, provides a more precise and reliable basis for performance analysis than traditional model data, which often suffers from uncertainties and approximations. The study analyzes actual performance metrics from a number of vessels, offering insights into fuel consumption and operational behavior under various conditions. Our results show a significant improvement in the accuracy of performance evaluations when based on measured data, highlighting the potential for more effective decision-making and optimization strategies for emission reduction in maritime operations. The findings emphasize the need for adopting real-time data-driven approaches to meet international sustainability goals and support the shipping industry's transition towards a greener future.

1. Introduction

Vessel performance plays a vital role in modern maritime operations, shaping fuel efficiency, operational costs, environmental impact, and regulatory compliance. As the shipping industry faces increasing pressure to reduce emissions and improve energy efficiency, operators depend on reliable performance assessments to guide technical decisions and optimize day-to-day operations. High-quality data is essential in this context: inaccuracies in speed, fuel consumption, or environmental conditions can lead to misinformed decisions, obscured performance trends, and ultimately higher operational and environmental costs.

To support global decarbonization goals, the IMO has introduced several regulatory indices, including the design-based Energy Efficiency Design Index (EEDI/EEXI) and the operational Carbon Intensity Indicator (CII). While these indices represent important steps toward standardizing vessel efficiency evaluations, they each have limitations. The EEDI/EEXI reflects design conditions at the time of delivery and provides only limited insight into how a vessel performs in-service, where factors such as hull and propeller fouling, degradation, and retrofitted energy-saving devices significantly alter actual performance. Similarly, CII is strongly influenced by operational choices such as speed, loading condition, route, port stays in addition to the weather, and therefore does not isolate the vessel's underlying technical condition.

DNV recently published a Recommended Practice (RP) (DNV 2023) introducing the Vessel Technical Index (VTI) as a complementary performance metric. The VTI isolates a vessel's technical performance by comparing it to a defined reference state—typically the newbuild condition—thereby removing the influence of external operational factors such as weather and loading conditions. The RP also presents a new method for evaluating key sources of uncertainty, enabling users to make informed decisions based on VTI results.

The VTI method focuses on how a vessel's propulsion power is distributed among various consumers.

Although propulsion power is primarily used to move the vessel, a significant share may be spent counteracting wave-induced resistance in adverse weather. The magnitude of wave-related power losses depends on wave height, direction, and period. When multiple wave systems are present, each contributes to the overall wave-induced power loss. In some cases, waves may even assist propulsion by contributing positively to forward motion.

Similarly, wind forces also have a substantial impact on vessels. Its effect depends on wind speed and direction and can be either beneficial or detrimental, as is evident in sailing vessels. For ships equipped with wind-assisted propulsion (WASP) technologies, wind effects are highly significant. VTI analyses for such vessels can offer valuable insights into WASP efficiency across different operational conditions. Although such studies are ongoing, they fall outside the scope of this paper.

Water depth is another factor that influences propulsion power demand. In shallow waters, generally defined as less than 5–10 times the vessel's mean draft, added power is required compared to deep-water conditions. Moreover, seawater temperature and salinity affect the resistance encountered by the hull, and thus the power needed to achieve a given speed through the water.

Power losses may also result from navigational actions, such as changes in speed or course. Adjustments in trim (the difference between forward and aft draft) may likewise affect the power needed for propulsion.

To determine a vessel's true technical performance, the effects of vessel loading condition, environmental conditions and navigational behavior must be removed. Weather-related effects are typically eliminated through a process known as weather normalization, for which the DNV RP defines a specific set of methods. The RP also provides normalization procedures for correcting the influence of water temperature and salinity.

Navigational effects can be filtered out by excluding periods in which the vessel is changing speed or course. Due to system response lag, it is often necessary to remove an additional time period (e.g. 30 minutes) after such changes to ensure the vessel has returned to a steady state condition. The VTI method is outlined in Fig.1.

A comprehensive VTI analysis relies on high-frequency, high-accuracy data to fully leverage the method. To ensure reliable and meaningful results, the following data inputs are required for calculating the VTI:

- Directional wave spectrum (alternatively wave height, period and direction)
- Speed Through Water (STW)
- Speed Over Ground (SOG)
- Vessel heading
- Wind magnitude and direction (corrected to 10 m height above sea level)
- Water temperature and (if available) salinity
- Water depth
- Shaft power
- Shaft revolutions
- Vessel draft

The VTI analysis can be conducted in real time when the vessel is equipped with an advanced performance monitoring system, such as the Miros VTI service, which combines high- frequency, high-accuracy sensor data from the Miros Wavex system with a complete implementation of the DNV Recommended Practice. When such real-time capabilities are not available, the VTI can alternatively be computed retrospectively using recorded data or model data.

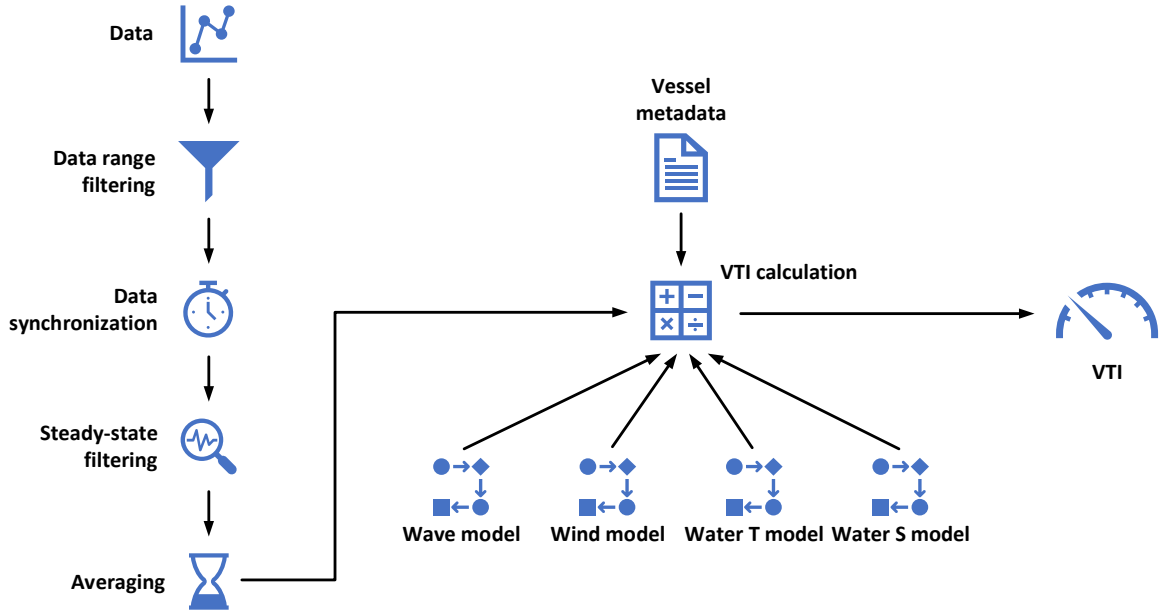


Fig.1: The VTI calculation process

Several applications of the VTI method have already been reported in the literature, *Bertelsen (2021), Guo (2021), Bertelsen (2022), Guo (2024), Prytz (2025). Gupta (2023a)* showed that the uncertainty associated with a VTI analysis can be significantly reduced when high-accuracy, directly measured sea state data is used instead of lower-quality measurements supplemented with hindcast data. Likewise, *Bertelsen (2022)* found a substantial improvement in accuracy when relying on high-quality measured data rather than hindcast estimates. Building on this work, *Gupta (2023b)* applied Principal Component Analysis (PCA) to further quantify and understand the sources of uncertainty in VTI calculations. *Prytz (2025)* showed how the VTI approach combined with accurate data could be used to rapidly identify the effect of propeller cleanings with high accuracy.

In this paper, the effect of propeller cleanings performed during two voyages is evaluated through a VTI analysis. The approach leverages accurate, high-frequency measurements of waves, speed-through-water (STW), and wind, combined with the latest weather-normalization techniques defined in the DNV RP, *DNV (2023)*. This enables the impact of the propeller cleaning to be isolated from external influences such as weather and navigational disturbances. In addition, VTI analyses from two long voyages without propeller cleanings is also presented. In practice, the method effectively performs a virtual sea trial, allowing a direct and rapid investigation of vessel performance during normal vessel operation.

2. Measuring waves, currents and STW based on X-band radar imaging

Radar-based sea state measurement technologies have advanced significantly over the past decade, *Gangeskar (2017,2018a,b,c,2019,2021).* Modern solutions are now capable of accurately measuring ocean waves, surface currents, and STW under a wide range of environmental conditions, and they deliver high levels of availability and reliability.

The Miros Wavex system derives its measurements from radar images that cover a localized area of interest, typically extending a few hundred meters ahead of the vessel. These images are processed using specialized algorithms to produce real-time wave spectra, integrated wave parameters, surface current vectors, and STW data, *Prytz (2019), Bertelsen (2020).* The analog video signal from a marine navigation X-band radar is digitized, enabling the extraction of high-resolution radar images. For radars that provide digital data output (IP-based radars), the images can be acquired directly. In addition, Wavex makes use of metadata from onboard sensors such as GPS and gyro compasses to support accurate processing. Further details on the core components and configuration of a Wavex system can be found in *Prytz (2019)* and *Bertelsen (2020).*

3. VTI analysis with measured vs. model input

This study evaluates the influence of data quality on the accuracy of VTI analyses by conducting parallel analyses using both high-frequency, accurately measured wave, wind and STW data and model-based wave, wind and current data.

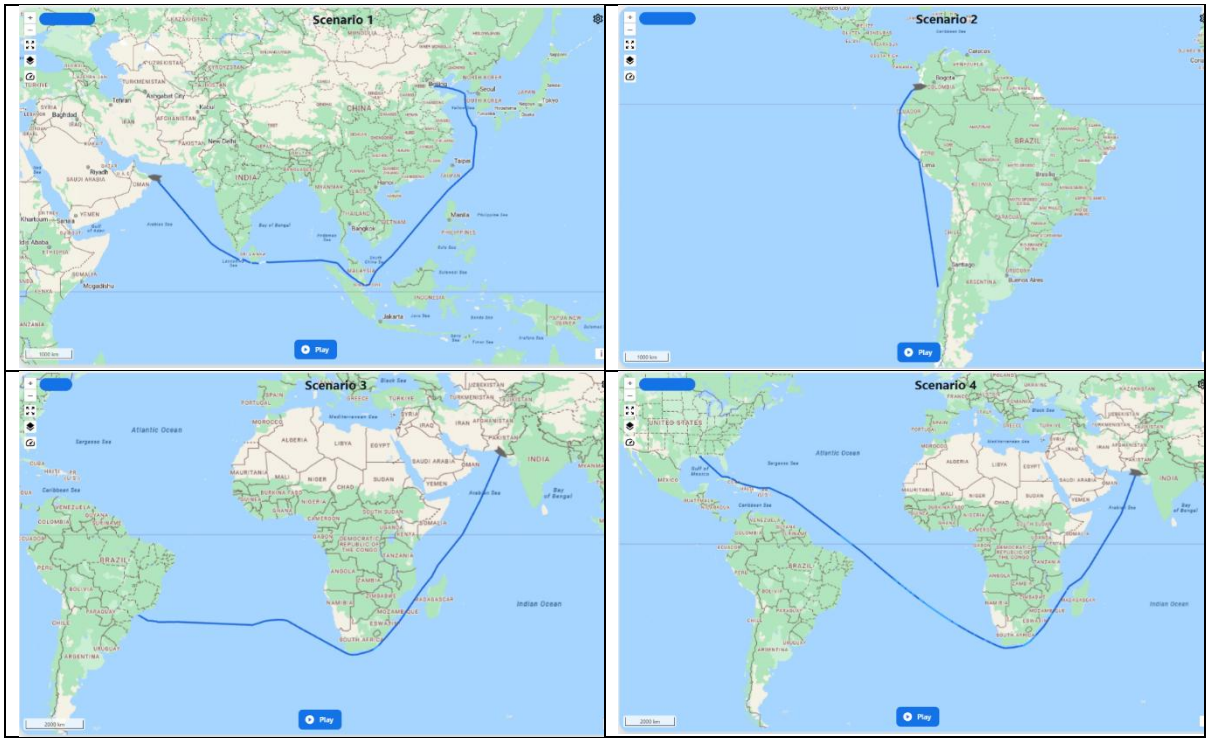


Fig.2: The 4 scenarios analyzed in this paper. Scenarios 1-2 included a propeller cleaning during the voyage while scenarios 3-4 were voyages with no propeller cleaning.

The VTI analysis was conducted for three cargo vessels with the use of the following equipment and services:

- Miros Wavex for measurement of directional wave spectra and speed through water
- Miros Edge platform for collection of propulsion data
- Miros VTI service for automatic calculation of the VTI

The Miros onboard system also collected wind data, vessel position and heading. Hindcast water temperature data from the EU Copernicus database was collected by the Miros VTI service. Water salinity was not used in the analysis. Finally, water depth data was also collected from the EU Copernicus service to filter out shallow water situations.

3.1. Scenario descriptions

A total of four scenarios were analyzed as shown in Fig.2:

- Fig.2a: An LNG carrier on a voyage from Tianjin, China to Qalhat, Oman with a propeller cleaning during a stopover in Singapore.
- Fig.2b: A bulk carrier on a voyage from San Vincente, Chile to Buenaventura, Colombia with a propeller cleaning during a stopover in Callao, Peru.
- Fig.2c: A bulk carrier on a voyage from Santos, Brazil to Karachi, Pakistan. There was no propeller cleaning during this voyage.
- Fig.2d: A bulk carrier on a voyage from New Orleans, USA to Karachi, Pakistan. There was no propeller cleaning during this voyage.

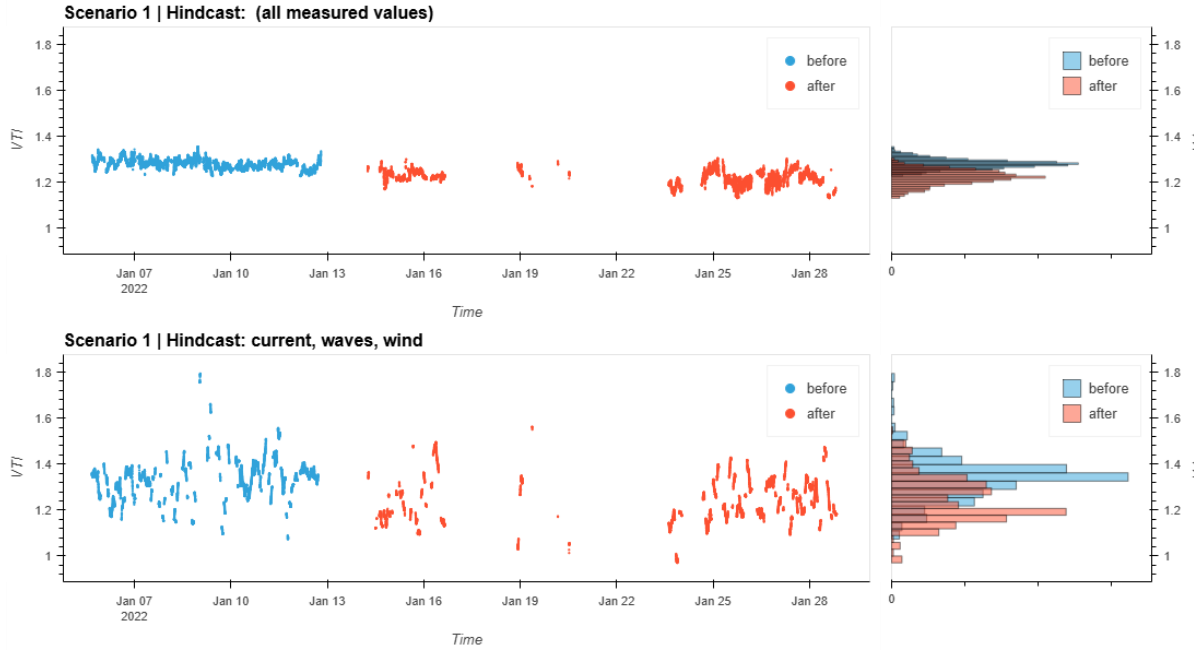
3.2.1. Scenario 1

This voyage consisted of two parts, one 8-day first leg, followed by a propeller cleaning, and then finally a 15-day second leg. The VTI for the complete voyage duration was calculated both using measured data and model data. The results are shown in Fig.3. It can clearly be seen that the VTI is significantly smoother and more stable when using measured data as compared to model data. The distributions of VTI values before/after the propeller cleaning are narrow when using measured data and considerably wider when using model data.

This is also demonstrated by the associated statistics. The standard deviation in the VTI is considerably higher when using model data. It is also worth noting that the change in performance as a result of the propeller cleaning can only be determined by using measured data. When using model data, the change in performance is not statistically significant.

3.2.2. Scenario 2

This voyage consisted of two parts, one 5-day first leg, followed by a propeller cleaning, and then finally a 4-day second leg. The VTI for the complete voyage duration was calculated both using measured data and model data. The results are shown in Fig.4. As for scenario 1, the VTI is considerably smoother when using measured data. The statistics paint a similar picture as in scenario 1. The effect of the propeller cleaning can only be found with statistical confidence when using measured data as input to the VTI.



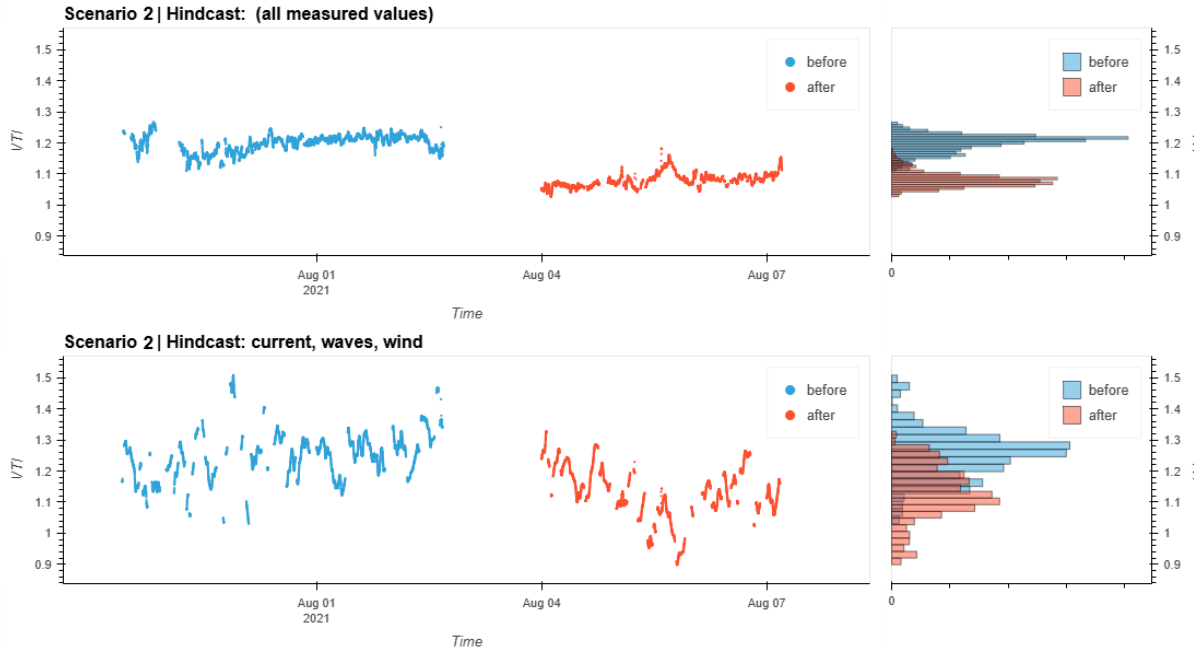
VTI statistics Scenario 1							
Input data	Before (8 days)		After (15 days)		Difference		Significant change
	Avg	Std	Avg	Std	Avg	Std	
Measured	1.28	±0.02	1.22	±0.03	0.06	±0.03	Yes
Model	1.33	±0.09	1.24	±0.10	0.09	±0.10	No

Fig.3: Results from the VTI analyses including VTI distributions before/after the propeller cleaning in scenario 1 with measured data (top) and model data (middle) used as input. Statistics for the two analyses are shown at the bottom.

3.2.3. Scenario 3

The third scenario was a 30-day voyage. There was no propeller cleaning performed during the voyage and so the performance is expected to be relatively constant. The VTI for the complete voyage was calculated both using measured data and model data. The results are shown in Fig.5.

As for the earlier scenarios, we can see that the VTI is considerably smoother when using measured data. Both VTI time series contain some variations on a time scale of hours and days. This is not investigated further in this paper, but it is likely that such variations are due to effects that are not captured by the VTI method. The statistics again show that the results using measured data as input are considerably more accurate than when using model data. However, the lower-frequency variations seen in the VTI across the voyage result in higher inaccuracies for both approaches.



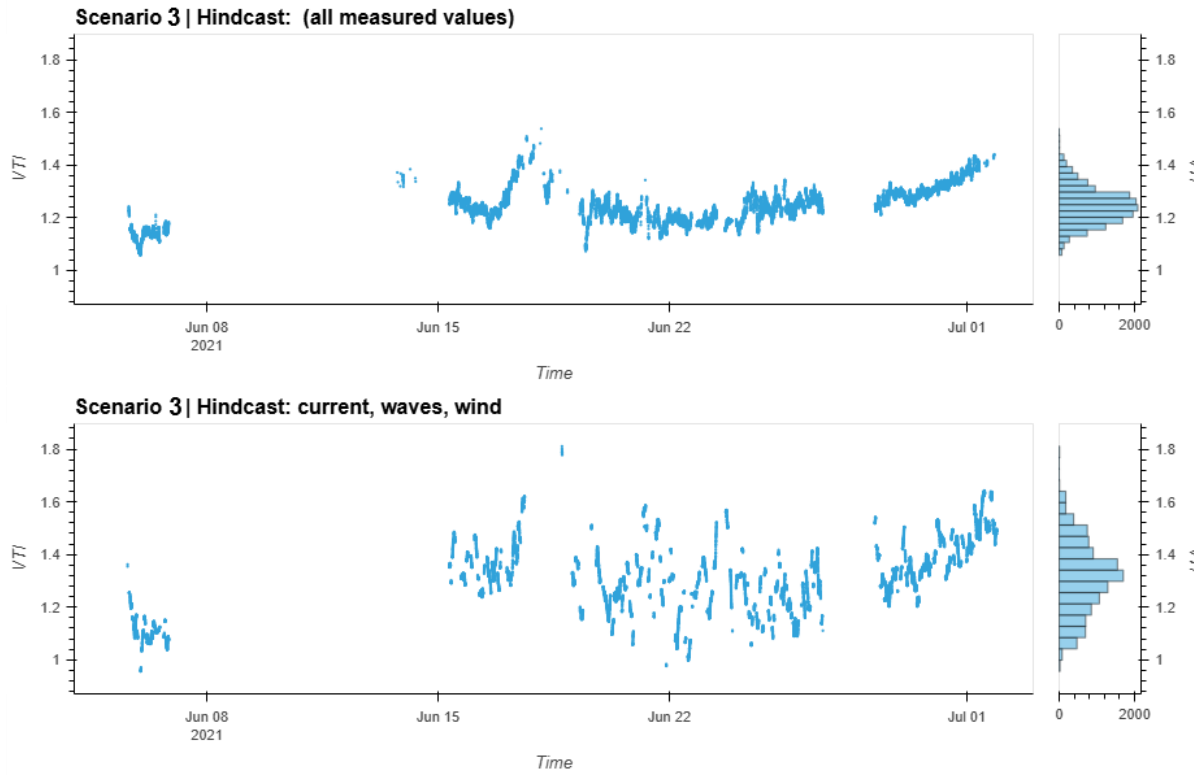
VTI statistics Scenario 2							
Input data	Before (5 days)		After (4 days)		Difference		Significant change
	Avg	Std	Avg	Std	Avg	Std	
Measured	1.20	±0.03	1.08	±0.02	0.12	±0.03	Yes
Model	1.25	±0.08	1.13	±0.09	0.12	±0.09	No

Fig.4: Results from the VTI analyses including VTI distributions before/after the propeller cleaning in scenario 2 with measured data (top) and model data (middle) used as input. Statistics for the two analyses are shown at the bottom.

3.2.4. Scenario 4

The third scenario was a 50-day voyage. There was no propeller cleaning performed during the voyage and so the performance is expected to be relatively constant. The VTI for the complete voyage was calculated both using measured data and model data. The results are shown in Fig.6.

As for the earlier scenarios, we can see that the VTI is considerably smoother when using measured data. As in scenario 3, both VTI time series contain some variations on a time scale of hours and days, although not as prominent as in voyage 3. The statistics again indicate that the results using measured data as input are considerably more accurate than when using model data.



VTI statistics Scenario 3		
Input data	Full voyage (30 days)	
	Avg	Std
Measured	1.24	±0.07
Model	1.31	±0.13

Fig.5: Results from the VTI analyses of the 30-day voyage in scenario 3 with measured data (top) and model data (middle) used as input. Statistics for the two analyses are shown at the bottom.

4. Discussion

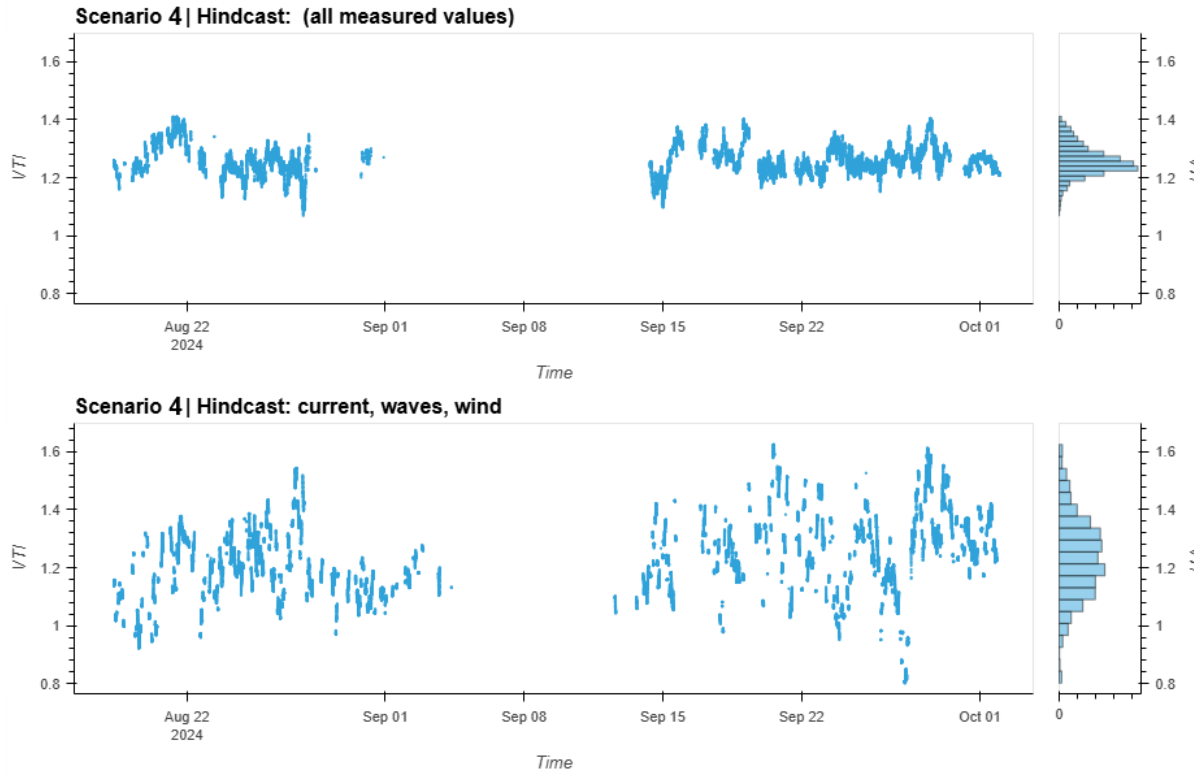
This paper has investigated the technical performance of cargo vessels in four different scenarios. Scenarios 1 and 2 analyzed vessel performance before and after propeller cleaning events by applying a VTI analysis method. Scenarios 3 and 4 analyzed vessel performance during long voyages.

In all cases it was clear that using accurately measured data gave results with considerably higher accuracy as compared to using model hindcast data. In scenarios 1 and 2, the effect of the propeller cleanings could only be determined with statistical significance when using accurately measured data. The inaccuracy of the VTI analysis was 3-4% in these two scenarios when using measured data. When using model data, the inaccuracy increased to 9-10%.

Two long voyages were analyzed in scenarios 3 and 4. Although the inaccuracies were higher in these scenarios, the pattern with considerably better results when using measured data was repeated. While the VTI analysis using measured data yielded inaccuracies of 5-7%, using model data increased the inaccuracies to 13-14%.

The long voyages in scenarios 3 and 4 displayed some variations in the VTI that are likely due to effects that are not handled by the present VTI approach. Such effects can be due to shortcomings in the normalizations procedures for the environmental effects, e.g. wave effects that are not captured. In

addition, there may be other effects like vessel rolling and pitching combined with rudder movement coming into play. A study of such effects is beyond the scope of this paper. However, work is ongoing to further develop the DNV method to address such issues.



VTI statistics Scenario 4		
Input data	Full voyage (50 days)	
	Avg	Std
Measured	1.26	±0.05
Model	1.34	±0.14

Fig.6: Results from the VTI analyses of the 50-day voyage in scenario 4 with measured data (top) and model data (middle) used as input. Statistics for the two analyses are shown at the bottom.

A change in vessel technical performance as described by the VTI can be converted to a change in fuel consumption by using a vessel fuel table. Such a dynamic fuel table, taking into account both the VTI of the vessel and the environmental conditions experienced by the vessel is an integral part of the Miros VTI service. In scenario 1, the propeller cleaning resulted in a theoretical fuel saving (assuming a 31-day voyage at 16 kn speed and a fully laden vessel) of 78 tons or 46 500 USD per month (assuming a fuel cost of 600 USD/ton). In scenario 2, the propeller cleaning resulted in a theoretical fuel saving (assuming a 31-day voyage at 13 kn speed and a fully laden vessel) of 65 tons or 39 000 USD per month (assuming a fuel cost of 600 USD/ton).

This clearly shows that doing a propeller cleaning can have a significant impact on the vessel's fuel consumption and corresponding running cost. An accurate VTI analysis performed in real-time provides a good decision support tool in order to perform the right type of maintenance at the right time. Furthermore, doing a combined hull and propeller cleaning for a significantly fouled vessel can rapidly unlock a savings potential of several hundred tons of fuel per month and corresponding savings beyond 100 kUSD per month.

In this paper we have compared using high frequency measured data versus hindcast model data for the analysis of vessel performance. In many situations, vessel performance calculations are instead based on noon report data, which is significantly worse when it comes to determining the actual efficiency of a vessel. Furthermore, with an inaccurate vessel performance system, shipowners and charterers may tend to postpone cleaning operations till they have more data to verify the fouling condition of a vessel, i.e. they can easily perform 1-2-3 voyages before they actually clean the vessel. Thus, the lack of an accurate vessel performance analysis method can lead to considerable losses in fuel and cost.

5. Conclusion

This paper has demonstrated that the VTI method can rapidly identify the performance of a vessel with high accuracy when using accurate, high-frequency measured data as input. The accuracy of these vessel performance analyses was considerably higher when using accurately measured data as input as compared to using model data. This approach can therefore be used in both the planning and evaluation of hull maintenance activities. The demonstrated accuracy enables a new level of condition-based maintenance within hull and propeller management. Furthermore, this approach can be used to accurately assess the effectiveness of energy saving devices and to define better performance-based contracts.

The analysis of the four scenarios clearly shows that the accuracy of the input data is a critical determinant of the reliability of the resulting VTI estimates. Analyses based on accurately measured waves, wind, and STW data consistently yielded significantly better VTI results. This enables ship owners and operators to efficiently unlock large fuel saving potentials while meeting sustainability goals and addressing new regulations.

References

BERTELSEN, V.S.; GANGESKAR, R.; PRYTZ, G.; SCHMIDT, M. (2020), *Accurate Voyage Sea State and Weather Measurements Improve Performance-Based Vessel Management*, 5th HullPIC Conf., Hamburg, pp. 157-171

BERTELSEN, V.S.; BRUSET, T.; GANGESKAR, R.; PRYTZ, G.; SCHMIDT, M. (2021), *Accurate Speed Through Water Measurements Enable Accurate Vessel Performance Management*, 6th HullPIC Conf., Pontignano, pp.112-123

BERTELSEN, V.S.; CARELLA, A.; GANGESKAR, R.; PRYTZ, G.; SCHMIDT, M. (2022), *Vessel Technical Performance Analysis using Accurate Weather and Speed Through Water Data*, 7th HullPIC Conf., Tullamore, pp.260-274

DNV (2023), *Recommended Practice - Technical ship performance*, DNV-RP-0675, DNV, Hovik

GANGESKAR, R. (2017), *Automatically calibrated wave spectra by the Miros Wavex system – Accuracy verified*, Miros, Asker

GANGESKAR, R.; PRYTZ, G.; BERTELSEN, V.S. (2018a), *On-Board Real-Time Wave & Current Measurements for Decision Support*, 3rd HullPIC Conf., Redworth, pp.223-233

GANGESKAR, R. (2018b), *Verifying High-Accuracy Ocean Surface Current Measurements by X-Band Radar for Fixed and Moving Installations*, IEEE Trans. Geosci. Remote Sens. 56/8, pp.4845-4855

GANGESKAR, R. (2018c), *Surface Current Measurements from Moving Vessels by Wavex*, Miros, Asker

GANGESKAR, R. (2019), *Measuring the speed through water by Wavex*, Miros, Asker

GANGESKAR, R. (2021), *Automatic Radar Data Quality Control – Applying Deep Learning, AI to Ocean Wave, Current Measurements*, Sea Technology 62/2, pp. 23-27

GUO, B.; ROGNEBAKKE, O.; TVETE, H.A.; ADAL, C.; STORHAUG, G.; SCHMIDT, M.; BRUSET, T.; PRYTZ, G. (2021), *Setting the standard for evaluation of in-service technical ship performance*, 6th HullPIC Conf., Pontignano, pp.88-105

GUO, B.; TVETE, H. A.; VARTDAL, B. J.; WANG, S. (2024), *Piloting of Vessel Technical Index,* 43rd Int. Conf. Ocean, Offshore and Arctic Engineering (OMAE), Singapore

GUPTA, P.; GUO, B.; STEEN, S.; TVETE, H.A. (2023a), *Reliable Hull Performance Analysis using Vessel Technical Index*, 15th HIPER Symp., Bernried

GUPTA, P.; STEEN, S.; GUO, B.; TVETE, H.A. (2023b), *Evaluating vessel technical performance index using physics-based and data-driven approach*, Ocean Eng.

PRYTZ, G.; GANGESKAR, R.; BERTELSEN, V.S. (2019), *Distributing Real-Time Measurements of Speed Through Water from Ship to Shore*, 4th HullPIC Conf., Gubbio, pp.114-127

PRYTZ, G.; CARELLA, A.; SVANES BERTELSEN, V.; SCHMIDT, M. (2025), *Accurate Analysis of the Effect of Propeller Cleaning using the Newly Established Vessel Technical Index,* 10th HullPIC Conf., Mulheim, pp. 236-244

Cost-Benefit Analysis of Hull Coating Technologies for Swire Shipping & Bulk Vessels: A Comparative Study of SPC and FRC Performance

Ashwani Gore, Swire Shipping, Singapore, ashwani.gore@swireshipping.com
David Pang, Swire Shipping, Singapore, david.pang@swireshipping.com

Abstract

This study conducts a practical cost-benefit analysis of self-polishing copolymer (SPC) and foul-release silicone (FRC) hull coating technologies across Swire Shipping and Swire Bulk vessel fleets operating in diverse trading routes. Using the ISO 19030 framework and dynamic speed-loss tables derived from quality-controlled noon reports, the analysis evaluates the performance degradation from the paints, meanwhile including post-cleaning recovery effectiveness and lifecycle economic impacts. The study encompasses 8 bulk carriers and 8 liner vessels with coating ages ranging from 6 months to 5 years, operating in mixed exposure conditions. Results indicate that FRC coatings generally demonstrate better performance stability compared to SPC coatings under expected operated conditions. Premium FRC and SPC coatings can reduce excess fuel consumption due to hull inefficiency significantly despite higher capital costs. The analysis reveals that coating selection impacts operational expenditure, with fuel penalties constituting up to 60% of total lifecycle costs.

1. Introduction

The accumulation of marine organisms on ship hulls, known as biofouling, represents one of the most significant operational challenges facing modern shipping fleets. The frictional resistance caused by hull roughness increases fuel consumption and greenhouse gas emissions, while substantially raising operational costs for vessel operators, *Schultz (2007)*. The choice of appropriate hull coating technology has become a strategic decision for shipping companies including operational efficiency, revenue performance as well as environmental compliance. In the last decades, self-polishing copolymer (SPC) coatings have been widely used due to their established biocidal efficacy and performance properties.

Self-polishing copolymer (SPC) coatings have traditionally dominated the maritime industry due to their proven biocidal efficacy and performance characteristics. These coatings function through controlled release of copper-based biocides that prevent marine organism attachment, hereby maintaining a smooth surface profile throughout the coating lifecycle, *Kwon et al. (2020)*. However, there are increasing concerns on the biocide discharge by SPC coatings and the growing emphasis on vessel efficiency has thus sparked interest in alternative technologies.

Foul-release coatings (FRC), predominantly silicone-based systems, represent a non/low-biocidal alternative that leverages ultra-low surface energy to prevent organism adhesion. FRC coatings create a surface to which organisms cannot effectively attach, allowing self-cleaning through vessel sailing movement through waters, *Lagerström et al. (2022)*. Despite higher initial costs and sensitivity to mechanical damage, FRC coatings offer potential advantages in reducing environmental impact and ensuring hull efficiency.

The selection of optimal coating selection for diversified fleets presents substantial complexity. Factors including trading route characteristics, vessel operational profiles and environmental regulatory requirements heavily influence paint performance and economic returns, *Kim et al. (2022)*. Moreover, the interaction between coating type and maintenance strategy also creates an additional layer of decision complexity that is beyond simplified analysis.

In response to these challenges, the ISO 19030 standard enables a suitable framework for measuring changes in vessel-specific performance over time, *ISO (2016)*. By establishing baseline conditions and tracking speed-loss relative to defined reference states, ISO 19030 enables objective comparison of

coating technologies across different vessel types and operating environments. This methodology moves beyond generic manufacturer claims to provide vessel specific performance data.

This paper employs methodologies adopted by Coach Solutions to evaluate SPC and FRC coating performance across Swire Shipping's liner fleet and Swire Bulk's bulk carrier fleet. Through analysis of quality-controlled noon report data, vessel specific dynamic speed-loss profiles are developed which capture loss in hull performance. These technical performance metrics are then translated into economic terms through fuel penalty quantification analysis.

The analysis addresses three primary questions. One, how do SPC and FRC coating technologies compare in vessel performance degradation across containerships and bulk carriers? How do various SPC and FRC coatings perform relative to one another? Two, what are the lifecycle economic implications of coating selection when incorporating capital expenditure, fuel penalties, and maintenance costs? Finally, how can the fleet performance team develop a simple lifecycle costs analysis that optimize coating selection for each ship type? Through addressing these questions, this study aims to provide actionable insights in hull coating technology selection.

2. Methodology

2.1. ISO 19030 Performance Measurement Framework

The ISO 19030 standard establishes a possible methodology for monitoring vessel performance changes over time, providing the analytical foundation for this study. The framework defines performance through the speed, power, and displacement relationship normalized to reference conditions to isolate hull and propeller performance from environmental and operational variables, *Oliveria et al. (2022)*. The core metric employed is speed-loss percentage (SL%), calculated as:

$$SL\% = \frac{V_0 - V_m}{V_0} \times 100\%$$

V_0 represents the baseline speed at reference conditions such as post-drydocking, while V_m represents the measured speed at equivalent power and draft conditions at time m. This normalization is critical for paint performance assessment, since it removes the confounding effects of varying operational parameters and solely measures the impact of hull condition on vessel efficiency.

The standard mandates quality control filtering of operational data to ensure valid performance comparisons. As such, noon reporting data points are excluded based on conditions that fall outside acceptable ranges. Coach Solutions has developed their speed-loss percentage metric to access vessel performance, which is defined as the ratio between obtainable vessel speed with fouling and the obtainable vessel speed without fouling, with both conditions at engine Continuous Service Rating, *Coach Solutions (2017)*. In the form of speed percentage terms, the ideal condition for a vessel at newbuild condition will be determined at 100%, which clearly shows the relationship of the speed percentage term.

$$Speed\ Percentage = \frac{vessel\ speed\ including\ fouling}{vessel\ speed\ excluding\ fouling}$$

Speed-loss tables were constructed by binning validated data points according to required power and draft conditions at even keel. This discretization recognizes that paint performance degradation patterns vary with draft, since vessels operating at higher loading condition may experience accelerated fouling in specific regions.

2.2. Vessel Population and Cost-Benefit Methodology

The study encompasses 16 vessels across two distinct operating segments, providing representation of both paint types. Table I summarizes the vessel population characteristics.

Table I: Vessel Population and Coating Specifications

Vessel Class	Count	Coating Type	Application Period
Bulk Carriers (Handysize)	8	SPC (4) / FRC (4)	2020 – 2024
Liners (Trans-Pacific voyages)	4	SPC (2) / FRC (2)	2022 – 2023
Liners (Feeder 2-class)	4	FRC (4)	2025

The bulk carrier fleet provides the longest performance history, with paint ages ranging from 1.5 to 5 years at the analysis date. This temporal spread enables analysis of degradation trends across the later coating lifecycle. These bulk carriers, unlike their high-frequency liner vessels counterparts, operate under irregular sailing patterns and have extended port stays. These conditions require coating solutions engineered for durability in static environments rather than optimized for hydrodynamic performance during high-speed, continuous voyages. On the other hand, the Swire Shipping liner fleet offers comparative data across similar trading routes with different coatings, and also different FRC paints across similar class vessels.

The economic evaluation integrates three cost categories, from capital expenditure (CAPEX) for coating application to operational expenditure (OPEX) for fuel penalties and estimated hull cleaning costs for operational flexibility. Conventionally for fuel penalty calculations, speed-loss percentages are usually translated to fuel consumption impacts through the cubic relationship between speed and power. The fuel penalties are then usually derived by applying this power increment to operating hours, specific fuel oil consumption (SFOC) characteristics, and prevailing fuel prices. However, using a faster and realistic simulation approach to attain fuel penalties, this study will use vessels relevant speed-fuel tables evaluated by Coach Solutions to derive more accurate fuel performance values via a prevalent sailing speed and loading condition or draft. Correspondingly, the obtained main engine fuel consumption values at the assumed conditions will be applied with reference to the pertinent speed percentage values. These simulated values are also taken at calm sea conditions, removing any deviations in fuel performance that may be affected by weather.

Cleaning cost mainly includes hull cleaning events. For SPC coatings, cleaning frequency and aggressiveness are often tracked as operational variables, while FRC coatings are more evaluated on their ability to maintain performance without mechanical intervention. In this study, hull-cleaning expenses are derived from a generalized average cost calculated across multiple commercial invoices, while deliberately excluding any estimated off-hire opportunity costs for analytical simplicity.

3. Performance Analysis Results

3.1. Bulk Carriers Speed-Percentage Loss Trends and Excess Consumption

The objective of the bulk carriers' study is to compare the fouling behaviour of two types of SPC coatings and two types of FRC coatings used on bulk carriers. Applying Coach Solutions speed-loss percentage trends highlights distinct degradation patterns between SPC and FRC technologies. Fig.1 illustrates the 2025 speed-percentage trends for the selected bulk carriers coated with FRC systems. The optimal speed percentage or referred as the "Expected Performance", represents the vessel's ideal hull condition and is projected at 100% immediately after dry-docking. This performance then gradually declines according to the predicted speed-loss associated with hull coating degradation over the specified year. Comparably, the fuel consumption values at assumed conditions of the bulk carriers uniform average sailing speed and draft have been plotted in Fig.2 with reference to the pertinent speed percentage values. To compare the fuel performance differences on a fair basis, the excess consumption between simulated fuel consumption and expected performance consumption (assuming that bulk carriers have spent ~50% of their operational time sailing) is used as the metric to display the impact of fouling on fuel efficiency.

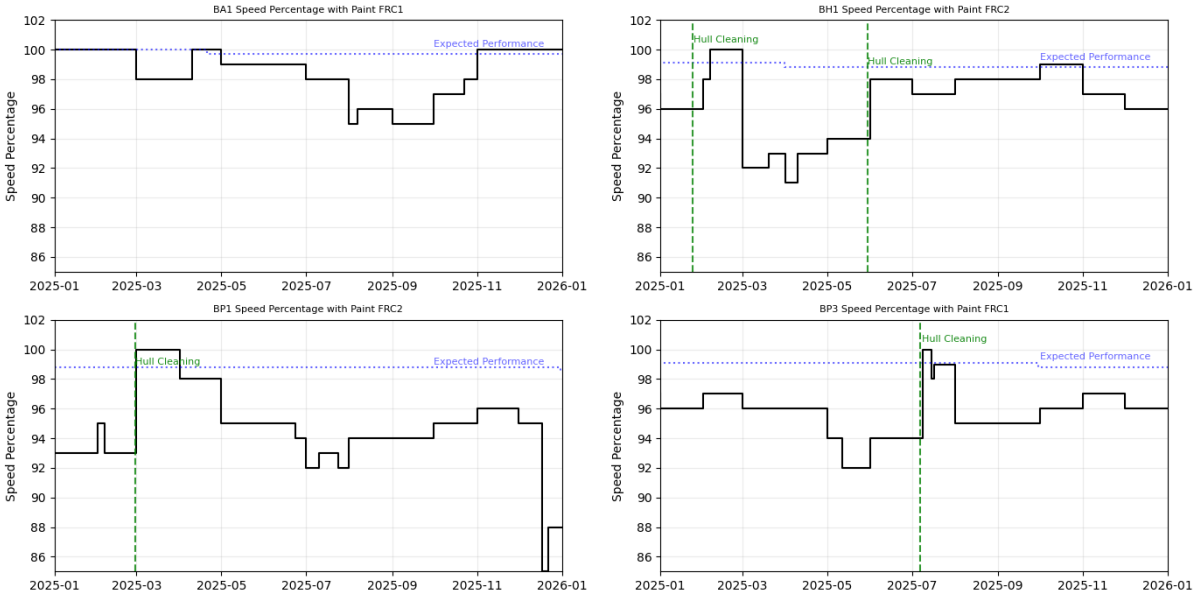


Fig.1: Bulk carriers with FRC paints speed-percentage loss trend

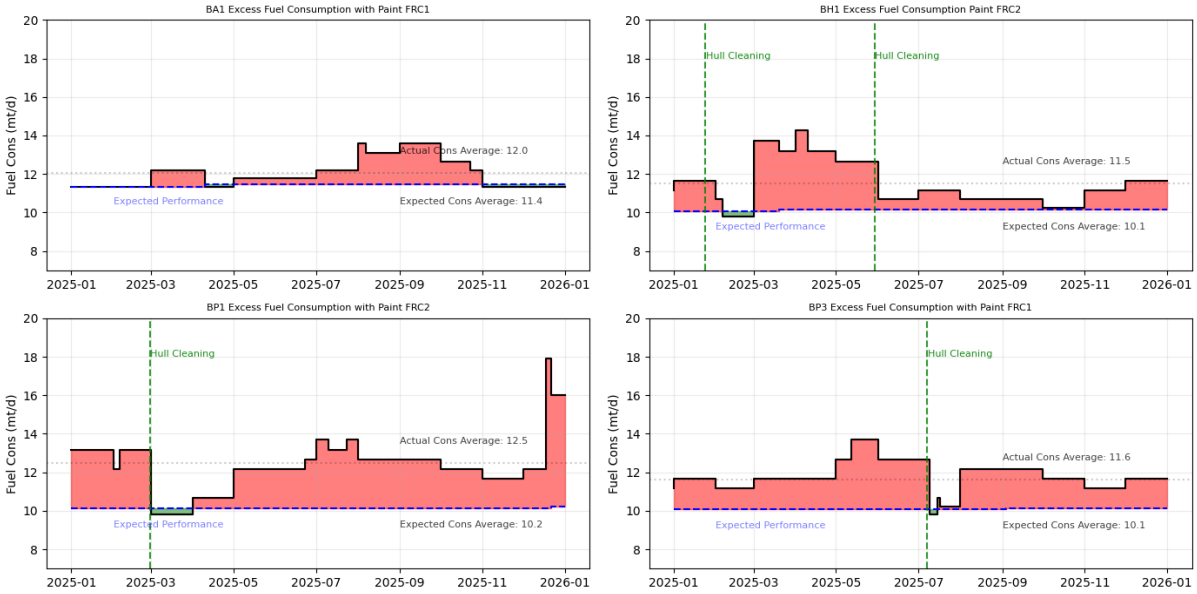


Fig.2: Bulk carriers with FRC paints excess consumption

Figs.3 and 4 are also plotted in similar context for the bulk carriers which are painted with SPC paint.

Table II compiles the bulk carriers paint type and age, together with the speed-percentage losses and derived excess fuel consumption, together with the exposure category.

Table II: Speed-loss performance by bulk carriers in 2025

Vessel	Paint	Age (Months)	Speed-Loss %	Excess Cons. (mt)	Exposure
BA1	FRC1	8 – 20	5.5 %	114.5	Mixed
BH1	FRC2	44 – 56	13.6 %	252.4	Mixed
BK1	SPC2	44 – 56	5.3 %	99.1	Mixed
BK2	SPC2	34 – 46	1.4 %	25.4	Mixed
BL1	SPC1	46 – 58	24.6 %	460.1	Mixed
BP1	FRC2	48 – 60	23.0 %	426.3	Mixed
BP2	SPC1	39 – 51	32.5 %	604.0	Mixed
BP3	FRC1	38 – 50	15.3 %	282.0	Mixed

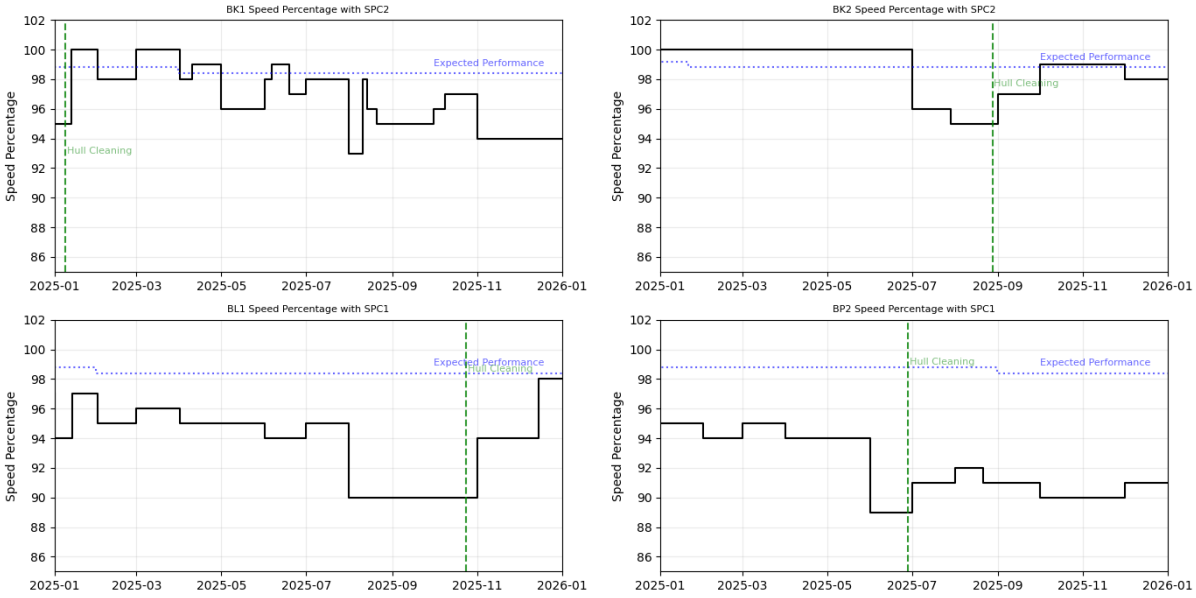


Fig.3: Bulk carriers with SPC paints speed-percentage loss trend

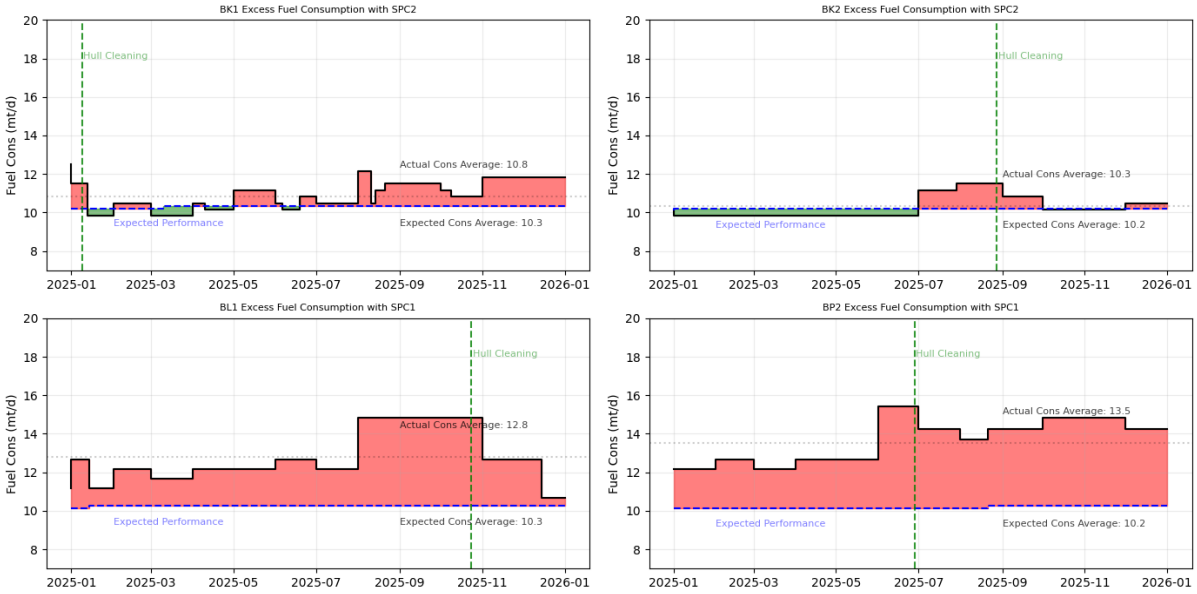


Fig.4: Bulk carriers with SPC paints excess consumption

FRC coatings demonstrate markedly different degradation characteristics. Both FRC1 and FRC2 showed speed-loss around 15% at around 48 months (4 years), with the exception of BP1. The 23% speed-loss at 48 to 60 months reveals critical vulnerability in FRC2 technology. Another possibility is that the paint has been exposed to mechanical damage. The hull cleaning inspections conducted in March 2025 confirmed that vessel BP1 had sustained mechanical damage to its coating, Fig.5.

These severe fouling events may require aggressive cleaning, generating costs exceeding typical SPC lifecycle expenditure. Furthermore, it underscores that some FRC paints performance is highly dependent on operational diligence and paint effectiveness. In addition, the observed speed loss at BH1 and BP3 appears higher than the expected performance for FRC paints.

For SPC coatings, SPC2 exhibited notably stable performance, showing only a 1–5% speed loss at nearly four years in service. In contrast, SPC1 coatings recorded a significantly higher speed loss of 25–33% between four and five years, highlighting the superior durability of SPC2.

Notably, SPC2 continues to perform well on bulk carriers despite their prolonged port-stay periods in different waters, which typically accelerate fouling.

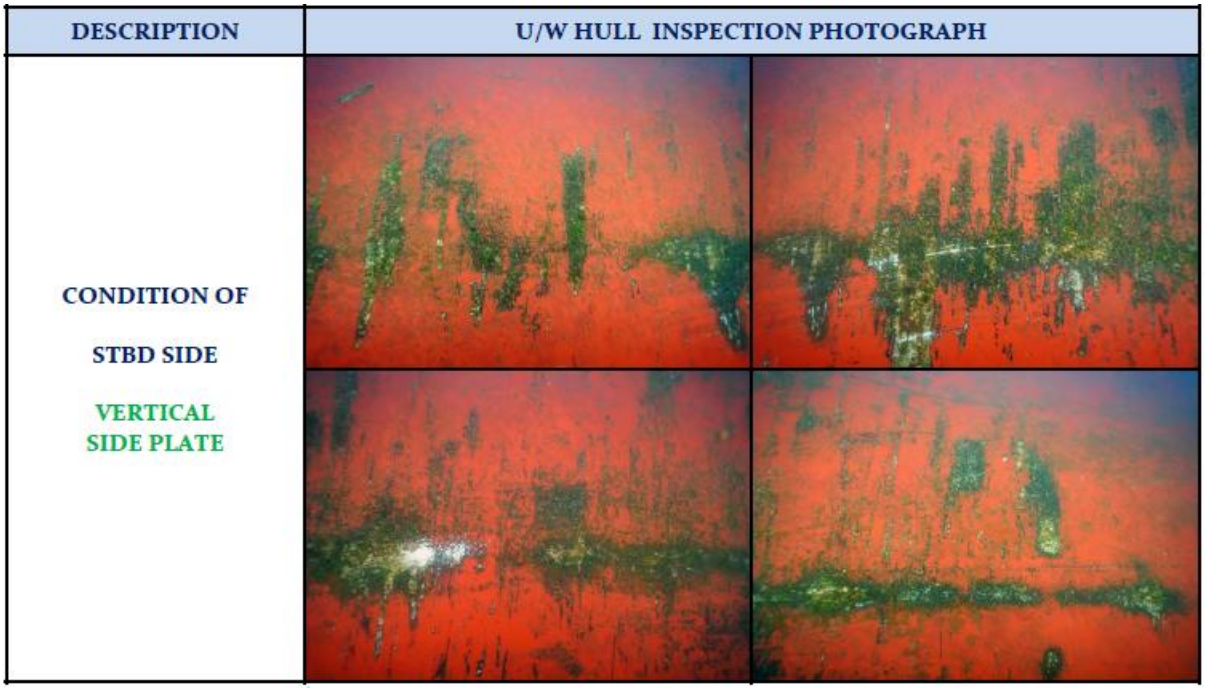


Fig.5: BP1 verticals paint damage March 2025

3.2. Containerships Speed-Percentage Loss Trends and Excess Consumption

On the containership analysis, two study groups were defined. The first group comprises of four vessels with near similar size and paint age operating along the Pacific trade routes in 2025. Two are coated entirely with SPC paint and the other two with FRC paint. The second group consists of four sister containerships that recently received different FRC coatings between early and mid-2025, providing only a six-month post-drydock observation window. The objective is to determine whether any of the applied FRC coatings exhibit noticeably poorer performance within the available evaluation period. As with the bulk-carrier case study, speed-percentage trends for the containerships are plotted alongside the corresponding fuel-consumption estimates, derived under average sailing speed and draft conditions with respect to the observed speed-percentage values, and also assumed sailing 65% of their total operating time. The figures are tabulated in Table III.

Table III: Speed-loss performance by containerships

Vessel	Paint	Age (Months)	Speed-Loss %	Excess Cons. (t)	Exposure
CW1	SPC3	25 – 37	3.0 %	64.3	Trans-Pacific
CW2	SPC3	21 – 33	2.1 %	94.9	Trans-Pacific
CM1	FRC1	15 – 27	1.9 %	100.9	Asia/Pacific
CN1	FRC1	14 – 26	2.8 %	72.3	Asia/Pacific
CC1	FRC2	4 – 10	4.4 %	73.2	Tropical
CH1	FRC1	3 – 9	1.6 %	27.3	Tropical
CG1	FRC3	2 – 8	14.2 %	238.1	Tropical
CP1	FRC1	1 – 7	2.9 %	49.4	Tropical

The containerships results reveal comparatively mild fouling effects for most coating systems, though notable variations still exist across paint types and operating regions. The SPC3 coated vessels (CW1 and CW2) show particularly stable performance, with speed-loss values remaining low at 2 to 3%

despite more than two years in service on the Trans-Pacific route. This indicates that SPC3 maintains a consistent antifouling performance under cooler waters and high-activity operational profiles.

FRC1 coatings applied to CM1 and CN1 similarly demonstrate strong resistance to fouling, with speed-loss values below 3% with around 2 years of service. Their performance suggests that FRC1 provides reliable fouling control in the Asia-Pacific environment, where vessels frequently encounter temperature and biofouling variability.

In contrast, performance differs significantly among the containerships which mainly operate in tropical waters. While CC1, CP1 and CH1 maintain modest speed-loss values between 1.6% and 4.4%, vessel CG1 which is coated with FRC3, has shown a pronounced 14.2% speed-loss at only 8 months of age. This high degradation indicates substantial vulnerability of the FRC3 coating in warm, high-fouling tropical environments and suggests possible early-stage coating failure. The elevated excess consumption value reinforces the severity of this deterioration.

4. Economic Impact and Life Cycle Analysis

4.1. Fuel Penalty and OPEX Impact

The bulk carriers and containerships case studies have translated speed-loss to fuel penalties or excess fuel consumption which reveals the potential economic impact of coating selection. To estimate the OPEX impact associated with fuel penalties, the excess consumption values are multiplied by the average LSFO price of $535 for 2025, *Ship & Bunker (2026)*.

For the bulk carrier examples, SPC-coated vessels (BL1 and BP2) incurred annual fuel penalties of $250,000 to $325,000 at 3 to 5 years post application, representing 25 to 33% excess consumption versus expected performance. FRC-coated vessels (BH1 and BP3) maintained penalties of $125,000 to $140,000, or 14–15% excess at equivalent ages, with the exception of BP1 where operational factors drove penalties to $225,000 or 23% excess.

The liner segment showed similar patterns at reduced absolute values due to vessel characteristics such as higher sailing frequency and less idling or port stays. The trans-Pacific SPC-coated vessels incurred 2.1 to 3.0% speed-loss in a 2 to 3 years frame, translating to $50,000 to $75,000 annual fuel penalties. The Asia-Pacific FRC-coated vessels achieved 1.9 to 2.8% loss in 14–26 months with penalties of $40,000 to $55,000.

Table V: Annual Estimated Excess Operating Cost Comparison (based on 2025)

Paint	Coat/s	Age Yrs)	Vessel Type	Fuel Excess	Cleaning	Total
FRC1	2	3-4	Bulk Carrier	$ 150,000	$ 0	$ 150,000
FRC1	2	1-2	Bulk Carrier	$ 60,000	$ 0	$ 60,000
FRC1	2	1-2	Containership	$ 45,000	$ 0	$ 45,000
FRC2	1	4-5	Bulk Carrier	$ 200,000	$ 40,000	$ 240,000
FRC2	2	0-1	Containership	$ 50,000	$ 0	$ 50,000
FRC3	2	0-1	Containership	$ 170,000	$ 0	$ 170,000
SPC1	1	3-4	Bulk Carrier	$ 350,000	$ 20,000	$ 370,000
SPC2	2	3-4	Bulk Carrier	$ 21,500	$ 20,000	$ 41,500
SPC3	2	2-3	Containership	$ 55,000	$ 20,000	$ 75,000

Taken together, these findings highlight that coating performance on bulk carriers carries more consequences on the fuel penalties due to the trading nature and the environmental conditions. SPC 1, SPC3, FRC1 and FRC2 consistently deliver stable results across diverse trades, whereas certain newer FRC systems, such as FRC3, may require closer monitoring when deployed in tropical regions where fouling pressure is significantly higher. Table V illustrates the expected total OPEX impact for 2025, including

fuel penalties and annualized hull-cleaning costs, where an average global hull-cleaning cost of roughly $20,000 per event is applied as a simplifying assumption.

The results from the tropical-waters containerships case study should be interpreted with caution, as the coating had been in service for less than one year and the analysis period covered only six months. Nevertheless, the findings have highlighted that fouling-release coatings do not perform uniformly, since FRC3 exhibited five to six times higher excess fuel consumption as compared to FRC1 and FRC2 on a similar class of vessels.

4.2. Capital Expenditure Comparison

Comparing capital expenditure across different paint types is essential because the choice of hull coating involves not only performance considerations but also significant upfront investment from shipowners. High-quality coatings may deliver substantial long-term operational savings, yet these benefits must be evaluated against the initial cost required at each dry-docking cycle. Coating application costs also vary significantly by technology, surface area, number of coatings and specification complexity. Table VI details the mentioned paints CAPEX for representative applications.

Table VI: Coating Application Costs

Paint	Coat/s	Vessel Type	Year	Actual Costs ($)	Present Costs ($)
FRC1	2	Bulk Carrier	2021	170,000	183,000
FRC1	2	Bulk Carrier	2024	175,000	179,000
FRC1	2	Containership	2025	195,000	195,000
FRC2	1	Bulk Carrier	2021	210,000	226,000
FRC2	2	Containership	2025	225,000	225,000
FRC3	2	Containership	2025	215,000	215,000
SPC1	1	Bulk Carrier	2021	145,000	156,000
SPC2	2	Bulk Carrier	2021	N.A.	320,000
SPC3	2	Containership	2023	N.A.	250,000

The present costs values are adjusted by applying a 2% annual cost escalation from the original year to reflect price appreciation over time. While the cost data for most FRC applications are based on verified quotations from 2021 to 2025, the values for SPC2 and SPC3 rely on estimated present-day prices for vessels of similar size, as actual quotes could not be obtained.

4.3. Lifecycle Cost Analysis

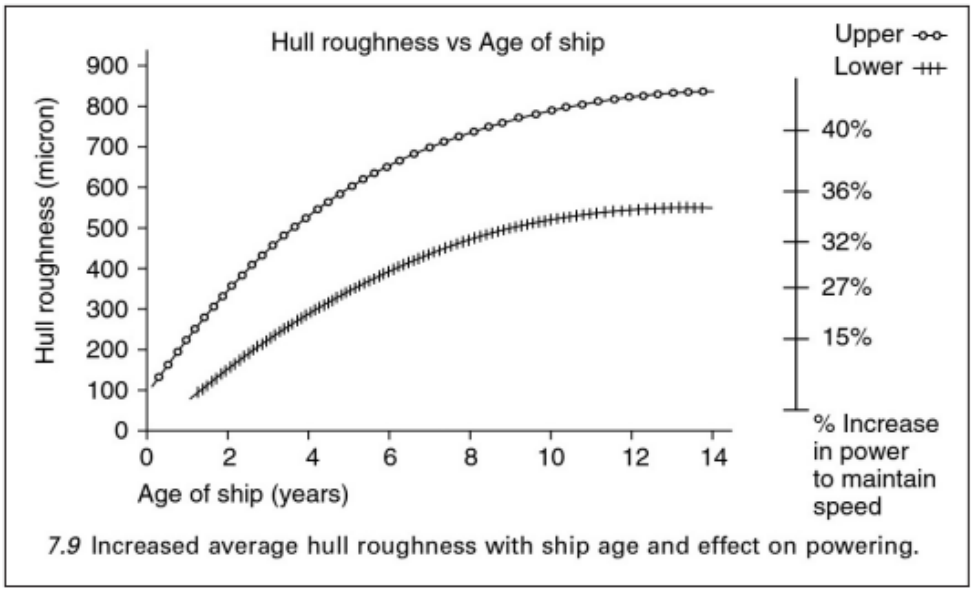
A structured comparison of capital expenditure provides an initial view of financial risk and payback potential, helping determine whether a coating aligns with the vessel's operational profile. When combined with estimates of fuel penalties and hull-cleaning costs, a 5-year lifecycle analysis can identify which coating delivers the most favourable overall performance. However, the earlier case studies only quantify fuel penalties for 2025. To obtain a complete 5-year OPEX profile, fuel penalties for the remaining years must be reconstructed. This can be achieved by applying a general hull-fouling degradation curve to back-cast or forecast the intermediate years using 2025 as the anchor point. Since fouling and coating degradation are inherently non-linear, they are expected to follow a logarithmic growth pattern in biofouling-induced resistance, consistent with the findings of *Munk et a. (2009)*, who reported a logarithmic increase in average hull roughness with vessel age, as illustrated in Fig.6.

Given that the fouling penalty starts at zero at drydock and then grows logarithmically,

$$F(t) = F_{known} \times [\frac{\ln(1+t)}{\ln(1+\ t_{known})}]$$

F(t) is the estimated fuel penalty or excess fuel consumption at year t after drydocking, F_{known} is the known 2025 fuel penalty, t is the number of years after drydocking and t_{known} refers to the number of

years after drydocking that corresponds to the 2025 data points. With the full 5 years of OPEX information available, Tables VII and VIII present the 5-year lifecycle costs analysis for the bulk carriers and containerships respectively as of 2025.



7.9 Increased average hull roughness with ship age and effect on powering.

7 Ibid, p 31.
8 International Marine Coatings, "Hull Roughness Penalty Calculator: The economic importance of hull condition," Akzo Nobel, 2004.

Fig.6: Increased average hull roughness with ship age, *Munk et al. (2009)*

Table VII: 5-year lifecycle economics (Bulk carrier scenario)

Paint	Coat	CAPEX	Cumulative Fuel Penalty ($)	Cumulative Cleaning Cost ($)	Total ($)
FRC1	2	180,000	420,000	80,000	680,000
FRC2	1	226,000	700,000	80,000	904,000
SPC1	2	145,000	1,000,000	120,000	1,265,000
SPC2	2	320,000	85,000	80,000	485,000

Table VIII: 5-year lifecycle economics (Containerships scenario)

Paint	Coat	CAPEX	Cumulative Fuel Penalty ($)	Cumulative Cleaning Cost ($)	Total ($)
FRC1	2	195,000	195,000	80,000	470,000
FRC2	2	225,000	320,000	80,000	625,000
FRC3	2	215,000	1,040,000	140,000	1,395,000
SPC3	2	250,000	170,000	80,000	500,000

The 5-year lifecycle assessment for the bulk carrier scenario highlights clear economic differences among the coating systems. Although SPC2 emerges as the most cost-effective option, the high capital layout may be a showstopper for shipowners. The traditional SPC1 coating results in the highest total cost of $1,265,000, driven primarily by its large cumulative fuel penalties despite having the lowest CAPEX. The FRC coatings fall between two extremes, where FRC1 provides a balanced outcome at $680,000, while FRC2 is notably more expensive at $904,000 because of its high fuel-penalty contribution. The containerships scenario is more straightforward, showing that FRC1 is the dominant option. Overall, the results demonstrate that coating selection cannot be based on CAPEX alone, as long-term operational behaviour like fouling-related fuel penalties will play a dominant role in determining lifecycle economics.

4.4. Sensitivity and Risk Analysis

Sensitivity and risk analysis are also considered to evaluate the robustness of the lifecycle cost results under key economic and operational uncertainties. Fuel price volatility represents the most influential parameter, given that fuel penalties constitute the dominant share of 5-year OPEX across both fleet segments. The sensitivity analysis indicates that fuel-price variation between $400 and 900 per ton has a pronounced impact at the upper bound, where total costs rise by over 25% for all coatings other than SPC1. At the lower bound of $400, the magnitude of cost variation remains comparatively limited.

Beyond fuel price movements, exposure related risks such as extended idle periods, tropical port calls, and irregular sailing schedules will pose performance uncertainty, particularly for FRCs that depend on vessel motion for self-cleaning. Mechanical damage risk is an important factor, as evidenced in BP1 where coating abrasion or impact damage resulted in disproportionately high fuel penalties. Cleaning related risks also contribute to lifecycle uncertainty, since aggressive in-water cleaning may accelerate coating wear, while insufficient cleaning can lead to biofouling accumulation and rapid speed-loss escalation. Coating application variability introduces an additional technical risk. Differences in surface preparation, required environmental conditions during application, and yards workmanship can generate performance deviations that exceed those predicted by the modelling framework. Regulatory risks must also be considered. Tightening restrictions on biocidal discharges such as emerging EU directives, can possibly elevate compliance costs for SPC coatings over the next few years, while FRC systems may face increased scrutiny in the near future.

This emphasizes that although FRC systems may demonstrate economic robustness across both fleets, their performance advantage is also largely dependent on diligent operation, coating integrity management and active vessels disposition. To protect anticipated lifecycle benefits, these risks underscore that coating selection decisions should be properly integrated with vessel type, route planning, hull maintenance strategies and performance monitoring regimes.

5. Strategic Recommendations

The study supports adopting mainly FRC coating strategies across both fleet segments. For the Swire Shipping liner fleet operating in tropical or Asia-Pacific trades, FRC coatings like FRC1 and FRC2 are recommended. The combination of high fuel costs over long ocean passages and the importance of schedule reliability all favour the stable performance of silicone based FRCs. Unfortunately, FRC3 has shown poor performance despite the recent implementation on CG3, hence more data is required to conclude its overall performance. Although SPC3 works well for CW1 and CW2, its effectiveness is strongly supported by the colder Pacific Ocean waters and long-distance voyages, both of which help limit fouling and preserve coating efficiency. However, unless these vessels are consistently assigned to Trans-Pacific routes, adopting FRCs would offer more robust and reliable performance across varying operating conditions.

As regards to the Swire Bulk fleet, a similar approach is advisable although SPC2 has shown its superiority in Table VII. High utilization bulk carriers operating on consistent routes can benefit from FRC1 coatings, despite the increased fouling risk if they suffer any mechanical damage because the associated fuel savings can be substantial. Although SPC2 coatings provide clear operational economic benefits, their high application cost requires significant upfront investment from shipowners. SPC2 also provides a much shorter fouling-free idle period than most FRCs and usually performs optimally only within demanded speed ranges, which may not align with the highly variable operational profile of bulk carriers.

For newbuildings and retrofit planning within a five-year drydock cycle, FRC1 may offer compelling economic advantages across both segments. Although they require approximately 20% higher capital expenditure than SPC1, they deliver more than 60% reduction in operating costs, resulting in a payback period within the few years of service. To maximise these benefits, coating specifications should emphasise application quality and include warranty provisions for mechanical damage protection.

6. Uncertainty Factors and Limitations

There are also several factors which introduce uncertainty into the cost-benefit analysis outcomes. The methodology which employs Coach Solutions dynamic speed-loss tables based on ISO 19030 framework, are estimating fuel consumption from modelling based on prevalent sailing conditions with reference to actual parameters. This estimation method has been perceived as a more reliable and fair methodology compared to directly comparing actual fuel consumption from the vessels, since actual fuel consumption will vary vessel to vessel due to external factors like weather and operational differences like speed. Modelling estimates eliminate these inconsistencies and provide a more unbiased approach using identical simulated conditions. However, modelling still relies on theoretical assumptions like calm sea weather in the case studies, or have inherent uncertainties in the modelling, which may not reflect actual reality, hence the conclusions may be seen as speculative rather than empirically validated.

The excess fuel consumption attributable to hull-coating degradation is estimated only for a defined period, such as a single year of 2025 or half a year period, thus requiring the remaining periods to be extrapolated using the logarithmic curve. Ideally, the full trend of speed-percentage loss since coating application should be used. However, this method has only been implemented by both Swire Shipping and Swire Bulk since 2024, resulting in an incomplete representation of total fuel penalties.

Coating performance is influenced by application quality factors not fully captured in this analysis. Thickness variation and surface preparation standards can cause performance variation exceeding the technology-specific differences identified. The study assumes uniform application quality across the population, which may not reflect actual yard performance variation.

Lastly, the cost-benefits analysis did not include possible financial considerations such as speed-loss in charter party penalties due to vessel inefficiency, or regulatory costs from new compliance requirements such as Fuel-EU or EU-ETS. The speed-loss caused by biofouling would be hard to be defined from simulation models analysis, and similarly the penalties from FuelEU or EU-ETS have not been imposed fully since setup, making it difficult to determine a possible penalty cost for the inefficiency.

7. Conclusion

This study presents a cost-benefit analysis of SPC and FRC technologies applied across the Swire Shipping and Swire Bulk fleets. The analysis uses Coach Solutions dynamic speed-loss tables and speed percentage losses measurement framework to ensure consistent evaluation. The results show that FRC technology may provide clearer performance and economic benefits when used under most or all operational conditions across both fleets.

The findings demonstrate that FRC coatings have overall stronger and consistent performance stability. The paint FRC1 has shown a speed-loss range of 2 to 5% over at the 2 to 3 years range. The economic results also favour FRC coatings. Although FRC systems cost 20% more at the start, they reduce operating costs by 60%. This comes from lower fuel consumption due to better hull efficiency and fewer cleaning events. FRC coatings however can be sensitive to operating conditions. They perform well when vessels remain in continuous service especially in containerships as seen from Swire Shipping. Long idle periods in warm, high-fouling waters can lead to sudden fouling growth. This may reduce the economic benefits of FRC, especially when they are exacerbated by mechanical damage to the paint itself. Operational vigilance and maintenance schedules must consider this risk.

Choosing SPC coatings usually leads to higher long-term operating expenses, resulting in the total cost of ownership becoming less favourable. SPC1 has shown a higher speed loss range over the same period as compared to the FRCs coatings. Nevertheless, the study indicates that some SPC coatings such as SPC2 are capable of surpassing FRC performance and offering better economic benefits, but these gains are typically limited to particular operational environments and requiring much higher capital layout.

In conclusion, coating selection should consider the entire lifecycle and other possible operational requirements. Shipping companies should focus on performance-based specifications and improved monitoring systems. This will help verify actual coating performance using similar ISO 19030 standards. Future work should examine new coating technologies such as biocide-free hybrid systems and graphene-based surfaces. Studies that observe coating performance over 5–10 years will also help define long-term degradation patterns and determine optimal recoating intervals.

References

COACH SOLUTIONS (2017), *Coach Technical Documentation*, Rev. 1, Coach Solutions

ISO (2016), *ISO 19030 — Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

KWON, S.H.; LEE, I.; PARK, H.; LEE, S.G. (2020). *Decomposition mechanisms of self-polishing copolymers for antifouling coating materials through first-principles approach*. Progress in Organic Coatings 138, 105406

OLIVEIRA, D.R.; LAGERSTRÖM, M.; GRANHAG, L.; WERNER, S.; LARSSON, A.I., KIM, K.; LEER-ANDERSEN, M.; WERNER, S. (2022), *A study on the effect of hull surface treatments on ship performances*, 9th Conf. Computational Methods in Marine Engineering

LAGERSTRÖM, M.; WRANGE, A.L.; OLIVEIRA, D.R.; GRANHAG, L.; LARSSON, A.I.; YTREBERG, E. (2022), *Are silicone foul-release coatings a viable and environmentally sustainable alternative to biocidal antifouling paints*, Science of the Total Environment 838, pp.156–168

MUNK, T.; KANE, D.; YEBRA, D.M. (2009), *The effects of corrosion and fouling on the performance of ocean-going vessels: A naval architectural perspective*, Advances in Marine Antifouling Coatings and Technologies, pp.148–176

SCHULTZ, M.P. (2007), *Effects of coating roughness and biofouling on ship resistance and powering*, Biofouling 23(5), pp.331–341

SHIP & BUNKER (2026), Average VLSFO price in 2025 falls to US$535 per metric ton, Port News, https://en.portnews.ru/news/386533/

YTREBERG, E. (2022), *A novel tool for cost and emission reduction related to ship underwater hull maintenance*, J. Cleaner Production 356

CO_2 Concentration and Mass Determination on a Seagoing Vessel

Uwe Altenbach, Hoppe Marine, Hamburg/Germany, u.altenbach@hoppe-marine.com

Abstract

This paper presents the test results from a pilot project on onboard CO_2 concentration measurement and mass determination on a seagoing vessel, highlights the lessons derived from commissioning and operational use, and finally addresses the regulatory challenges associated with direct GHG emission measurement.

1. Introduction

While Continuous Emissions Monitoring Systems (CEMS) have long been a de facto standard in land-based applications, they have increasingly been implemented onboard seagoing vessels in recent years. In particular, scrubber systems utilize such measurement technology to determine the SO_2/CO_2 ratio downstream of the exhaust gas treatment process.

For Hoppe Marine, the monitoring of exhaust gas emissions in general has opened an entirely new chapter, as the application of this technology was previously beyond our company's core expertise. This challenge required deep knowledge and brought us together with AVL List GmbH, the world's largest independent company for the development, simulation, and testing of powertrain systems.

The deployment of testing and sensing equipment originally developed for automotive powertrain development — typically operated in controlled laboratory or test bench environments — into the significantly harsher operating conditions of a seagoing vessel's engine room introduced a range of technical challenges. These included exposure to vibration, temperature fluctuations, humidity, corrosive atmospheres, exhaust gas with a high flue gas content and demanding EMC conditions. The nature and mitigation of these challenges are examined in more detail in the following chapters.

Ultimately, the project did not only result in the successful installation of a CO_2 and SO_x gas sampling system combined with exhaust mass flow determination but also demonstrated the technical robustness of the overall setup. The installation delivered highly accurate and reliable measurement results under operational conditions, confirming both the functionality of the system design and the suitability of the selected measurement approach.

Test results have limited validity without appropriate reference values and verification procedures. Although marine main engines typically operate in steady-state conditions, it must nevertheless be ensured that the quality and plausibility of the recorded data are demonstrably validated.

For the trial installation, we had the opportunity to integrate the CEMS on a fully instrumented vessel equipped with Coriolis-type mass flow meters and a shaft power measurement system. This configuration enabled cross-verification of fuel consumption, engine load, and calculated emission factors.

The CEMS was installed in the exhaust stack of the main engine of a medium-sized LPG tanker, powered by a MAN Energy Solutions 6S35ME-B9 (two-stroke combustion principle).

Beyond assessing the quality of the test results alone, it is necessary to conduct a more comprehensive evaluation of the added value and revenue potential associated with directly measuring greenhouse gas emissions—focused in this trial on CO_2—within the exhaust gas of the respective combustion units. This assessment should explicitly consider the comparatively slightly higher installation effort and increased maintenance requirements relative to a solution based solely on mass flow meters.

Finally, the regulatory framework governing the operation and application of this measurement method for CO_2 reporting will be examined, particularly with regard to its acceptance and integration within both CII and EU MRV reporting requirements.

2. Continuous Emission Test System - Pilot Installation on a Gas Tanker

2.1. CEMS Pilot System Functional Design

The measuring system was designed to determine greenhouse gas emissions—limited in this trial to CO_2—directly within the exhaust gas stream of the respective combustion units. The system comprised a heated exhaust gas sampling probe, a cold sampling line (NDIR applications do not require heated sample lines since gas condensation is not an issue) and a gas conditioning unit for filtration moisture removal, and pressure stabilization. CO_2 concentration was measured continuously using an NDIR analyzer equipped with automatic zero and span calibration. Measurement data were recorded via a central data acquisition system and processed to calculate CO_2 mass emissions in combination with exhaust flow data.

The test setup involved installation of the sampling probe at a defined position within the exhaust duct to ensure homogeneous flow conditions and representative gas composition.

Fig.1: Heated probe for marine onboard ship application (©mc-techgroup)

For the gas mass flow determination, a differential pressure sensing devices was applied. The method is based on the relationship between flow velocity and pressure drop across a defined flow restriction. When exhaust gas passes through a primary element installed in the duct, a differential pressure (Δp) is generated. This pressure difference is proportional to the square of the flow velocity.

The calculation follows these steps:

1. Measurement of differential pressure (Δp)
2. Determination of gas velocity based on the Bernoulli equation
3. Calculation of volumetric flow from velocity and duct cross-section
4. Conversion to mass flow by using gas density
5. Gas Temperature (correction)

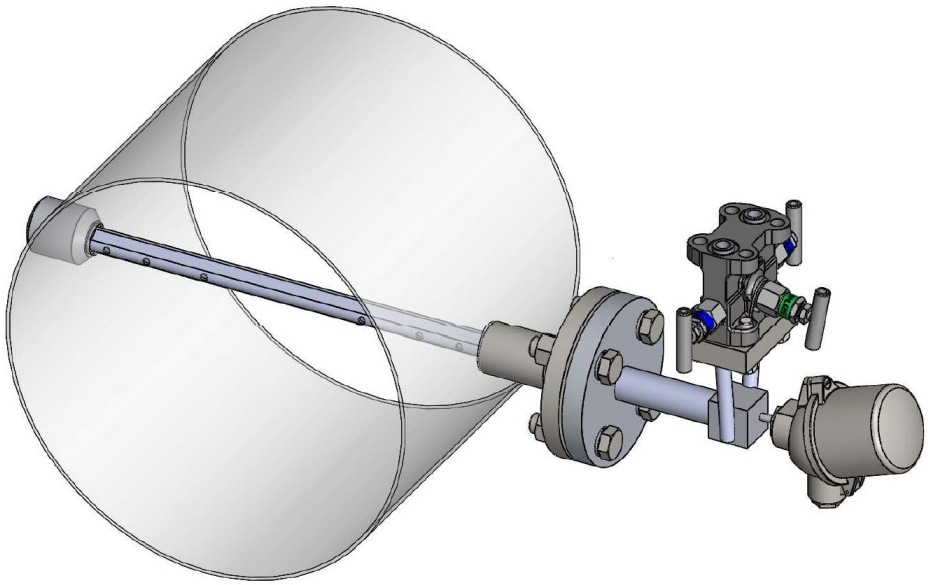


Fig.2: Differential pressure probe for mass determination

To determine mass emissions (e.g., CO_2) from measured concentration and volumetric flow, the procedure can be described as follows:

Step 1 – **Determination of the volumetric flow rate**

Obtain the exhaust gas volumetric flow rate from the flowmeter: $\dot{V}[\mathrm{m}^3/\mathrm{h}]$. The flow rate must refer to the same reference conditions (temperature, pressure, dry/wet basis) as the gas concentration measurement.

Step 2 – Measuring the CO_2 concentration

Measure the CO_2 concentration by the NDIR gas analyzer as volume fraction (e.g., % vol or ppm). Convert percentage to a decimal fraction:

$$C_{vol} = \frac{\mathrm{CO_2(\%)}}{100}$$

Step 3 – Converting concentration to mass concentration

For gases, volumetric fraction equals molar fraction under identical conditions. Using the ideal gas law:

$$\rho_{CO_2} = C_{vol} \cdot \frac{P \cdot M_{CO_2}}{R \cdot T}$$

P= absolute pressure [Pa]
T= absolute temperature [K]
M_{CO_2}= molecular weight of CO_2 = 0.044 kg/mol
R= universal gas constant = 8.314 J/(mol·K)
Result: $\rho_{CO_2}[\mathrm{kg/m}^3]$

Step 4 – Calculating mass emission rate

Multiply the CO_2 mass concentration by the volumetric flow rate: $\dot{m}_{CO_2} = \rho_{CO_2} \cdot \dot{V}$

Combining all steps, we derive the compact final formula: $\dot{m}_{CO_2} = \dot{V} \cdot C_{vol} \cdot \frac{P \cdot M_{CO_2}}{R \cdot T}$

CO_2 Mass Emission Rate = CO_2 Concentration (kg/m³) × Flow Rate (m³/h)

Fig.3: CEMS system cabinet arrangement

The main components of an NDIR sensor are an infrared (IR) source (lamp), a sample chamber or light tube, a light filter and an infrared detector. The IR light is directed through the sample chamber towards the detector. In parallel, there is another chamber with an enclosed reference gas, typically nitrogen. The gas in the sample chamber causes absorption of specific wavelengths according to the Beer–Lambert law, and the attenuation of these wavelengths is measured by the detector to determine the gas concentration. The detector has an optical filter in front of it that eliminates all light except the wavelength that the selected gas molecules can absorb. In particular, lower initial investment costs speak in favor of this measuring principle, capable of monitoring a of the most relevant components, offering the option to modularly add hardware analyzers to expand the number of measured exhaust gas components.

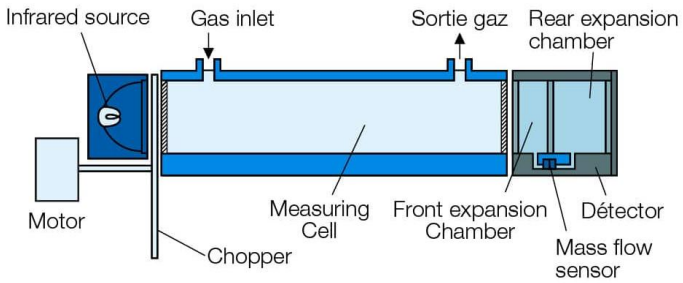


Fig.4: NDIR Sensing Principle

2.2. Technical challenges and lessons learned

As anticipated, several technical challenges emerged during the test phase. One of the most significant challenges proved to be the only partially predictable operating profile of our test vessel. The trading pattern—highly typical for gas tankers—was often determined at very short notice. This confronted the technicians of our cooperation partner, AVL, for the first time with the reality that, unlike land-based test facilities, measurement systems in maritime applications are inherently mobile and not tied to a fixed location.

A second, equally significant challenge was the high flue gas and particulate content in the sample gas. This resulted in increased maintenance requirements, particularly with regard to filter replacement, which initially proved to be impractical for routine operation. In response, the filter configuration was adapted to better reflect the actual operating conditions. With the optimized setup now in place, filter replacement is required only approximately once per month.

Further optimizations were implemented with respect to the sampling lines and the software interface to the Hoppe data logger. The initial version of the sampling lines, which had an insufficient chromium content, led to corrosion-related issues and ultimately to the failure of a sample gas pump. The material specification was subsequently upgraded to ensure long-term durability.

In addition, the software interface to the Hoppe data logger was refined to improve data stability, synchronization, and overall system reliability.

2.3. Testing Results and comparison to secondary monitoring alternatives

The CEMS measurement results were benchmarked against both, the Coriolis fuel flow measurement systems installed onboard and the installed shaft power meter. Both reference systems had been in operation for several years prior to the commissioning of the CEMS, consistently delivering accurate and reliable measurement data. As part of the CEMS commissioning process, these systems were additionally verified and calibrated to ensure a sound and traceable reference basis for comparison.

The measurement method used for comparison, based on fuel consumption monitoring, follows the requirements set out in *IMO (2022)* and in *EU (2015)*.

In accordance with these regulatory provisions, CO_2 emissions are determined by multiplying the monitored fuel consumption by the applicable fuel-specific emission factor.

$$C_F = \frac{t\ Co2}{t\ Fuel} = 3.206\ for\ Diesel\ /Gas\ oil$$

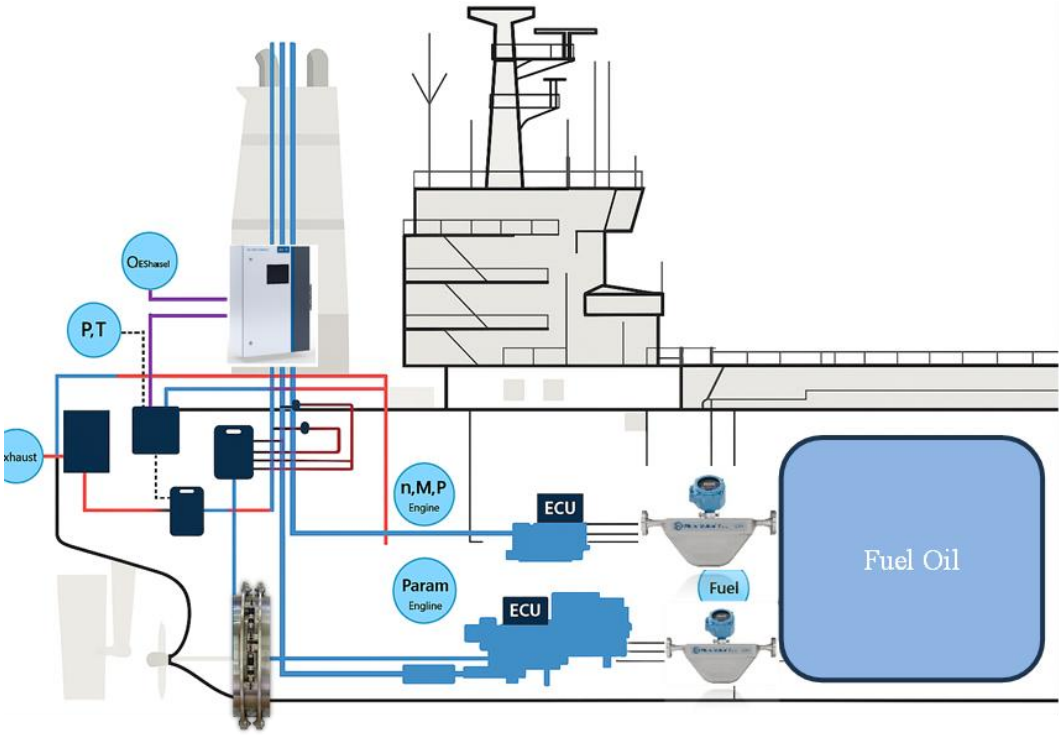


Fig.5: Pilot vessel's sensor setup

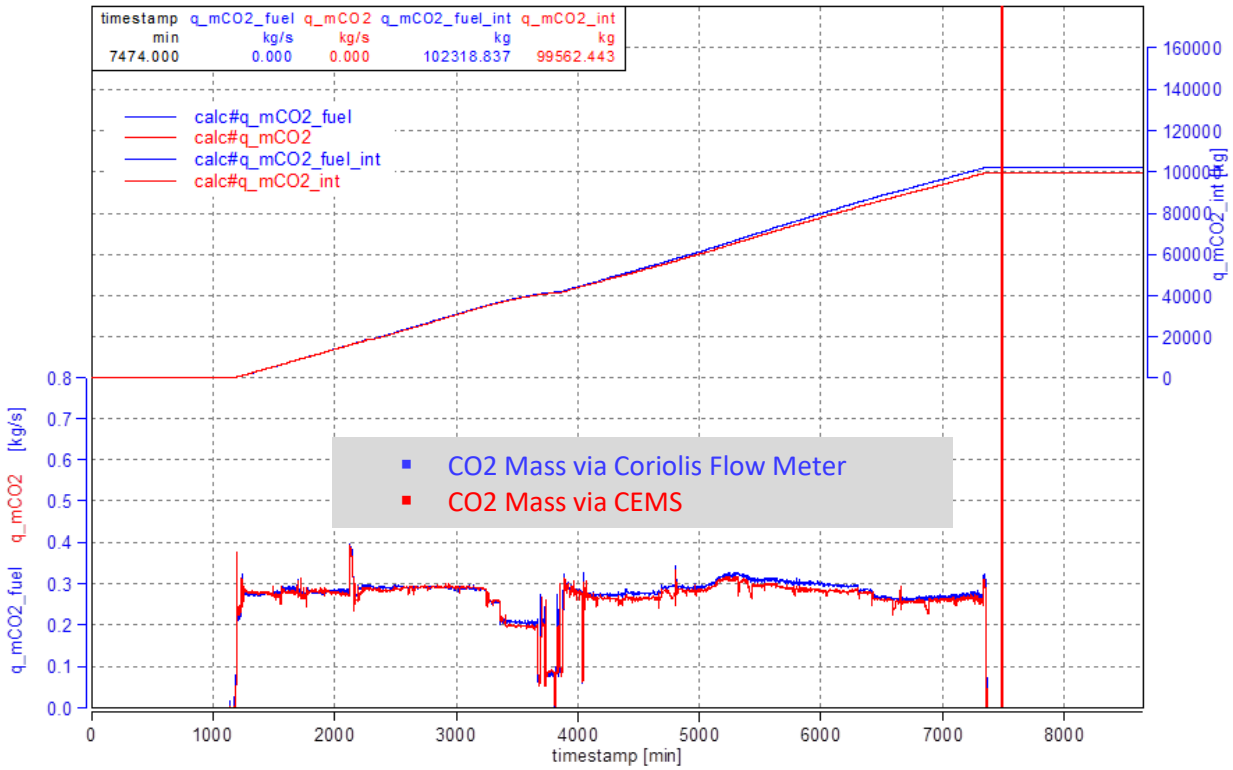


Fig.6: Correlation CO_2 mass CEMS vs. Coriolis flowmeter (applying the CO_2 correlation factor)

Fig.6 illustrates the correlation between monitoring Method C (fuel consumption × emission factor) and Method D (direct CO_2 measurement) over the course of a four-day voyage. The comparison is limited exclusively to emissions from the main engine.

The correlation between the two monitoring methods is shown to be exceptionally strong. No significant deviations or abnormal measurement outliers were detected not reduced to this single voyage, but during the whole evaluation period.

In particular, the exhaust gas mass flow determination using the differential pressure probe demonstrated outstanding performance and significantly exceeded expectations.

Especially at low engine loads—such as during canal transits—potential inaccuracies had been anticipated due to reduced exhaust gas velocities and the associated lower differential pressure signals. However, these concerns were not confirmed during the evaluation period. The system maintained stable, accurate, and reliable measurement performance even under low-flow operating conditions.

On average, the CO_2 emission values determined by the CEMS were approximately 2.7% lower than those derived from fuel consumption data obtained via the Coriolis fuel flow meter. This deviation remained consistent throughout the monitoring period and within the expected uncertainty range of the respective measurement methods.

Fig.7 presents a scatter plot illustrating the correlation between the raw signal of the exhaust gas differential pressure probe and the fuel consumption of the main engine over a monitoring period of several days.

Even without applying signal filtering, flow profile corrections, or additional compensation factors, a clear and consistent correlation between the differential pressure signal and the engine fuel consumption is evident. The observed relationship confirms the stable response behavior of the differential pressure probe across varying engine load condition

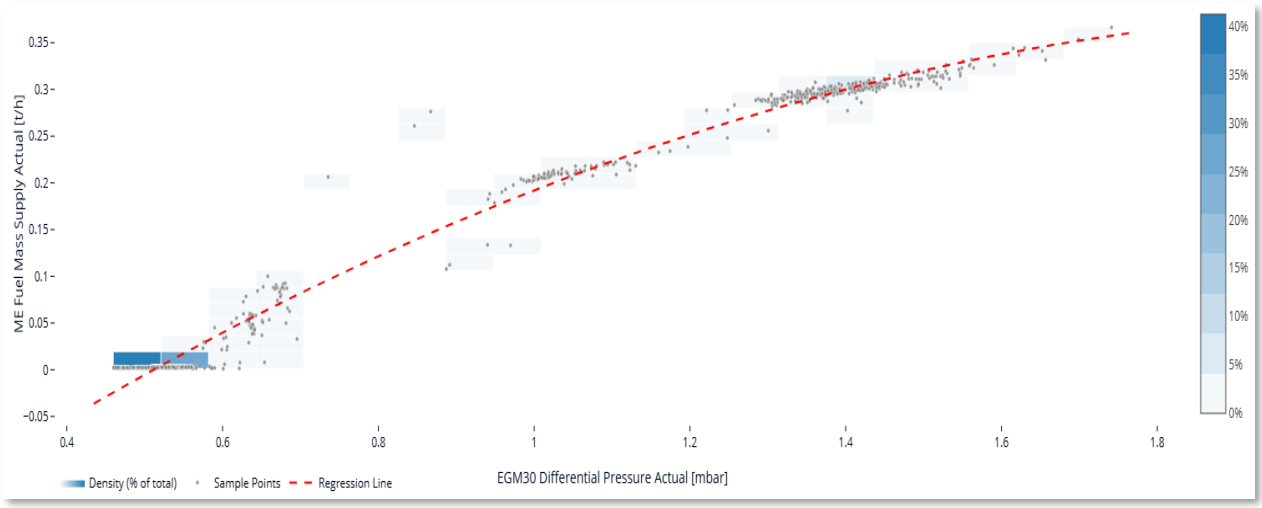


Fig.7: Correlation Exhaust Gas Differential Pressure vs ME Fuel Supply

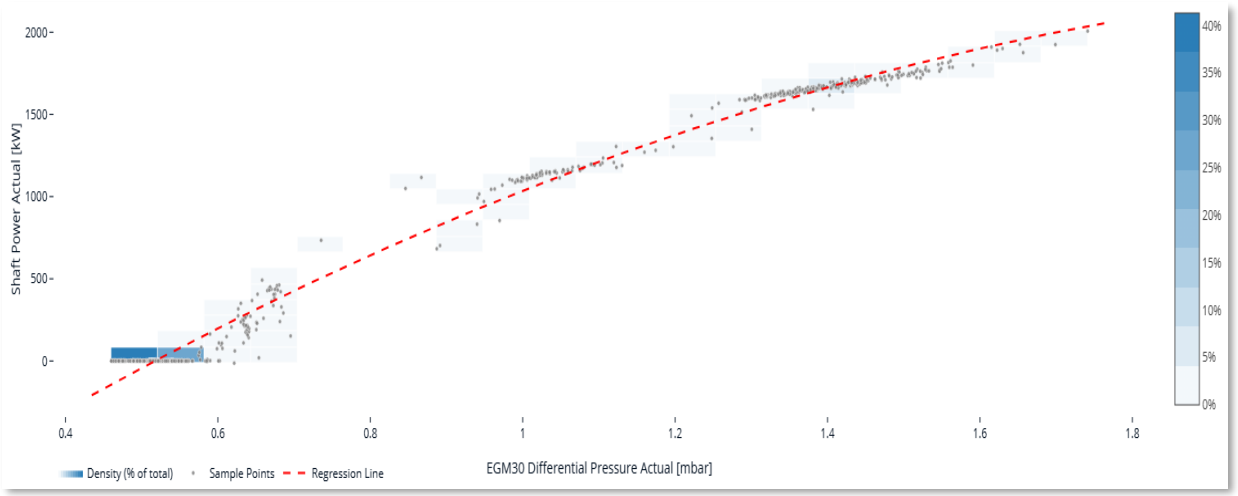


Fig.8: Correlation exhaust gas differential pressure vs shaft power

A very similar pattern is observed when analyzing the correlation between the raw signal of the exhaust gas differential pressure probe and the engine power measured by the shaft power meter.

Again, no irregularities or abnormal deviations are evident. The consistently strong correlation further confirms the high level of agreement between the fuel-based monitoring system and the engine power measurement system (MAIHAK). This consistency supports the overall plausibility of the exhaust gas flow determination and demonstrates coherent system behavior across independently measured operational parameters.

3. Evaluation of Measurement Results and estimated Savings

The observed deviation between the above-mentioned monitoring methods is not unprecedented; rather, it confirms findings that have been reported in previous technical evaluations. The primary cause of the deviation lies in the fact that the total fuel mass does not participate in combustion with 100% completeness. More specifically, not all hydrocarbons are fully oxidised to CO_2. Incomplete combustion results in the formation of carbon monoxide (CO), soot particles, and other partially oxidised carbon compounds.

Under Method C, however, the entire carbon content of the consumed fuel is stoichiometrically attributed to CO_2 formation. This methodological assumption inherently assigns the total fuel carbon mass to CO_2 emissions, thereby neglecting the fraction of carbon present in other exhaust components. As a result, a systematic difference arises between fuel-based calculations and direct exhaust gas measurements.

The magnitude of this deviation can vary depending on engine design, combustion concept, maintenance condition, operating profile, and fuel quality. In the present pilot project, the assessment was conducted on a comparatively small, well-maintained engine operating exclusively on marine diesel fuel. Under such favorable conditions, combustion efficiency is expected to be high, limiting the discrepancy between the two methods.

For engines operating on heavy fuel oil (HFO), particularly in combination with higher power output a greater deviation between fuel-based and direct measurement methods can reasonably be expected.

Vessel Type	Bulker (Handy Size)	Oil Tanker (VLCC)	Container Vessel (20.000 TEU)
Daily Fuel Oil Consumption per day	20 t	150t	200t
Days at Sea	250	230	280
Yearly CO_2 Emission (@CF = 3.206)	16,030t	104,400t	179,540t
2,7% Deviation	433t	2,819t	4,847t
Percentage of Relevance to EU/ETS	45%	45%	45%
Cost Savings @85 US$ per ton	**$16,600**	**$107,800**	**$231,800**

Fig.9: Estimated Annual Cost Savings under the EU ETS Framework

The calculation example presented above in Fig.9 is based on the deviation identified in our pilot project between the fuel-consumption based approach and the directly measured CO_2 emissions (Method C vs. D).

Although this evaluation supports the direct measurement approach for CO_2 in exhaust gas, a comprehensive cost assessment must also take into account the additional expenditures associated with this method. These include the initial investment and installation effort, ongoing operating costs (e.g., consumables such as filters), as well as servicing requirements. Accordingly, any economic evaluation should balance the potential savings in EU ETS allowance surrender obligations against the total cost of ownership of the direct measurement system over its operational lifetime.

4. Regulatory Framework

Given the unexpectedly strong performance of the first prototype and the extensive dataset obtained during the pilot phase, the next logical step was to evaluate the suitability of the CEMS for statutory greenhouse gas (GHG) reporting purposes. To support its potential use within a regulated framework, it was necessary to examine the applicable regulatory provisions and to assess whether the CEMS could be justified to an accredited verifier as an acceptable monitoring methodology.

Accordingly, an in-depth literature review was conducted, accompanied by a detailed analysis of the regulatory requirements set out in *EU (2015)* on the monitoring, reporting and verification (MRV) of greenhouse gas emissions from maritime transport, as well as MEPC.346(78), the 2022 Guidelines for the Development of a Ship Energy Efficiency Management Plan (SEEMP). Subsequently, attention was directed to MEPC.402(83) in order to assess whether the evolving regulatory approach to load-

dependent methane slip factors and associated spot-check methodologies may provide insights into the future regulatory acceptance of continuous emissions monitoring systems (CEMS). The objective was to understand whether these developments indicate a broader shift towards direct measurement-based approaches within international GHG reporting frameworks.

4.1. General Regulation for direct CO2 Measurement

Within the scope of the pilot installation, the measurement system was deployed exclusively for the determination of CO_2 mass emissions. The system was specifically designed for application on an internal combustion engine operating solely on diesel fuel.

For CO_2 emissions, the regulatory provisions governing direct measurement are clearly defined. The applicable framework is set out in *IMO (2022),* which is applied in the context of determining the Carbon Intensity Indicator (CII). In addition, the requirements are established under *EU (2015)* on the monitoring, reporting and verification of carbon dioxide emissions from maritime transport. Both regulatory frameworks explicitly recognize direct exhaust gas-based measurement methods for CO_2 determination, provided that appropriate and validated measurement technology is used, and that the methodology is properly documented within the approved Monitoring Plan. Both regulatory frameworks also permit the application of different monitoring methodologies within the same vessel. For example, the direct exhaust gas measurement method may be applied to the main engine, while fuel consumption–based monitoring may be used for auxiliary engines.

However, neither document refers to a dedicated analytical procedure, nor do they specify the measurement setup in detailed technical terms. Accordingly, the responsibility for defining and justifying the applied measurement concept—including system design, calibration procedures, and uncertainty assessment—rests with the operator and must be accepted by the verifier.

Methods / Features	Bunker Delivery Note A	Bunker Tank Monitoring B	Flow Meter (Coriolis) C	Direct CO2 Measurement D
Accuracy	Low	Medium	High	Highest
Consideration of Residuals	No	No	Conditionally	Full
Consideration of actual fuel quality (heat value and carbon intensity)	No	No	Conditionally	Full
Consideration of fuel consumption over individual legs of journey	No	Yes	Yes	Yes
Consideration of the specific combustion unit (ME, AE, Boiler)	No	No	Yes	Yes
Added Value (Assessment of Engine Operating Condition	No	No	Conditionally	Yes

Fig 10: Methods for IMO Data Collection System (DCS) and *EU (2015)*

4.2. Rules comparable to MEPC.402(83)

Although MEPC.402(83) does not explicitly mandate the use of Continuous Emissions Monitoring Systems (CEMS), since it is mainly focused on Methane (CH_4, applicable for dual fuel engines) And / Or Nitrous Oxide (N_2O) Emissions on it signals a broader regulatory trend towards:

- Greater accuracy in emission quantification
- Consideration of real operating conditions

- Increased acceptance of measurement-based approaches

The recognition of load-dependent methane slip factors and associated verification mechanisms suggests a growing openness within the IMO framework to more advanced monitoring methodologies, including direct exhaust gas measurement systems.

In this context, CEMS can be viewed as a technically robust solution capable of providing continuous, load-resolved emission data, thereby aligning with the underlying regulatory objective of improved emission transparency and accuracy.

MEPC.402(83) can be regarded as a relevant reference framework for FuelEU Maritime and EU MRV reporting. The *Guidelines for reporting and verification of actual methane slip tank-to-wake emission factors from marine diesel engines under the scope of the FuelEU Maritime Regulation*, published by the European Commission on 8 October 2025, explicitly state that the provisions of MEPC.402(83) may be applied for FuelEU reporting purposes. This explicit cross-reference demonstrates regulatory alignment between IMO and EU frameworks and confirms that methodologies developed under MEPC.402(83) are considered suitable for application within the European regulatory context. Consequently, the resolution may serve as a technical benchmark when evaluating measurement-based approaches, including Continuous Emissions Monitoring Systems (CEMS), for use in FuelEU and EU MRV reporting.

The Regulation (EU) 2023/1805 (FuelEU Maritime) explicitly provides, under Article 10(5), that individual measured methane slip factors may be applied instead of default values. The above-mentioned guidance document clarifies the methodological framework under which such individual slip factors can be determined and verified.

In addition, Regulation (EU) 2015/757, Annex I, Part A, Point 2, establishes a regulatory linkage by referring to Article 10(5), thereby reinforcing the acceptance of measurement-based emission factors within the EU monitoring and reporting system.

MEPC.402(83) itself builds upon the NOx Technical Code 2008 and adapts selected provisions to accommodate methane spot-check measurements. These adaptations primarily address sampling methodology, measurement accuracy, and verification procedures to ensure the reliability of load-dependent methane slip determination.

Given that methane slip spot checks are accepted within this regulatory framework under defined technical requirements, it can be reasonably argued that the technical expectations for a continuous exhaust gas measurement system should not exceed those applicable to validated methane slip spot-check procedures. Accordingly, a system that demonstrably fulfils the requirements set out in MEPC.402(83), including those derived from the NOx Technical Code, should in principle be considered technically suitable for application as a Continuous Exhaust Gas Measurement System (CEMS), subject to verification and approval within the applicable regulatory framework.

5. Conclusion

The pilot installation of a Continuous Emissions Monitoring System (CEMS) for direct CO_2 measurement onboard a seagoing LPG tanker has demonstrated that exhaust gas–based greenhouse gas quantification is technically feasible, operationally robust, and reliable under real maritime conditions. The successful transfer of measurement technology from controlled engine test environments to the marine engine room confirms that, with appropriate system adaptation and integration, high-precision emission monitoring can be achieved at sea.

The comparative assessment between direct exhaust gas measurement (Method D under Regulation (EU) 2015/757) and the conventional fuel consumption–based approach (Method C) showed strong correlation across all operating profiles, including low-load conditions. The observed average deviation

of approximately 2.7%, with CEMS consistently reporting lower CO_2 values, remained stable and within the expected uncertainty range of both methodologies. This difference can plausibly be attributed to the assumption of complete carbon oxidation inherent in fuel-based calculations, whereas direct measurement reflects actual combustion conditions. The consistency of deviation supports the reliability of the measurement system.

Particular emphasis should be placed on the performance of the exhaust gas mass flow determination via differential pressure measurement, which remained stable and accurate even under low exhaust gas velocities without additional maintenance effort. The strong correlation between exhaust differential pressure, fuel consumption, and shaft power confirms the coherence of independently measured parameters and underlines the technical plausibility of the overall concept.

From an economic perspective, even modest systematic differences between monitoring methodologies can result in significant financial effects under the EU ETS when applied to large annual fuel consumptions and prevailing CO_2 allowance prices. While potential reductions in surrender obligations provide a measurable incentive, these must be balanced against capital expenditure, operating costs, maintenance, and verification requirements. The total cost of ownership remains decisive for large-scale implementation.

Regulatorily, both Regulation (EU) 2015/757 and MEPC.346(78) permit direct exhaust gas–based CO_2 measurement, provided it is documented within the Monitoring Plan and accepted by the verifier. Although detailed technical prescriptions are not defined, the operator must demonstrate validity and uncertainty control. Recent developments, particularly MEPC.402(83) and the integration of measurement-based methane slip factors under Regulation (EU) 2023/1805, indicate a broader shift toward real-time, load-dependent emission quantification.

In conclusion, the pilot project confirms that continuous CO_2 measurement in marine exhaust gas is technically achievable, economically relevant, and regulatorily compatible. Given the ongoing tightening of decarbonization requirements, CEMS represents a credible and forward-looking alternative to purely fuel-based emission calculations and may become increasingly significant in future maritime GHG compliance strategies.

As a direct outcome of the successful pilot phase, the system will enter series production in 2026, strategically expanding and completing the Hoppe Marine Performance and Energy System portfolio. All technical improvements and operational insights gained during the trial installation have been fully incorporated into the final product architecture, resulting in a mature, field-proven solution ready for large-scale deployment.

The production-ready system will not be limited to CO_2 monitoring but will be capable of measuring all greenhouse gas–relevant components. It therefore represents a comprehensive solution for exhaust gas–based GHG monitoring in maritime applications, supporting both regulatory compliance and performance optimization objectives.

Acknowledgement

I would like to express my sincere appreciation to my esteemed colleague, Leonard Krause, for his valuable contributions to the regulatory assessment and for his outstanding technical expertise throughout the project. I also extend my cordial thanks to our partners at AVL for their excellent cooperation and dedicated support during the pilot phase. Their professional assistance and constructive input during the preparation of this paper are gratefully acknowledged.

References

EU (2015), *Regulation (EU) 2015/757 on Monitoring, Reporting and Verification of CO2 emissions from ships*, European Commission

IMO (2022), *2022 Guidelines for the Development of a Ship Energy Efficiency Management Plan (SEEMP)*, MEPC.346(78), Int. Mar. Org., London

IMO (2025), *2025 Guidelines for Test-Bed and Onboard Measurements of Methane (Ch4) and/or Nitrous Oxide (N2o) Emissions from Marine Diesel Engines,* MEPC 402 (83), Int. Mar. Org., London

EU (2025), *The EU ETS and MRV - Maritime General guidance for shipping companies,* European Commission, https://climate.ec.europa.eu/document/download/31875b4f-39b9-4cde-a4e2-fbb8f65ee703_en?filename=policy_transport_shipping_gd1_maritime_en.pdf

ZHOU, S.; ZHOU, Y.; WANG, Y. (2022), *Combustion and Emission Characteristics of Marine Diesel Engines Using Alternative Fuels: A Review*, Applied Energy 310

Data-Driven Trim Optimization in Fleet Operations

Federico Franceschini, GNV, Genova/Italy, Federico.Franceschini@gnv.it
Alida Abbo, GNV, Genova/Italy, Alida.Abbo@gnv.it
Lorenzo Figoli, GNV, Genova/Italy, Lorenzo.Figoli@gnv.it
Ivana Melillo, GNV, Genova/Italy, Ivana.Melillo@gnv.it

Abstract

This paper presents the implementation of a data-driven strategy aimed at improving energy efficiency across a maritime fleet. The initiative involved installing automatic data acquisition systems onboard to continuously collect information on energy performance and operational procedures. The data are processed ashore to develop mathematical models that enable objective performance evaluation and support continuous improvement. Sensors such as torquemeters, flowmeters, inclinometers, and draft meters were installed, along with integrated automation, navigation and weather data. Individual performance models now allow targeted monitoring of key efficiency factors, including hull paint degradation, main engine and genset usage, and trim optimization. A dedicated trim model, validated using low-frequency data, demonstrated energy savings peaking 5% (depending on operating conditions). The initiative has proven effective and is being considered for fleet-wide expansion, highlighting the value of data-driven optimization and crew collaboration.

1. Introduction

IMO has set ambitious targets to achieve significant reductions in carbon emissions by 2030 and 2050. These targets are crucial for driving the maritime industry towards more sustainable practices and ensuring compliance with global environmental standards.

In this context GNV, founded in 1992, and now part of the MSC Group started a path toward decarbonization. GNV is one of the leading shipping companies operating in the coastal and passenger transport sector in the Mediterranean and is undergoing a modernization program which entails the delivery of four new Ro-Pax vessels with high energy efficiency standards as long as refitting and advanced energy monitoring and management on the existing ones.

Optimizing energy efficiency in maritime operations requires accurate onboard systems data measurement and analysis, therefore, as reported in a previous paper, SERTICA Performance software was installed onboard of both units (as well as most of the fleet) enabling real-time data acquisition from existing connected systems, aggregating figures in 5-minute intervals. These data are then processed ashore to develop mathematical models that support objective performance assessment and continuous improvement. The final part of the problem is then carried out having in mind ISO 19030 and its approach to performance monitoring.

The decision to implement automated data acquisition was driven by the need for objective, continuous performance monitoring. By installing sensors and integrating navigation data, we aimed to identify inefficiencies and optimize configurations. The selection process prioritized measures that could provide actionable insights into ship energy performance, with a focus on reliability and scalability. This leads to first install sensors and monitor EnPIs on Delivered Power (PD), Main engines Power (ME_P) Auxiliary Power (Aux_P) and some proxy data such as dynamic trim and SFCs.

Roughly 85% of GNV's total power demand is dedicated to propulsion, while the remaining 15%–25% supports auxiliary systems. Because propulsion dominates overall energy consumption, the company is investing significant effort in monitoring hydrodynamic performance and propulsion behavior. By prioritizing propulsion-related monitoring, GNV focuses on the area where the potential return on optimization is greatest.

1.1. Newbuilding specs and first month operations

GNV Polaris is the first of four new Ro-Pax vessels (Polaris, Orion, Virgo, Aurora) designed to enhance the efficiency and sustainability of the GNV fleet. Built by Guangzhou Shipyard International (GSI) in China, GNV Polaris and all his coming sisterships, incorporates advanced technologies for emission reduction, energy efficiency, and passenger comfort.

With a GT of approximately 47,000 tons, a length of 218 m, and a width of 29.60 m, GNV Polaris, can reach speeds of more than 25 kn and accommodates 240 cabins, supporting up to 1,500 passengers and 3100 lane meters. She took service on November 2024: 6 months later GNV Orion, her sistership with increased pax/cabins capacity (GT abt.52000), was delivered and took service on the same route between Genova and Palermo.

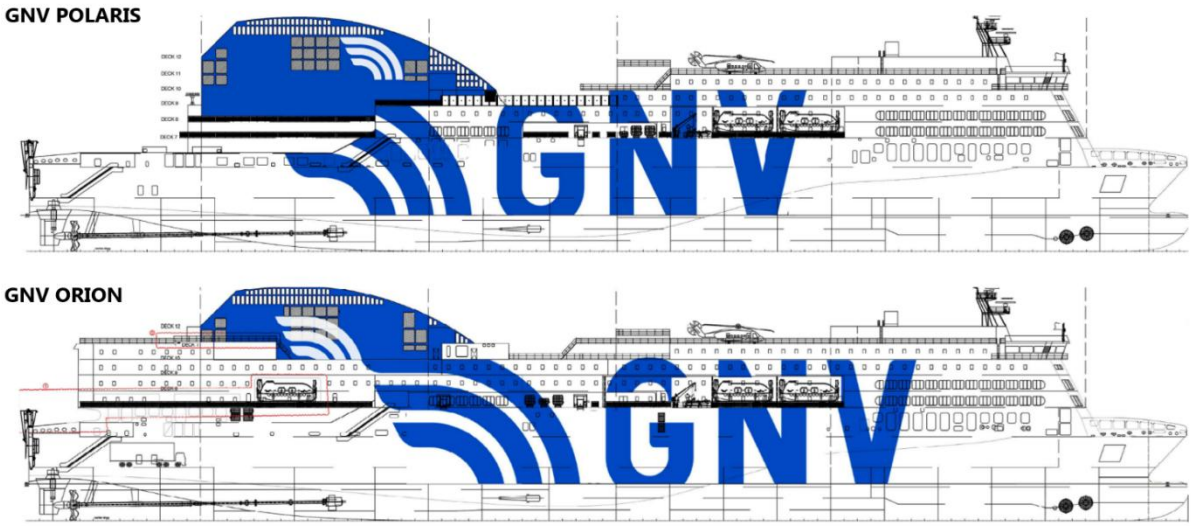


Fig.1: GNV Polaris and GNV Orion

The two units have the same propulsion and auxiliary power generation layout as described below:

- Main propulsion engines (abt. 10'000 kW/each)
- Shaft generators (2500 kW each)
- Auxiliary generators (2*2400 + 2*1600 kW)
- 1 Waste heat turbine generator (abt. 900kW)

After some weeks of parallel operations in similar speed conditions higher consumption on Orion was evidenced. The below table shows relevant sailing conditions parameters the two vessels had in the period between 15/06/2025 and 23/9/2025 on the Genova – Palermo route.

Table I: Polaris and Orion average consumption and operative profile from 15/6/2025 to 23/9/2025 (*All data above are recorded manually at Departure/Arrival)

	GNV POLARIS	GNV ORION	Δ
Cons	0.262 t/nm	0.286 t/nm	9%
Vs	23.9 kn	24.0 kn	1%
Draft	6.32 m	6.53 m	3% (~800t)
Trim	0.24 m	0.35 m	

Added weight from superstructures and the associated increase in hotel-load power demand were initially examined as potential contributors; however, neither factor proved sufficient to account for such a significant deviation.

In fact, the additional hotel-load consumption has been quantified at ~250 kW, corresponding (at 23.5 kn) to roughly 0.003 t/nm, and even less at higher speeds. This represents about 1% of total consumption under the most pessimistic assumptions.

Also, by considering the model-basin-derived delivered-power curve, Fig.1, the effects of speed and draft differences were also evaluated; however, taken together, these factors account for no more than 5–6% (~0.1 t/nm) of the variation in power absorption and fuel consumption.

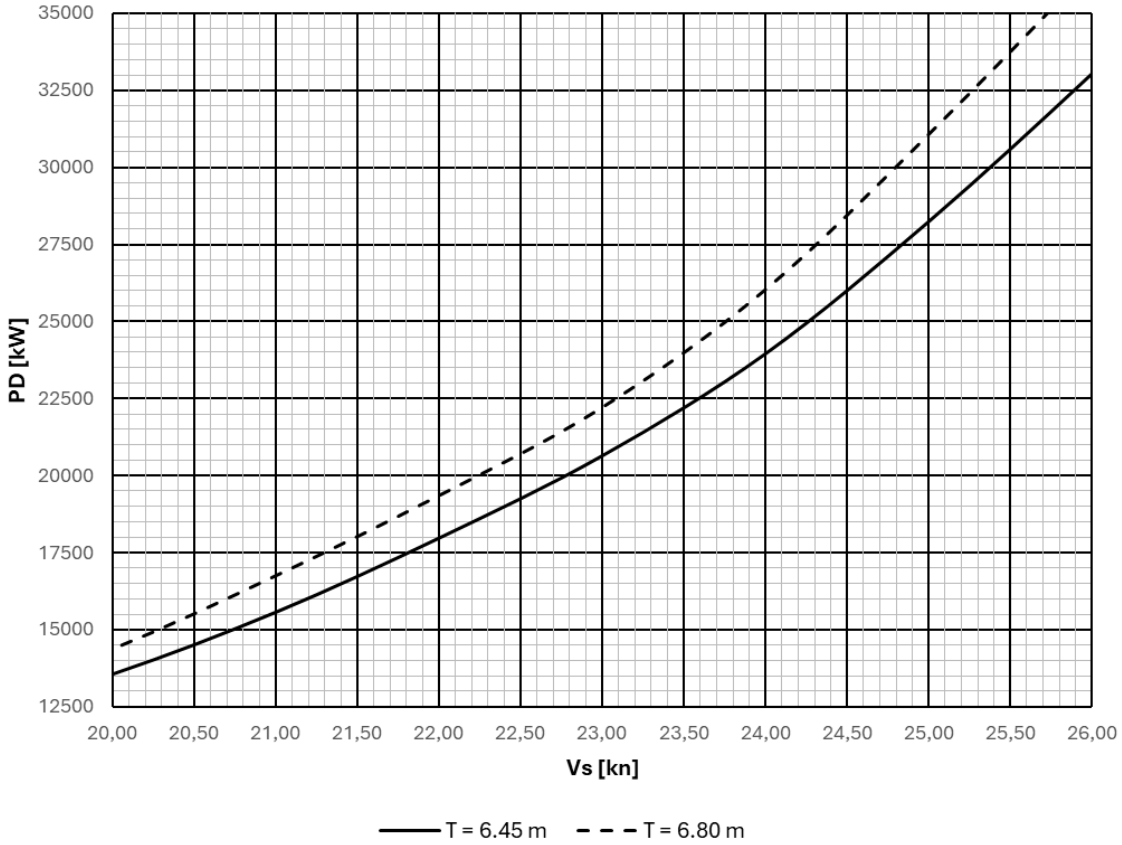


Fig.1: Model basin derived PD curves

Therefore, a performance gap between the two units remains evident, and the influence of trim on delivered power has been further investigated.

2. Selection of efficiency measures

2.1. Monitoring set-up

A suite of sensors—including torquemeters, flowmeters, inclinometers, and draft meters—was installed to quantify energy absorption. These measurements were complemented by navigation data such as vessel position, speed over ground (SOG), and prevailing weather conditions.

The analysis relies on SERTICA Performance as data collector. Developed by RINA, SERTICA Performance is a web-based platform tailored to the maritime sector, supporting Fleet Managers in optimizing vessel and fleet efficiency. The system acquires real-time data from onboard automation and monitoring systems, processes it using machine-learning and predictive analytics, and presents the results through a customizable dashboard. By integrating onboard measurements with external datasets, the software delivers a comprehensive view of operational performance.

Key benefits include enhanced fleet optimization, improved energy efficiency, real-time alerts,

reduced maintenance costs, and strengthened operational safety. All data are also fully available from shore operators for further analysis.

2.1.1. Data collection

The automatic data-acquisition system installed on GNV vessels consists of integrated hardware and software modules capable of continuously collecting data from navigation and automation systems, as well as directly from dedicated sensors. The system processes this information in real time, providing actionable insights to both onboard personnel and shore-based teams for performance optimization and safety monitoring.

Within this architecture, SERTICA Performance system, Tablee II, Fig.2, includes an industrial PC installed onboard and connected to the ship's network. This unit interfaces with navigation and automation systems or directly with high-precision sensors such as inclinometers, torque meters, and flow meters.

Table II: Installed hardware - Sertica performance hardware kit

1	Industrial PC
2	MOXA Nport IA-5250 2-port RS-232/422/485 serial device server, 10/100MBaseT(X) (RJ45)
3	2x MOXA Nport 5250A 2 port device server, 10/100M Ethernet, RS-232/422/485, DB9 male, 0.5KV serial surge, 12~48VDC, 0~60°C
4	EDS-208A Unmanaged Ethernet Switch with 8 10/100BaseT(X) ports, -10 to 60°C
5	ioLogik E1240 Remote Ethernet I/O, 8AI, 2-port Switch and SEIKA inclinometer
6	DIN Rail Mounting Kit 35mm, for DE-311/211, NPort 5200/5400, NPort W2250/2150
7	DC/DC 150W
8	Thermoplastic box IP56 240*190*160 with DIN rail
9	AC/DC 30W MOXA power supply
10	Thermoplastic box IP66 300*250*150
11	Cable RS 485 BUS LD 1X2X0,22 (20m)
12	Cable for inclinometer 4*0.75 grey (10m)
13	LAN cable cat 6/FTP 24awg 10m
14	Connector DB9F

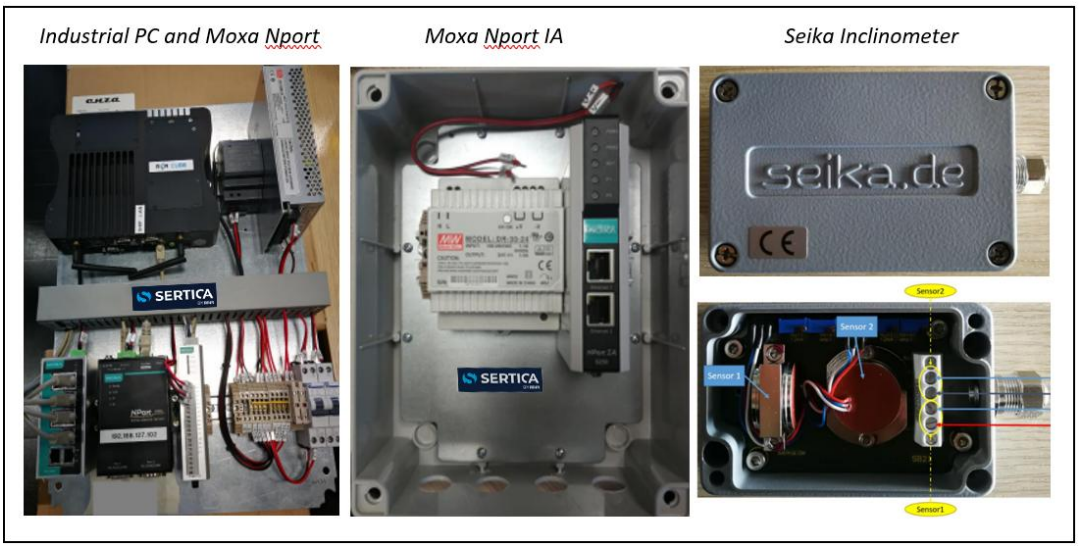


Fig.2: SERTICA Performance Hardware Kit

As illustrated in Fig.3, the onboard data collector acquires measurements at configurable sampling frequencies, performs initial filtering and processing, and stores the results in a database with a standard 5-minute aggregation interval. The processed data is then streamed in real time to shore,

where information from the entire fleet becomes immediately accessible for monitoring and advanced analysis.

In addition to raw sensor and automation data, the onboard system enriches the dataset using external sources. This includes weather information from predictive models and manual crew inputs such as drafts, displacement, and other operational parameters not captured automatically.

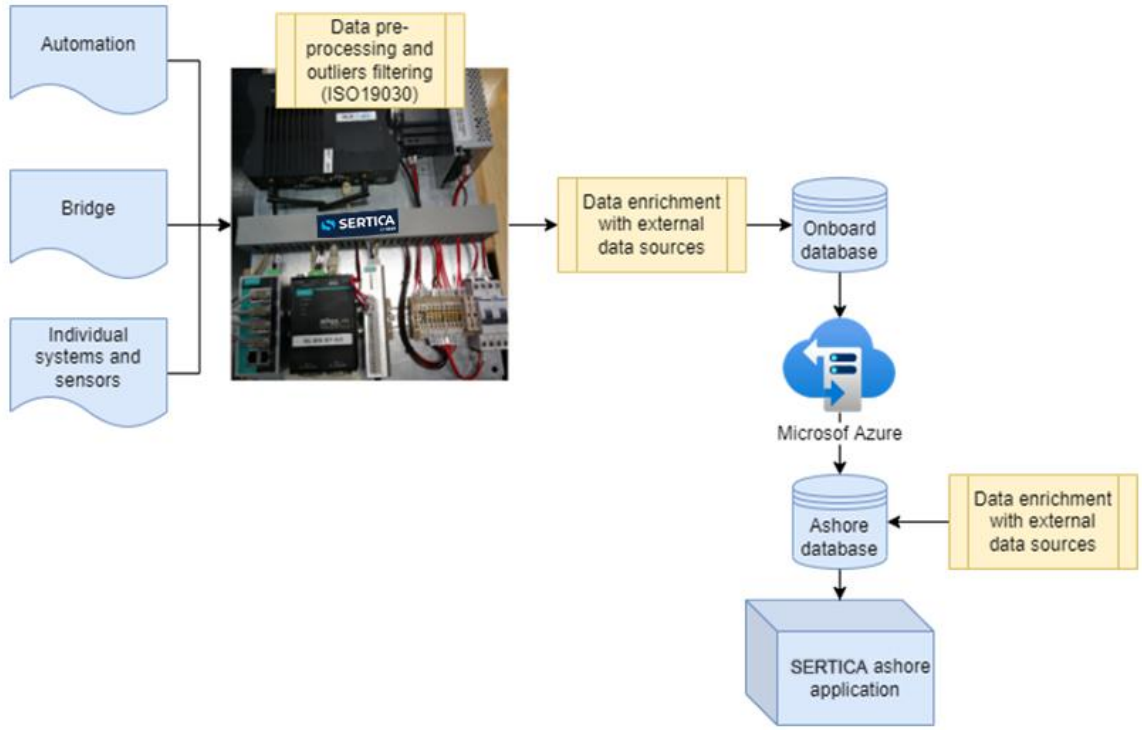


Fig.3: SERTICA PERFORMANCE data acquisition process

2.1.2. Data filtering process

A key challenge in the analysis was the reliability of onboard signals, which necessitated the development of a standardized data-cleaning process on top of SERTICA provided data. This step was essential to ensure the development of accurate and comparable reference models. The objective was twofold: to create data clusters that were as consistent as possible across operating conditions, and to reduce the number of regressors to a minimal yet meaningful set.

After preliminary screening, the dataset was further refined by removing outliers and retaining only observations recorded under acceptable “good enough” weather conditions, thereby minimizing noise and improving signal quality. To achieve this, a three-month sailing period—including all voyages—was filtered according to the constraints given in Table III.

Table III: Dataset constraints

	Min	Max
Speed	17.5 kn	25.5 kn
Sea state	0	2
dyn. Trim	-0.4 m	0.4 m
dyn. Draft	6.40 m	7.00 m

References

ISO (2016), *ISO 19030 — Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

MELILLO, I. (2025), *Monitoring and analyzing to enhance the efficiency and sustainability onboard a Ro-Pax vessel*, GNV, Genova

Physics-Informed AI for Vessel Performance Modeling: Marketing Term or Measurable Advantage?

Casimir Morobé, Toqua, Gent/Belgium, casimir@toqua.ai
Michaël Deschoolmeester, Toqua, Gent/Belgium, michael@toqua.ai
Oran Casimiro, Toqua, Gent/Belgium, oran@toqua.ai

Abstract

This paper proposes a practical taxonomy—black-box, physics-constrained, and physics-informed—based on where physics enters the modelling process. We evaluate these approaches through targeted stress tests focused on underrepresented conditions, including (i) a main-engine power → fuel case and (ii) a speed–power case at an underrepresented draft. Results show that physics constraints can improve average fit, but may still impose structurally biased behaviour. Physics-informed models, when designed correctly, consistently produce more physically credible extrapolation and improved predictive performance where data are thin. Crucially, the benefit is not only lower error: the physics-informed models extrapolate in a more physically consistent way in under-sampled regimes, producing more reliable sensitivities where black-box and shape-restricted models can drift into plausible-looking but wrong behaviour. Finally, we argue that the central challenge in shipping is not choosing between "data" and "physics," since both are imperfect: operational data are noisy and biased, and physical formulas are approximate. Physics-informed modelling is therefore a balancing problem under uncertainty—harder to design, but critical for reliable performance predictions needed for robust operational decisions and decarbonisation.

1. Introduction

Vessel performance models are a key instrument in shipping and their accuracy underpins critical decisions such as maintenance, warranty setting, voyage optimization, retrofit investments, and emissions forecasting. Recent machine-learning models can fit large operational datasets well, but as black boxes they may extrapolate erratically and violate basic hydrodynamic expectations outside the conditions they have seen. A growing response is to "add physics" to learning, yet that phrase now covers very different ideas. Some methods are physics-constrained, using physics as guardrails (bounds, monotonicity, or fixed curve shapes) that discourage implausible behaviour while leaving the model mostly data-driven. Other methods are genuinely physics-informed, where physical relationships enter as a structural prior, embedded in the loss, parameters, or architecture, so the model must negotiate data and laws together. This distinction is not semantic: it changes how models behave under sparse data, biased operations, and rare drafts or sea states. This paper makes the categories explicit and shows, through vessel-level case studies, how the degree of physical integration governs generalisation and operational reliability.

Even within the physics-informed family there is wide design freedom: different priors, different levels of enforcement, and different ways to allocate what is learned versus what is prescribed. There are many ways to inject physics; which is 'right' can only be answered by validation; how well the model predicts on unseen, reliable, out-of-distribution data.

2. Theoretical Framework

We frame vessel-performance learning methods along a single question: how does physical knowledge influence the model's behaviours? This yields three practically distinct classes: black-box, physics-constrained, and physics-informed. The classes are not about whether "physics is mentioned," but about where physics enters (features, hypothesis space, objective/architecture, etc.).

2.1 Black-box models

A black-box model learns an input–output mapping purely from data with no explicit physical structure. Physics may be implicit only to the extent it is represented in the training distribution. Typical examples are tree ensembles (e.g., XGBoost), neural networks, or Gaussian-process regressors trained directly on operational inputs (speed, draft, weather, etc.) to predict power or fuel.

2.2 Physics-constrained models

Physics-constrained models incorporate physical knowledge as constraints on admissible behaviour, but typically in a coarse or partial way. The goal is to prevent obviously implausible outputs, not to represent the full causal structure of hydrodynamics.

This class includes two common mechanisms:

1. Output/behaviour guardrails (soft constraints)
   Physics appears as penalties or checks (bounds, monotonicity, convexity, feasibility filters, or "physics losses") that discourage violations but do not fully determine the solution.

2. Hypothesis-space restriction (shape-constrained / parametric).
   The model is forced to live inside a prescribed functional family (e.g., fixed polynomials, P=cVnP=cV^nP=cVn, or limited semi-empirical forms). This can enforce the shape you expect, but it can be crude if the true relationship deviates or if key state variables (draft, waves) alter the shape in ways the chosen family cannot express.

Physics-constrained models are best seen as regularised black-boxes or shape-restricted regressors: they still learn primarily from data, but with guardrails that reduce extreme failures.

2.3 Physics-informed models

Physics-informed models embed physical relationships as structural priors that shape the learning problem more fundamentally, through the objective, parameterisation, or architecture, so the model must jointly satisfy data evidence and physical structure. The key idea is not merely preventing nonsense, but improving generalisation and reliability when data are biased, sparse, or missing key regimes.

Table I: What's behind every approach

Approach	What is it	How physics enter	Enforcement
Black-box (purely data-driven)	Learns mapping purely from data; physics only appears if present in training distribution	Nowhere explicitly (at most via feature choices)	None
Physics-constrained	Adds coarse physical restrictions to reduce implausible behaviour while remaining mostly data-driven	(a) Soft guardrails: penalties/checks (bounds, monotonicity, plausibility) (b) Shape restriction: fixed functional families / limited hypothesis class	Soft (penalty/ regularisation) and/or hard shape restriction (restricted function class)
Physics-Informed	Physics shapes the learning problem so the model must negotiate data + physical structure	Physical principles act as a prior that shapes the solution space and guides training, while data determine the remaining degrees of freedom.	Moderate to strong: physical priors shape learning throughout, learns physically consis-tent relationships even when data are sparse or biased.

Table II: Implications for usability

Approach	What it's good at	When it fails
Black-box (purely data-driven)	High in-sample accuracy when data are dense and representative; easy to train.	Behave erratically outside the training envelope, and even in-distribution may fit patterns that violate basic physical relationships.
Physics-constrained	Reduces "crazy" predictions; stabilises training under noise; improves plausibility within modest shifts	Can be too crude: the imposed structure can be misaligned with reality, producing wrong responses and materially reducing accuracy (even in-distribution); partial constraints also fail to prevent some out-of-envelope failures.
Physics-Informed	Matches or outperforms black-box accuracy where data are rich, while delivering stable, physics-consistent behaviour under sparse, biased, or unseen operating conditions.	Requires careful design: choosing priors/what is residualised and to what extent, requires managing competing goals (fit vs physical correctness) and the right compromise is application- and vessel-dependent, so it takes experience and careful validation.

3. Methodology

We adopt a targeted case-study evaluation rather than a broad fleet-wide benchmark. We select a small number of vessel cases designed to isolate specific physical mechanisms: draft sensitivity and main engine power–fuel modeling. For each case, we compare modelling approaches side-by-side using both predictive accuracy on hold-out data and diagnostic checks of physical behaviour, with visual overlays used to make differences interpretable.

Our methodology is a set of variable-specific stress tests. Instead of averaging results over many vessels, we hand-pick cases where a specific physical effect is expected to dominate and where distribution shift is likely. For each case, we report standard hold-out accuracy metrics (MAPE and bias), but we also evaluate regime-specific generalisation in regions with limited data support. Concretely, we compute diagnostics on underrepresented regions, and we compare curve behaviour visually to detect physically implausible extrapolation that may not be reflected in global averages. This produces clear, interpretable contrasts in both physical correctness and out-of-sample accuracy.

4. Case Study 1: Power-Fuel model

To evaluate whether "adding physics" is a marketing label or a measurable advantage, we model main-engine fuel consumption (visualized as SFOC) as a function of engine power, and compare three approaches: a purely data-driven black-box, a physics-constrained model with an imposed curve family, and a physics-informed estimator that embeds the expected SFOC structure while retaining flexibility. Each model is trained on the same set of training data, and then evaluated on the same set of hold-out test data. Fig.1 overlays each model against the test data.

- Black-box (purely data-driven)
  The black-box model fits the operational data statistically well in the region where data is most dense, but overall the fit is clearly suboptimal. This shows up in two ways. First, the curve extrapolates erratically at lower and higher loads, producing a physically completely incorrect relation that contradicts basic engine behaviour. Second, even the error on the test data is

considerably higher than on the training data. This is the classic black-box failure mode: good apparent fit "where the data are dense," but unreliable behaviour when the operating point shifts.

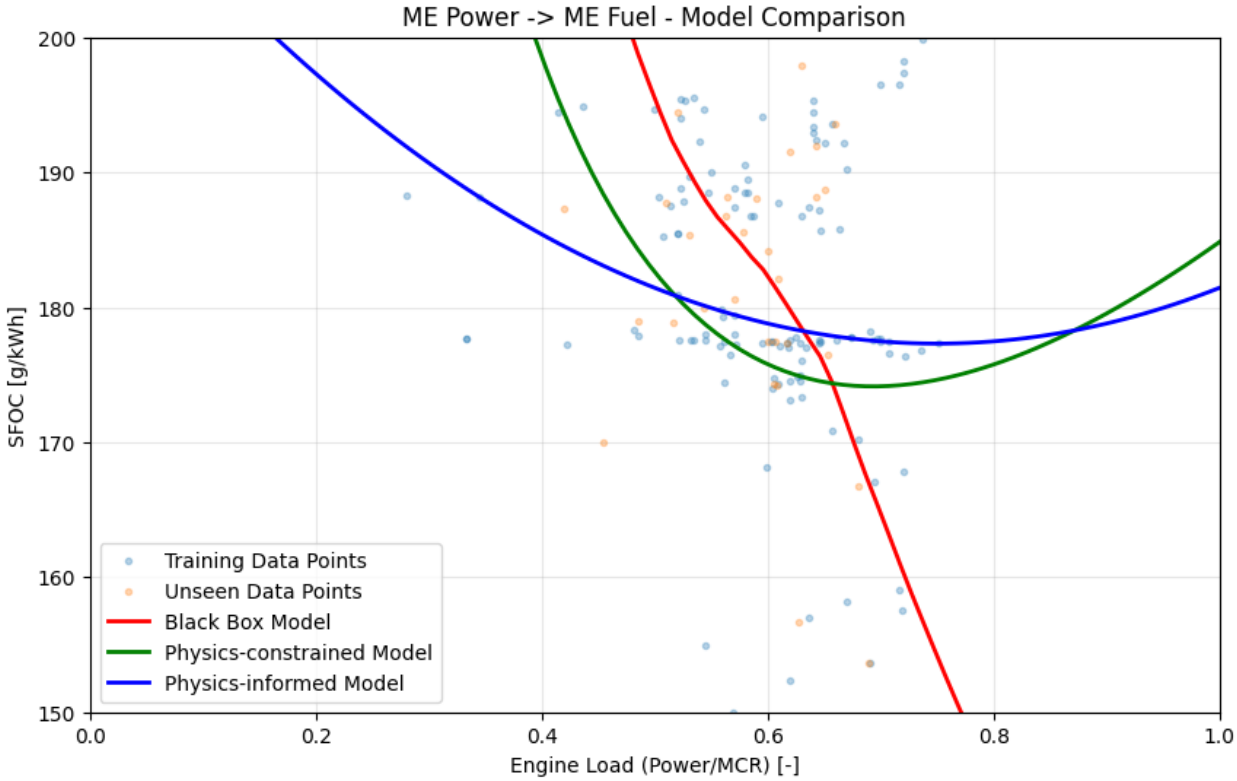


Fig.1: Power to fuel modeling using 3 different modeling approaches on limited data

- Physics-constrained
  Adding physics as constraints improves the fit on the operational dataset: performance improves to MAPE = 5.30% with a reduced bias =-2.10%, and the model no longer exhibits the most blatantly non-physical behaviour seen in the black-box curve at low load. However, this gain comes with a structural cost: the enforced functional form is still too restrictive. In Fig.1, the constrained curve rises too sharply as load decreases, implying disproportionately high fuel demand (and, by extension, an inflated SFOC once normalised by power) in the low-load regime where data coverage is sparse. The issue is visible through the model's shape: the constraint prevents "nonsense", yet it can still impose a wrong kind of plausibility: a systematic curvature bias that makes extrapolation conservative and potentially misleading precisely in the operating regions where robustness matters most.
- Physics-informed
  The physics-informed estimator achieves the best balance: it improves operational accuracy further (MAPE = 4.69%, bias=-1.18%) while preserving credible curve shape across the load range. The qualitative result in Figure 1 matches the quantitative outcome: physics is not merely acting as a post-hoc filter, but as a structural prior that guides learning toward physically consistent extrapolation without over-constraining the response.

Looking only at operational test performance, adding physics as constraints may seem adequate: the constrained model improves on the black-box model (MAPE 5.46%→5.30%) and resolves the most blatant non-physical behaviour. However, Fig.1 highlights the remaining risk: a fixed or overly restrictive curve family can still impose unrealistic curvature in sparsely observed regimes (notably at low load), creating a structurally biased extrapolation even when aggregate error improves. The physics-informed approach delivers the strongest overall result on the operational test set – lowest MAPE (4.69%) and lowest bias (-1.18%) - while also maintaining a more credible curve shape. In this sense,

physics-informed modelling is not just “guardrails,” but a way to improve both accuracy and reliability by balancing data fit with physically plausible behaviour.

Table III: Operational test-set performance summary (MAPE and bias)

	Test MAPE	Test Bias
Black-Box	5.46%	1.59%
Physics-constrained	5.30%	-2.10%
Physics-informed	4.69%	-1.18%

5. Case Study 2: Draft Effect

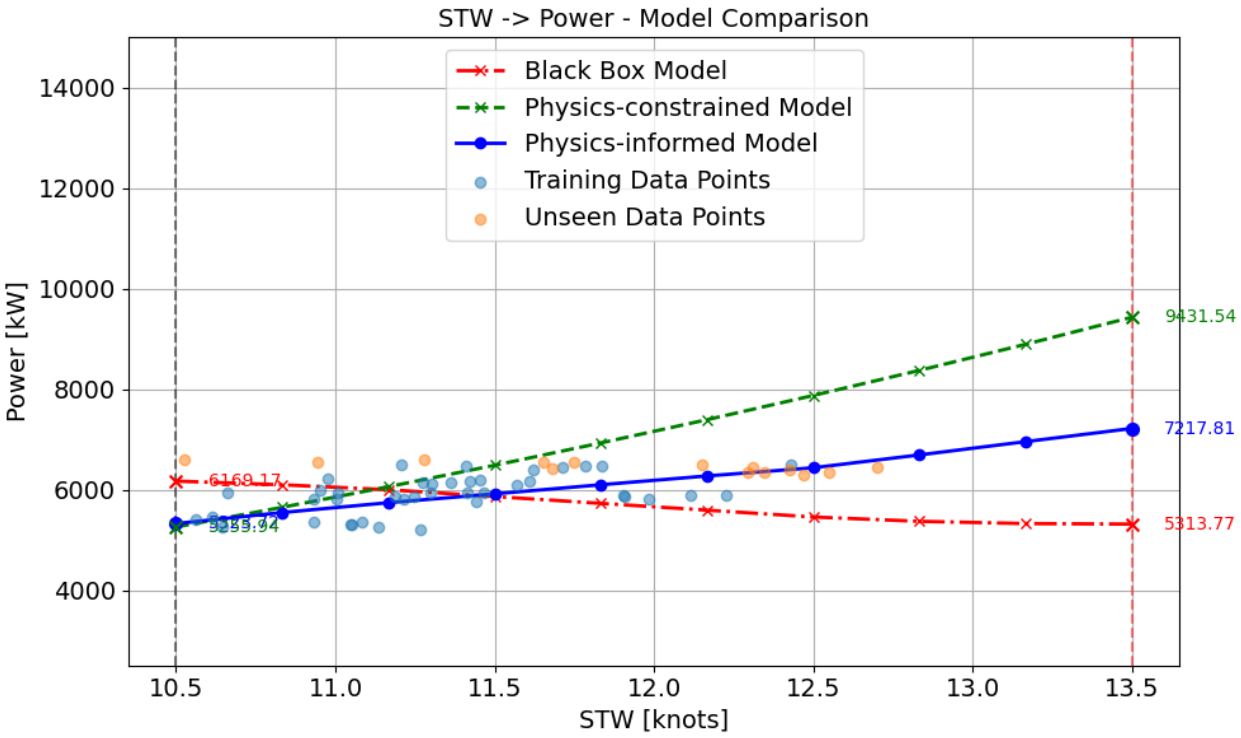


Fig.2: Speed-power modeling for uncommon draft conditions

A recurring challenge in vessel performance modelling is learning how draft affects the speed–power relationship. Higher draft typically increases resistance, so achieving the same speed should require more power (and therefore more fuel), but exactly by how much is not always clear. In operational datasets, draft coverage is often uneven: vessels operate predominantly at a few typical loading conditions, resulting in dense observations around those drafts and sparse coverage elsewhere. This case study is designed to stress-test that exact situation.

Fig.2 illustrates the problem on one representative vessel: we compare predicted power curves versus speed at a target draft (here shown at 12 m), even though the model has seen very limited training data at that exact draft. The question is not whether the model can interpolate where data are dense, but whether it can infer the draft effect correctly when the draft regime is underrepresented.

- Physics-constrained approach
  The physics-constrained model improves stability by restricting the hypothesis space to a “reasonable” family of speed–power curves and discouraging obviously implausible shapes. This typically prevents a lot of blatantly incorrect behaviour already, but it can still fail in many other ways: it may produce a curve that is smooth and monotonic, yet misestimates the draft penalty: either exaggerating or dampening the effect too much. The constraint enforces plausibility of the curve shape, but it does not necessarily provide a mechanism for learning how the curve should move with draft when the data do not explicitly show it.

- Physics-informed approach
  The physics-informed model retains enough flexibility to fit the data, but it is regularised toward physically plausible draft dependence. This matters when the evaluated draft has few samples: instead of defaulting to the behaviour learned at the most common drafts, the model produces a curve that both respects the sparse points and remains consistent with the expected direction of the draft effect. In Fig.2, this is visible as a more realistic shift of the speed–power curve with draft, compared with the constrained model's form-driven response.

This case study highlights a key distinction: when draft coverage is sparse, average fit metrics can look similar, but the operational risk lies in whether the model learns a realistic draft sensitivity. Physics-informed modelling provides a measurable advantage here, not by just smoothing the curve, but by improving the reliability of predictions in under-sampled drafts.

When draft coverage is sparse, average fit metrics can mask large differences in reliability. Here, the physics-informed model provides a measurable advantage, achieving 9.09% MAPE with -8.63% bias, versus 14.42%/-14.45% for the black-box model and 23.83% /+20.76% for the physics-constrained model. More importantly, the physics-informed speed–power curve extrapolates with a more physically plausible shape, preserving the expected direction and smoothness of the draft effect in the under-sampled regime.

Table IV: Operational test-set performance summary (MAPE and bias)

	Test MAPE	Test Bias
Black-Box	14.42%	14.45%
Physics-constrained	23.83%	20.76%
Physics-informed	9.09%	-8.63%

6. Discussion

6.1 Not all "physics in the loop" is the same

Our case studies show that the label "physics-based" hides materially different modelling choices. Physics can enter as (i) guardrails on outputs, (ii) restrictions on allowable curve families, or (iii) a deeper structural prior that shapes what the model can represent and how it learns. These choices are not interchangeable: they can produce similar average errors while behaving very differently in the regimes that matter most operationally: under-sampled drafts, rare sea states, and off-design engine loads.

A key takeaway is that physics-constrained approaches often succeed at preventing the most obvious failures (e.g., implausible slope or oscillatory curve shapes), but can still introduce a quieter failure mode: structural bias. If the chosen functional family is too restrictive, the model may remain smooth and monotone yet still be wrong in a systematic way, especially when extrapolating into low-data regions.

6.2 Why average metrics can hide the real risk

In operational datasets, the average error is dominated by the regimes where data are abundant (e.g., typical drafts and operating speeds). Standard aggregate metrics such as global MAPE therefore weight the "common regime" heavily and can understate model risk in low-support regions.

This is precisely where the distinction between physics-constrained and physics-informed approaches becomes visible. When the evaluated conditions have limited support, the model is forced to rely on its inductive biases: what it assumes about how performance should change across conditions. Two models with similar average MAPE can be separated by whether they extrapolate in a physically credible way

where the data provide little guidance. For various shipping applications, this reliability in sparse regimes is often more important than marginal improvements in average fit.

A practical implication is that validation should not stop at a single overall accuracy score. It should include "stress-tests" that isolate specific high-risk cases for underrepresented data.

6.3 The real-world constraint: neither data nor physics is a ground truth

A defining challenge in shipping is that operational data quality is often poor and unreliable. At the same time, physics-based formulas and semi-empirical relations are approximations with known limits, vessel specificity, and sensitivity to unmodelled effects. In practice, neither "trust the data fully" nor "trust the formula fully" is a safe strategy.

This is why physics-informed AI is challenging: the design task is to allocate what should be learned from data versus what should be imposed as structure, while acknowledging uncertainty in both. Doing this well requires (i) explicit validation under distribution shift, (ii) diagnostics of physical behaviour, and (iii) careful calibration of how strongly physical priors are enforced. When that balance is achieved, the payoff is significant: models that are not only accurate in the common regime, but reliable for sparse regimes that drive operational risk and decarbonisation decisions.

7. Conclusion

This paper argues that "physics-informed AI" should be evaluated as a measurable modelling choice rather than a marketing label. We make three contributions.

First, we clarify a practical taxonomy of vessel-performance learning approaches - black-box, physics-constrained, and physics-informed - based on how physical knowledge affects the model.

Second, we propose a validation mindset aligned with real deployment: performance must be judged not only by average error, but also by error in underrepresented conditions, where models are forced to extrapolate and where operational risk concentrates.

Third, through targeted vessel-level stress tests we show that these model categories differ meaningfully in practice. Physics-constrained models can reduce obvious failures and often improve aggregate metrics, but may still suffer from structurally biased extrapolation when their enforced functional form is misaligned with the true relationship, particularly in underrepresented conditions. Physics-informed models not only improve accuracy for underrepresented conditions, they also better preserve physically credible behaviour overall.

Finally, we stress that "physics-informed" is not a single method and not a guarantee: it is a design space, and outcomes depend on how well the model is designed. On top of that, physics-informed modelling in shipping is inherently a balancing act: operational data are often unreliable, and physical formulas are approximate. Achieving robust performance therefore requires expertise in both domains and careful validation, but the reward is reliable vessel performance understanding, which is essential for trustworthy fuel and emissions predictions and for accelerating decarbonisation decisions.

Enhancing Shipboard Decision-Making and Fuel Efficiency through AI-Assisted Voyage Briefings

Herman Malmsten, Wärtsilä, Helsinki/Finland, herman.malmsten@wartsila.com
Teemu Kuusisto, Wärtsilä, Helsinki/Finland, teemu.kuusisto@wartsila.com
Morgan McCall, NCLH, Miami/USA, momccall@nclcorp.com

Abstract

This paper describes a collaborative initiative between Norwegian Cruise Line Holdings and Wärtsilä Voyage Advisory to develop AI-Assisted Voyage Briefings, a Human-AI framework to enhance operational performance and shipboard team engagement in voyage optimization by delivering actionable voyage insights directly to onboard officers. Starting with manual retrospective reports to capture expert heuristics, evolving through AI-integration, the AI agents now support experts in preparing Voyage Briefings, improving productivity, and enriching analysis. The paper details the iterations, final architecture, and results of our AI-assisted Voyage Briefings, highlighting AI's transformative potential while preserving expert oversight to drive fuel savings and sustainable operations.

1. Introduction

Artificial Intelligence (AI), particularly Generative AI enabled by Large Language Models (LLMs), has been shown in empirical studies to enhance productivity and improve decision-making across a range of industries. In sectors as diverse as healthcare, *Singhal et al. (2023),* finance, *Yang and Huang (2020),* and computer science, *Chen et al. (2021),* AI systems are sifting through vast data sets to uncover patterns, forecast outcomes, and assist human experts with timely insights. The maritime industry, traditionally reliant on seasoned human judgement and characterized by a cautious approach to technological change, is increasingly exploring AI-driven decision support to address mounting regulatory, operational, and sustainability pressures. Such systems offer measurable potential to reduce costs, enhance safety, and support ambitious decarbonization targets aligned with global climate goals. Crucially, these technologies, at current stage, are most effective not as replacements for human decision-makers, but as augmented intelligence, providing real-time analytics and recommendations that enhance human expertise. This blend of AI and human judgement is particularly valuable in maritime operations, where complex variables require both data-driven analysis and contextual understanding for sound decision-making.

2. The shift to a human-AI Framework

In response to growing pressure and interest to improve fuel efficiency and operational sustainability, Wärtsilä Voyage Advisory and Norwegian Cruise Line Holdings (NCLH) initiated a collaborative effort to develop a new kind of operational support service for shipboard team as an Insight-as-a-Service: the Voyage Briefings. The vision was to create a concise, data-informed report that offers retrospective, targeted insights to support decision-making onboard and encourage fuel-efficient practices. From the beginning, there was interest in exploring whether such briefings could, over time, be reliably generated automatically using AI. However, the initial focus was on validating the concept through a fully human-cantered manual process, ensuring the briefings were valuable, actionable, and well-received by their intended users.

This manual phase served a dual purpose. First, it allowed experts to refine the content, tone, and structure of the briefings based on real-world feedback from shipboard teams. Second, it provided a rich dataset of expert-generated insights and heuristics that could later be used to inform LLM models on common insights related to marine operational efficiency. As the programme expanded across the NCLH fleet, the limitations of a fully manual process became increasingly evident, particularly in terms of scalability and consistency. While automation was initially considered to accelerate production, it became clear that fully automating the briefings would compromise the quality and contextual nuance

required for operational decision-making. As a result, the team adopted a Human–AI framework in which AI supports experts by generating draft insights, while expert validation remains central to the process.

3. Design of the Voyage Briefings

The design of the Voyage Briefings was guided by a user-centred design (UCD) approach, emphasising early and deep involvement of end-users, onboard and partly onshore personnel, to ensure the product met their needs. According to human-centred design principles, *ISO (2019),* this approach calls for active user participation and a clear understanding of the users' tasks and context of use. In practice, we employed several qualitative research methods commonly recommended in design research to gather insights:

- Interviews with onboard personnel: We conducted semi-structured interviews, *Adams (2015),* with onboard officers, both in person and remotely via video calls. Interviews are a core UCD method for eliciting user requirements and preferences. Through these discussions, officers shared what information they currently use in voyage planning, what challenges they face in optimizing fuel efficiency, and what kind of feedback or report format would be most helpful to them.

- Contextual inquiry: In addition to interviews, field observations of bridge teams during their normal work were performed. This shadowing of officers on the bridge allowed us to see first-hand how voyage information is used and what the workflow looks like in context. In a contextual inquiry, the researcher observes users in their real environment and simultaneously discusses the ongoing activities with them, *Holtzblatt and Beyer (2014).* By taking on an observer role during actual voyage planning meetings and monitoring how officers interacted with existing tools and data, this helped validate insights from the interviews.

By using these human-centred research methods, interviews to gather user needs, and contextual inquiry to understand real-world usage, the design was grounded in actual user behaviour and preferences. User insights shaped the Voyage Briefings' language, format, length, and delivery schedule. This careful attention to user-centred design established a strong foundation for the success of the Voyage Briefings.

4. AI in Voyage Briefings

LLMs have demonstrated strong capabilities in analysing structured data formats such as charts and tables, making them particularly well-suited for applications like the AI-assisted Voyage Briefing tool. Studies show that LLMs can accurately interpret chart images, *Wu et al. (2024),* summarise patterns, and detect anomalies in tabular data*, Lu et al. (2024).* Techniques such as in-context learning*, Dong et al. (2024)*, and chain-of-thought prompting, *Wei et al. (2024)*, further enhance their reasoning capabilities. In the context of voyage briefings, where experts routinely examine fuel consumption charts and leg routes to derive actionable insights, these capabilities could allow LLMs to automate repetitive analytical tasks with high accuracy.

To scale the creation of the Voyage Briefing reports and because prior research suggest it is viable to do so, an AI authoring tool was created for Wärtsilä experts. The AI authoring tool was conceived as a human-in-the-loop web application that accelerates, but never replaces, expert analysis. Its core purpose is to help experts compose Voyage Briefings that are accurate, concise, and contextually aligned with onboard decision-making, while preserving clear human accountability for every report delivered to the bridge team. In keeping with both human-centred design principles and emerging AI governance guidance, the tool is intentionally separated from the end-user i.e. the shipboard team: shipboard officers do not interact with the AI; they only receive the final report, which carries the expert's name and responsibility for its content.

To guide the tool's design, we distilled several design principles informed by literature on LLM-assisted analytics workflows, *Inala et al. (2024), Guo et al. (2024), Weng et al. (2024), Drosos et al. (2024)*. One key insight from these studies is that LLM tools should blend into experts' existing workflow and software environment. Prior studies note that AI adoption falters if the tool is disjointed from users' normal processes, so we designed the prototype to feel like a natural extension of Wärtsilä's Fleet Optimisation Solution (FOS) platform by mimicking the elements, visuals and structure of FOS.

Given LLMs' tendency to hallucinate, the tool must make it easy for experts to trace and validate the source of each insight. This led us to ensure that all AI-generated content is in vicinity of the data it is based on, allowing experts to cross-check against the original data.

Finally, task-specific guidance rather than relying on a single general-purpose chatbot was adopted. The system was implemented as a collection of specialized agents, each configured with a dedicated set of instructions aligned with a particular analytic sub-task in the Voyage Briefing workflow, Fig.1. Every agent is provided with targeted context, a clearly bounded objective, examples and instructions tailored to the insights commonly expected within its respective section of the briefing.

By assigning each analytic responsibility to an individually instructed agent, the system leverages the repetitive and rule-based nature of much of the analysis. Experts are spared from crafting new prompts for each case, while each agent remains grounded in the exact domain assumptions required for its task. This design ensures that the model approaches each component of the Voyage Briefing with the appropriate domain lens, leading to more stable outputs and reducing the cognitive overhead typically associated with manual prompt engineering.

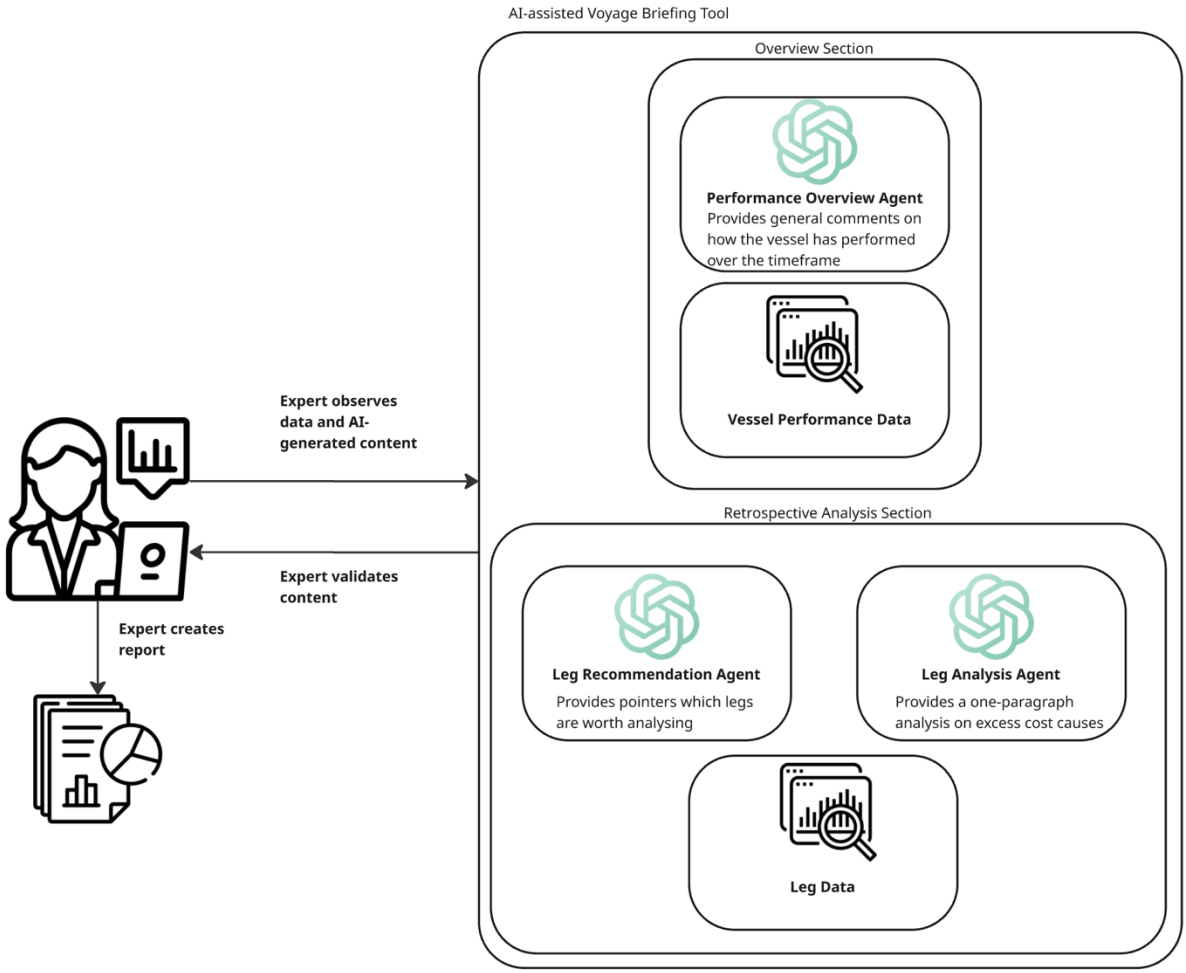


Fig.1: Expert Workflow

The UI of the AI tool is organised into task-oriented panels aligned to the real briefing sections that officers value:

- Overview Summary of the analysed period: Generates a synopsis of performance and factors that drive excess fuel costs.
- Analysis of sailed legs: Produces draft narratives for individual voyage legs, highlighting anomalous leg and generating automated analyses of selected legs, accounting for the sailed leg's excess fuel consumption due route, speed, engine and trim decisions made onboard.
- Upcoming Itinerary: Renders the scheduled route and port calls and allows the expert to preview forward-looking remarks.

The workflow embodies "high automation and high human control" (Shneiderman, 2020). Experts initiate a briefing, review AI-generated sections, and accept, modify, or discard suggestions and even provide guidance for AI on what to focus on when analysing a specific leg. This ensures accountability and that AI-generated content is validated.

5. Results

To assess the impact of the AI-assisted Voyage Briefing tool on expert analyst productivity, we conducted a controlled evaluation with expert users. The study involved a within-subjects design: 4 experts completed two batches of 12 reports, 2-4/expert, first using the conventional manual workflow, and then using the AI tool. The reports were drawn from real-world vessel data and matched for complexity across both conditions.

The results showed a substantial improvement in efficiency. On average, the time required to complete a report decreased by 55% when using the AI-assisted workflow. Importantly, the tool preserved expert control and accountability, with experts reviewing and editing all AI-generated content before finalization. Notably, from the end-user (shipboard team) perspective, the resulting Voyage Briefings were indistinguishable from manually produced reports. These findings suggest that the AI tool not only accelerates the reporting process but also supports a more scalable approach to delivering high-quality Voyage Briefings across the fleet.

6. Operational efficiency as a decarbonization lever

From Wärtsilä's perspective, a core objective of the Voyage Briefings initiative has been to support NCLH on their decarbonisation journey by helping shipboard teams identify and act on focused opportunities to reduce excess fuel consumption. By delivering targeted, data informed recommendations directly to vessel officers, the briefings translate complex performance data into practical operational decisions that reduce emissions. In doing so, the initiative addresses one of the most immediate levers for decarbonisation: operational efficiency. Even marginal improvements in voyage execution, when sustained across multiple vessels and itineraries, can result in meaningful cumulative emissions reductions over time.

At scale, the voyage briefings have the potential to drive behavioural change and operational optimization. By reinforcing best practices and highlighting actionable insights, the Voyage Briefings have contributed to a broader culture of energy awareness and continuous improvement. As the tool development evolves, AI-assisted workflows are expected to further enhance the consistency, scalability, and timeliness of these insights. At the fleet-wide level, the impact is visible in improved alignment toward performance baselines and reduced excess consumption; at the vessel level, the value lies in how these recommendations are embedded into daily decision-making onboard.

7. Benefits beyond fuel savings

From NCLH's perspective, the value of the Voyage Briefings extends beyond fuel analytics. The initiative represents a structured approach to bridging data, expertise, and onboard execution. While advanced performance dashboards were already in place, the structured briefing format has strengthened how insights are translated into operational actions. By distilling complex performance data into concise, voyage-specific narratives, the briefings support officers in identifying and addressing avoidable excess fuel consumption without increasing workload or reporting obligations. Although external factors such as weather and itinerary constraints continue to influence fuel use, the structured briefings have contributed to post-voyage analysis. This has helped embed fuel efficiency considerations more consistently into voyage planning operations.

The phased Proof-of-Concept approach was instrumental in ensuring operational relevance. Trial vessels were selected carefully, briefing templates were agreed jointly, and frequent feedback loops were established to refine both content and format before broader scaling. This iterative development process ensured that the AI-assisted briefings reflected real operational conditions rather than theoretical assumptions. As a result, the final product was not perceived as an external reporting layer, but as a practical decision-support tool shaped by operational teams input.

Equally important has been officer reception. Feedback indicates that the briefings are valued for their clarity and practical relevance, particularly when insights are directly tied to specific voyage legs. The collaborative development process and ongoing expert validation have ensured operational credibility. For NCLH, the Voyage Briefings demonstrate how a carefully developed Human–AI workflow can enhance engagement, improve consistency in operational practice, and support sustained progress toward decarbonisation without disrupting established bridge routines.

8. Limitations

While the AI-assisted Voyage Briefing tool has shown promising results in early testing, several limitations should be acknowledged. Firstly, no formal technical evaluation has yet been conducted to quantify the AI's accuracy specifically on the types of charts and maps used in this maritime context. As a result, we currently rely on expert validation to assess the correctness and relevance of AI-generated insights. Secondly, the evaluation was conducted with a relatively small user pool and a limited batch of reports. Although the participant group included all available experts with relevant experience, the sample size nonetheless constrains the generalizability of the findings. Resource limitations also prevented a larger-scale or longitudinal study at this stage. Finally, while the observed 55% reduction in reporting time is encouraging, it is difficult to isolate how much of this improvement can be attributed solely to the AI component. The tool incorporates a few quality-of-life enhancements which may also have contributed to the efficiency gains. Further studies are needed to disentangle these effects and better understand the specific value added by the AI.

9. Future Directions

While the current AI-assisted Voyage Briefing tool operates within a human-in-the-loop framework, our long-term ambition is to explore the feasibility of full or near-full automation of the briefing process. At present, expert validation remains essential to ensure the accuracy, contextual appropriateness, and operational relevance of AI-generated insights. This safeguard is particularly important given the safety-critical nature of maritime operations and the known limitations of LLMs, such as occasional hallucinations or misinterpretation of domain-specific data. Looking ahead, our focus is on systematically reducing the dependency on manual oversight by improving the reliability, transparency, and domain alignment of the AI system. A key research question moving forward is: how close can we get to full automation without compromising safety, trust, or analytical quality? Answering this will require a combination of technical benchmarking and iterative design.

Ultimately, we envision a future where Voyage Briefings can be generated at higher frequency and scale, with minimal expert input, while still upholding the standards of quality and accountability that shipboard teams rely on. Achieving this will not only enhance operational efficiency but also unlock new opportunities for real-time decision-making support and continuous performance monitoring across the fleet.

Acknowledgements

We would like to express our sincere thanks to the NCLH Decarbonization & Energy Performance team for their collaboration and support throughout this work. We are also grateful to the trial-ship personnel who shared their time, insights, and practical experience during the interviews. Special thanks also go to Anshul Tuteja for his early support and encouragement, which made this work possible. We extend our appreciation to Mona Kanafi and Cedric Deymier for their guidance, encouragement, and continuous feedback during the project. Finally, we wish to thank the Wärtsilä Voyage Briefing experts, whose contributions were instrumental in shaping both the project and the evolution of the Voyage Briefings.

References

ADAMS, W.C. (2015), *Conducting semi-structured interviews*, Handbook of Practical Program Evaluation, pp.492-505

CHEN, M. (2021), *Evaluating large language models trained on code*, arXiv:2107.03374

DONG, Q.; LI, L.; DAI, D.; ZHENG, C.; MA, J.; LI, R.; XIA, H.; XU, J.; WU, Z.; CHANG, B.; SUI, Z. (2024), *A survey on in-context learning*, Conf. Empirical Methods in Natural Language Processing, pp.1107-1128

DROSOS, I.; SARKAR, A.; XU, X.; NEGREANU, C.; RINTEL, S.; & TANKELEVITCH, L. (2024). *"It's like a rubber duck that talks back": Understanding Generative AI-Assisted Data Analysis Workflows through a Participatory Prompting Study*, 3rd Annual Meeting of the Symp. on Human-Computer Interaction for Work, pp.1-21

GUO, J.; MOHANTY, V.; PIAZENTIN ONO, J.H.; HAO, H.; GOU, L.; REN, L. (2024), *Investigating Interaction Modes and User Agency in Human-LLM Collaboration for Domain-Specific Data Analysis*, CHI Conf. on Human Factors in Computing Systems, pp.1-9

HOLTZBLATT, K.; BEYER, H. (2014), *Contextual design: evolved*, Morgan & Claypool Publ.

INALA, J. P.; WANG, C.; DRUCKER, S.; RAMOS, G.; DIBIA, V.; RICHE, N.; ... GAO, J. (2024), *Data Analysis in the Era of Generative AI*, arXiv:2409.18475

ISO (2019), *ISO 9241 - Ergonomics of Human System Interaction,* Int. Standard Org., Geneva

SHNEIDERMAN, B. (2020), *Human-centered artificial intelligence: Reliable, safe & trustworthy*, Int. J. Human–Computer Interaction 36(6), pp.495-504

SINGHAL, K.; AZIZI, S.; TU, T.; MAHDAVI, S. S.; WEI, J.; CHUNG, H. W.; ... & NATARAJAN, V. (2023), *Large language models encode clinical knowledge*, Nature 620(7972), pp.172-180

WEI, J.; WANG, X.; SCHUURMANS, D.; BOSMA, M.; XIA, F.; CHI, E.; ... ZHOU, D. (2022), *Chain-of-thought prompting elicits reasoning in large language models*, Advances in Neural Information Processing Systems 35, pp.24824-24837

WENG, L.; WANG, X.; LU, J.; FENG, Y.; LIU, Y.; CHEN, W. (2024), *InsightLens: Discovering and Exploring Insights from Conversational Contexts in LargeLanguage-Model-Powered Data Analysis*, arXiv:2404.01644

YANG, Y.; UY, M.C.S.; HUANG, A. (2020), *Finbert: A pretrained language model for financial communications*, arXiv:2006.08097

2,000 Digital Ships: A Scalable Method for Emissions Benchmarking and Independent Validation

Erik Verboom, AlbatrosDigital, Rotterdam/The Netherlands, erik@albatros.digital
Nico van der Kolk, AlbatrosDigital, Rotterdam/The Netherlands, nico@albatros.digital
Joris de Vroom, Hermess, Emmeloord/The Netherlands, joris.devroom@hermess.nl
Jos de Laat, Royal Netherlands Meteorological Institute, The Netherlands, jos.de.laat@knmi.nl

Abstract

This paper presents a scalable method for constructing large numbers of digital ship models from public specifications and AIS-derived operating profiles. Models are compared against EU MRV data to assess the match for predicting fuel use and CO_2 output. We also investigate correlations with satellite-based emission estimates, offering pathways for independent verification. Combining naval architecture, big data, and remote sensing, the approach supports benchmarking, compliance, and technology evaluation. Results prove a good match between simulation results and MRV reporting at only 2.9% median discrepancy with a 1σ spread of 31.9% on an individual ship-level. Ship-type and age-specific correlations are explored as further methods for reducing the spread.

1. Introduction

The objective of this study is to evaluate the consistency of physics-based ship performance modelling when scaled from single-vessel applications to fleet level. While individual ship performance analysis is well established, its industrialisation across thousands of vessels introduces challenges in data quality, automation robustness, and methodological comparability. This paper investigates how such a pipeline performs when applied systematically to multi-year operational histories across a heterogeneous fleet.

We present a modular modelling workflow that integrates AIS-derived vessel tracks, publicly available ship particulars, and hindcast metocean data to estimate fuel consumption over multi-year periods. The pipeline is designed for reproducibility and minimal ship-specific tuning, enabling large-scale application without manual intervention per vessel.

To evaluate modelling behaviour at scale, results are benchmarked against EU THETIS-MRV data. MRV reporting is treated as a standardized regulatory reference dataset rather than absolute ground truth. The purpose of the comparison is therefore not to certify individual ship emissions, but to assess systematic agreement patterns, bias structure, and dispersion characteristics across a statistically meaningful fleet sample.

The analysis focuses on fleet-level accuracy characteristics over a six-year period (2018–2023), covering more than 1600 ship-years of simulation. By examining aggregate bias, spread, and structural correlations with ship type and age, the study provides insight into where large-scale physics-based modelling performs consistently and where systematic deviations remain.

The contribution of this work lies not in replacing regulatory reporting, but in demonstrating that fleet-scale digital ship modelling can produce statistically coherent results when benchmarked against standardized emissions data. The findings inform both the robustness of large-scale modelling approaches and the limitations that must be considered when interpreting such comparisons.

2. Data sources

To enable the complete modelling pipeline, several external data sources will be used. Where possible, open-source datasets were used, but when necessary, commercially available datasets were used as a more complete fallback option.

- EU THETIS-MRV: ship-year CO_2/fuel, distance, time at sea/berth, transport work; used strictly for external validation.
- AIS: Historic location and navigation status data per ship; used to infer operational states and reconstruct voyages.
- Technical registers: Main particulars of vessels. Used for vessel modelling and scaling from similar known models.
- Metocean: Global hindcast data on wind, waves and currents. Used to model voyages and estimate the fuel consumption at location – timestamp combinations.
- Port locations: An open-source list of 4000 ports worldwide with their locations.

2.1. EU THETIS-MRV

THE¬TIS-MRV is the EU system/platform run by EMSA (European Maritime Safety Agency) for Monitoring, Reporting, and Verification (MRV) of fuel consumption and CO_2 (and other GHGs) emitted by large ships calling at EU/EEA ports. It supports companies (shipowners/operators), verifiers, and administering authorities in fulfilling their legal obligations under the EU MRV Regulation (Regulation (EU) 2015/757) and its amendments. It involves Monitoring Plans, Emission Reports, verification, and public reporting. THETIS-MRV reporting is mandatory since 2018 for each vessel that has had at least one port call in a given year. Originally, this was for all vessels larger than 5000 GT, as per 2025 some vessels between 400 and 5000 GT are also required to report their emissions.

THETIS-MRV data is reported to EMSA on a yearly basis. Reporting of fuel consumption and emissions data within this framework is done after third-party validation. Nevertheless, one long term goal of this research is to enable independent validation of reported emissions data, for example by comparing the reporting to satellite observations. Relevant to this study are reporting details such as how a voyage is defined (port-to-port), which voyages are reported (at least one port inside the EU), which ports count as EU ports, how to handle voyages that start in one year and finish in another, etc. Section 3.2 provides further details on the methodology used enable replication of MRV reported data.

2.2. AIS Tracks

AIS (Automatic Identification System) is a vessel tracking technology that was introduced to improve maritime safety and collision avoidance. The system is mandatory for all passenger ships and all cargo ships greater than 500 GT on all voyages. In practice, this means that all vessels of interest to this study have AIS transceivers installed.

AIS messages are transmitted by the ship using VHF radio and can be picked up (and forwarded) by other ship- and shore station (Terrestrial AIS) as well as satellites (Satellite AIS). Combined, a global location history of a vessel can be reconstructed, with a new location every few minutes on average. Commercial datasets with vessel histories going back to the early 2000s for the entire merchant fleet are available and have been used for this study. Their completeness and correctness have been evaluated based on individual sample ships in the dataset of this study.

An AIS message contains several fields regarding vessel information: it's current voyage and navigational status. It is worth noting that commercial providers of AIS datasets will often divide the complete set of transmitted data into several packages, to be sold separately. Another drawback of AIS is that the data is self-reported, and for some fields manually entered, leading to inaccurate values and significant differences between individual ships.

For this study, the fields of relevance are: Latitude, Longitude, Timestamp, Speed, Draught, and Navigational status code.

Within the context of this study, AIS data has been used to construct a history for each vessel within the timeframe of interest. Speed and navigational status have been used to estimate when separate voyages start and end. Furthermore, the draught field has been used to correctly model each voyage and provide accurate fuel consumption estimates.

2.3. Vessel Data

One of the goals of this study is to get an idea of vessel modelling accuracy based on a minimal set of vessel-specific information. The vessel specific datapoints that have been used for this study are: Length over all, Beam, and Draft.

The rest of the required modelling is obtained and scaled within the AlbatrosDigital modelling engine from comparable ships of the same type.

2.4. Metocean data

The metocean data considered for this study consists of wind data, wave data and current data. For each, different models have been used to provide the most accurate hindcast weather.

2.4.1. Wind / Waves

The global hindcast was generated by Oceanum with the third-generation spectral wave model WAVEWATCH III, *Tolman (1991),* with forcing the ERA5 dataset to produce a global wave hindcast of unprecedented quality. The hindcast was run using the recently released version 6.07 of WAVEWATCH III. ST4 source terms were used *Ardhuin et al. (2010)*, which have been established over the past decade as the most suitable for global application, e.g. *Stopa et al. (2015)*. The model uses ERA5 hindcast wind-data obtained from ECMWF.

In addition to the global model several regional hindcasts are generated several regional grids have been generated to ensure high quality data that covers the globe (https://datasets.oceanum.io/). The regional hindcasts are generated with SWAN ST6 with a resolution of 0.05° x 0.05° and are calibrated and validated against satellite measurements. The models are forced with ERA5 wind fields.

The regional SWAN models are nested in our 0.5° global WaveWatch III model. This way the swell from the oceans is properly taken into account. The specifications for regional SWAN models are:

- WW3 Spectral boundaries
    - 0.5°
    - ERA5 forcing
    - Extremes correction
- Spectral output typically at 0.05° x 0.05° resolution
- Validated against satellites

2.4.2. Currents

Total surface current data are obtained by linearly superimposing tidal current components on residual ocean currents. For tidal data, the global Oceanum tidal dataset with a spatial resolution of 400 m is used. Here, the tidal constituents have been downscaled from TPXO9 global tidal solution using the OTIS model of barotropic ocean tides, *Egbert and Erofeeva (2002)*. Current residual velocities, sea surface elevation, temperature and salinity will be extracted from GLORYS, a global 1/12 ° reanalysis product released by the EU-funded Copernicus Project. Atmospheric forcing data, including 10 m wind speed, mean sea-level pressure, precipitation, temperature, humidity and solar radiation will be sourced from the ERA5 reanalysis product.

Non-tidal (residual) currents and water level will be obtained from the Mercator 9 km Global reanalysis. The reanalysis includes daily means of temperature, salinity, currents, sea level, mixed layer depth and ice parameters from the top to the bottom over the global ocean. It is based largely on the current real-time global forecasting CMEMS system. The model component is the NEMO platform driven by ECMWF winds. Observations are assimilated by means of a reduced-order Kalman filter. Along track altimeter data (Sea Level Anomaly), Satellite Sea Surface Temperature, Sea Ice Concentration and in-situ Temperature and Salinity vertical Profiles are jointly assimilated. Moreover, a 3D-VAR scheme provides a correction for the slowly evolving large-scale biases in temperature and salinity.

2.5. Port locations

An extensive and open-source list of about 4000 worldwide port locations has been used for voyage identification. The complete list can be found on https://github.com/tayljordan/ports.

3. Methodology

The modelling for this study takes place in multiple steps. An individual model of each ship will be made. This model will contain a fuel table detailing the expected – time-averaged – fuel consumption in every possible operational state (speed, draft, metocean parameters).

With these individual vessel models, voyages are simulated over an analysed timeframe of 6 years (2018 to 2023). The historical vessel tracks are divided into separate voyages. For each voyage, the fuel consumption at each track point is calculated and added up to aggregate over desired timeframes.

At last, aggregated numbers on distances sailed, days at sea and fuel consumption can be computed. Great care is taken to follow the methodology as prescribed by the EU regulations to ensure that results can be reliably compared to reported in the EU THETIS-MRV datasets.

3.1. AlbatrosDigital vessel modelling

AlbatrosDigitals software is a ship simulation tool based on a performance prediction program. Around that, a cloud-based modelling suite of naval architecture and marine engineering tools has been developed aggregating vessel information, metocean data and vessel history. At its core, the software uses a constrained-optimization solver that computes a ship's response to maintain a steady course in 4-dof under given conditions such as draft, ship speed, wind conditions and wave conditions. An adjustable level of simulation physics is available to match the modelling purpose, or available ship information. The AlbatrosDigital software is highly customizable, allowing users to easily include their own specific models and requirements.

For the presented study, the challenge was to work with a minimal set of ship-specific information but keep the modelling accuracy as high as possible. This has been done by working with a total of 48 reference ships of various ship types and sizes. Detailed ship characteristics such as calm water resistance curves, propulsive efficiencies and engine/fuel types for each individual model have been taken from the closest reference ship and scaled according to the main dimensions where appropriate. A sufficient similarity score between reference ship and individual model has always been ensured.

3.2. EU THETIS-MRV definitions and details

To be able to compare simulation results to the reported MRV data, it is essential to follow the methodology prescribed by the regulation. This encompasses the definition of exactly which fuel consumption and emissions are reported, the definition of a “voyage”, which voyages do count for the reporting, and many more details.

3.2.1. Voyage definition

A “voyage” according to EU MRV regulation is always berth to berth. Each voyage will start and end with a port call. Only port calls where cargo is loaded/unloaded and/or where passengers embark/ disembark are counted. Events like anchoring, refuelling, pilot boarding, changes of crew etc do not break a voyage, even if this happens in a port. Anchoring time, while it does not break a voyage, is not counted as “hours at sea”, a metric that is also reported on a in the MRV reports, https://www.dnv.com/maritime/insights/topics/mrv/FAQs-EU-MRV/.

A voyage is always reported entirely in the year in which the start date of the voyage falls. For example, a voyage that starts in December 2018 but ends in January 2019, should be reported in 2018.

3.2.2. Voyage reporting

Voyages must be reported if the start and/or end port lie within EU territory. EU territory is defined as ‘mainland’ EU and several overseas territories: Açores, Madeira, Canarias, Guadeloupe, French Guyana, Martinique, Mayotte, Saint Martin and Reunion. Next to that, also Norway (except Svalbard) and Iceland ports are also included in the EU MRV reporting.

If a voyage starts in one year and finishes in the next, it shall be reported in the first year, the year of departure, https://www.lr.org/en/services/statutory-compliance/eu-ets-and-eu-mrv/eu-mrv-regulation/.

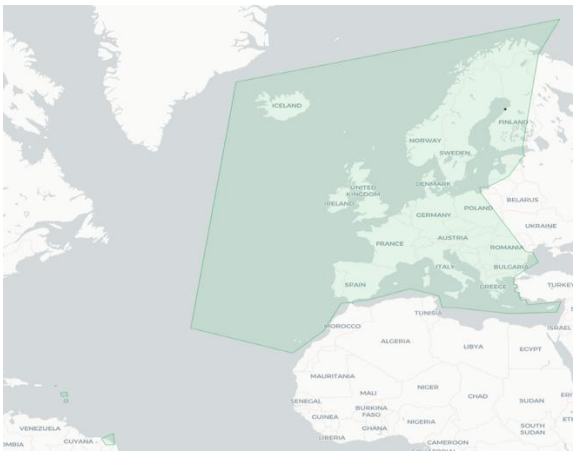


Fig.1: EU geofence for identifying EU-connected voyages. EU mainland and overseas territories in the Caribbeans

3.2.3. Fields of interest and definitions

The public MRV reports have multiple fields that are of interest in this study, and a few assumptions need to be made to be able to make sensible comparisons to simulations:

- Total fuel consumption [t]: This is the total fuel that is consumed within the Thetis MRV jurisdiction (EU connected voyages + EU ports)
- Annual average Fuel consumption per distance [kg/nm]
- Total CO_2 emissions [t]: Total emissions: in EU ports + EU connected voyages
- CO_2 emissions which occurred within ports under a MS jurisdiction at berth [t]: in-port emissions

It is important to note that both the fuel consumption and CO_2 fields refer to the total amounts for the entire ship, including main engine and any generators that might be operating. To make the simulation vs reporting comparison that this study tries to achieve, the total distance sailed annually is calculated based on the first two fields.

$$Distance\ sailed\ [nm] = \frac{FC\ total\ [m\ tonnes]}{FC\ per\ nm\ [\frac{m\ tonnes}{nm}]}$$

In a second step, to estimate the fuel consumption attributed to voyages only, we look at the CO_2 emissions in-port vs total:

$$FC\ at\ Sea\ [m\ tonnes] = FC\ total * (1 - \frac{In\ Port\ CO2\ [m\ tonnes]}{Total\ CO2\ [m\ tonnes]})$$

This second step makes the assumption that the emissions per kg of burned fuel are identical in port and at sea, which is a fairly dramatic simplification glossing over things like shore power, dual-fuel systems and much more. However, for the goals of this study, this crude approximation should serve well.

3.3. AIS tracks

One of the main challenges for this study is to correctly interpret a vessels history from its AIS data. Specifically, the AIS track needs to be separated into voyages, voyages need to be categorised into 'intra-EU', 'entering EU' and 'leaving EU'. Each voyage can be simulated in the AlbatrosDigital Virtual Voyage in order query the correct historic metocean data and calculate aggregated variables such as fuel consumption and voyage duration.

3.3.1. Down-sampling

To keep the total amount of points manageable and to keep the required compute within the budget, all AIS tracks have been down sampled to one point per hour. Larger gaps that occurred within tracks were also filled up with one point per hour, assuming a straight-line trajectory of the vessel.

3.3.2. Identifying voyages

The first step is to divide the AIS track into separate voyages. Several methods have been investigated to determine whether a new voyage has been started. Looking at the reported speed, destination, changes in draft and navigation status, or any combination of the above have been tried out. It turns out that the accuracy of reported (AIS) data can be very questionable, Fig.2.

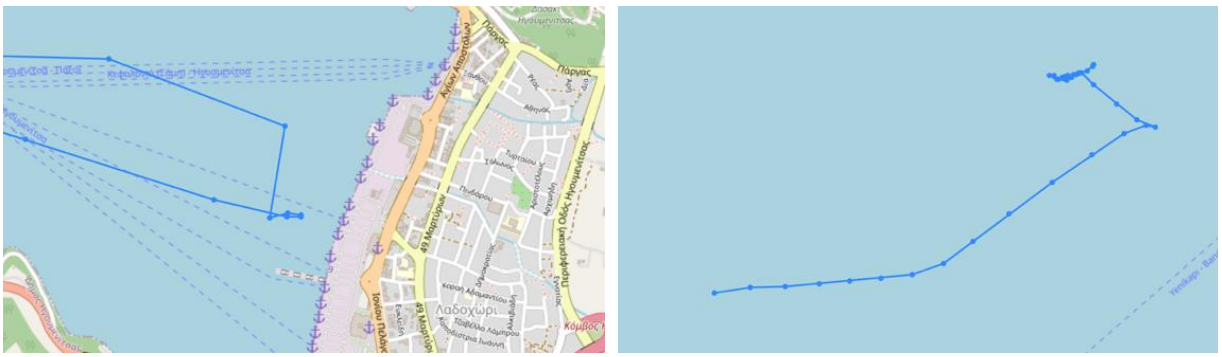


Fig.2: Discrepancies in navigation status: A ferry berthing for about 1.5 h, but never reporting status 5 (left), 1 h of sailing while the status says "anchoring". (right)

In the end, to keep the algorithm simple, fast, and applicable to datasets from various sources, we focussed on a combination of navigation status, reported speed and harbour proximity. To validate the voyage identification, resulting yearly sailed miles have been compared to the MRV reports. The mismatch between yearly reported distance and yearly sailed distance can be seen below. This has been used as one of the main criteria to decide which vessels to include in the results.

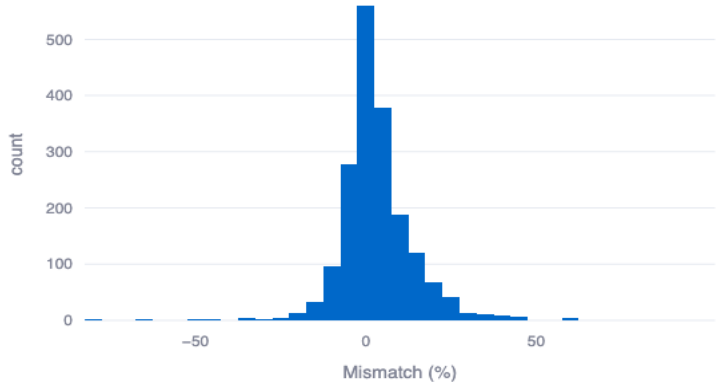


Fig.3: Matching sailed miles to reported miles per ship over a 6-year period (2018 to 2023). Y-Axis counts ships, X-Axis the percentage difference between reported and simulated voyages. Half of the vessels (959 / 1848) are below 5% mismatch.

4. Vessel Selection

While the title of the study says 2000 digital ships, the reality is a bit more complicated than that. Starting from 21.063 ships that are present in the MRV reporting data, through multiple steps and due to various boundary conditions, the set of ships has been reduced to 270 ships that have been modelled for the complete 6-year period. The objective of this selection process was to keep the ship set representative of the MRV fleet while complying with hard technical constraints and minimising the influence of data or pre-processing errors. This way the study can provide a complete and fair indication of the match that can be obtained between simulation data and reported data.

4.1. Ship type

To be able to create a model with only publicly available data of a ship, the Digital Shipyard needs to take the ship type into account. Reality is messy and the ship types reported in the Thetis MRV dataset will contain exotic values. Also, it turns out that – at least according to the reporting – ships seem to change type rather frequently. One year a vessel might be a general cargo ship, and next year it suddenly becomes a Bulk carrier. After cleaning up the ship type categorization, a complete table of ship types and the number of times they occur in the dataset can be found in Table I.

Table I: Ships in MRV dataset, by ship type

Ship type	**Number**	**Normalized**	**To Model**
Bulk carrier	7.365	35 %	Yes
Oil tanker	3.388	16 %	Yes
Container ship	2.967	14 %	Yes
Chemical tanker	2.110	10 %	Yes
General cargo ship	1.691	8 %	Yes
Gas carrier	626	3 %	Yes
Vehicle carrier	615	3 %	Yes
LNG carrier	492	2 %	Yes
Ro-pax ship	468	2 %	Yes
Passenger ship	290	1 %	No
Ro-ro ship	288	1 %	Yes
Refrigerated cargo carrier	200	1 %	No
Container/ro-ro cargo ship	70	0 %	Yes
Combination carrier	15	0 %	No
Passenger ship (Cruise Passenger ship)	10	0 %	No
Other ship types	468	2	No
Totals	**21.063**		**20.080**

About 983 ships, or about 4.6% of the ships from the dataset are excluded due to a ship type that can't easily be created with AlbatrosDigitals fully automated modelling approach. For the modelled fleet that will be modelled, we will try to approach the percentages per ship type found in Table II.

4.2. Reporting periods

A second factor for vessel selection is the total number of sailed miles that a vessel has reported and the number reported data points. It is relevant to have a sufficiently long timespan to compare simulation to reported data for each ship. This is to ensure that reporting errors and data anomalies will be averaged out instead of skewing the study results. For this reason, ships that have many miles reported are preferred.

The choice has been made to only consider ships that reported to MRV-THETIS every single year between 2018 and 2023. These will be ships that have a large reported milage on one hand, and a total of six datapoints for comparison (one for each year). This criterium leaves 4.219 ships for this study.

4.3. Reference models

To be able to create a detailed and accurate model of each individual ship based off only main particulars, AlbatrosDigital works with a system of reference ships. To ensure model accuracy, a similarity score between the ship to model and the reference ship is calculated based on the main particulars. If this similarity score is below a certain threshold, no model will be created. This restriction further limits the number of ships suitable for this study. Especially small general cargo ships show a wide variance in design choices, making them less suitable for this approach. In combination with the desire to keep ship sizes and ship-types representative of the entire fleet, this limits the ships suitable for this study to 1.848.

4.4. Sailed distance

A check is done comparing the reported miles sailed in each year to the calculated miles sailed each year for each ship. Due to the specific definitions and reporting rules of THETIS-MRV and inaccuracies in the AIS data and the voyage identification algorithm, a good match on sailed distance was not always possible.

Since a mismatch in sailed distance will inevitably lead to differences in fuel consumption, a threshold has been set at a maximum of 5% discrepancy over the 6-year period. If this threshold is exceeded, the ship has been excluded from this study. This leaves 959 vessels for this study.

4.5. Simulation time

Finally, the compute budget for the study is limited, making it impossible to simulate all 6 years for all 959 vessels. The selection of ships and/or simulation timeframe needs to be reduced further. In the end the choice has been made to limit the simulations to 270 vessels but keep the entire 6-year period, covering over all 1620 ship-years of tracks.

5. Results

As mentioned before, the simulation runs for this study cover 1620 ship-years of sailing. For each ship-year there are two fuel consumption estimates available: the yearly MRV reported value and the value that results from the simulation. This study does not aim to take any of these sources as the "ground truth". Nevertheless, the good match – looking at mean and median – suggests that both fuel consumption estimates yield plausible results. As expected for such a large sample size and real-world data, there is a decent spread in data-match. The results presented in this paper will focus on analysing the mean, the spread and various interesting correlations that can be found in the data.

5.1. Data match

The first thing for analysing the results of this study is to have a look at the over-all match between reported and simulated fuel consumption in Figure 4. The simulation results match the reported data well: The simulation mean predicts a 5.18% higher fuel consumption than reported, with the median discrepancy being 2.89%. Except for some explainable outliers, all datapoints lie within the bandwidth of ±50%. It is worth noting that the spike at 100% is caused by several outliers. This bin collects anything > 100% error. Outliers that are collected in this bin and that skew the mean can be explained mostly by clear misreporting over one specific year (see section 6.3.2 for an example).

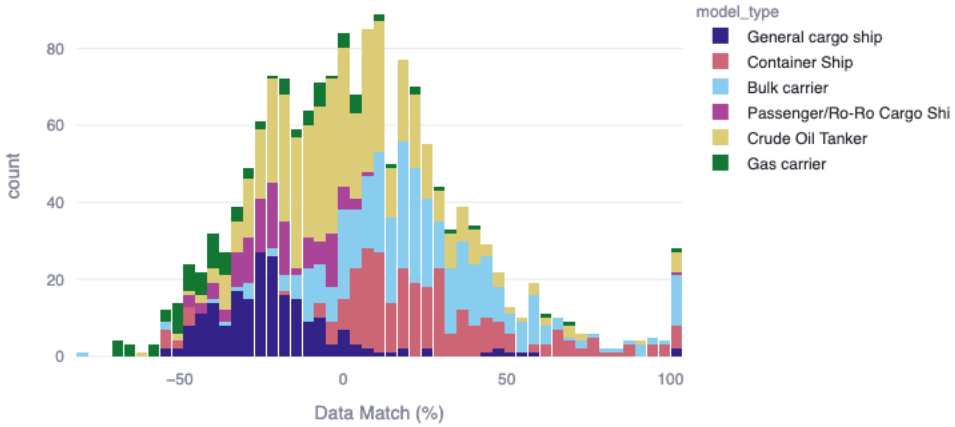


Fig.4: Match between simulated and reported fuel consumption. Median deviation is 2.89% while the 1σ spread is 31.94%.

5.2. Correlations

To understand the source of this spread further and potentially allow further analysis and improvement of modelling, correlations between various variables and the fuel consumption have been analysed. The most clear and strong correlations are discussed below.

5.2.1. Sailed distance

As expected, the mismatch in sailed distance between simulation and reported data translates pretty much on-to-one to a mismatch in predicted fuel consumption mismatch, accounting for about 10% of the observed spread in the data-match.

Fig.5: Correlation between sailed distance error and fuel consumption mismatch

5.2.2. Ship type

As can be seen in Fig.4, both the mismatch in fuel consumption and the spread between individual ships strongly depends on the ship type. The simulation seems to underpredict for Bulk carriers and Container vessels, while it overpredicts for General cargo ships, Ro-Ro ships and Gas carriers. Digging into the cause of this effect deeper is a clear recommendation for further research. Currently, the data does not provide satisfactory explanation for these trends. Some hypothesis might be that there are variations in hotel-load that are not modelled correctly. Another hypothesis is that the chosen hull shapes that represent each ship type differ quality.

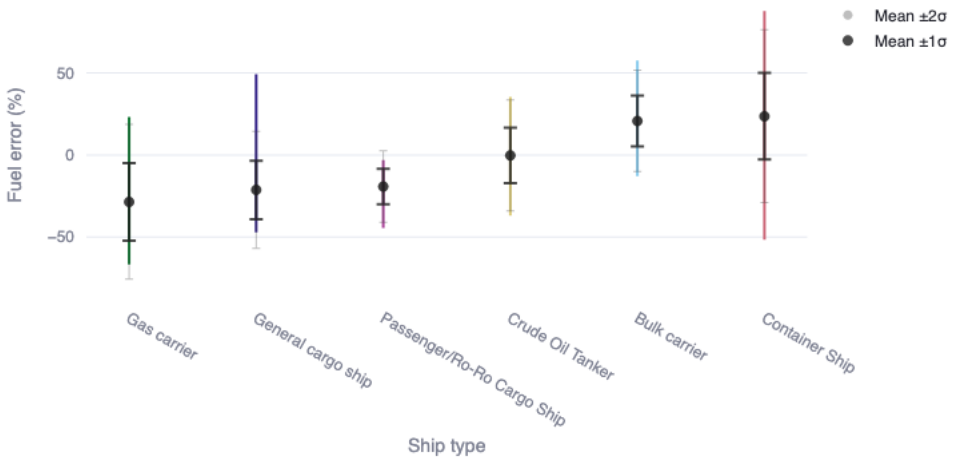


Fig.6: Spread and mean error per ship type

5.2.3. Year of build

Another Interesting correlation is the year of build compared to the fuel consumption mismatch. The modelling did not consider any penalty for older ships. This shows that – on average – newer ships are indeed more efficient than older ships. This effect seems to be quite significant over longer time periods.

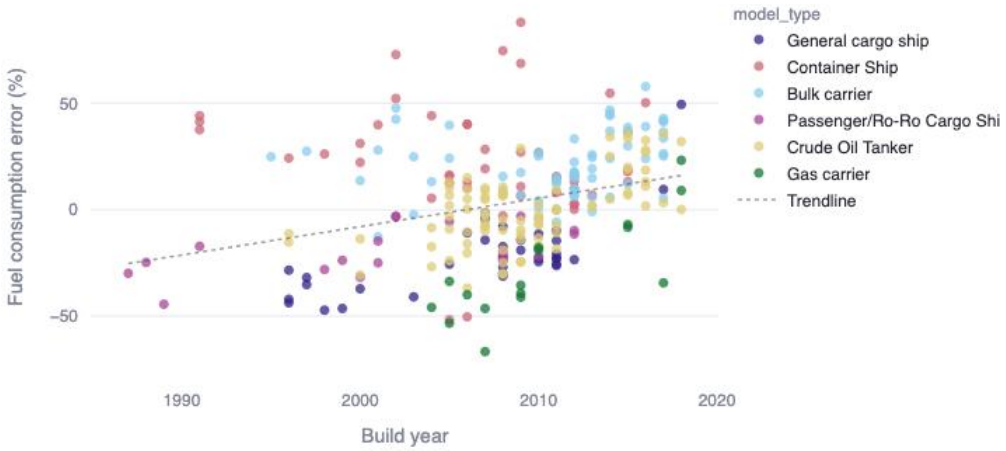


Fig.7: Correlation between year of build and data match. The Pearson correlation is about 0.3, indicating weak to moderately linear relationship.

5.3. Outliers

Apart from the discussed structural conclusions drawn from this study, it is worth to have a quick look at extreme outliers. These cases provide some interesting insights into the data quality, and where we might find which failure modes exist for such a large-scale study. While the provided plots are

extreme examples, various more subtle cases and outliers are found in the AIS dataset and reporting. Such extreme outliers could be filtered out by the data-pipeline before any computationally heavy simulation steps are taken. Meaning that these kinds of outliers do not skew the over-all results of the fuel consumption comparison between simulations and MRV reporting.

5.3.1. Misreporting

Below is the track of a single vessel for one year. According to the MRV reporting data. This vessel only sailed 400 nm of voyages that started or ended inside the EU. This kind of under-reporting was found in a hand full of cases. Due to the vessel selection criteria only looking at the total reported mileage, not on a per-year level, this specific case, as well as a handful of similar cases are still part of the reported results and explain the >100% discrepancy bin in Fig.4.

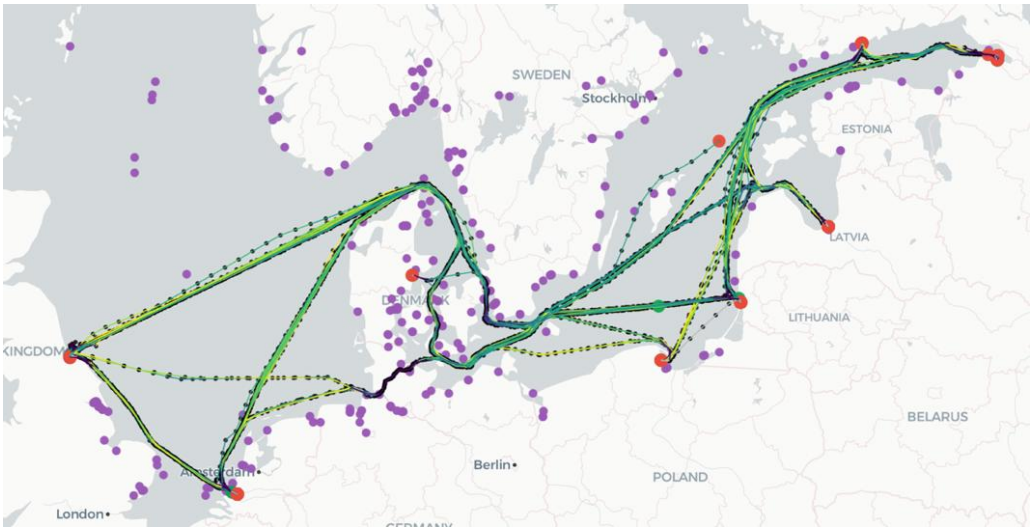


Fig.8: An underreported track: This is the 2018 track of a ship that reported only 400 nm sailed in that year.

5.3.2. Faulty AIS signals

It should come as no surprise that the AIS sailed distance of this vessel is far larger than the reported mileage. In fact, over all six years, the difference amounts to about 300%. Such extreme cases were filtered out of the reported results due to the large discrepancy in reported vs AIS track miles.

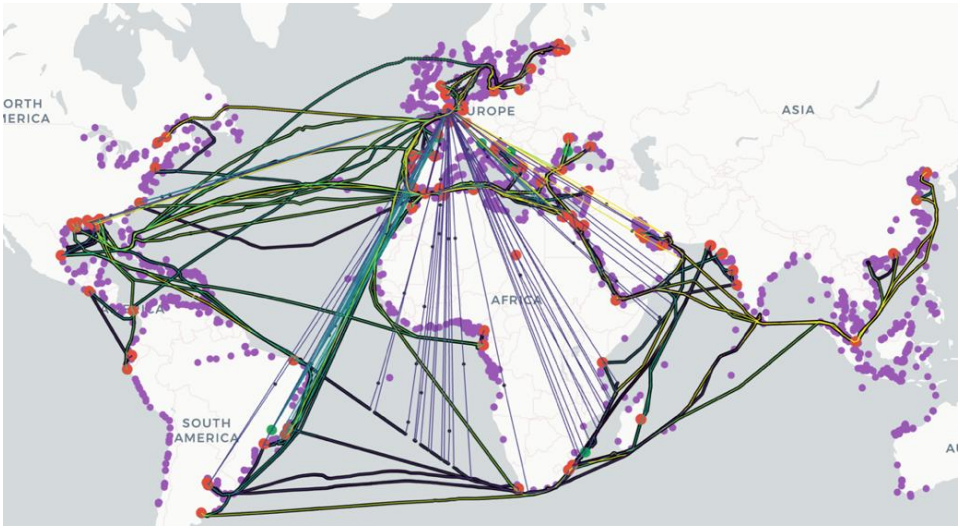


Fig.9: Faulty location signals lead to largely overestimated sailed distances based on AIS tracks

6. Satellite measurements

Satellite observations of nitrogen dioxide (NO2) have a proven history of detection of shipping lanes, and more recent advances in earth observation – in particular the ESA Sentinel-5p satellite and its TROPOMI instrument - have brought detection of emission plumes from individual (large seafaring) ships within reach, Fig.10, see also *Kurchuba et al. (2024)*. As shipping NO_2 emissions from Sentinel-5p are a reliable proxy for shipping CO_2 emissions this opens the envelope for monitoring ship performance and regulation compliance. Monitoring currently consists of incidental daily snap shots when a ship emission plume is sampled under the right viewing conditions. That allows for building a large database of ships and emissions, which can be resampled according to certain ship specifications for which – due to regulations - differences in emissions are expected, *Riess et al. (2025)*. Lack of such differences strongly indicates that regulations have not been effective.

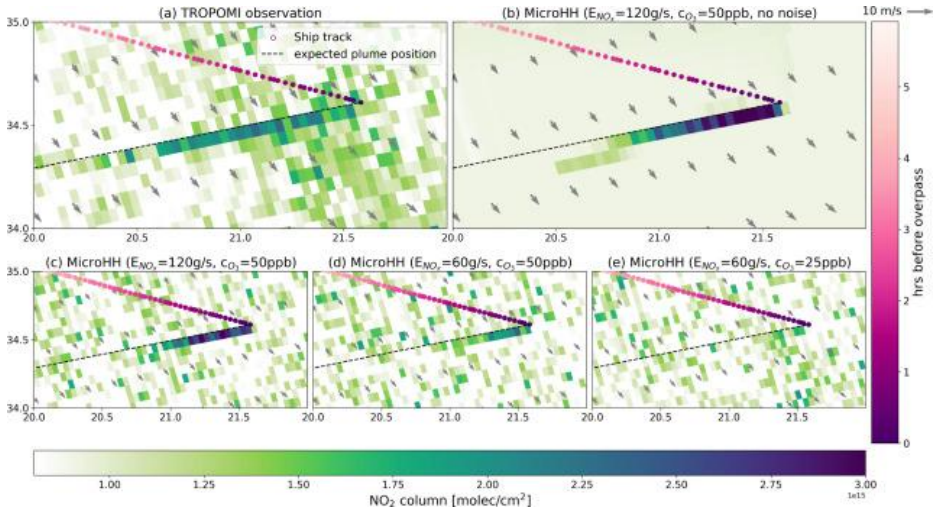


Fig.10: NO_2 plume of a ship sailing in the Mediterranean on 5 June 2019. (a) Observed TROPOMI data, (b) MicroHH simulation resampled at TROPOMI grid using an emission strength of 120 g s^{-1}, (c) like (b) but with realistic instrument noise, (d) like (c) but with half the emission flux and (e) like (d) but with half the background ozone. The pink dots indicate the AIS position of the ships in the three hours before overpass at 11:42 UTC as indicated by the right color bar, the black dashed line shows the wind-shifted track, which is the ship location displaced the 10 m wind and time difference between the TROPOMI measurement and ship location. The choice for halving the emission is motivated by the emission limit of Tier III ship, the choice for halving the background ozone is motivated by the difference in ozone levels between the eastern Mediterranean and the North Sea. The NO_2 emissions given here and reported elsewhere in this manuscript are given as mass of NO_2, meaning that an emission of 46 g s^{-1} equals 1 mol s^{-1}. Source: *Riess et al. (2025).*

7. Conclusions and recommendations

In conclusion, this study has evaluated more than 1600 ship-years of ship histories, a total of about 82 thousand voyages and 73 million nm of sailed distance. The first conclusion that can be drawn is that the quality of various data sources – AIS tracks, metocean hindcast, vessel particulars, etc. – is sufficient for such large scale and multi-year simulations. The data pipeline from raw data, automated cleaning and filtering to the final, processable vessel histories yields a dataset that is representative of a heterogeneous fleet across many ship types, operating conditions and years.

Secondly, studying physics-based ship modelling on a fleet level leads to very usable results. The average and median of computed fuel consumption lie very close to what is being reported, suggesting good high-level assumptions, good average match between assumed and actual vessel characteristics and good modelling methodology.

Finally, it can also be concluded that the fleet-wide spread is still significant. While averages and means match well, the 2-sigma level boundary does approach 50% mismatch between simulation and reporting. To reduce this spread, this study provides multiple correlations that can serve as a starting point to further improve fleet-wide modelling. Looking closer at ship-type specific factors such as hotel-load, hull shape and engine characteristics should enable a better match for individual ship types. Introducing a built-year factor to account for newer ships becoming more efficient should further reduce the fleet-wide spread. These are clear recommendations for follow-up research to significantly reduce the spread in fleet-wide studies.

Another question is how accurate the two discussed methods for estimating vessel efficiency are. The fact that the discussed methods match well hints at a decent accuracy, however both – simulations and reporting – come with their own set of challenges, and each can be susceptible to different kinds of errors and biases. In this respect, a very promising recommendation for further work is to look into measuring ship emissions using Satellite observations as discussed in section 6. This could provide a third point of calibration, enabling well-founded conclusions on over-all biases between different efficiency estimation methodologies.

References

ARDHUIN, F.; HERBERS, T.H.C.; JESSEN, P.F.; O'REILLY, W.C. (2003), *Swell transformation across the continental shelf. Part II: validation of a spectral energy balance equation*, J. Phys. Oceanogr. 33, pp.1940-1953

EGBERD, G.D.; EROFEEVA, S.Y. (2002), *Efficient Inverse Modeling of Barotropic Ocean Tides*, J. Atmospheric and Oceanic Technology 19, pp. 183-204

KURCHABA, S.; et al. (2024), Sensitivity analysis for the detection of NO_2 plumes from ships using TROPOMI, Remote Sensing of Environment

RIESS, T.C.V.W.; et al. (2025), Emissions of individual ships from TROPOMI NO_2 plumes, Remote Sensing of Environment

STOPA, J.E.; ARDHUIN, F.; BABANIN, A.; ZIEGER, S. (2015), *Comparison and validation of physical wave parameterizations in spectral wave models*, Ocean Model 103, pp.2-17

TOLMAN, H.L. (1991), *A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents*, J. Phys. Oceanogr. 21, pp.782-797

VAN DER KOLK, N.; TARDIF, T.; STELLA, F.; JACOBI, G. (2025), *Aero-hydrodynamic inter-actions for wind powered ships: retrofit of a high-speed mega RoRo vessel*, HIPER Symp., Tullamore

Data-Driven Emission Reporting Framework

Aleksandra Hein, Enamor, Gdynia/Poland, aleksandra.hein@enamor.pl
Dominik Leśniak, Enamor, Gdynia/Poland, dominik.lesniak@enamor.pl

Abstract

The article presents an innovative approach to automate and validate the manual entries submitted by ship crews for mandatory emission reporting, fulfilling IMO and EU regulatory requirements. It emphasizes the use of high-frequency data collection from various sensors, such as GPS receivers, fuel oil flowmeters, combined with additional data sources including AIS services and hindcast weather data, to enhance the accuracy and reliability of emission reports. This framework leverages multi-source spatiotemporal data fusion and big marine data analytics to simplify the reporting process while ensuring truthful and transparent monitoring of shipborne emissions. Through integration of real-time and historical data, the approach aims to minimize manual errors, streamline compliance workflows, and improve environmental impact assessments in maritime operations. This paper demonstrates how data-driven automation can support regulatory demands and promote sustainable shipping practices accommodate the capabilities of crew personnel by providing a more robust, verifiable emission reporting system.

1. Introduction

Despite postponed adoption of the Net-Zero Framework decided during MEPC ES.2 – Second extraordinary session of Marine Environment Protection Committee, the requirements of mandatory emission reporting still apply to the global fleet. Credibility of provided emission reports is thoroughly validated by the accredited bodies and frequently cross-validated with use of data collected from AIS transceivers and documents in form of cargo information samples, bunker delivery notes samples stating proofs of data compliance. The requirements of data emission reports data quality are constantly increasing often leading to many non-conformities between reported data and supporting documents of proof, ship's operation detection based on AIS data, operation naming conventions between organizations and stakeholders. The source of non-conformities is often related to human error, lack of data validation or lack of system supporting reporting operations. According to *Bhatia et al. (2024),* seafarer's exceeding 72 working hours per week limit mentioned by *ITF (2016)* was reported by 53.3% of respondents.

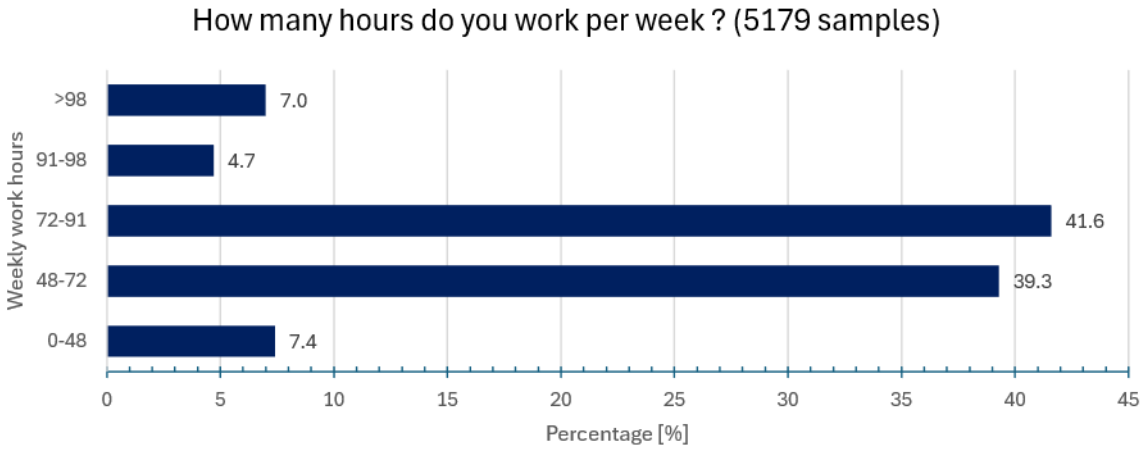


Fig.1: Seafarers Weekly Working Hours *Bhatia et all (2024)*

Increased workload and fatigue of the seafarers lead to more errors in daily tasks including ship operation reporting. Meanwhile ship's efficiency operational report errors may lead to false conclusions and inefficient operation or inadequate maintenance decisions, the errors of emission reports cause further consequences during the validation and emission compliance pipeline leading to further escalation and financial consequences related to emission trading system EU-ETS allowances settlements. Shipping

companies that fail to comply with the emission reporting requirements may face legal consequences including vessel's flag state detentions as stated by *EU (2015)*. The importance of emission reporting and consequently increasing crew fatigue elements require use of appropriate measures to tackle the emission data reporting challenge. Use of high frequency, automatic data collecting systems is one of the possible solutions. Its implementation helps to ensure emission data flow, data verification with automatic error detection and supports crew members to fulfill reporting requirements. Use of automatically recorded data ensuring less workload, more situation awareness and streamlines data quality and flow. Even relatively small amounts of high-frequency automatically collected data from GPS receivers can significantly improve the data quality and decrease risk of errors related to the ship's operation classification. Furthermore, the emission reporting can be improved with extended types of systems incorporating more data sources, that can help to validate reported fuel oil consumptions on different levels.

2. Emission Data Quality Challenges

From the crew perspective record of the emission related data is highly dependent from the available onboard equipment supporting operation monitoring functions and its operational status. The certified monitoring plans include indication of adopted emission monitoring methodology related to the technical feasibilities of onboard emission monitoring. Basing on the publicly available Thetis MRV system reports for year 2024 issued by *EMSA (2024),* most of the vessels adopted use of A-method for greenhouse gas emission determination in the emission monitoring plans.

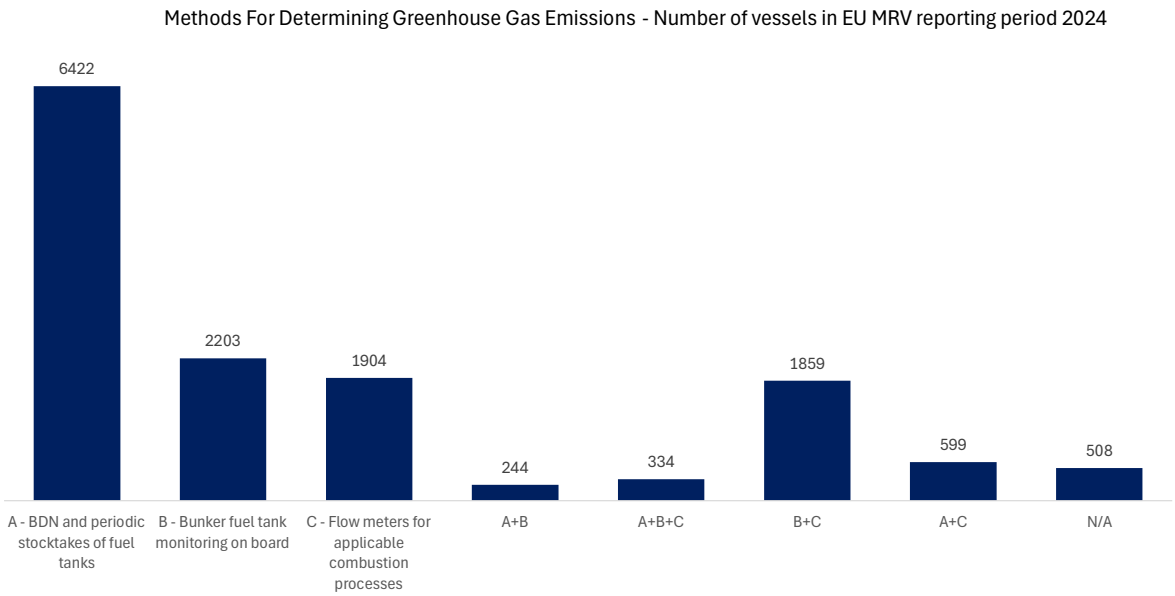


Fig.2: Number of vessels per GHG determination method for reporting year 2024, *EMSA (2024)*

Second most common method for GHG determination was the method B - onboard bunker fuel tank monitoring. Most common GHG determination methods solely relay on the human actions related to the fuel tank sounding, that further requires digitalisation of the sounding results. Most used combination of the methods, that increase but not fully guarantee the correct measurement is the combination of fuel tank monitoring and fuel oil flowmeter readings. Additionally for second, most common monitoring method, the daily tank readings are obliged when vessel is at sea according to *EU (2015)* which may lead to discrepancies and high reading uncertainties in some situations.

For example: when ship is facing heavy weather inducing severe rolling motion and the dip tapes are used for fuel oil tank sounding, the measurement will be indicating highest level recorded while measurement, which might be affected by the tank sloshing. In such situation error mitigation by statistics observation of measurement is not possible to be conducted as the free surface is not visible for the

observer during the measurement and only highest values of the tank levels will be indicated on the measuring tape. Furthermore, use of automatic data collection from tank level sensors, flowmeters or other sources still requires data quality assessment and periodic data cross-check and eventually monitoring device maintenance.

Each GHG reporting shipping company responsible for vessel's operation shall carry out a risk assessment to identify sources of errors in the data flow as stated by *EU (2015)*. Data quality control activities and quality assurance measures shall minimise the associated data quality error risks identified in the risk assessment. Moreover, the data review, corrective actions, record keeping & document management processes shall be established. Despite those activities and procedures, the data quality errors may result from various sources. They depend on the unique onboard monitoring capabilities linked to the ship's specific installations, conditions and sensor's reliability. Basing on sources of data the emission report quality problems might be divided into following sections describing root causes of the quality problems:

- Onboard sensors (sensors might be faulty and provide inaccurate data)
- Onboard installations operation (some installation might be used not as intended for example: flowmeters are bypassed during port operations if emergency FO supply line is used)
- Human error (data reported manually can include typing errors, identification errors)

Each of these sources can be further analysed and investigated. However, each potential reporting error risk independent of its source can be mitigated with help of expert system supported by the high frequency data collection and appropriate user interface enhancing user's capabilities of data quality monitoring and error detection response. Even despite latest technology advancements, user interaction cannot be fully eliminated from the loop due to need to maintain the data consistency and reports continuity even if high-quality redundant sensors are available. There's still potential risk of such monitoring system failure, therefore it is highly important to ensure proper redundancy, automatic data collection integrity against power supply losses, implementation of sensor failure detection or lack of data alarms, that will alert the user and impose awareness of current automatic data availability when reporting.

2.1. Sensor data quality monitoring & sensor failure detection

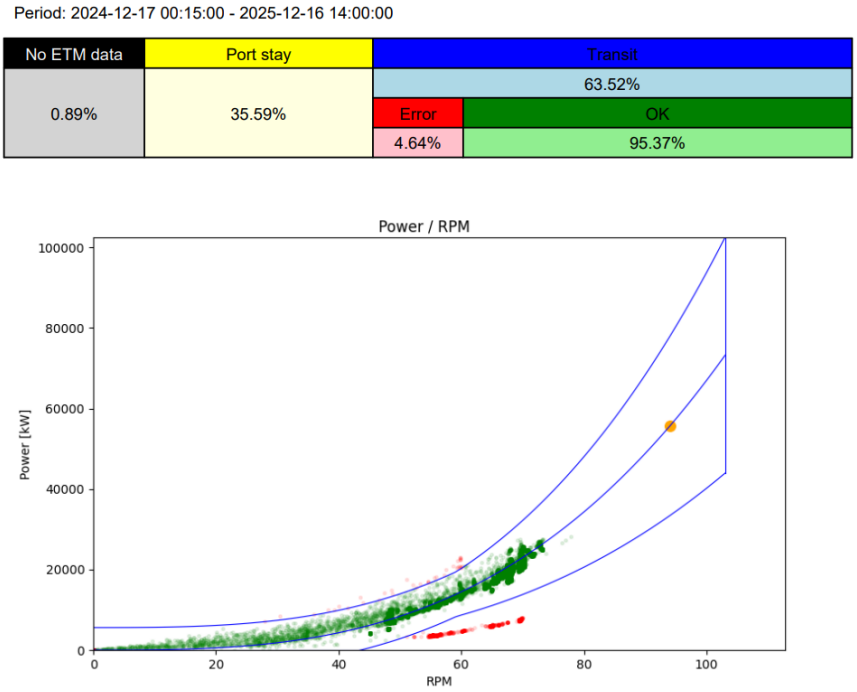


Fig.3: Torquemeter data quality check report example

Even in case of high-quality monitoring system implementation, the risk associated with potential sensor failure cannot be fully eliminated. When implementing automatic data collection, it's vitally important to also ensure automatic data quality verification procedures and algorithms to inform the users about potential data quality problems and impose necessity of manual data provision until the data quality alarm is not fully rectified. Various techniques can be implemented for data quality assessment and sensor failure detection. One of the most important is the capability of torquemeter data validation, which has a significant impact on possibilities of main engine consumption cross validations and other vessel-performance critical information. The torquemeter readings provide direct information about propulsion power in time, that combined with the specific fuel oil consumption characteristics of the main engine used for propulsion purposes can provide important cross validation of the reported fuel oil consumption either by comparing the expected consumption ranges with manually reported consumption or with the direct measurement of the fuel oil consumption based on the flowmeters reading.

2.2. Data Availability

To provide support for new reports and data quality enhancement as well as the data quality management and reporting error detection various data sources might be used. AIS trans receivers-based data sources frequently fail to record entire voyage period to further precisely analyse the covered distance or properly analyse and classify the vessel's operation. The distance covered deviations can be erroneously evaluated when vessel is operating outside of the vicinity of shore-based stations and in regions with high traffic congestion where data processing bandwidths are limited and thus limited amount of data are available in the AIS transponder data providers.

Fig.4: GPS receiver (left) and AIS Transponder (right) data with ship's position and speed over ground

2.3. Human Errors

When listing reports from group of anonymous 12 vessels with different crew members employed into reporting process, repetitive patterns of the reporting error sources can be distinguished. Nature of possible error sources are heavily dependent on use of reporting system data verification, guardrail mechanisms and automatic data check. To understand most common types of human errors only low-level questionaries for data without implemented data compliance verifications were considered. Following main reporting human error types can be distinguished:

- Time of reported operation synchronization (UTC time misunderstood with local time, not precise time assignment). When reporting different vessel operation, it's frequent that crew operating in local time regime will incorrectly identify proper UTC-based report time and thus incorrect identification of under way start and end period can occur leading to incorrect data reported in time regime. Also, it's worth to highlight that expected fuel oil consumption records and thus emissions will be heavily dependent on the type of operation that vessel has been engaged into at different time periods.
- Incorrect identification of reported fuel oil type in use. With limited data related to fuel oil identification, the incorrect fuel oil type can be reported during the consumption or bunker report.

- Omitting operation report (Anchor and Port stay without passage from anchorage location) Depending on the vessel's operation profile the lay-days defined in the charterparty or carriage will impose the vessel prior arrival at the anchorage even if berthing location is not yet available. After leaving the anchorage a short passage to the berthing location is usually made. This passage is often deemed as non-relevant and thus omitted in the reports.
- Type of operation classification (under way instead of adrift) Depending on the vessel's operation profile, it is common for some situations to remain in the adrift state for example when pilotage services waiting time is prolonged due to heavy traffic and anchorage operations are not intended. The adrift operation is often classified by the under-way vessel's status and omitted in the reporting leading to further data inconsistency.
- Misaligned remaining onboard and consumed fuel oil consumption values. When two independent values need to be provided manually, typing errors might occur, that will further cause reported data consistency.

2.4. Onboard Systems Exploitation

Even competent crews with fully operational sensors integrated with an automatic data collection system can face challenges. These can arise from specific onboard practices or unusual system configurations. Such situations can compromise the quality of emission reporting data. Errors often result from the complexity of fuel oil supply systems. They can also be related to procedures for preparing fuel across different operational scenarios. One example is the use of the emergency gravity fuel supply line. This occurs when diesel generators are running in port or anchorage during limited electric power supply demand periods. Emergency lines typically lack flowmeters. This design prevents flow restriction during emergency situations. Their primary function is to prioritise vessel safety over precise emission data. However, when used beyond emergencies, automatic fuel monitoring will fail to provide correct data for emission reporting. There are several methods that can detect such inaccurate records.

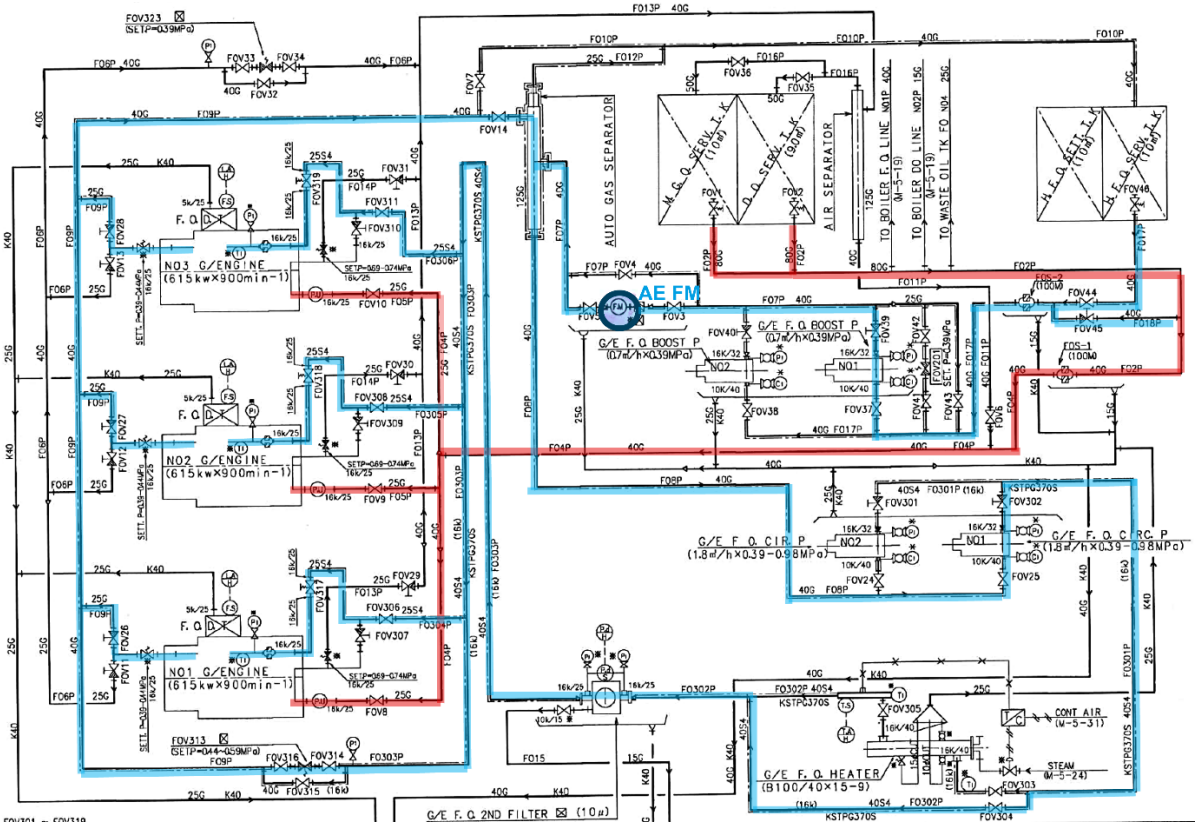


Fig.5: Standard (blue) and emergency (red) Generator Engines Fuel Oil Supply system

Firstly, consumption should be validated against SFOC characteristics. This method is effective when generator load is measured via power sensors and SFOC data is available. A second approach uses vessel operation detection. When the vessel is anchored, moored or adrift, the generators supply critical equipment throughout time subjected for report. Exceptions occur when moored vessels use shore power. In these cases, separate energy reporting is required from shore facilities, however such case has to be separately reported with amount of the energy received from shore-based facility. When implementing the automatic data collection systems for emission monitoring or using the data from onboard sensors and indicators it is necessary to understand the potential risks and limitations of the crucial parameters recording. The data collection systems shall facilitate the methods for automatic detection of such situations and alert the users that recorded information is not according to the expected consumption rates that shall be observed in present vessel operation status and machinery exploitation.

3. Methods of Emission Data Error Mitigation

Previously identified data quality error risk can be minimised with various techniques of data availability and data operations implemented through reporting process.

3.1. Ship Operation Classification

Incorrect ship operation time or incorrect ship operation type is one of the common human errors reported in the emission reporting systems. Such error can be often identified in situations where prior to port call vessel remained in the anchorage in the port of call vicinity. Short under way status report is often omitted during the manual reporting phase leading to data inconsistency especially for fuel oil consumption related to the propulsion system operation. To mitigate this error risk, the automatic ship operation classification techniques can be utilized. Observations based on high-frequency GPS position data can provide information about the ship's motion characteristics. Different motion patterns can be detected during the anchorage periods, where long-term motion paths are influenced by the length of the anchor chain, anchor location and environmental parameters like current speed & direction, wind speed and wind direction. Similar to the anchorage periods, vessel's secured at berths with use of mooring lines will have minimal motions recorded mostly related to the GPS receiver static position determination accuracy.

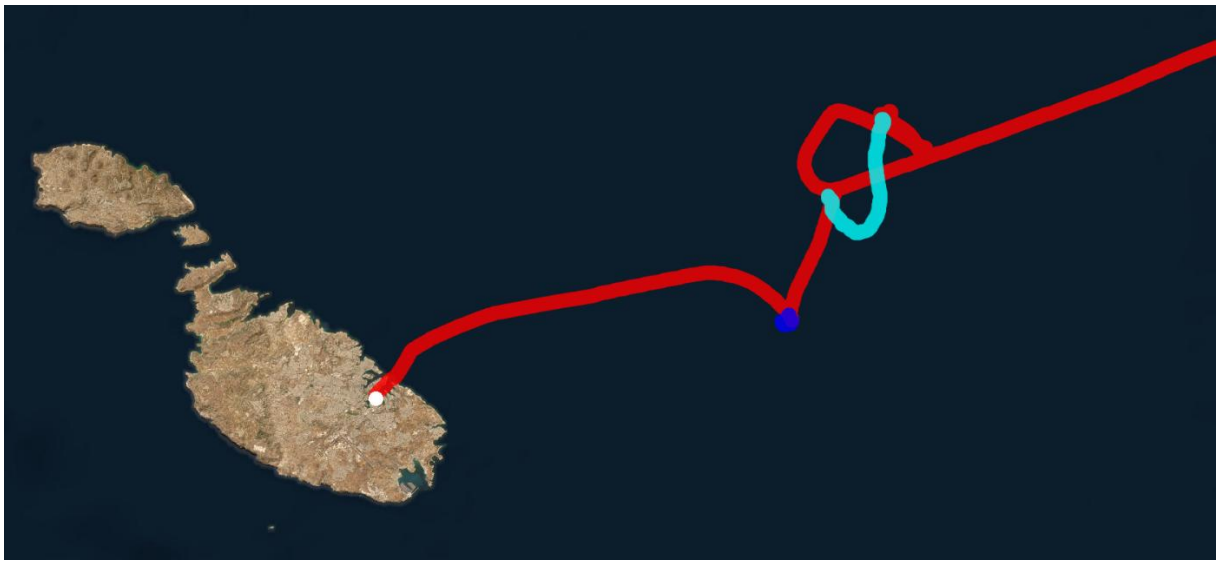


Fig.6: Example of vessel operation classification (path: red-under way, white - In port, dark blue - anchor, light blue - adrift)

Vessel's in adrift situation will be characterized by the motion paths directed along the direction of the external environment parameters, like the anchorage periods but with different characteristics. The motion paths of the vessels in adrift conditions will not be congested in one location in long-term in contrast

to the anchoring periods. Various features of the motion paths can be used to classify the ship's operation condition, one of the simple, yet very effective is the position spread parameter, which can be described as distance between first and last position of the vessel in given time span window assessed in the moving window regime. Such approach will ensure that the anchoring positions spreads can be distinguished from adrift, under way or in port operations.

Only basic ship's operation classification can be supported, further detailed situation status description requires manual interference from the user and identification for example of the bunkering operation which may happen during anchorage or at berth. The same applies to repair or docking operations, distress or involvement in the search and rescue operations. Further adoption of the vessel's major operation classification automatically informs the crew about ship's operational status in given period synchronized with the satellite navigational system high frequency data. The need for report preparation and report time scope is also highlighted for the reporting crew members ensuring, that there will be no gaps or omissions in statements of the vessel's operational status, even if they were maintained for relatively short time periods.

3.2. Data Integration Levels and Data Quality Cross Validation

To support and enhance the reporting accuracy meanwhile addressing different needs and possibilities of the sensor integrations and investments into the complex data collection systems, the subsequent data integration levels are proposed and considered by the authors. Higher data integration levels will ensure lower risk of the emission reporting errors; however, they will require more sensors to be integrated with the data collection systems. Nevertheless, some information vital for determination of vessel operation state is difficult to quantify with sensors either due to technical or economic reasons.

3.2.1. Level 1 Minimal required data integration

Systems equipped only with basic GPS receiver can still support the reporting data verification in significant way leading to further reduction of the error risk probability. Despite missing information about the measured fuel oil consumption the vessel's particulars including Main Engine specific fuel oil consumption and speed to power model can further increase the accuracy based on early error detection in reported consumption values based on the report time span, vessel's operation type and speed over ground profile giving first base estimation of the expected main engine fuel oil consumption range.

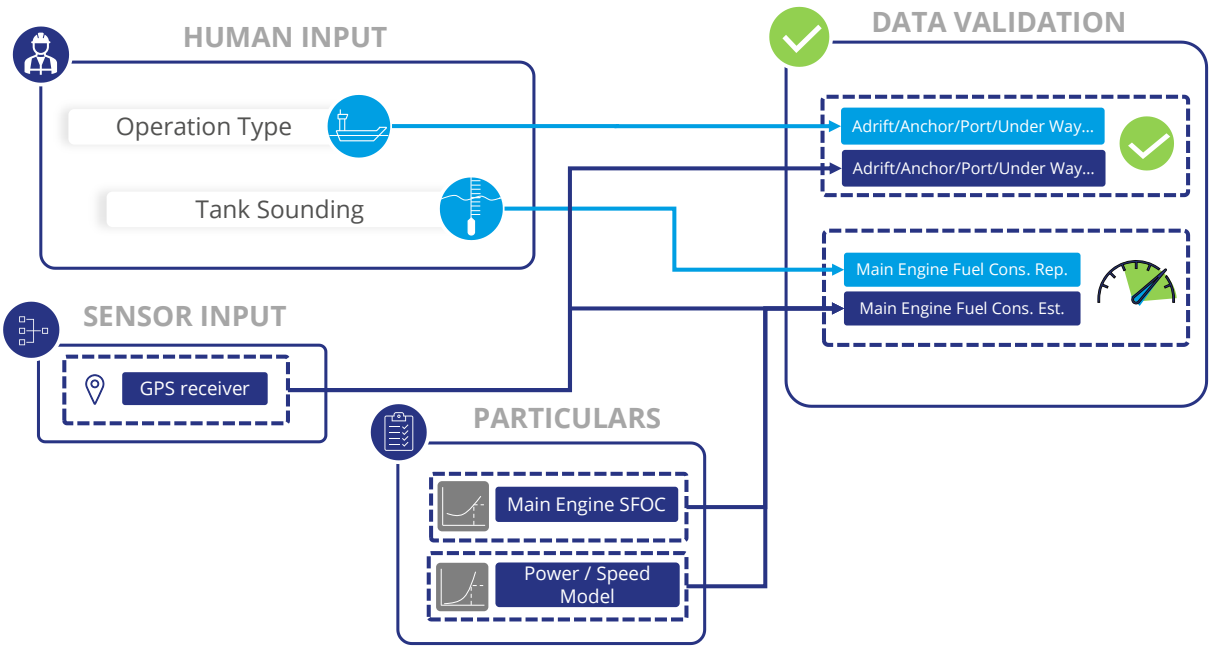


Fig.7: Level 1 enhanced data integration for validation purposes

3.2.2. Level 2 Enhanced by propulsion power monitoring

With integration of the shaft power meter and diesel generators power monitoring further data validation enhancement can be achieved. The shaft line load monitoring and generator power monitoring can be realized with lower investments into the system's sensor integration than in case of complex fuel oil supply system and implementation of the fuel oil flowmeters. This integration gives further extension for the main engine expected fuel oil consumption range in more accurate way and also including the expected diesel generator set consumption range, which is not available in the GPS-only based integration with ship's particulars information.

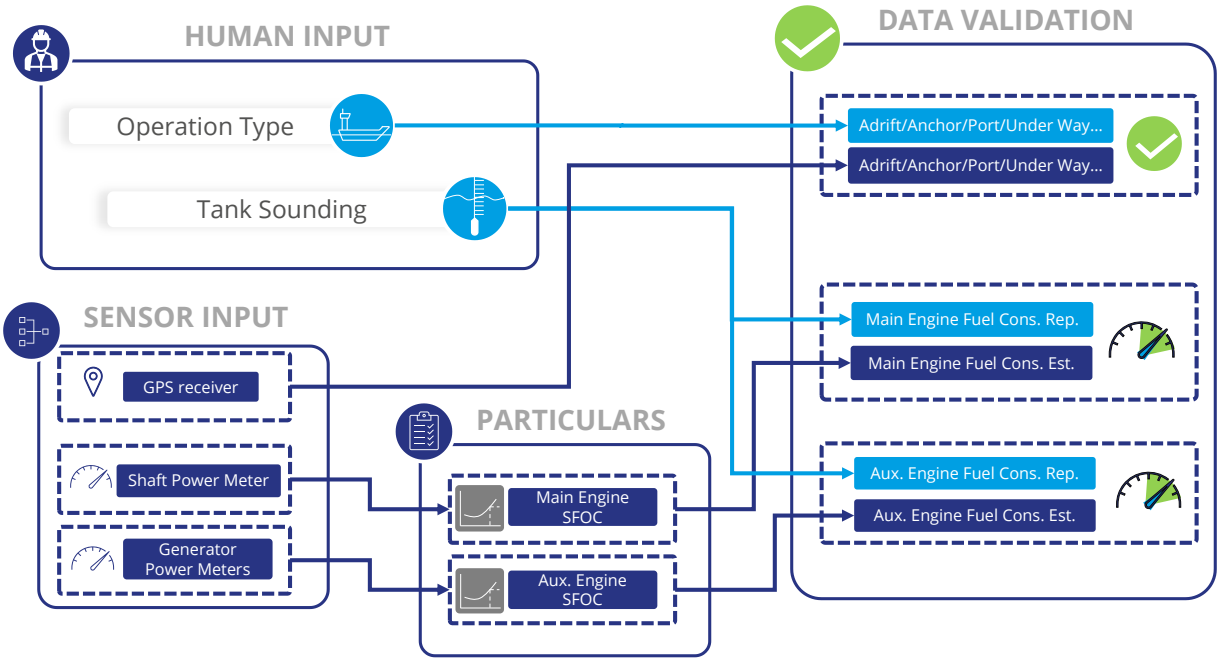


Fig.8: Level 2 enhanced data integration for validation purposes

3.2.3. Level 3 Enhanced by direct Fuel Oil Consumption Monitoring

The level 3 integration incorporates direct measurement of the fuel oil consumption with use of the flowmeters. Due to different arrangements of the fuel oil supply pipelines, different monitoring capabilities and accuracy can be achieved on different vessel installation types. Due to relatively high uncertainty and impact of single fuel oil flowmeter failure or accuracy drift on the entire fuel oil consumption monitoring capabilities it is highly valuable to supplement the data validation with cross check based on the fuel oil consumption estimation from the level 2 presented before. This cross validation allows to indicate correct operation of the entire fuel oil monitoring system consisting of many components which have no redundant counterparts incorporated into the system.

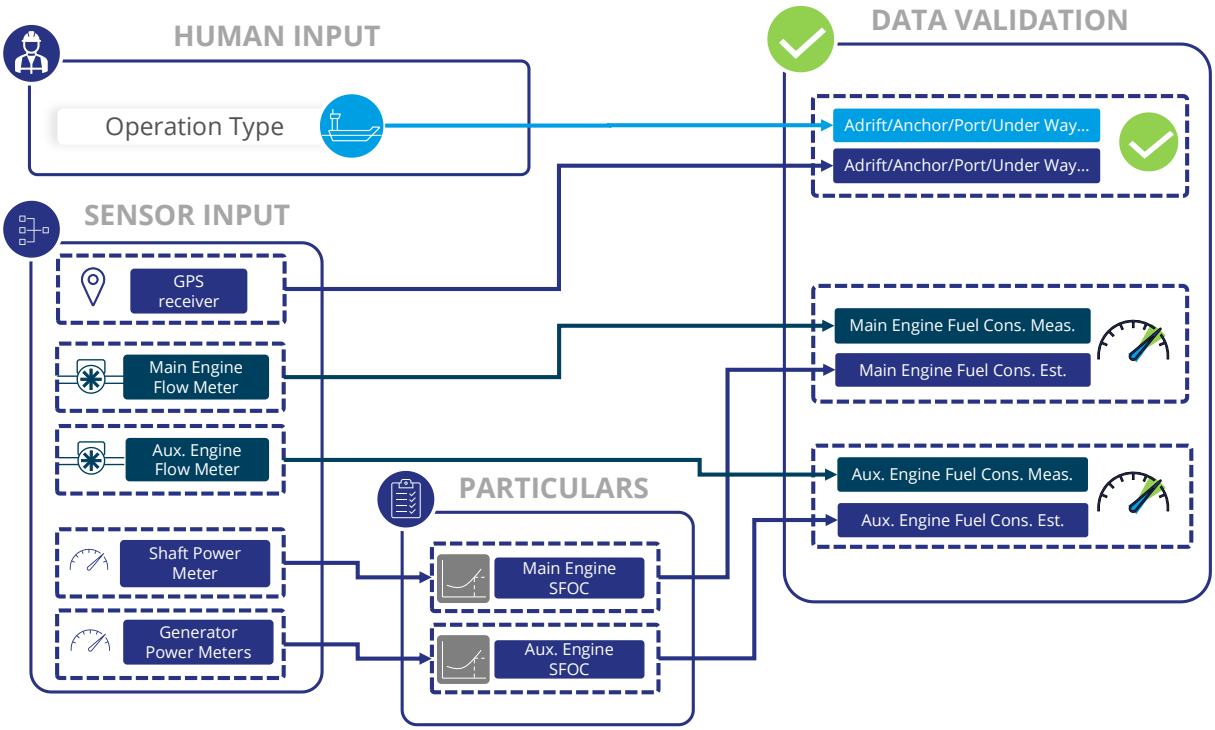


Fig. 9 Level 3 enhanced data integration for validation purposes

3.3. Supportive methods for cargo reporting on different vessel types

Transport work assessment in the MRV reporting system is based on information about the Cargo transported by the vessel. However, this cargo definition depends on the vessel type and reporting requirements often requires detailed understanding of the formula. To simplify this requirement, appropriate data flow preparation can be arranged in order to assess the cargo according to its definition with maintaining the requirement of minimal information input from the crew's perspective. Example of such approach can be presented in relation to the general cargo vessels where cargo shall be reported as deadweight carried according to *EU MRV (2017)*. The deadweight carried can be expressed as:

$$DWT\ carried\ = V \cdot \rho\ -\ ship's\ lightweight\ -\ fuel\ weight\ [MT]$$

ρ – water density t/m^3
V – volume displacement m^3

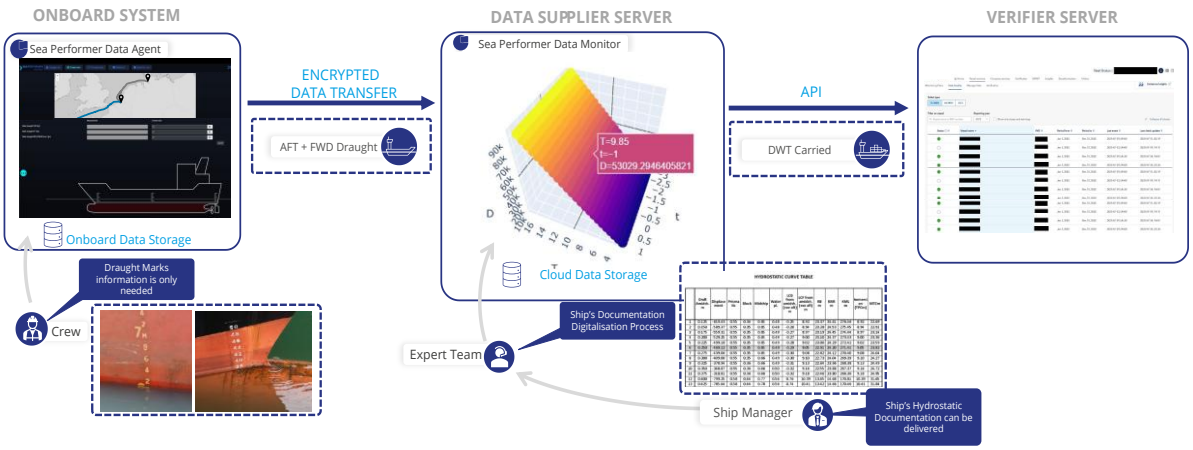


Fig.10: Cargo reporting data flow example for deadweight carried cargo reporting method

To prepare the data crew members would have to be obliged to enter the data from loading computer for each report where cargo change has been reported. To simplify this process, the vessel's hydrostatic data can be imported into the ship's characteristics database stored in the data supplier server, hindcast weather data enriched with water density information. The final value of deadweight carried can be calculated based on minimal input from the crew related only to the current ship's draught parameters. Draft readings can be also automated if draught sensors are integrated with automatic data collection system.

3.4. Fuel Oil type Identifications for Consumption & Bunkering operations Reporting

As stated by *Barsotti (2025),* with new rules imposed by Fuel EU regulation, each fuel bunker delivery shall be classified and identified due to the encountered emissions associated with the fuel oil production. Moreover, when considering biofuels, it shall be noticed, that even two biofuels of the same fuel category can have different sustainability factors or different composition. To evaluate each possible operational case, it is worth to note that some of the fuel blends can be created by addition of the blend component to existing onboard fuel composing new, unique fuel type that shall be identified as new fuel bunker available onboard since composition resulting in unique mix creation.

The complexity of biofuel reporting varies from vessel to vessel. It depends on several factors. These include the vessel's operational profile, its characteristics, and the company's biofuel policy. Users of the emission reporting systems need accessible data interfaces to report the use of biofuels correctly. These should facilitate precise and unambiguous fuel type and fuel parcel identification. They must also accommodate different monitoring methods, each according to adopted fuel reporting needs and monitoring method indicated in emission monitoring plan.

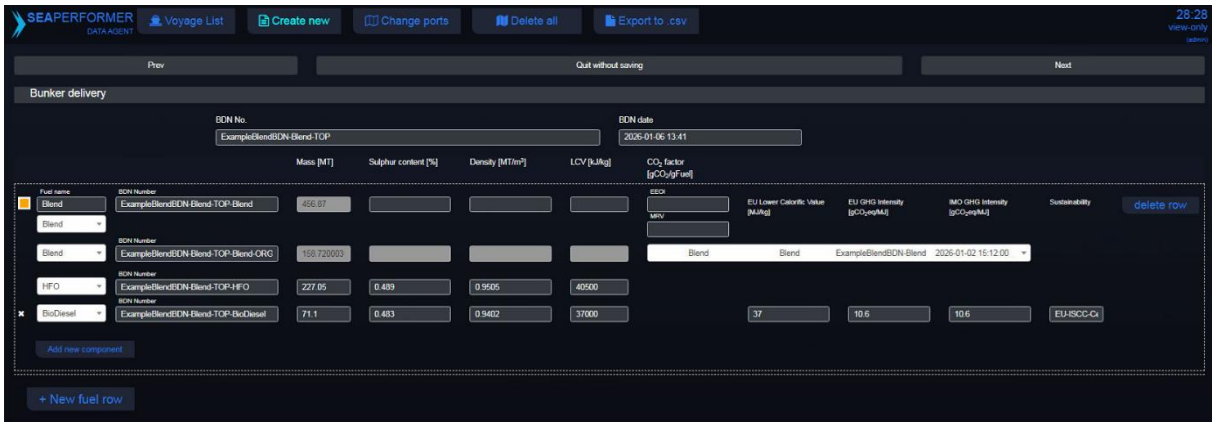


Fig.11: Bunker delivery reporting interface with Blend fuel components indication

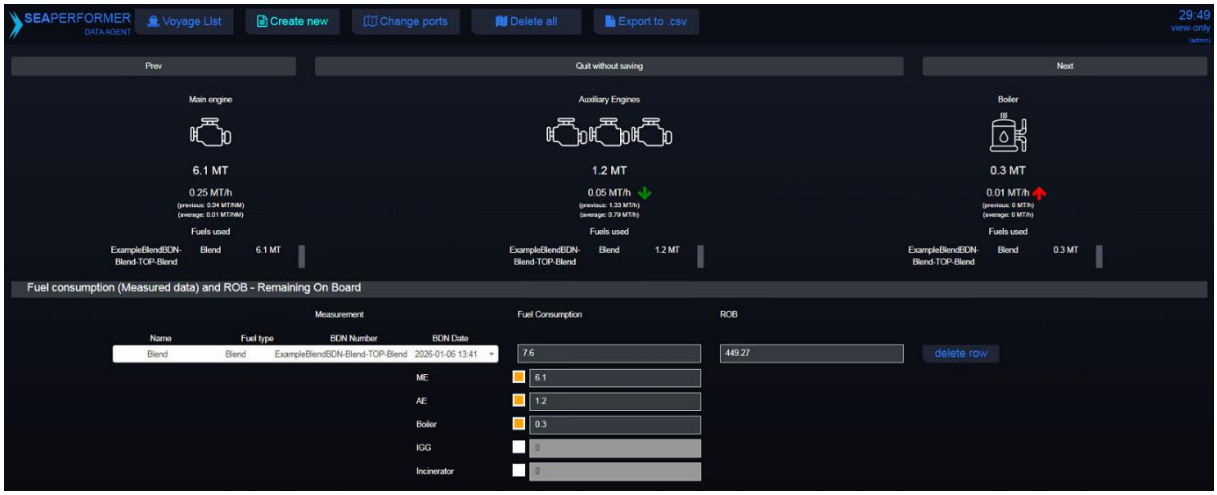


Fig. 12 Fuel Oil consumption reporting with blend fuel oil type bunker identification

Some crews have advanced fuel consumption monitoring systems. These can provide consumption rates over specific time periods. Others rely on periodic tank sounding methods or tank level indicators readings. For them, reporting based on Remaining On Board (ROB) is most straightforward method. A simple user interface helps to ensure consistency. Regardless of the reporting method, it preserves aligned ROB and consumption values allowing the crew members to report in selected, preferred method. It also retains fuel parcel information when available.

4. Emission Data Pipeline

In order to fully support the data availability and communication between the stakeholders, automatic data transfer is maintained in the entire dataflow of the emission reporting framework. Even though automatic data quality check and high frequency sensors information is utilised, some cases still need broader explanation in communication between shipowner and data verifier. This process can be simplified thanks to the data platform, where various levels of information related to the ship operation parameters and ship operation environment are combined and visualized. Selected data verification cases can be further analysed with help of the experts supporting shipowner management tools with advisory on how to utilise the available data to present relevant explanations to the data verifier thought the reporting process.

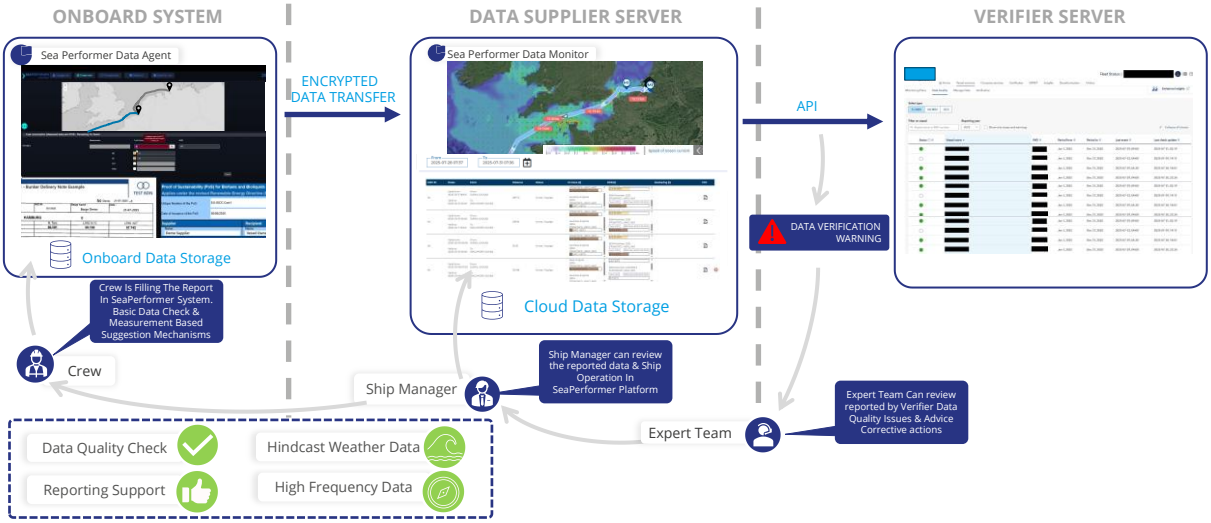


Fig.13: Emission Report Data Flow Pipeline

5. Supported Verification Cases

Even with assurance of the correct data monitoring the questions during the verification process can arise related to the vessel operation.

Fig.14: Case Identified as port stay by verifier relaying on the AIS-based data (vessel anchoring on river anchorage in port vicinity)

Basing on the experience of the verifier's interaction, states of the vessel operation might require detailed explanation with full operation case scenario including information from the hindcast weather data used for example to justify the speed reduction and to justify the correct under way operation identification. Moreover, other cases where verification process relied on the AIS-based data of limited quality and availability, some anchorage operations in port's vicinity were identified by verifier as the port stay. Identification misinterpretation can be however easily clarified taking advantage of the high frequency GPS-based data to justify the correct operation classification.

6. Conclusions and future development paths

With ambitious goal of achievement of carbon-neutrality for maritime shipping by the 2050 the emission reporting accuracy becomes very important task and every reporting year more resources are submitted towards increasing reporting accuracy and data validation. Assurance of the emission reporting data inconsistencies and errors can be achieved with implementation of various supportive data analysis practices combined with the data collection pipeline designed with priority towards user experience support in the data validation task, data quality assessment, minimal required effort and experience of the crew. Various techniques for data processing presented in this paper are expected to be further enhanced thanks to the possibility of integrating more accurate, more reliable sensors allowing for detailed representation of current vessel operation profile and details of information crucial for emission-reporting tasks. It's expected that number of crew members will be further decreased in the future due to progressing automatization of various onboard maintenance and watchkeeping tasks. The data collection, verification and emission reporting systems must also follow this trendline ensuring that even newcoming crewmembers will be able to provide minimal required emission reporting data input, monitor the system health status and react to the sensor failure events. With properly aligned emission data reporting frameworks, the requirements towards mandatory emission reporting and monitoring can be fulfilled in efficient manner realizing subsequent steps towards carbon-neutral shipping industry.

References

BARSOTTI, M. (2025), *Bunker Parcel Data Collection: Enhancing Emissions Reporting*, 10th HullPIC Conf., Mulheim, p.15-18, http://data.hullpic.info/HullPIC2025_Mulheim.pdf

BHATIA, B.S.; BAUMLER, R.; CARRERA-ARCE, M.; MANUEL, M.E.; BARTUSEVICIENE, I. (2024), *Quantifying an inconvenient truth: revisiting a culture of adjustment on work/rest hours*, World Maritime University, Malmo

EMSA (2024), *2024-v180-26022026-EU MRV Publication of information*, European Maritime Safety Agency, Lisbon, https://mrv.emsa.europa.eu/#public/emission-report

EU (2015), *Regulation (EU) 2015/757 on the monitoring, reporting and verification of carbon dioxide emissions from maritime transport, and amending Directive 2009/16/EC*, European Parliament

EU MRV (2017), *Guidance Document The Shipping MRV Regulation – Determination of cargo carried*, European Sustainable Shipping Forum Sub-group on Shipping MRV Monitoring

IMO (2019), *Guidelines on Fatigue, MSC.1/Circ.1598*, Int. Mar. Org., London

ITF (2016), *An ITF Guide for Seafarers to the ILO Maritime Labour Convention*, International Transport Workers' Federation

Comparing AI and Ship Hydrodynamics for Ship Performance Estimation

Pauline Røstum Bellingmo, SINTEF Ocean, Trondheim/Norway, pauline.bellingmo@sintef.no
Thor Albrektsen, SINTEF Ocean, Oslo/Norway, thor.albrektsen@sintef.no
David Odd Erik Fürst Standal, NTNU, Trondheim/Norway, david.standal@ntnu.no
Severin Sadjina, SINTEF Nordvest, Ålesund/Norway, severin.sadjina@sintef.no
Endre Sandvik, SINTEF Ocean, Trondheim/Norway, endre.sandvik@sintef.no

Abstract

This paper evaluates hydrodynamic, data-driven, and hybrid approaches for modeling ship performance across varying operating conditions. Using full-scale measurements from a 200 m deep-sea bulk carrier, shaft power is estimated with an empirical resistance-based method, machine learning models, and physics-informed machine learning that integrate calm water resistance from the Hollenbach method. All predictions methods are assessed under different speeds, power levels, and weather conditions. Results show that hybrid models outperform purely empirical and data-driven approaches. In particular, the physics-informed neural network achieves the highest accuracy and robustness, performance across regimes and demonstrating resilience in sea states.

1. Introduction

The IMO aims to reduce emissions by 30% from the global fleet by 2030 relative to 2008, a daunting goal for the global shipping industry. A key factor for achieving this goal is fast and accurate ship performance predictions, crucial for developing effective operational decision support tools. Accurate performance estimation in varying operation conditions is essential for both operational energy efficiency measures, *GCMD (2024)*, e.g. voyage optimization, and optimal design. Suboptimal estimates will yield a high uncertainty related to the potential benefits, especially if the level of uncertainty of the estimate approaches the potential for energy savings. Studies have shown that the potential emission reductions from voyage optimization on its own are in the order 10%, *Sung et al. (2022), Wang et al. (2021), Moradi et al. (2022),* up to 30% according to ship operators, *NAPA (2020)*. If addressed properly, this may bring the industry closer to the 2030 goals.

In this study, ship performance is evaluated using the shaft power, which is essential parameter in voyage optimization. There exist different approaches to predict the shaft power, including hydrodynamic, data-driven, and hybrid methods.

Hydrodynamics

The estimation of ship performance has a long history, beginning with towing tanks and model tests. Modern hydrodynamics is generally divided into experimental and numerical approaches. Experimental methods rely on scale model tests in towing tanks, ocean basins, or cavitation tanks, and typically use non-dimensional coefficients to represent different components of total resistance. Most methods split the total resistance into calm water resistance (≈80%) and added resistance, with the latter primarily accounting for environmental effects such as waves and wind. Numerical methods, e.g., potential theory and CFD, require detailed ship hull geometry, while the experimental methods require a build ship model.

Empirical and semi-empirical methods have been developed to predict these resistance components without full-scale experiments, instead relying on results from previous experiments. Examples include formulations for residual resistance, *Hollenbach (1998),* and the ITTC frictional resistance method. Modern semi-empirical approaches such as SNNM and STAWAVE combine regression on experimental data with heuristic or blending techniques to estimate added resistance in waves. The SNNM method established a semi-empirical formula for added resistance in waves of arbitrary headings by combining a regression analysis of experimental data with a heuristic approach for following waves. Specifically, the resistance from wave diffraction is calculated with an updated,

approach, *Skjørdal and Faltinsen (1980),* while radiation effects are determined by regression on the experimental data, *Liu and Papanikolaou (2020).* The STAWAVE combined method integrated the CTH and L&P approaches using a hyperbolic tangent function to smoothly blend motion-induced and reflection-induced resistance, *Liu and Papanikolaou (2020).* These methods provide efficient and reasonably accurate predictions, particularly for large cargo ships.

Data-driven methods

By leveraging operational data, AI has emerged as a promising approach for predicting shaft power. Multiple studies have shown that purely data driven models, like Artificial Neural Networks (NNs), can be used to predict shaft power or related variables like fuel usage, *Ghani et al. (2025)*. However, relying solely on data without incorporating domain knowledge can limit model reliability. Recent studies have shown that purely data-driven models often struggle to generalize and can produce spurious correlations that defy physical laws, *Fan et al. (2022), Nguyen et al. (2023).* This can be made even worse if the data is noisy or inconsistent, which real world data often is.

In addition, data-driven models can be influenced by dynamic operational effects that temporarily decouple shaft power from steady-state performance. Sudden changes in weather or heading may introduce transient behavior, where shaft power does not correspond directly to vessel speed due to acceleration or deceleration. This effect is particularly relevant for large cargo ships, that can spend hours accelerating or decelerating *Cai et al. (2024), Coraddu et al. (2019).*

Hybrid methods

To address the limitations, hybrid models that combine AI techniques with physics have emerged as a promising approach. By integrating the strengths of both methodologies, hybrid models aim to improve accuracy while maintaining physical consistency. Such approaches have shown promising results when applied in tasks related to ship performance, including energy usage prediction and evaluation of vessel technical performance index, *Guo et al. (2023), Jørgensen et al. (2022).*

What constitutes a hybrid method is not clearly defined, and multiple different approaches have been tried with success. One approach is to use a physical model to generate training data for a machine learning model, then using transfer learning to refine the machine learning model on real world data, *Mavroudis and Tinga (2025).* Another approach is to incorporate physical knowledge directly into the model, either through modifying the loss function or the structure of the model itself, *Altan et al. (2025), Bourchas and Papalambrou (2025).*

Despite these developments, accurately predicting shaft power across varying operational conditions remains challenging. Environmental effects such as waves, wind, and current, along with hull fouling and changing loading conditions, introduce complex dynamics that are difficult to capture using either purely hydrodynamic or purely data-driven approaches alone. This motivates further investigation into hybrid methods that can leverage physical insight while adapting to real operational data.

1.1. Contribution

This paper builds upon previous hybrid modeling methods by enhancing machine learning models with physics-informed input features, more specifically calm water resistance. The hybrid method is compared with a hydrodynamic model and purely data-driven models for estimating shaft power of a ~200 m deep-sea bulk carrier during transit. Furthermore, an in-depth analysis of model performance across different operating conditions, including speed, power, and weather, is conducted.

1.2. Assumptions

To make the task tractable, a set of assumptions are made. Firstly, the model is developed for transit operation, defined as where the ship is sailing in steady-state behavior, excluding maneuvering operations. Since the model is developed for transit, we assume deep water conditions for the modelling. Additionally, to make a model that can be used with limited ship information, we assume

that only main ship and propeller dimensions are available, in addition to operational data logged from the ship and weather data.

2. Method

This section presents the data and its processing, followed by the implementation details of the hydrodynamic, data-driven, and hybrid methods.

2.1. Data description

This study investigates the ship performance of a deep-sea bulk ship. The data consists of three data sources: 1) onboard ship data, 2) maintenance logs, and 3) weather data. Onboard measurements from the ship were logged from April 2020 to November 2025. The ship data is sampled at 15 minutes frequency and contains navigation data, loading conditions (draft and trim), and propulsion data. The onboard ship data is augmented with hindcast weather data provided by the Copernicus Marine Service, collected using their database and the GPS location of the ship. Table I gives an overview of the data variables in the complete dataset. Not all available data from onboard ship measurements are listed here, only the essential data for this study.

Table I: Overview of main variables in dataset

Onboard ship data	**Maintenance logs**	**Weather data**
Speed through water (STW)	Dry docking event	Wave height
Draft	Propeller polish event	Wave peak period
Trim		Wave direction
Shaft power		Wind speed eastward
		Wind speed northward
		Current speed eastward
		Current speed northward

2.1.1. Data processing

Ship data logging varies in quality and accuracy, not only between vessels but also across time and different sensors on the same ship. Furthermore, some measured or externally obtained data may not be optimally formatted or adequately filtered for use in training models. To ensure high-quality input data, the following processing steps were applied:

- Data validation
  - Single variable validations: Each variable was examined independently to identify anomalies or outliers.
  - Multivariable validations: Relationships between relevant variables were checked for consistency, for example speed through water against longitudinal sea water velocity, drift angles against lateral sea water velocities, or the dependence of speed-power curves on environmental loads.
- Identification of transit phases
  - Two sliding window filters are applied in sequence.
  - First, a simple filter is applied to identify static transit conditions based on minimum mean and maximum mean absolute deviation of STW.
  - Secondly, a steady state filter proposed by *Dalheim and Steen (2020)* was applied on shaft power, propeller RPM and STW to ensure that only pure transit data were selected for model training.
- Feature derivation
  - Relative wind speeds: Longitudinal and transverse wind components relative to the ship were derived from the northward and eastward wind hindcast.
  - Relative wave direction: Computed from true wave direction and ship heading.

- Trim and draft: Continuous sensor logs are strongly affected by noise from weather and hull-dynamics. To ensure stability, static fore and aft draft values were first established for each transit using departure noon reports. From these fixed values, the mean midship draft and trim were derived. While this choice neglects changes in draft due to fuel burn and varying ambient conditions, it stops variations in trim and draft from introducing noise into the training data and weakening correlations with shaft power.

2.2. Hydrodynamic modelling

In this study, shaft power is estimated using hydrodynamic modeling based on empirical methods, due to the limited availability of ship-specific data (only main ship and propeller dimensions). Hydrodynamic methods compute ship resistance, which can be estimated via empirical, numerical, or experimental approaches. Numerical methods require detailed hull geometry, while experimental methods require physical ship models, which were not available. For the empirical methods, calm water resistance is estimated using the *Hollenbach (1998)* formulation suitable for large cargo ships. Added resistance in waves is calculated using the SNNM method, *Liu and Papanikolaou (2020)*, which covers all wave directions and is recommended in ISO 15016:2025. Added resistance due to wind is estimated following *ITTC (2021)* guidelines.

To relate resistance to the required shaft power, effective power (P_E) is first defined as the power required to move the hull at a given speed (V) without propeller action: $P_E = R_T \times V$. To determine the shaft power (P_S) or delivered power (P_D), $P_E$is divided by the propulsive efficiency (η_D): $P_D = \frac{P_E}{\eta_D}$. The shaft power can be derived from the delivered power through the transmission efficiency (η_S): $P_S = \frac{P_D}{\eta_S}$. For this study, a constant efficiency provided by the ship operator is used for simplicity to find the shaft power from the resistance.

2.3. ML modelling

Two different ML approaches were implemented in this study; an NN and XGBoost.

2.3.1. NN

Neural networks (NNs) are well established machine learning models in modern applications. Attempting to recreate real neural networks they consist of interconnected neurons that are trained to predict one or multiple target values. In this paper, a feed forward neural network is used. This architecture is a NN where the first layer of neurons is the input, which then flows to the output through a series of layers, outputting a prediction at the final layer. It is trained in a series of epochs, where each time the weights on the connections are updated with the error in prediction. The NN is implemented and trained through the Pytorch package in Python.

2.3.2. XGBoost

Extreme Gradient Boosting (XGBoost) is an advanced form of decision trees, *Chen et al. (2026).* It builds upon decision trees to make predictions, where each tree partitions a feature space by making a series of binary splits that minimize a specified loss function. Trees are created sequentially using gradient information from the feature space, with a focus on creating relatively shallow trees combined with regularization to avoid overfitting. This approach allows XGBoost to efficiently produce highly accurate models while maintaining computational efficiency.

2.4. Hybrid modelling

In this paper, hybrid modelling refers to the combination of hydrodynamic and machine learning

methods. As illustrated in Fig.1, a sequential integration approach is adopted to establish physics-informed NN and XGBoost models. The Hollenbach empirical method is employed to estimate calm water resistance, which is then provided as an additional input feature to the models alongside the standard operational and environmental variables. Incorporating this physical baseline helps guide the learning process, improving convergence and overall predictive performance compared to purely data-driven models.

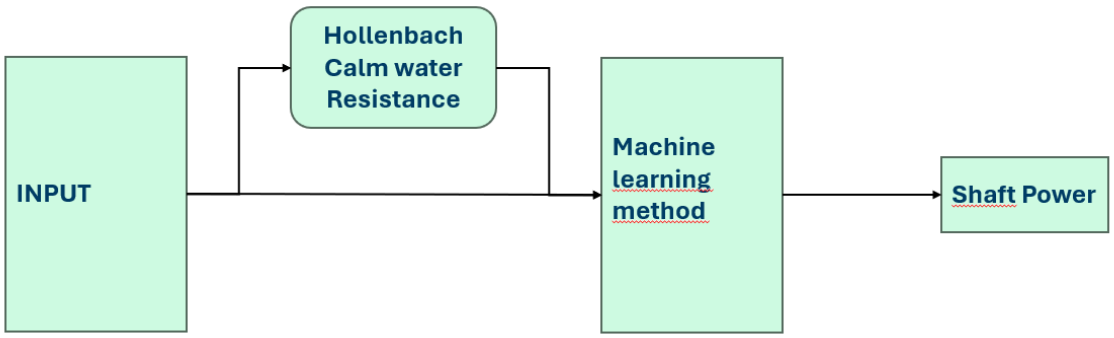


Fig.1: Hybrid model implementation

3. Results and discussion

3.1. Data

The dataset was divided into training, validation, and test subsets based on time. The first three years of data were used for training, while the remaining two years were split into validation and test sets (one year each). The validation dataset was used for model selection and hyperparameter tuning, whereas the test dataset was retained as a holdout set and remained unseen until the final evaluation. The number of samples and corresponding time periods for each dataset are listed in Table II.

A time-based split was chosen instead of a random split to reduce the risk of overfitting. Since the data were filtered to include only steady-state transit conditions, consecutive observations are often highly similar, as environmental conditions typically change slowly over 15-minute intervals. A random split would therefore distribute nearly identical observations across the training, validation, and test sets, allowing models to memorize operational patterns rather than learn generalizable relationships between inputs and shaft power. By separating the data chronologically, this risk is mitigated, while also providing a simple and realistic evaluation of model performance on unseen future conditions.

Table II: Number of samples in each dataset

Dataset	Size (number of samples)	Date
Train	60000	2020-2023
Validation	13000	2024
Test	12000	11/2024 - 11/2025
Test-reliable	6000	11/2024 - 4/2025

During the evaluation of the model, a portion of the test data was found to deviate significantly from the rest of the dataset. From mid-2025 onwards, a noticeable drop in shaft power values was observed relative to other related parameters, such as STW and fuel consumption. This decrease could not be explained by external factors such as hull cleaning or weather conditions. The change in shaft power is clearly illustrated in Fig.2, which shows shaft power vs. fuel consumption, two parameters known to be strongly correlated. Data from May 2025, highlighted in red, demonstrates a clear deviation in the fuel-power relationship compared to earlier observations. As no logical explanation for this shift in shaft power was identified, the latter portion of the test data was excluded due to this inconsistency.

Specifically, all data collected after 15 April were removed. The resulting cleaned test dataset, referred to as test-reliable, reduces the original test set by approximately half to 6000 samples, Fig.2, consequently limiting the range of operational conditions represented. Hereafter, test-reliable is used as the test dataset.

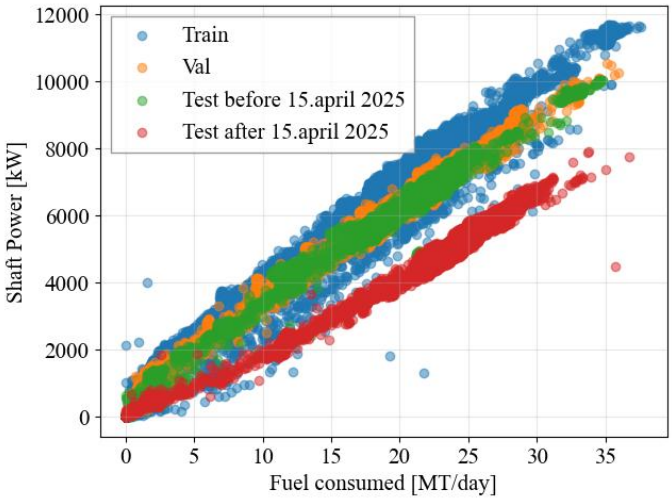
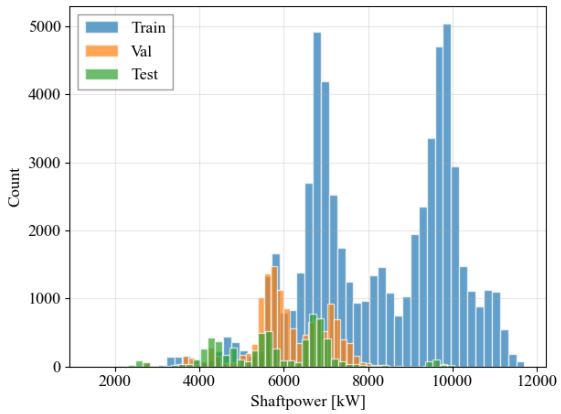
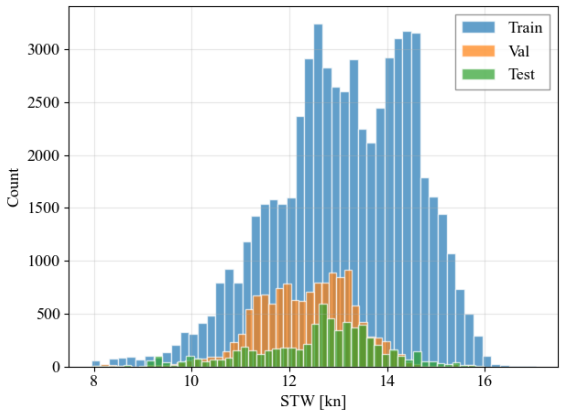


Fig.2: Fuel consumption vs. shaft power

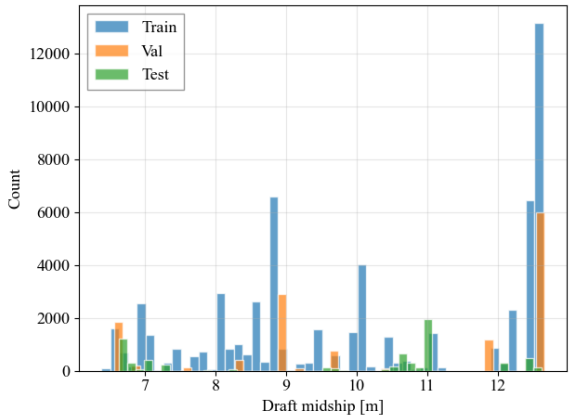
Fig.3 shows the data distribution across the datasets for some selected features. The dataset is processed as described in Section 2.1.1 and only contains the identified transit operational phases, and with unreliable data filtered out.



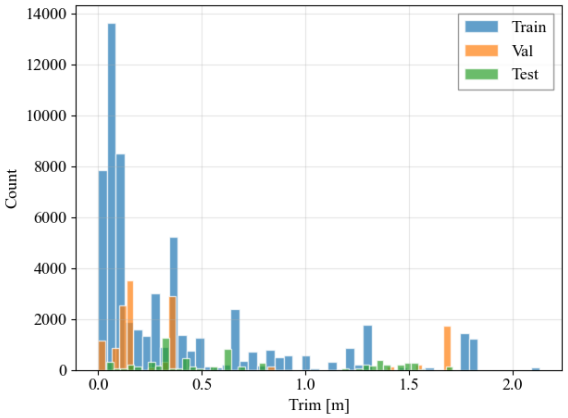
(a) Shaft power



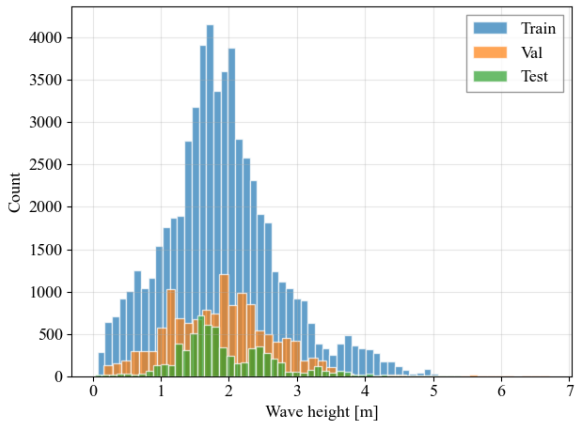
(b) STW



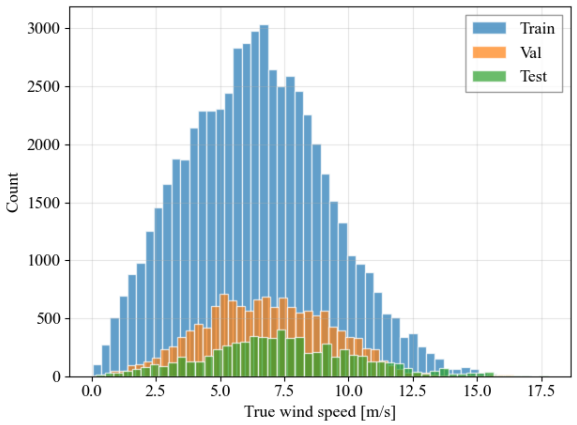
(c) Draft



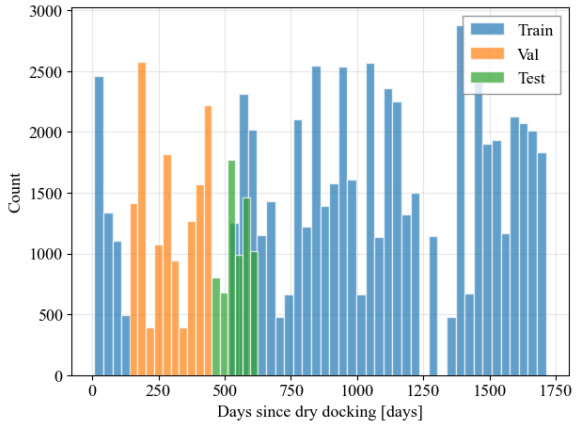
(d) Trim



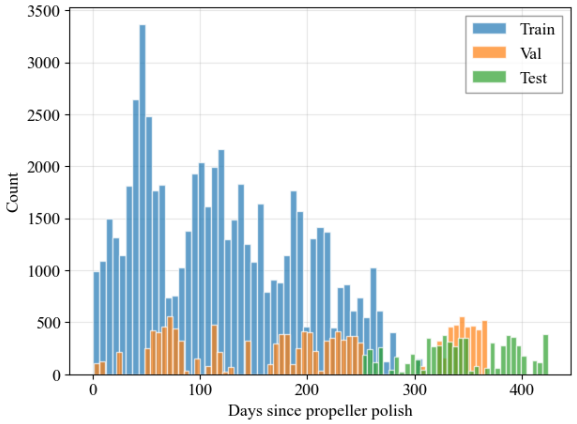
(e) Wave height



(f) True wind speed

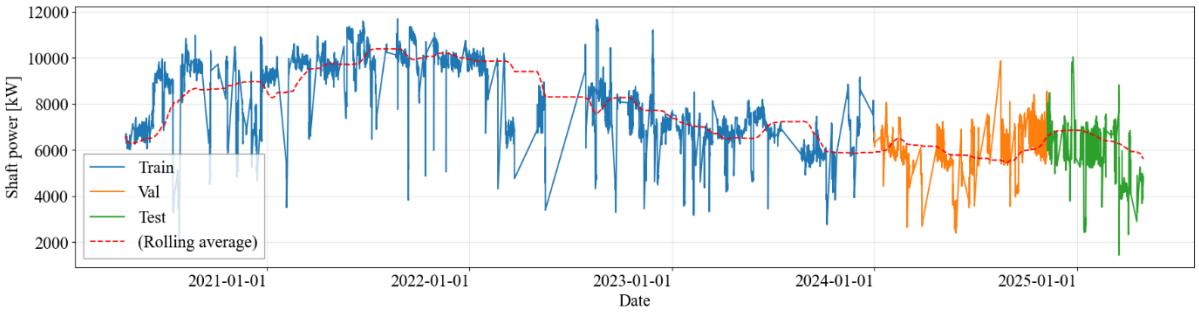


(g) Days since dry docking

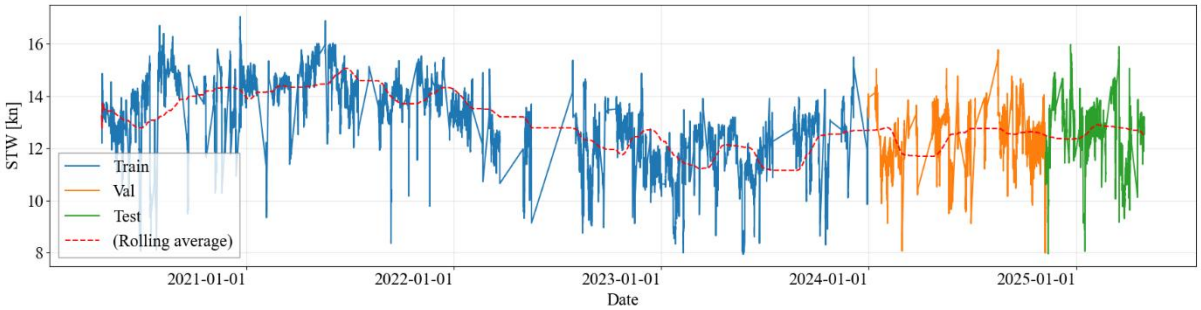


(h) Days since propeller polish

Fig.2: Data distribution for training, validation, and test datasets



(a) Shaft power over time



(b) STW over time

Fig.3: Shaft power distributed in time for different datasets, with a gaussian rolling average across the entire dataset.

The distributions across the three datasets (training, validation, and test), Fig.3, are similar for several variables, including STW, draft, trim, wave height, and wind speed. These comparable distributions ensure that model performance can be evaluated across all encountered environmental conditions and most STW ranges, providing a foundation for the subsequent analysis of different operating conditions, including speed, shaft power, and weather. For draft and trim, Fig.3c and Fig.3d), the training dataset spans a wider range of configurations than the validation and test datasets. This is expected, as draft and trim remain relatively constant throughout a voyage; therefore, a dataset covering a longer period naturally captures more individual vessel configurations.

The maintenance-related variables are more unevenly distributed, with no dry docking or propeller polishing events occurring during the test period. The shaft power distribution differs notably between the datasets. The training data contains many high-power values (above 8000 kW) that are rarely observed in the validation and test sets. Conversely, lower power values (below 5000 kW) are largely absent from the training data but constitute a substantial portion of the test data. This trend is also visible in Fig.4, where higher STW and shaft power values occur predominantly in the earlier part of the dataset (i.e., the training period) and are less common in the later validation and test periods. These differences in shaft power and STW distributions highlight the need to evaluate model performance across different operating conditions, which is the focus of the following analysis.

3.2. Feature and model selection

Features were selected based on domain knowledge of factors influencing shaft power and insights from prior studies on shaft power prediction. An initial screening on various feature subsets was performed to verify the feature selection. The resulting set of features is summarized below, with base features used for all machine learning models and additional features included for the physics-informed models. All models share the same target variable.

Target:

- Shaft power

Base features:

- Speed through water
- Static draft
- Static Trim
- Wave height
- Wave period at spectral peak
- Relative wave direction
- Relative wind speed longitudinal
- Relative wind speed transverse
- Time since dry docking
- Time since propeller polishing

Additional features (used in hybrid models)

- Calm water resistance

3.2.1. Model selection

A description of the selected models for this study is provided in Table III. Building on the features described in the previous section, the validation dataset was used to perform a hyperparameter sweep, evaluating multiple parameter combinations to identify the most promising configurations. For the NN-based models, both the input features and target are scaled with standard scalers. A complete description of hyperparameters for the selected models is provided in Appendix A.

Table III: Description and hyperparameters of models

Model	Description	Hyperparameters
NN	NN with base features. Log transformed target.	hidden_layer_sizes: [164, 108]
PINN	NN with base features and physics informed with Hollenbach calm-water resistance. Log transformed target.	hidden_layer_sizes: 177
XGBoost	Trained base features	n_estimators: 200 max_depth: 2
PI_XGBoost	XGBoost trained with based features and physics informed with Hollenbach calm water resistance.	n_estimators: 200 max_depth: 2
Empirical method	Hollenbach calm-water resistance and empirical methods for added wave and wind. Constant propulsion efficiency.	

3.3. Overall model performance

The objective of the models is to predict shaft power, with onboard shaft power measurements serving as the ground truth. The models were evaluated with the following performance metrics; Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R^2 Score. All metrics were calculated using the Scikit Learn package in Python, *Pedregosa et al. (2011).* The performance of the models for the validation and test datasets is shown in Table IV.

Table IV: Model performance comparison for the validation and test data

Model	MAE [kW] (mean ± std)		MAPE % (mean ± std)		R2	
Dataset	**Val**	**Test**	**Val**	**Test**	**Val**	**Test**
NN	505 ± 320	896 ± 377	8.7 ± 6.3	16.6 ± 9.8	0.62	0.50
PINN	441 ± 376	750 ± 369	7.3 ± 6.2	13.7 ± 8.1	0.64	0.63
XGBoost	727 ± 461	1085 ± 747	13.3 ± 10.8	22.1 ± 20.2	0.20	0.09
PI XGBoost	614 ± 380	888 ± 625	11.0 ± 8.8	17.6 ± 15.7	0.44	0.38
Empirical method	1543 ± 970	1056 ± 826	24.8 ± 14.4	18.1 ± 13.1	-2.6	0.05

Table IV shows that the best performing model, with lowest MAPE and highest R2 score, is the PINN. For the validation data, the PINN has high accuracy, with MAPE is 7.3% and a standard deviation of 6.2%, indicating some spread in the predictions. For the test data, the accuracy is lower, with MAPE of 13.7%, but the standard deviation remains comparable to the validation data. Compared to the purely data-driven NN model, incorporating calm water resistance into the PINN improves overall model performance. Among the tree-based models, XGBoost does not reach the same level of accuracy as the NN-based models. However, informing XGBoost with physics (PI_XGBoost) improves its predictions, though it still remains inferior to NN-based models. The empirical model shows similar performance to XGBoost on the test data, but exhibits significantly higher error on the validation data, with a MAPE of 24.8%, compared to the data-driven models.

All data-driven models perform better on the validation dataset than on the test dataset, whereas the empirical model shows the opposite trend: its accuracy improves on the test dataset and becomes comparable to the data-driven models. Part of this effect is due to hyperparameter tuning, which was performed on the validation data and can enhance performance specifically for this dataset. Another contributing factor is the difference in data distributions between the training, validation, and test datasets. The validation data contains more points overlapping with the densely distributed regions of the training data (e.g., higher shaft power), whereas the test data requires more extrapolation outside the confident training range, Fig.3a.

Fig.5 shows the measured and predicted shaft power for the PINN model. The model tends to slightly overestimate shaft power in the test dataset, particularly at lower power values, consistent with the

observed decrease in accuracy when predicting outside the main training data distribution.

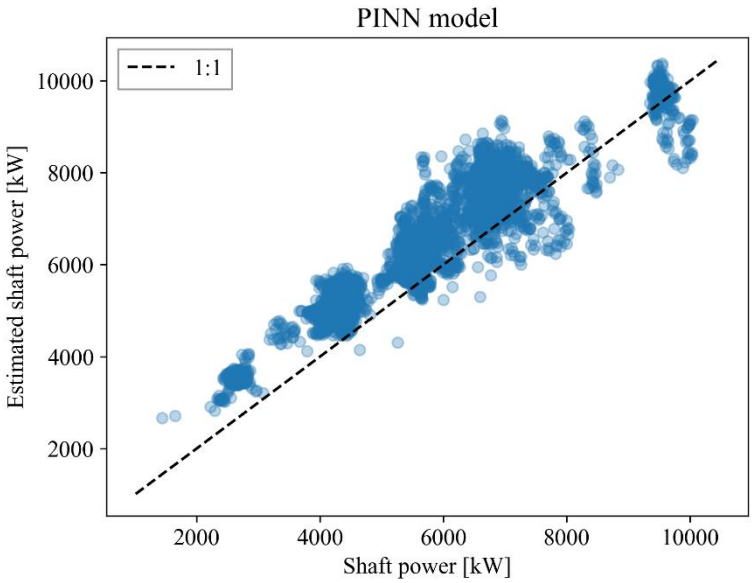


Fig.4: Measured and estimated shaft power with PINN for test data

While these overall metrics provide a general assessment of model accuracy, they do not capture how performance varies across different operational conditions. The following sections therefore analyze model performance as a function of key factors such as speed, shaft power, and environmental conditions, providing a more detailed understanding of model reliability in real-world scenarios.

3.3.1. Performance in different operational conditions

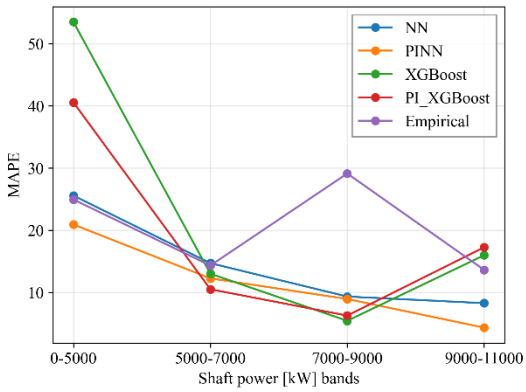
To further assess model robustness, performance is evaluated across different operating regimes, including variations in shaft power, STW, and sea state.

Shaft power

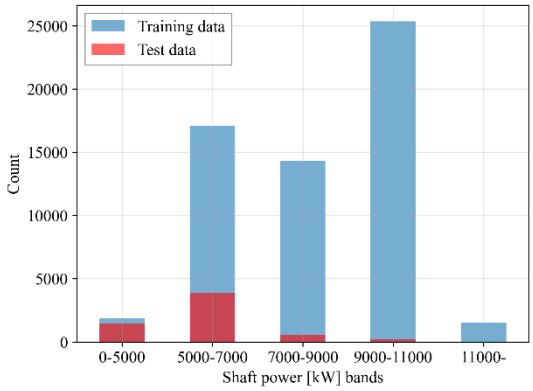
Model performance is first analyzed across five shaft power ranges. Fig.6 shows performance and sample distribution within each shaft power band for the test and validation data.

As shown in Fig.6a and Fig.6c, all data-driven models exhibit reduced accuracy in the low-power conditions (0-5000 kW). As previously noted, this condition is underrepresented in the training dataset, Fig.6b, limiting the models' ability to learn and generalize behavior in this regime. In the medium shaft power range (5000–9000 kW), the data-driven models achieve relatively high accuracy, with errors typically between 5–15% for both validation and test data. These operating conditions are well represented in the training dataset, indicating a clear relationship between model performance and data availability. For the high-power range (9000–11000 kW), model accuracy remains comparable to the mid-range conditions, although variability between models increases. However, this range contains relatively few samples in both validation and test datasets, reducing the statistical robustness of the observed performance. The highest power condition (above 11000 kW) is present only in the training data and absent from both the validation and test data. Consequently, performance cannot be evaluated for this operating regime.

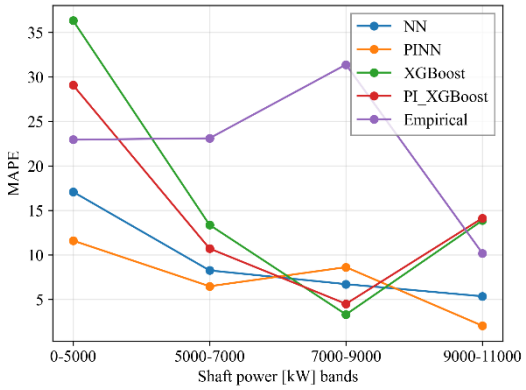
The empirical model exhibits a different performance pattern across shaft power conditions. It shows consistently higher errors for power levels below 9000 kW but achieves accuracy comparable to the data-driven models in the 9000–11000 kW range. This trend is particularly pronounced in the validation dataset.



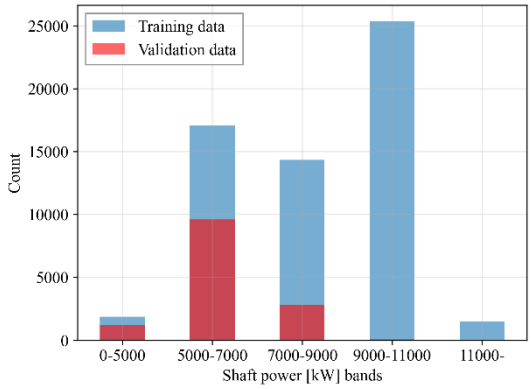
(a) MAPE at test data



(b) N samples at test data

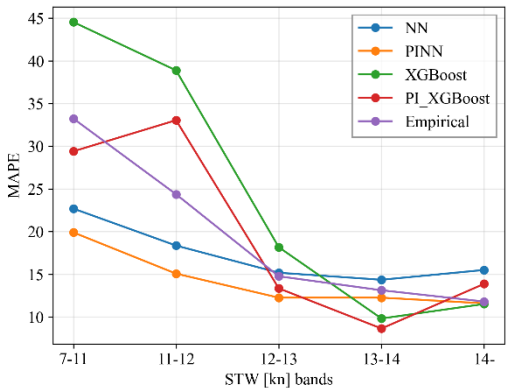


(c) MAPE at validation data

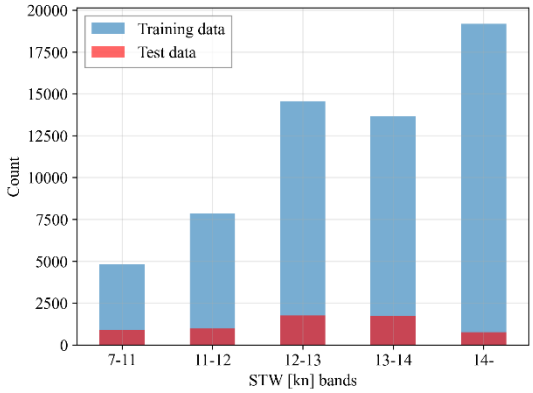


(d) MAPE at validation data

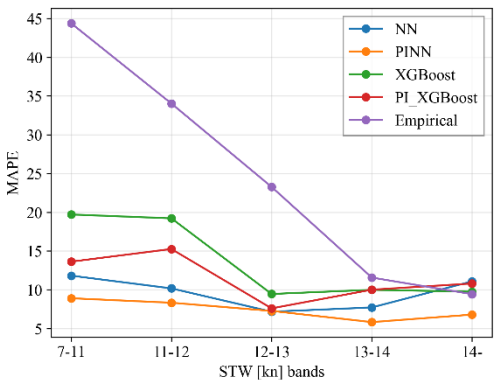
Fig.5: Models evaluated at different shaft power conditions



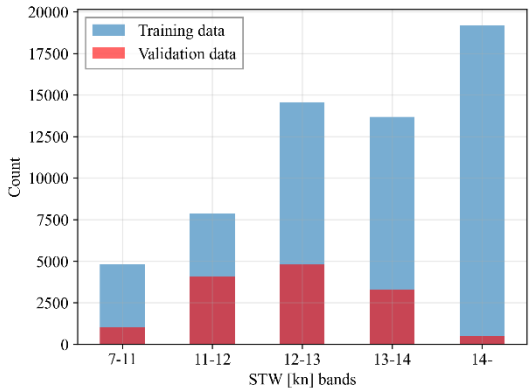
(a) MAPE at test data



(b) N samples at test data



(c) MAPE at validation data



(d) N samples at validation data

Fig.6: Models evaluated at different STW conditions

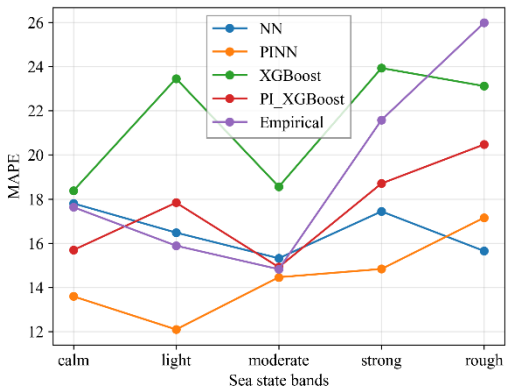
Speed Through Water

Secondly, the models are evaluated across different STW conditions. STW is divided in the following five bands: 7-11 kn, 11-12 kn, 12-13 kn, 13-14 kn, and above14 kn. Fig.7 presents the model performance and distribution of samples within each STW band for the validation and test data.

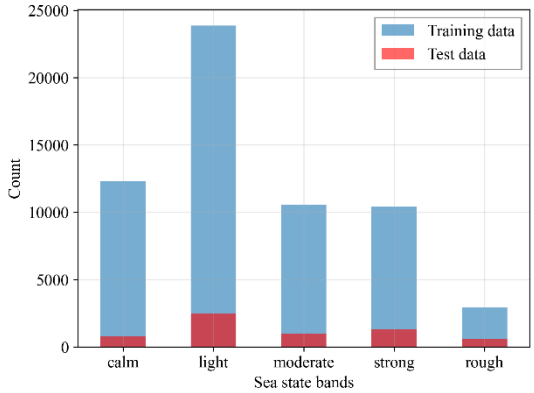
As shown in Fig.7a and Fig.7c, model accuracy generally improves at higher speed conditions compared to lower speeds. This trend is particularly evident in the test data, Fig.7a. The improvement with increasing speed is most pronounced for the empirical model, which exhibits relatively high errors (15–45%) at speeds below 13 kn. At speeds above 13 kn, however, the empirical model achieves accuracy comparable to the data-driven models. This behavior is expected, as the empirical model is based on model test data typically conducted at higher operating speeds, often about 13 kn and above. For both the test and validation datasets, the distribution across different speed conditions is relatively even and comparable to that of the training data, Fig.7b and Fig.7d, suggesting that the observed performance trends are not driven by data imbalance but reflect genuine differences in model behavior across speed regimes.

Sea State

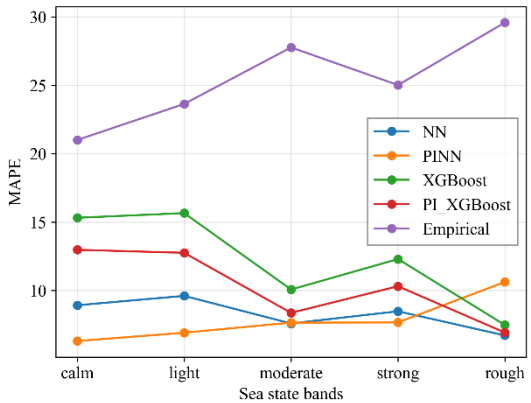
Lastly, the models are evaluated at different sea state conditions. Sea state bands are defined as a combination of true wind speed and significant wave height, Table V. Fig.8 shows model performance and distribution of samples within each sea state band for the validation and test data.



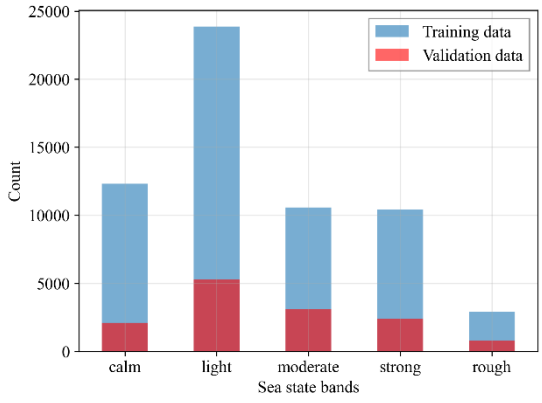
(a) MAPE at test data



(b) N samples at test data



(c) MAPE at validation data



(d) N samples at validation data

Figure 7: Models evaluated at different sea state conditions

As shown in Fig.8a, model performance on the test data varies considerably across different sea states, with substantially greater variability than observed in the validation dataset, Fig.8c. In particular, the XGBoost model performs poorly in most sea state conditions in the test data and exhibits large fluctuations in accuracy across weather categories. In contrast, the data-driven models

show more consistent performance across sea states in the validation dataset. This difference may partly be explained by the smaller size of the test dataset compared to the validation dataset, which can lead to greater variability in performance estimates.

Table V: Definition of the sea state bands used for accuracy analysis across weather conditions

Sea state bands	Condition given by true wind speed [m/s] and significant wave height (Hs) [m]
Calm	(wind_speed_true <= 4.5) & (Hs <= 2)
Light	[(wind_speed_true <= 4.5) & (2 < Hs < 3)] or [(4.5 < wind_speed_true < 9) & (Hs <= 2)]
Moderate	(4.5 < wind_speed_true < 9) & (2 < Hs < 3)
Strong	[(wind_speed_true >= 9) & (Hs < 3)] or [(wind_speed_true < 9) & (Hs >= 3)]
Rough	(wind_speed_true >= 9) & (Hs >= 3)

Among all models, the PINN generally achieves the highest accuracy under varying weather conditions. Compared to the purely data-driven NN model, the PINN performs better in calm and light sea states but shows comparable or slightly lower performance in rougher conditions. This suggests that incorporating calm water resistance improves model predictions under lighter environmental loads. However, in rough sea states, the PINN's accuracy declines slightly relative to other data-driven models, indicating some difficulty in capturing extreme conditions. Similarly, the empirical model shows reduced accuracy in harsher conditions, despite including empirical formulations for added resistance due to waves and wind. This suggests that the empirical approach does not fully account for the ship's behavior under more severe sea states. The distribution of samples across sea state bands is relatively even in both the validation and test datasets, indicating that the observed trends primarily reflect genuine differences in model robustness rather than data imbalance. However, slightly fewer samples are available in the rough weather category, particularly in the validation dataset, which may influence the robustness of the results in this regime.

Overall performance

Overall, the model evaluation across different operating conditions indicates that the PINN consistently outperforms other models in both validation and test datasets, achieving high accuracy across shaft power, ship speed, and sea state conditions. The PINN performs well in most conditions, only degrading performance slightly in rough weather and low shaft power values. As expected, the data-driven models generally perform well in regions well represented in the training data. In contrast, the empirical model shows strong dependence on vessel speed and tends to underperform in low-power or low-speed regimes. A key limitation of the empirical model is its inability to account for fouling. In comparison, the data-driven models incorporate dry docking and propeller polish events as input features, enabling them to model the effects of fouling. This may have a significant impact on performance, as fouling has shown to increase power demand by 10-18%, *Schultz (2007)*. Further investigation would be needed to quantify the effect of this limitation on the model performance in this case study.

In this study, we have evaluated the model performance for different operational conditions individually. However, isolating the effect of individual variables can be challenging. For example, variations in speed through water may be unevenly distributed across the other analyzed variables, i.e., shaft power or sea state, meaning some observed differences in performance could be influenced by speed rather than sea state or shaft power alone. This may particularly affect the empirical model, given its strong dependence on operating speed. Such interactions could also influence the observed performance across shaft power ranges, highlighting the importance of considering combined effects of multiple operational factors when interpreting model results.

Taken together, these analyses highlight both the overall strengths and the limitations of the evaluated models under different operational conditions. While hybrid models, particularly the PINN, demonstrate robust performance across a wide range of shaft power, ship speed, and sea states, the results also reveal areas where model accuracy is affected by data sparsity or interactions between

variables. These findings emphasize the conditions where model performance is degraded and highlight the importance of analyzing the combined effects of multiple operational factors.

5. Conclusion

This paper investigated hydrodynamic, data-driven, and hybrid methods for predicting the shaft power of a bulk ship under varying operating conditions. The evaluated models included NN, PINN, XGBoost, PI-XGBoost, and an empirical approach. Overall, the PINN demonstrated the best performance on both validation and test datasets. The models were further evaluated across different operational conditions, including shaft power levels, ship speed, and sea state. Across different operational regimes, the PINN consistently showed superior accuracy, particularly in mid-to-high shaft power and speed conditions. Model performance decreased under low-power, low-speed, or rough sea state conditions, highlighting areas where additional data or model refinement could be beneficial.

While the physics-informed approach demonstrated strong performance, several limitations should be noted. The models were developed for a single bulk ship and may not perform equally well when applied to other vessels. Only basic ship and propeller dimensions were assumed available, limiting hydrodynamic modeling to empirical methods that do not require more detailed ship information. Additionally, the empirical model did not account for fouling. Consequently, the accuracy of hydrodynamic predictions may be lower than what could be achieved with more detailed methods requiring additional ship-specific information. These assumptions should be considered when applying the models in practice. Despite these limitations, the approach provides a robust framework for predicting shaft power in operational planning and energy efficiency assessments, offering a practical tool for ship operators.

5.1. Further work

Generalization: The results presented here reflect the performance for a single vessel and may be sensitive to this ship and its measurements. Future work should evaluate the methods on multiple vessels to provide a more general understanding of performance across different ships.

Data availability: The models were trained on all available data, consisting of three years of operational measurements. Future studies should investigate how model performance depends on the quantity and quality of training data, and whether similar accuracy can be achieved with more limited datasets. More work should also be done on the splitting of datasets, employing some method to not only rely on a simple temporal split, but rather ensure an even distribution of operating conditions across the three datasets.

Combined operational effects: Future work could investigate the influence of combined operational factors, such as speed and sea state, on model performance, to better understand and mitigate performance degradation under complex conditions and explore the effects of collinearity on model performance.

Hybrid model: Further research could explore alternative hybrid model architectures or physics-informed features to improve performance, particularly under low-power, low-speed, or rough sea conditions.

Empirical model: Rather simple empirical resistance methods were used to establish the empirical model. To improve the empirical model, the effect of hull fouling could be added, as it has been documented have a large impact in power demand, *Schultz (2007)*. Several hull correction formulae have been developed, with the industry adopting common standards as the preferred framework for monitoring these changes in hull and propeller performance using automated shaft power and speed data, *ISO (2016)*.

Acknowledgement

This study is funded by the research project KSP FUSE supported by the Research Council of Norway (project number 352781).

References

ALTAN, D.; MOHAMMED, H.H.; LINES, G.T.; MARIJAN, D.; MARESSA, A. (2025), *Physics-Guided Neural Network-Based Shaft Power Prediction for Vessels,* arXiv:2512.20348

BOURCHAS, O.; PAPALAMBROU, G. (2025), *Physics Informed Neural Networks for Vessel Main Engine Power Prediction*, Ocean Eng. 332

CAI, Z.; LI, L.; YU, L.; LI, C.; SUN, M. (2024), *Diversity, Quality, and Quantity of Real Ship Data on the Black-Box and Gray-Box Prediction Models of Ship Fuel Consumption*, Ocean Eng. 291

CHEN, T.; HE, T.; BENESTY, M.; et al. (2026), *Xgboost: Extreme Gradient Boosting*, https://github.com/dmlc/xgboost.

CORADDU, A.; ONETO, L.; BALDI, F.; CIPOLLINI, F.; ATLAR, M.; SAVIO, S. (2019), *Data-Driven Ship Digital Twin for Estimating the Speed Loss Caused by the Marine Fouling*, Ocean Eng. 186

DALHEIM, Ø.Ø.; STEEN, S. (2020), *A Computationally Efficient Method for Identification of Steady State in Time Series Data from Ship Monitoring*, J. Ocean Eng. and Science 5(4), pp.333-345

FAN, A.; JIAN, Y.; YANG, L.; VLADIMIR, N.; WU, D. (2022), *A Review of Ship Fuel Consumption Models*, Ocean Eng. 264C

GHANI, A.A.; ZOOLFAKAR, M.R.; OSNIN, N.A. (2025), *Harnessing Artificial Neural Networks for Optimizing Marine Design: A Comprehensive Review of Applications and Innovations*, Marine Systems & Ocean Technology 20(4)

GCMD (2024), *Pay-as-You-Save. Closing the Data-Financing Gap to Turbocharge Maritime Energy Efficiency Technologies (EET) Retrofits*, Global Center for Maritime Decarbonization, https://www.gcformd.org/_files/ugd/4bf1f4_9edc6b4db49347c6a48f76bc1322b46c.pdf

GUO, B., GUPTA, P.; STEEN, S.; TVETE, H.A. (2023), *Evaluating Vessel Technical Performance Index Using Physics-Based and Data-Driven Approach*, Ocean Eng. 286

HOLLENBACH, K.U. (1998), *Estimating Resistance and Propulsion for Single-Screw and Twin-Screw Ships*, 10th ICCAS Conf.

ISO (2016), *ISO 19030-1:2016*, Int. Standard Org., Geneva

ITTC (2021), *ITTC Recommended Procedures and Guidelines. Preparation, Conduct and Analysis of Speed/Power Trials*, Int. Towing Tank Conf.

JØRGENSEN, U.; BELLINGMO, P.R.; MURRAY, B.; BERGE, S.P.; POBITZER, A. (2022), *Ship Route Optimization Using Hybrid Physics-Guided Machine Learning*, SINTEF, Trondheim, https://sintef.brage.unit.no/sintef-xmlui/handle/11250/3013472

LIU, S.; PAPANIKOLAOU, A. (2020), *Regression Analysis of Experimental Data for Added Resistance in Waves of Arbitrary Heading and Development of a Semi-Empirical Formula,* Ocean Eng. 206

MAVROUDIS, S.; TINGA, T. (2025), *Application of Transfer Learning on Physics-Based Models to Enhance Vessel Shaft Power Predictions*, Ocean Eng. 323

MORADI, M.H.; BRUTSCHE, M.; WENIG, M.; WAGNER, U.; KOCH, T. (2022), *Marine Route Optimization Using Reinforcement Learning Approach to Reduce Fuel Consumption and Consequently Minimize CO2 Emissions*, Ocean Eng. 259

NAPA (2020), *Survey: Owners and Charterers Believe That Voyage Efficiency Savings of 28% Are Possible*, NAPA, https://www.napa.fi/news/survey-owners-and-charterers-believe-that-voyage-efficiency-savings-of-28-are-possible/

NGUYEN, V.G.; RAJAMOHAN, S.; RUDZKI, K.; et al. (2023), *Using Artificial Neural Networks for Predicting Ship Fuel Consumption*, Polish Maritime Research 30(2), pp.39–60

PEDREGOSA, F. ; VAROQUAUX, G.; GRAMFORT, A. ; et al. (2011), *Scikit-Learn: Machine Learning in Python*, J. Machine Learning Research 12(85), pp.2825–2830

SCHULTZ, M.P. (2007), *Effects of Coating Roughness and Biofouling on Ship Resistance and Powering*, Biofouling 23(5), pp.331–341

SKJØRDAL, S.O.; FALTINSEN, O.M. (1980), *A Linear Theory of Springing*, J. Ship Research 24(2), pp.74-84

SUNG, I., ZOGRAFAKIS, H.; NIELSEN, P. (2022), *Multi-Lateral Ocean Voyage Optimization for Cargo Vessels as a Decarbonization Method*, Transportation Research Part D: Transport and Environment 110

WANG, H.; LANG, X.; MAO, W. (2021), *Voyage Optimization Combining Genetic Algorithm and Dynamic Programming for Fuel/Emissions Reduction*, Transportation Research Part D: Transport and Environment 90

Appendix A

Model	Description	Hyperparameters
NN	NN with base features. Log transformed target.	activation: tanh alpha: 0.001 hidden_layer_sizes: [164, 108] learning_rate_init: 0.0001 loss function: huber
PINN	NN with base features and physics informed with Hollenbach calm water resistance. Log transformed target.	activation: leaky_relu alpha: 8.284e-06 hidden_layer_sizes: 177 learning_rate_init: 0.00981 loss function: huber
XGBoost	Trained base features	n_estimators: 200 max_depth: 2 learning_rate: 0.05 subsample: 0.6 colsample_bytree: 0.4 random_state: 42 min_child_weight: 10 gamma: 0.5 reg_lambda: 5 reg_alpha: 0.5
PI_XGBoost	XGBoost trained with based features and physics informed with Hollenbach calm water resistance.	n_estimators: 200 max_depth: 2 learning_rate: 0.05 subsample: 0.6 colsample_bytree: 0.4 random_state: 42 min_child_weight: 10 gamma: 0.5 reg_lambda: 5 reg_alpha: 0.5
Empirical method	Hollenbach calm water resistance and empirical methods for added wave and wind. Constant propulsion efficiency.	

Call for Evidence-Based Performance Claims in the Marine Coatings Industry

Morten Sten Johansen, Jotun, Sandefjord/Norway, msj@jotun.com
Manolis Levantis, Jotun, Athens/Greece
Olav Rognebakke, DNV, Hovik/Norway
Anna-Louise Meyer Wangen, DNV, Hovik/Norway

Abstract

Speed loss (ISO 19030) has become a widely used Key Performance Indicator (KPI) for assessing hull and propeller performance. However, the use of unverified and non-transparent speed loss figures may turn this metric into a marketing numbers game rather than a reliable performance measure as originally intended when the ISO standard was published. Without transparent and substantiated evidence—such as standardized data collection, validated analysis methods, and consistent baselines—claims of "low speed loss" risk misleading shipowners and operators and distorting competition. This paper calls for an industry-wide commitment to evidence-based reporting, emphasizing the need for documented and verifiable data before performance claims can be made.

1. Introduction

Hull and propeller performance plays a critical role in vessel energy efficiency, fuel consumption, and greenhouse-gas (GHG) emissions. Regulatory frameworks such as the IMO Carbon Intensity Indicator (CII) and FuelEU Maritime place increasing pressure on operational efficiency, and reliable performance metrics are essential, *IMO (2023), EU (2023)*. Among these metrics, speed loss has emerged as a key indicator of hydrodynamic degradation over time. Speed loss represents the reduction in achievable vessel speed at constant delivered shaft power relative to a defined reference condition. ISO 19030 was developed to eliminate ambiguity in performance assessment and provide a globally consistent framework for evaluating hull and propeller efficiency degradation. Standardized through ISO 19030, speed loss was intended to provide a consistent and comparable methodology for quantifying changes in vessel's hull performance under real operating conditions, *ISO (2016)*. It included standardized sensor requirements, defined data filtering and correction procedures, environmental normalization (wind, waves, water temperature), transparent baseline selection and repeatability and verifiability. Crucially, the standard emphasizes that performance results should be traceable and reproducible, ensuring that speed loss remains a technical KPI rather than a subjective interpretation.

2. Deviation from the Standard: How Speed Loss Became a Numbers Game

ISO 19030 was developed to establish a common, objective framework for quantifying changes in hull and propeller performance under real operational conditions. Its intent was not merely to define a calculation method, but to ensure comparability, repeatability, and verifiability of performance results across vessels, *ISO (2016)*. Despite this, a growing gap can be observed between the principles of the standard and current industry practice. Speed loss, originally conceived as a technical KPI, is increasingly presented as a simplified commercial figure, often without sufficient transparency regarding data sources, assumptions, or analytical methodology.

Oversimplified communication of results has become increasingly prevalent. Single-value claims such as "2% speed loss over 60 months" are often presented without disclosure of data, baseline definition, uncertainty range, or operational context. Such representations may be suitable for marketing material but are inconsistent with good scientific practice and risk being interpreted as measured, long-term performance outcomes rather than indicative or model-based estimates, *Carlton (2019)*.

One of the most common deviations relates to baseline definition. ISO 19030 emphasizes the importance of a stable and representative reference condition, typically established once the vessel has

reached steady operational performance following coating application, *ISO (2016)*. In the absence of such reference, baselines are sometimes selectively defined or retrospectively adjusted, for example by excluding early operational periods. Such practices may reduce reported speed loss values but undermine the comparability and scientific validity of the results, *Schultz (2007)*. Another significant issue is the increasing reliance on proprietary or non-transparent analytical models. While the use of advanced data analytics is not inherently problematic, results derived from so-called "black-box" models − where filtering criteria, correction factors, and assumptions are not disclosed − cannot be independently verified or reproduced. This contradicts a fundamental requirement of scientific measurement and conflicts with the intent of ISO 19030, which assumes that methodologies can be documented and, in principle, replicated by third parties, *ITTC (2021)*.

Data quality and data sufficiency represent further sources of deviation. Reliable speed loss assessment depends on calibrated propulsion measurements, accurate speed data, and robust environmental corrections. Studies have demonstrated that insufficient data frequency, unvalidated sensors, or incomplete / unreliable corrections for wind, waves, water temperature, and shallow-water effects can lead to significant deviations in calculated performance trends, *Townsin (2003)*. In some cases, speed loss values are derived from datasets that do not meet the minimum data frequency by the standard, increasing the uncertainty of the speed loss values.

A related concern is the treatment of operational variability. Vessel performance is influenced by seasonal effects, trading patterns, loading conditions, and biofouling pressure. ISO 19030 accounts for this by requiring environmental normalization and consistent filtering criteria. However, selective exclusion of operating conditions or limited temporal coverage may mask underlying performance degradation and produce overly optimistic results, *Schultz (2007), DNV (2022)*.

Taken together, these deviations contribute to the transformation of speed loss from a standardized performance indicator into a competitive numbers game, where methodological rigor is subordinated to headline figures. This trend not only undermines the credibility of individual claims but also risks eroding confidence in speed loss as a KPI across the industry.

3. Implications for Shipowners and the Industry

The increasing use of non-transparent or weakly substantiated speed loss claims has direct and indirect consequences for shipowners, operators, and the wider maritime industry.

From a procurement perspective, speed loss figures are frequently used as a decisive input when selecting antifouling coatings. When such figures are derived from limited datasets, non-standardized baselines, or undocumented analytical methods, shipowners are exposed to significant decision risk. Even small deviations in reported speed loss − on the order of 1–2% − can translate into substantial differences in fuel consumption and operational expenditure over a docking interval, *MAN (2020)*. Consequently, procurement decisions based on optimistic but unverified claims may result in higher lifecycle costs rather than the expected efficiency gains.

As an illustrative example, consider a Suezmax bulk carrier operating at 14 kn with a 70% voyage factor (≈1278 days at sea over five years). A misleading deviation of 1 percentage point in reported speed loss may appear marginal, but operationally it can be significant. Given that propulsive power scales approximately with the cube of speed, a 1% shortfall in speed may require roughly ~3% additional power and fuel to maintain schedule speed, *Carlton (2019)*. At a fuel consumption of 50 t/day, a 3% penalty corresponds to 1.5 t/day, or approximately 1917 t of additional fuel over five years. At $550 per ton, this equates to roughly $1.05 million per vessel. For an owner operating 10 vessels, the cumulative exposure approaches $10 million over five years.

A particularly problematic development is the increasing number of newly launched antifouling products for which long-term speed loss claims are presented, often extending to 60 months or a full docking cycle, despite very limited time in service. In some cases, products introduced to the market less than

one or two years prior are marketed with speed loss values extrapolated to five years of operation. From a scientific perspective, such claims are inherently uncertain. Hull fouling dynamics, coating ageing, and biofilm development are strongly time-dependent and non-linear processes, *Schultz (2007)*, *Townsin (2003)*. Extrapolating early-life performance into long-term outcomes without sufficient empirical evidence introduces significant uncertainty and undermines the credibility of the reported KPI.

Beyond procurement risk, the misuse of speed loss figures has implications for regulatory compliance and ESG reporting. Performance metrics increasingly feed into compliance strategies related to CII, internal carbon pricing, and emissions-reduction roadmaps, *IMO (2023)*, *EMSA (2021)*. If speed loss data is overstated or inaccurately characterized, shipowners may overestimate future efficiency gains and underestimate compliance risk. This not only affects individual vessels but may also distort fleet-level decarbonization strategies.

For the Suezmax example above, the environmental implications are equally material. Using an emission factor of ~3.1 t CO_2 per tonne of fuel, the additional consumption represents approximately 5,900 tonnes of CO_2 per vessel, or nearly 59,000 tonnes of CO_2 across ten vessels over five years. Even small, unverified differences in reported speed loss can therefore translate into substantial financial and GHG consequences, underscoring the importance of transparent and evidence-based performance claims, *IMO (2023)*.

At an industry level, the situation creates a competitive imbalance. Suppliers that adhere strictly to ISO-compliant methodologies, disclose data limitations, and avoid extrapolating beyond available evidence may appear less competitive than those presenting simplified or overly optimistic figures. This dynamic discourages transparency and penalizes scientific rigor, ultimately weakening trust in performance-based decision-making across the sector, *Anderson (2018)*.

4. Industry alignment

Preserving the value of speed loss as a credible KPI requires coordinated industry action. Without alignment among suppliers, shipowners, and third-party stakeholders, speed loss risks losing its role as a reliable performance indicator and becoming merely a comparative marketing metric.

A first step toward alignment is the establishment of clear procurement requirements. Shipowners and operators can require that any speed loss claim:

- Is derived using ISO 19030-compliant methodologies,
- Is based on documented datasets covering the stated evaluation period,
- Includes transparent disclosure of assumptions, baselines, and data coverage.

Such requirements would reduce ambiguity and encourage suppliers to align marketing claims with verifiable evidence rather than extrapolated expectations.

Independent verification represents another critical element. Just as emissions data under MRV and CII frameworks increasingly rely on third-party verification, speed loss assessments could benefit from independent review by classification societies, research institutes, or trusted performance-monitoring providers, *EMSA (2021), IACS (2023)*. Verification does not necessarily require public disclosure of raw data, but it does require that methodologies and results can be replicated and audited under controlled conditions.

The issue of long-term performance claims for newly introduced products further highlights the need for common reporting principles. Where coatings have limited operational exposure, performance claims should be clearly framed as early-life observations, indicative trends, or model-based projections, rather than definitive long-term outcomes. Explicitly distinguishing between measured data and

extrapolated projections would align speed loss reporting with established scientific practice and reduce the risk of misinterpretation.

Five years is a long time to wait before a speed-loss number with high commercial impact is estimated and communicated. New antifouling products are introduced to meet evolving requirements, whether by reducing environmental impact, supporting ESG targets, lowering overall system costs, improving application efficiency, or delivering superior hull performance. The supplier needs to evaluate the effect of changes to hit the sweet spot for new products. Extensive studies are carried out to study relevant aspects; these studies produce evidence that an independent 3rd party may use for assurance of product claims. An established verification framework does not exist, and sound engineering judgement is required for a limited assurance.

A new product is normally a development of existing technology. If a verified speed-loss number exists for the original antifouling product, a relative assessment may be used as a basis for assurance of new performance claims. Evaluation of differences is still challenging. Some important elements are listed in the following. The hydrodynamic characteristics of the coating and resulting frictional resistance must be compared for the products as applied on the hull at acceptable conditions, as well as when ageing effects are considered. The antifouling characteristics must be assessed by exposing test panels to conditions covering relevant fouling scenarios. Deterioration and wear of the products should similarly be tested and quantified in a relative manner. The change in chemistry and composition may be used to explain expected differences or similarities, but likely a well-designed and substantiated accelerated process in the laboratory is required.

In parallel, there may be a need to further develop or clarify ISO 19030 guidance, particularly in light of increased digitalization, higher-frequency data availability, and the growing commercial importance of speed loss figures. Clarifications related to minimum data duration, acceptable extrapolation practices, and uncertainty reporting could strengthen the standard's robustness without fundamentally changing its structure, *ISO (2016)*.

Finally, an industry code of conduct for speed loss reporting—covering transparency, documentation, and responsible communication—could help restore confidence in the KPI. Similar approaches have proven effective in other areas of maritime performance reporting, including emissions monitoring and fuel-efficiency benchmarking, *EMSA (2021)*.

Credible performance claims ultimately depend on verifiability and reproducibility. Access to underlying datasets (subject to confidentiality), full documentation of analytical steps, and the possibility of third-party replication are essential. A performance result that cannot be independently reproduced does not meet basic scientific standards and should not be presented as an objective KPI, *ITTC (2021)*.

5. Conclusion

Speed loss, as defined in ISO 19030, was introduced to provide the maritime industry with a standardized, objective, and scientifically robust KPI for assessing changes in hull and propeller performance. When applied as intended, the metric offers valuable insight into hydrodynamic degradation, supports informed coating selection, and contributes to improved fuel efficiency and emissions reduction.

This paper discusses how current industry practice has increasingly diverged from the principles underpinning the standard. Simplified and non-transparent speed loss claims, selective baseline definitions, undocumented analytical models, and extrapolation beyond available operational data have contributed to a growing disconnect between reported figures and verifiable performance. In particular, the emergence of long-term speed loss claims for newly introduced antifouling products—despite limited time in service—highlights a critical methodological weakness that risks undermining the credibility of the KPI. Assurance of performance claims for new products may be possible based on a comparative study, if a verified base-line product exists. All aspects of the antifouling product should then be covered and accounted for in a conservative way.

The implications extend beyond technical debate. For shipowners and operators, reliance on weakly substantiated performance claims introduces financial, operational, and regulatory risk. At an industry level, inconsistent reporting practices distort competition and discourage transparency, ultimately eroding confidence in performance-based decision-making at a time when accurate metrics are increasingly required for decarbonization strategies and regulatory compliance.
Re-establishing speed loss as a credible KPI requires a renewed commitment to evidence-based reporting. This includes strict adherence to ISO 19030 methodologies, transparent documentation of data sources and analytical assumptions, clear differentiation between measured data and extrapolated projections, and verifiability through independent review. Procurement practices, third-party verification, and potential clarification of existing standards all have a role to play in driving alignment.

Until alternative, industry wide accepted standards are used, speed loss should remain a technical performance indicator, grounded in measured data and reproducible analysis. Only through collective industry action can the metric retain its intended value as a tool for fair comparison, informed decision-making, and continuous improvement in hull and propeller performance.

References

ANDERSON, C. (2018), *Performance Metrics and Market Transparency in Shipping*, J. Marine Engineering 44/33, pp.215–228

CARLTON, J. (2019), *Marine Propellers and Propulsion*, Butterworth-Heinemann, Oxford

DNV (2022), *Hull Performance Monitoring and Fouling Impact*, DNV Maritime Advisory Report, DNV, Høvik

EMSA (2021), *MRV and Performance Data Quality Considerations*, EMSA Report, European Maritime Safety Agency, Lisbon

EU (2023), *FuelEU Maritime Regulation (EU) 2023/1805*, European Union, Brussels

IACS (2023), *Unified Requirements for Digital Performance Monitoring*, Int. Association of Classification Societies, London

IMO (2023), *IMO GHG Strategy*, Int. Mar. Org., London

ISO (2016), ISO 19030:2016 - Measurement of changes in hull and propeller performance, Int. Mar. Org., Geneva

ITTC (2021), *Recommended Procedures and Guidelines – Speed and Power Trials*, Int. Towing Tank Conf.

MAN (2020), *Propulsion Efficiency and Fuel Consumption*, Technical Paper, MAN Energy Solutions, Augsburg

SCHULTZ, M.P. (2007), *Effects of Coating Roughness and Fouling on Ship Resistance*, Biofouling 23/5, pp.331–341

TOWNSIN, R.L. (2003), *The Economic Impact of Fouling on Ships*, Int. Biodeterioration & Biodegradation 52/2, pp.87–94

From Fouling to Fuel Savings: Accurate Fouling Avoidance by Implementing Performance Monitoring into Biofouling Management

Ditte Gundermann, **Andrea Farkas, Mads Bertelsen, Viktor Avlonitis**, Hempel, Lyngby/Denmark, digu@hempel.com

Abstract

This paper discusses performance monitoring as a practical detection layer for biofouling and outlines best practices for integrating it in biofouling management plans through clear KPIs, trigger thresholds and contingency actions. Our results support ship operators to increase operational efficiency and comply with the IMO Biofouling Guidance.

1. Introduction

Biofouling is a global challenge which affects both environment and economics. In shipping, the transfer of invasive aquatic species via ships is a significant driver of biodiversity loss. Organisms transported in ballast water or on ships' hulls can survive transit and establish viable populations in non-native environments, potentially displacing indigenous species, *IMO (2023b)*. This risk has intensified with the growth of global shipping traffic over recent decades, and rising seaborne trade suggests that the problem will persist without effective management.

Beyond biosecurity, biofouling on ship hulls increases ship resistance and alters propulsion characteristics, raising fuel consumption and emissions of air pollutants and greenhouse gases, *Farkas et al. (2020)*. Because fuel typically constitutes 60–70% of a vessel's total operating costs, even moderate biofouling growth on the hull or propeller can materially affect consumption. Understanding the degree of biofouling and its effect on ship performance has therefore become essential for shipping companies, *Park et al. (2018)*. Early detection of biofouling on hulls and propellers is critical for optimizing vessel performance, prevent possible functional problems and minimizing environmental impacts.

The IMO guidance on biofouling introduces the Biofouling management plan (BFMP) and the Biofouling Records Book (BFRB) as ship-specific documents that structure biofouling risk management. The BFMP defines preventive technologies (e.g. antifouling systems and marine growth prevention systems), operating practices, monitoring activities, survey schedule, in-water cleaning regimes, decision triggers and contingency plans. BFMPs often include a comprehensive mapping of all niche areas on the hull, providing critical input for planning activities for in-water surveys. The BFRB logs all relevant dry docking and in-service activities together with planned and unplanned conditions that may influence biofouling risk (e.g., idle periods) thereby serves as an audit trail for compliance.

Although the 2023 IMO Biofouling Guidelines recommend a "ship-specific contingency action plan based on specific triggers from monitoring of biofouling parameters should be described in the BFMP", *IMO (2023a)*, they stop short of prescribing how to implement a detection to action framework – including which methods to use, which KPIs to track, how to set quality controls, and what limits or triggers to set. This paper evaluates hull performance monitoring as a practical detection layer for biofouling, examines its limitations and benefits, and proposes best practices for integrating performance monitoring into Biofouling Management Plans (BFMPs) with specific emphasis on contingency planning and decision triggers.

2. Detection of biofouling

Early detection of biofouling can have environmental and economic benefits for ship operators. From a practical perspective, mitigation measures are most effective when implemented at the initial stage of biofilm formation. Biofouling accumulation on the hulls of in-service ships can be detected and

evaluated via direct observation (i.e., visual in-water inspections), and indirectly via monitoring methods. Advantages and challenges with common approaches are described in the following sections.

2.1. Visual inspections

Underwater inspections are recognized as an indispensable activity to identify the location, composition, and growth of biofouling, *Davidson et al. (2008)*. Visual inspections are the only way to directly observe the biofouling growth on ship hulls They can be classified into diver inspections, unmanned underwater vehicles inspections with hull cameras. Diver inspections represent the most common approach in the shipping industry nowadays and is typically accompanied by polishing of the propeller. Generally, after diving inspection is carried out, an underwater inspection report is prepared which includes some images as well as summary about fouling condition of the hull and propeller. However, reporting formats are not standardized, image interpretation can be subjective, and representativeness is limited by sampling locations and visibility. As a result, it can be challenging, based on the underwater inspection report alone, to make conclusions about financial feasibility of hull cleaning.

The IMO Biofouling Guidelines recommend visual inspections as key activity "if the monitoring of biofouling parameters identifies an indication of prolonged elevated risk", *IMO (2023a)*. Furthermore, the IMO Biofouling Guidelines draw a distinction on inspection frequency between ships undertaking performance monitoring and ships that do not. However, ships that rely only on scheduled inspections may miss early signs of biofouling accumulation that would otherwise be uncovered with hull performance monitoring.

Inspections are therefore the primary route to confirm need for IWC and compliance actions, while ROI of IWC is best assessed by combining inspection evidence with performance data trends and trade/idle exposure.

2.2. Other techniques

Besides visual inspections for biofouling detection, there are some other more complex techniques developed for this purpose. Biological indicators such as adenosine triphosphate (ATP) content and assimilable organic carbon (AOC) are commonly employed to evaluate biofouling potential, as ATP reflects cellular energy transfer and AOC promotes microbial growth and biofilm formation. However, these parameters require sampling and are therefore unsuitable for real-time observation. Alternative approaches include ultrasonic time-domain reflectometry for monitoring early biofilm development on polymeric surfaces and membrane fouling simulators, which can be integrated with high-precision monitoring systems, *Wu et al. (2023)*. Within an EU funded project, *BU (2012),* fouling detection using pulsed ultrasound was studied and sensors for fouling detection were developed and optimized. Transducers emit low-frequency waves to detect fouling by analyzing signal attenuation. This approach can identify even very thin layers, using an internal reference signal to ensure measurement accuracy.

2.3. Performance monitoring

In addition to the above mentioned more direct methods of biofouling detection, there are certain methods to detect biofouling indirectly. These entail the use of vessel performance monitoring systems. With these methods, Key Performance Indicators (KPIs) such as increased power consumption or reduced speed are inferred from in service data and monitored on continuous basis. Through proper choice and interpretation of such KPIs, the resistance increase caused by biofouling can be measured.

Vessel performance monitoring has evolved significantly since the earliest versions of commercial solutions emerged in the beginning of this century. The use of performance monitoring for biofouling detection gained momentum in particular after banning of TBT (Tributyltin) in antifouling coatings in 2008. While the effectiveness of TBT had reduced the problem with biofouling on ships hulls for many years, the ban of TBT meant that biofouling once again became a challenge for ship operators. Although

different coating technologies and alternative biocides existed and evolved, these were not as efficient at the time.

Over the past decade the performance-monitoring landscape has undergone several notable shifts. These include the broader implementation of high-frequency data collection as vessel technical capabilities have improved, the rapid expansion of digitalization supported by AI-driven analytics that enable more advanced performance insights, the widespread uptake of ship performance monitoring across the industry, heightened regulatory expectations prompting enhancements in monitoring practices, and a significant increase in the number of service providers offering specialized analytics and solutions.

The growing attention to biofouling risks has also reshaped how ship operators approach performance monitoring. While technologically advanced shipping companies have long relied on performance monitoring to optimize maintenance planning, the practice is now increasingly adopted by less technologically mature operators as the advantages of data-driven maintenance management become more widely recognized.

In spite of mentioned upgrades in analytical and technical capabilities, it remains a challenge to determine the true hull condition of a vessel based on performance analysis alone. This is mainly due to the following aspects:

- Many determining variables
  In order to assess the hull condition, it is necessary to eliminate the effects of non-relevant factors such as environmental and operational variations. Even with high frequency sensor data this represents a complex task.
- Complex relationship between variables
  The physics behind ship performance including ship resistance and propulsion, environmental effects, and ship operational profiles are interdependent and nonlinear. Changes in one variable (e.g., speed, trim, wave direction) can amplify or counteract the effect of others, meaning that identical hull conditions may manifest differently under different operating profiles.
- Uncertainty in data, both measured and external
  Data quality issues remain a major source of uncertainty. Measured onboard data can be affected by noise, sensor drift, calibration errors, and inconsistent logging practices. External datasets such as wave models, current fields, or wind re-analyses may have coarse spatial and temporal resolution, introducing additional uncertainty into performance corrections. These uncertainties accumulate throughout the analytical chain, reducing the confidence level in hull condition estimates and sometimes leading to misleading interpretations if not properly accounted for.
- Lack of good KPIs
  Despite significant effort across the industry, there is still no universally accepted KPI that reliably captures hull condition in a way that is consistent across vessel types, trade patterns, and sensor setups. Existing KPIs often blend effects from propulsion, weather, and operational choices, making them sensitive to factors unrelated to hull fouling or degradation. Without robust, standardized KPIs, benchmarking hull condition across time or fleets remains difficult, and conclusions drawn from performance curves can be highly dependent on the chosen methodology.

3. Coating performance guarantees and ISO 19030

The use of performance monitoring in biofouling management is reflected in the nature of coating guarantees offered by major coating manufacturers. The first coating guarantees were based on visual criteria, specifying allowable fouling size and coverage. With the emergence of performance monitoring, coating guarantees shifted from visual to performance based, specifying maximum permissible performance loss rather than fouling type and coverage.

The simultaneous increase in number of performance-monitoring providers in the 2010's, each using different methods and KPIs, together with a general industry shift towards performance-based guarantees and claims, created a need for independent and transparent assessment methods.

This led to the development of the ISO 19030, *ISO (2016),* standard, launched in 2016, which defines a simple and transparent framework for performance monitoring. Developed with input from a broad range of marine industry and academic stakeholders, the standard was expected to represent state-of-the-art practice. However, its practical applicability soon proved more limited than anticipated - primarily due to the extensive data filtering required. This often results in scarce usable data, lowering analytical reliability and weakening the basis for informed decision-making.

The averaging process in the ISO 19030 method plays a crucial role because it reduces the influence of speed and loading-condition dependencies on the performance values, *Farkas et al. (2025)*. It also helps compensate, to some extent, for limitations in environmental corrections, such as the absence of wave filtering, and makes it possible to calculate KPIs despite the inevitable gaps created by extensive data filtering.

Although ISO 19030 defines a short-term performance indicator, the "maintenance trigger," its practical use is often limited because the method filters out a large share of the datapoints needed for reliable short-term assessment. As a result, the ISO framework has been used mainly for coating guarantee evaluations, while most shipping companies continue to rely on alternative methods for more real-time performance monitoring.

Most coating companies nowadays still offer performance guarantees based on the speed loss KPIs as defined in ISO 19030. These performance guarantees do not fully meet the expectations or operational needs of modern shipping companies as they do not provide any guarantee on the level of biofouling at a specific point in time, but only guarantee average performance over the entire dry docking. There is hence a need for a new approach to performance guarantees in the coating industry, which better support the biofouling risk mitigation needs of today's shipping companies.

4. Recommended procedure for data processing

The following provides general recommendations for setting up a data processing framework which can be used for biofouling management. The main considerations and challenges related to data processing and analysis are discussed and recommendations for overcoming those challenges are presented.

4.1. Choosing a relevant Performance Value

An important component of the data processing procedure is the choice of a relevant Performance Value (PV). Here PV refers to the value of an individual data point after corrections. A wide range of PV's are used within the performance-monitoring industry, and it is not the purpose of this paper to evaluate all of them in detail. Instead, this work focuses on how effectively a PV can support assessment of hull fouling. In this context, some PVs are inherently better suited than others. Among the PVs most frequently encountered in practice are power deviation, fuel deviation, various forms of power index, speed deviation, speed index, and added resistance.

The usefulness of a PV for fouling detection depends primarily on how sensitive it is to operational and environmental conditions. Ideally, a PV should be minimally affected by variations in draft, engine setting, sea state, wind, and other external factors. When these influences dominate the PV, they can mask the underlying effect of fouling or introduce additional scatter, making it difficult to draw conclusions without averaging results over long time periods. Longer averaging windows, however, delay the ability to detect fouling and initiate corrective actions. This may affect voyage planning, fuel consumption, and even the ability to safely enter a scheduled port. The impact of operational parameters on commonly used PVs has been demonstrated, for example, by *Farkas et al. (2025).*

For practical purposes, power deviation (P_{dev}) relative to power used for propulsion for clean ship is useful (but not ideal) as PV as it is relatively easily calculated and easily converted to fuel. In addition, it also more directly translates into fouling than, for instance, speed deviation. Power deviation is calculated as:

$$P_{dev} = \frac{P_{meas} - P_{clean}}{P_{clean}},$$

Both P_{meas} and P_{clean} represent calm water powers at same speed and loading condition, i.e. corrected for environmental effects.

4.2. Establishing the baseline performance

Regardless of the specific PV chosen, the core purpose of performance analysis remains the same: to detect changes in the underlying speed–power relationship that can be attributed to the condition of the hull, propeller, or engine. It is hence not the PV itself that is relevant in this context, but rather the change in PV, i.e. the difference between current PV and a baseline. In the context of biofouling detection, the baseline condition should be the clean ship after coating application.

It is important to note that even under ideal post-dry-dock or post-cleaning conditions, operational and environmental factors introduce variability in the measured PV. Establishing a reliable clean-hull baseline therefore requires smoothing out these effects through appropriate averaging. Only with a stable and representative baseline value is it possible to meaningfully detect subsequent deviations caused by biofouling.

When choosing the length of the baseline/reference period, three things should be taken into account: the data quality, the operational profile and the coating. By data quality is meant: the number of (analyzed) data points available for averaging after filtering vs the amount of scatter (and uncertainty) in the data. The data quality will be a result of a combination of the frequency and reliability of the collected raw data, and the performance analysis method, including filtering criteria and applied corrections.

The combination of coating and operational parameters influence the optimal reference period length in the sense that some performance drop may be expected not long after application for some combinations of low-quality coating and/or challenging trade.

Six months will often be enough to establish a reliable baseline. However, a one-year reference period length is recommended to increase accuracy. Biofouling will usually not occur during the first year under normal trading conditions (no extensive idle) and decent antifouling protection.

4.3. How to establish “current” hull performance level

As is the case with the baseline performance, establishing the “current” performance level will require some averaging. Obviously, one is interested in using as short a time period as possible for the averaging in order for the results to be as up to date as possible. On the other hand, a shorter time period will mean higher uncertainty, so these two things need to be balanced to find the optimum time period. The data quality and operational profile play significant roles for this optimization: the poorer the data quality and lower the data frequency is, the longer time is needed to get a reliable average, and similarly, the lower the activity level and larger the operational variation (draft/trim and speed), the longer time period is needed.

4.4. Defining a biofouling KPI

When choosing a KPI to reflect the biofouling growth at any point in time, it has already been established that the comparison of the current performance level with the performance level for clean condition represented by the baseline level is required. However, there is a need to take one additional aspect

into account when assessing the biofouling condition from the averaged PVs. The additional aspect is the underlying performance deterioration over time in between dry dockings, which is caused by effects other than biofouling growth. These are mainly mechanical damages occurring during operation, some of which are permanent ("aging"), like plate dents and loss of shaft efficiency, and others, like large coating damages and scratches ("mechanical damages") can be retrieved in dry dock. Such effects are expected to slowly increase during a dry-docking interval. Since a useful KPI needs to reflect a certain level of biofouling isolated from the effect on performance of other factors, these mechanical damages and aging effects need to be subtracted from the "current" performance level. Figure 1 shows an illustration of the underlying mechanical damages and aging, and demonstrates that there will be a minimum performance loss over the DD period even in the case of a clean hull.

Using Power deviation as our PV, the recommended KPI is hence:

$$KPI = Performance\ level_{last\ 3\ months} - Performance\ level_{reference\ period} - Aging\ and\ mechanical\ damages$$

In Fig.1, the area below the clean hull performance line is illustrated with a grey color indicating that performance is expected not to reach below this line during a dry docking interval, although a 3 months moving average of performance might show performance even below the blue line for short periods of time due to scatter and operational variation. The extent of aging and mechanical damages will be ship specific depending on the operational profile and vessel type. It is challenging to obtain data on the magnitude of the performance degradation over time due to these underlying effects, but some indication is presented in *Gundermann et al. (2016)*. With lack of ship specific values, a general value of 0.8% per year is suggested in order to account for both permanent degradation (aging) and retrievable mechanical damages.

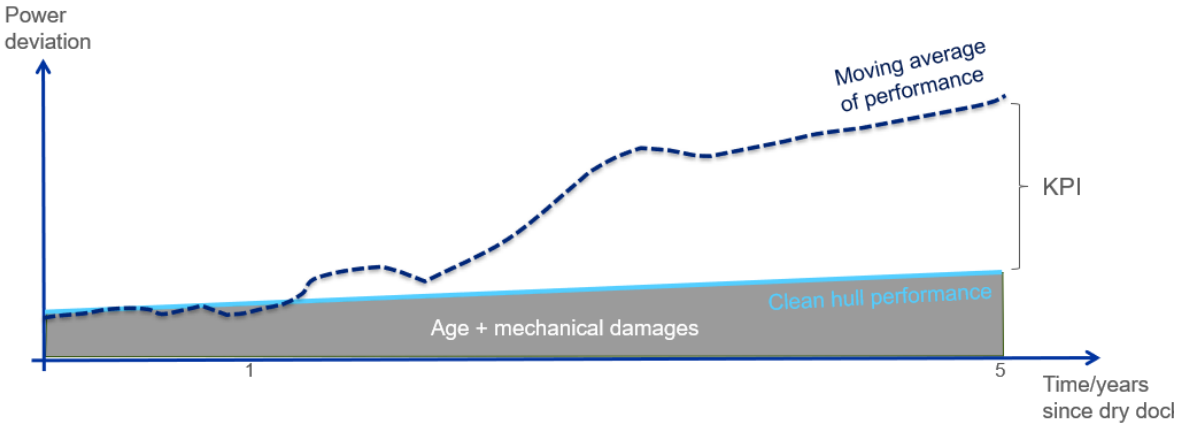


Fig.1: Performance degradation over the DD period. Light blue line represents clean hull performance, i.e. the minimum performance drop that can be expected during a dry-docking period. Grey area represents performance loss due to mechanical damages and aging of the vessel.

4.5. Data type and frequency

In order for the above mentioned KPI to be relevant for biofouling management, a time period of not more than 3 months is recommended for the "current" performance level assessment. This can only be achieved with high frequency data or with very good quality noon data and ideal operational conditions (high activity and low operational variation). A frequency of minimum 1 data point per hour is recommended. Conceptually, a KPI and contingency plan can be set up based on noon data alone, but one should always consider the uncertainty of the data as part of the action trigger level values (see later). Table I provides an overview of recommended data processing settings and options.

Table I: General recommendations for data processing

Variable	Recommended value	Comment
Minimum frequency of data	1/hour	Automatically collected
Length of reference/baseline period	1 year	Down to 6 months for certain ship types and operational profiles
Length of evaluation period ("current" performance level)	3 months	As short as possible but long enough to reduce uncertainty to acceptable level
Performance indicator	$Performance\ level_{last\ 3\ months} - Performance\ level_{reference\ period} - Aging\ and\ mechanical\ damages$	
Aging and mechanical damages	0.8% added power per year	This value may be higher or lower depending on operational profile and vessel type

5. Recommendations for setting up a contingency plan for biofouling management

After establishing a method for calculating relevant performance values and KPI, a contingency plan has to be set up. The contingency plan defines what actions should be taken and when. The following describes best practices for setting up a contingency plan for biofouling management using performance monitoring and outlines the considerations needed for the process.

5.1. Match antifouling coating type and contingency plan in BFMPs

The first step is to match the contingency plan with the antifouling coating or vice versa. Antifouling coatings (AFC) are the key prevention technology against biofouling on ship hulls. But not AFCs are created equally. The AFC prevention mechanism can fundamentally different and their response to in-water cleaning different too. For this reason, it is essential to take into account the AFC technology when setting the contingency plan in BFMPs, or to choose an AFC that matches a contingency action plan that the ship operated can commit to.

The three most common AFC types in the industry today are the following:

- Fouling Release Coatings (FRC): Silicone-based systems rely on a smooth, low-friction surface to shed biofouling. They typically require less frequent IWC and tolerate gentle cleaning methods well. They also have the ability to "self-clean" under the right circumstances, which means that biofouling can be removed by the shear forces of water during sailing. This is a result of the AFC surface properties which make it harder for biofouling to settle on the ship's hull.
- Self-Polishing Copolymers (SPCs): These AFC gradually wear away to expose fresh layers with active substances that can deter biofouling. They are more resilient to mechanical damages than FRCs and are hence less sensitive to IWC methods. However, frequent IWC can accelerate polishing thereby reducing the coating's effective lifetime.
- Hard Coatings (HCs): are coatings that do not have antifouling properties like the previous two categories. Those coatings are very durable but have no active substances to deter biofouling growth. As such, HC rely on very frequent IWC to achieve a clean hull. This means that the antifouling system becomes the combination of HC with very frequent IWC.

The type of technology and the tier of the coating should be considered when designing a biofouling management strategy. The expected lifetime of the coating and the remaining time until the next DD is another important factor to take into account together with the expected operational speed, activity and water temperature. The best biofouling strategy may hence change over time for a given vessel and coating as the above-mentioned factors change.

5.2. Types of actions

There are five actions to consider when setting up a contingency plan:

- Inspections (incl. niche areas)
- Propeller polishing
- In-water cleaning
- Dry-docking
- Sail/monitor

While the first four should be self-explanatory, the fifth requires some elaboration. This action covers situations where there is a slight decrease in performance, but not enough to trigger an inspection or IWC. The action may also cover a situation where some biofouling has been detected, but it is estimated that it may be released during sailing. This can be the case in particular for FRCs which have self-cleaning properties.

5.3. Trigger levels

Once the biofouling KPI is established, a set of thresholds have to be set for when the above-mentioned actions should be initiated. The thresholds mark the values of the KPI for which specific actions should be taken. Again, the relevant values of thresholds to be used in any given case will depend on various things among others trade. Thus, if the vessel will enter areas with biofouling management requirements such as Australia and New Zealand, the thresholds should be set correspondingly to minimize biofouling risk. Vessel type will also play a smaller role as biofouling will affect performance in different ways for different vessel types as a result of the portion of frictional resistance in the total resistance. A general set of recommendations for setting up trigger levels are presented in Fig.2 in the situation where power deviation or power index is used as performance value.

5.4. Contingency plan

The following provides recommendations for a contingency plan based on continuous performance monitoring using previously defined KPIs and trigger levels. Note that ships may set acceptable levels and thresholds differently in the BFMPs depending on the trade patterns and risk tolerance of the ship operator. The plan requires continuous performance monitoring and calculation of KPI, and the system should be set up to provide alerts when KPI reaches the predefined settings. Alternatively, a manual check of KPI level should be done on regular (not less than monthly) basis. Fig.3 shows a flow diagram of the process: performance monitoring (blue box) and KPI calculation (grey box) is done on continuous/regular basis, and KPI value against trigger levels, Fig.2, are checked. Based on the value of value of KPI, an action recommendation is triggered (green, yellow and orange boxes). The colors and labels of the boxes in Fig.3 corresponds to the colors and labels in Fig.2. For the example in Figs.2 and 3, we assume an FRC system is installed, but the framework can be used on SPC type coatings s well (although self cleaning step only applies to FRC coating). For hard coatings, a more scheduled cleaning approach is recommended.

The green area in Fig.2 represents performance which can be considered “good”, i.e. the level of fouling is at an acceptable level as set by the ship operator in the BFMP. This means that no action is needed as long as moving average of three months is in this range. When added power reaches above 4% on top of the clean hull performance line (blue), it is recommended to monitor more closely the performance trend. In Fig.2, this is represented by the yellow area with the “Monitor” label. The performance level may still improve over time due to self-cleaning properties in case of FR coatings. It is also possible that the decrease in performance is caused by uncertainties related to data or corrections, and in this case, it is also worthwhile to monitor before initiating an inspection or IWC.

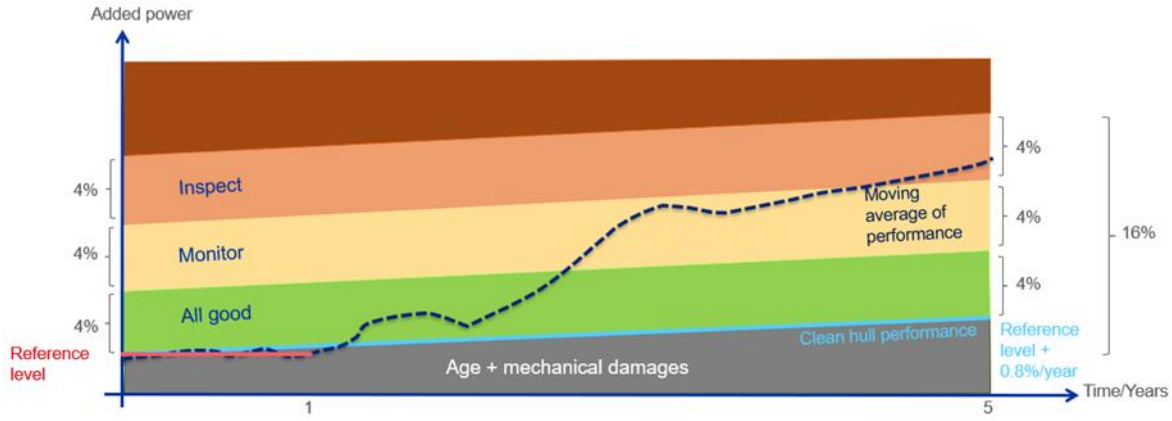


Fig.2: General recommendations for trigger levels when using power deviation as Performance Value. Dark blue line illustrates a 3-month moving average of power deviation. Light blue line is the previously defined clean hull performance line, and red line represents the first-year baseline performance level. The colors and labels indicate the recommended action which should be carried out when dark blue line crosses enter the areas.

However, if the added power reaches above 8% on top of the clean hull performance curve (orange area), this has to trigger action: inspection and/or propeller polishing. As stated in (*IMO, 2023*[a]), a consumption increase by >9% may indicate higher biofouling risk, i.e. presence of light biofouling. It is important to note that if not coated, then propeller has to be polished at least twice per year, ideally every six months. In a case where "inspect" trigger is reached and propeller is not polished in the last four months it is recommended to do propeller polishing along with hull inspection. It is always recommended to do a niche area inspection and cleaning (if needed) in combination with both propeller polishing and hull inspection, since biofouling in niche area will not be detected with performance monitoring.

If performance level has been in "Monitor" category for an extended time period, it can also be considered if an inspection should be carried out, or as above, if propeller polishing and hull inspection should be carried out. This is indicated by the dashed line in Fig.3.

Based on the visual inspection results, a decision must be made regarding whether hull cleaning should be carried out. This assessment is not always straightforward, as it can be influenced by a range of financial and operational factors. As the aim of this paper is to evaluate the potential of performance monitoring as a tool for detecting biofouling and to propose best practices for integrating performance monitoring into Biofouling Management Plans (BFMPs), the guidelines from IMO *(2023a)* are followed regarding the categorization of fouling based on visual inspections and (to some degree) recommended actions. Table II provides an overview of the *IMO (2023a)* rating scale.

Depending on the action plan written in BFMP, cleaning should be carried out according to the fouling rating in Table II. In a case where a HC with no antifouling properties is applied, cleaning should be already planned once "inspect" trigger is reached. On the other hand, in a case of either SPC or FR (silicone) technology, cleaning should be planned in the next possible opportunity provided that the presence of macrofouling on hull is confirmed (Fouling rating 2). The main aim of the contingency plan, as described in BFMP, is to keep fouling rating ≤ 1.

From performance point of view it is important to note that within BFMP fouling rating concept, there is no real distinguishment between animal and plant macrofouling, i.e. macrofouling is defined as large, distinct multicellular individual or colonial organisms visible to the human eye such as barnacles, tubeworms, mussels, fronds/filaments of algae, bryozoans, sea squirts and other large attached, encrusting or mobile organisms, *IMO (2023a)*. From ship resistance point of view however, it is quite different whether ship fouling is weed or small calcareous fouling, or barnacle fouling. In the famous paper by

Schultz (2007), equivalent sand grain roughness for weed or small calcareous fouling is 1000, while for heavy calcareous fouling (such as barnacles) it is 10000. Such differences in equivalent sand grain roughness corresponds to significant differences in ship performance.

Table II: Rating scale to assess the extent of fouling on inspection areas, *IMO (2023a)*

Rating	Description	Macrofouling cover of area inspected (visual estimate)
0	**No fouling** Surface entirely clean. No visible biofouling of surfaces	-
1	**Microfouling** Submerged areas partially or entirely covered in microfouling. Metal and painted surface may be visible beneath the fouling.	-
2	**Light macrofouling** Submerged areas partially or entirely covered in microfouling. Metal and painted surface may be visible beneath the fouling.	1 – 15% of surface
3	**Medium macrofouling** Presence of microfouling and multiple macrofouling patches.	16 – 40% of surface
4	**Heavy macrofouling** Large patches or submerged areas entirely covered in macrofouling.	41 – 100% of surface

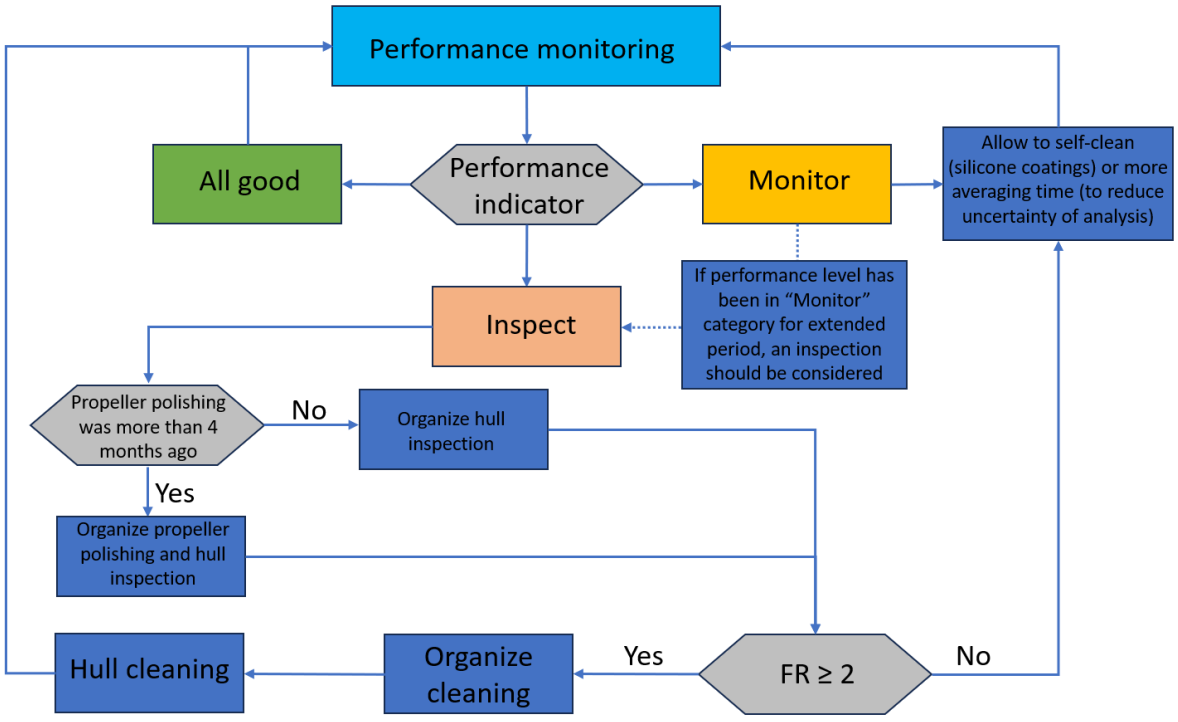


Fig.3: Flow chart of the proposed contingency plan framework

When an inspection indicates that plant fouling or small calcareous fouling covers less than 10% of the underwater surface, the impact on ship performance is generally expected to be limited, similar to vessels affected only by microfouling. In such cases, it can be concluded that the fouling itself is unlikely to be the primary cause of a reduction in performance, and that other factors, such as analytical uncertainty, may be more significant contributors.

The decision on whether to proceed with cleaning in this situation should be guided by both risk-based and financial considerations. The assessment should account for the potential for coating damage (as

both self-polishing and silicone coatings can be affected by cleaning), the level of uncertainty in the data and analysis, expected trading regions and related regulatory requirements, as well as the cost of cleaning and associated downtime relative to potential performance gains.

6. Discussion and conclusions

This paper demonstrates that performance monitoring offers significant value as a practical and informative method for detecting biofouling. It presents a structured approach to incorporate performance-based indicators and action triggers into BFMPs. Such approach strengthens adherence to IMO guidance, supports more efficient vessel operations, and ultimately contributes to wider maritime objectives, including enhanced energy efficiency and reduced emissions to air and water.

However, ships cannot rely solely on performance monitoring for biofouling detection. First of all, fouling in niche areas will for most cases not give rise to performance decreases and will hence not be detected via performance monitoring. A separate strategy is needed for niche area protection, one that involves more scheduled inspections. The separation between hull and propeller performance without thrust measurements is still quite difficult, which means that propeller fouling should always be considered a possible influence in results, and the best approach is to conduct regular propeller polishing if reliable thrust data are not available. Combining these propeller polishings with niche area inspections and cleanings if necessary is a practical way to minimize risk of fouling in those areas.

A persistent challenge related to fouling assessment from performance monitoring results is related to fouling distribution. As is discussed in detail in *Farkas et al. (2025)*, most of the typical performance values used in performance monitoring today are prone to loading condition dependency. In addition, fouling is almost always heterogeneously distributed on the hull. This will result in differences in measured KPIs for different loading conditions which might blur the picture and delay the trigger for action. The presented framework can be further improved by including trade parameters from AIS data as well as environmental data in a risk assessment. Data from previous underwater inspections can also be included in the assessment. Thus, a data-fusion framework that links in-water inspections, AIS-derived fouling exposure/idle risk, and performance-based indicators into a unified trigger and decision model for scheduled and contingency inspections, cleaning approvals, and compliance documentation would represent an ideal future framework. Such exploration represents an exciting area for further research that can inform industry and policy stakeholders.

References

BU (2012), *Final Report Summary - CLEANSHIP (Prevention and detection of fouling on ship hulls)*, Brunel Univ., London, Prevention and detection of fouling on ship hulls | FP7 | CORDIS | European Commission

DAVIDSON, I.C.; McCANN, L.D.; SYTSMA, M.D.; RUIZ, G.M. (2008), *Interrupting a multispecies bioinvasion vector: The efficacy of in-water cleaning for removing biofouling on obsolete vessels*, Marine Pollution Bulletin 56, pp.1538–1544

DAVIDSON, I.C.; BROWN, C.W.; SYTSMA, M.D.; RUIZ, G.M. (2009), *The role of containerships as transfer mechanisms of marine biofouling species*, Biofouling 25(7), pp.645-655

FARKAS, A.; SONG, S.; DEGIULI, N.; MARTIĆ, I.; DEMIREL, Y.K. (2020), *Impact of biofilm on the ship propulsion characteristics and the speed reduction*, Ocean Eng. 199

FARKAS, A.; GUNDERMANN, D.; MEHRI, S. (2025), *On the speed and loading condition dependency of the most common performance values*, 10th HullPIC, Mulheim, pp.19-33. Mulheim

GUNDERMANN, D.; DIRKSEN, T. (2016), *A statistical Study of Propulsion Performance of Ships and the Effect of Dry Dockings, Hull Cleanings and Propeller Polishes on Performance*, 1st HullPIC Conf., Pavone, pp. 282-291

IMO (2001), *International Convention on the Control of Harmful Anti-fouling Systems on Ships*, Int. Mar. Org., London

IMO (2004), *International Convention for the Control and Management of Ships' Ballast Water and Sediments (BWM)*, Int. Mar. Org., London

IMO (2012), *Guidance for minimizing the transfer of invasive aquatic species as biofouling (hull fouling) for recreational craft*, MEPC.1/Circ.792, Int. Mar. Org., London

IMO (2013), *Guidance for evaluating the 2011 guidelines for the control and management of ships' biofouling to minimize the transfer of invasive aquatic species*, MEPC.1/Circ.811, Int. Mar. Org., London

IMO (2023a), *2023 Guidelines for the control and management of ships' biofouling to minimize the transfer of invasive aquatic species*, Res. MEPC.378(80), Inr. Mar. Org., London

IMO (2023b), *Common Hull Fouling Invasive Species*, MEPC, Int. Mar. Org. London, Common Hull Fouling Invasive Species

IMO (2023c), *Marine Environment Protection Committee (MEPC 80)*, Int. Mar. Org., London, Marine Environment Protection Committee (MEPC 80), 3-7 July 2023

IMO (2024a), *2024 Guidelines for the development of a ship energy efficiency management plan (SEEMP)*, MEPC.395(82), Int. Mar. Org., London

IMO (2024b), *2024 Revised guidelines for the reduction of underwater radiated noise from shipping to address adverse impacts on marine life*, MEPC.1/Circ.906/Rev.1, Int. Mar. Org., London

IMO (2025), *Guidance on in-water cleaning of ships' biofouling*, MEPC.1/Circ.918, Int. Mar. Org., London

ISO (2016), *ISO 19030 - Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

KIM, B.C.; KIM, H.C.; HAN, S.; PARK, D.K. (2022), *Inspection of underwater hull surface condition using the soft voting ensemble of the transfer-learned models*, Sensors 22(12)

MALONE, J.A.; LITTLE, D.E.; ALLMAN, M. (1981), *Effects of hull foulants and cleaning/coating practices on ship performance*, Trans. SNAME 88, pp.75-101

NURIOGLU, A.G.; ESTEVES, A.C.C. (2015), *Non-toxic, non-biocide-release antifouling coatings based on molecular structure design for marine applications,* J. Materials Chemistry B, 3(32), pp. 6547-6570

PARK, J.; KIM, B.; SHIM, H.; AHN, K.; PARK, J.H.; JEONG, D.; JEONG, S. (2018), *Hull and propeller fouling decomposition and its prediction based on machine learning approach*, 3rd HullPIC Conf., Redworth, pp.20–26

SCHULTZ, M.P. (2007), *Effects of coating roughness and biofouling on ship resistance and powering*, Biofouling 23(5), pp.331-341

SUBSEA TECH SAS (2022), *Spectral Imaging Powered Ship Hull Biofouling Detection and Cleaning*, Spectral Imaging Powered Ship Hull Biofouling Detection and Cleaning | H2020 | CORDIS | European Commission

TEZDOGAN, T.; DEMIREL, Y.K. (2014), *An overview of marine corrosion protection with a focus on cathodic protection and coatings*, Brodogradnja 65(2), pp.49-59

UZUN, D.; DEMIREL, Y.K.; CORADDU, A.; TURAN, O. (2019), *Time-dependent biofouling growth model for predicting the effects of biofouling on ship resistance and powering*, Ocean Eng. 191

VAN BALLEGOOIJEN, E., HELSLOOT, T. (2019), *An approach to monitor the propeller separately from the hull*, 4th HullPIC Conf,, Gubbio, pp. 50–55

WOODS, C.M.; FLOERL, O.; JONES, L. (2012), *Biosecurity risks associated with in-water and shore-based marine vessel hull cleaning operations*, Marine pollution bulletin 64(7), pp.1392-1401

WU, D.; HUA, J.; CHUANG, S.Y.; LI, J. (2023), *Preventative biofouling monitoring technique for sustainable shipping*, Sustainability 15(7)

Navigating from Fluffy Data Clouds to Hard Data in the Cloud

Rune Freyer, Shipshave, Stavanger/Norway, rune.freyer@shipshave.no
Masaki Katafuchi, Shipshave, Stavanger/Norway

Abstract

Hull performance is influenced by hull roughness. Hull paint roughness is traditionally only measured in dry dock. And real hydraulic roughness is never measured. A sensor and software directly measure the hydraulic properties of paint and fouling during transit. By using big data and computer analysis it can answer questions like: Is premium paint worth the added cost in the real world? How do I optimize the cleaning operation? How frequent shall I groom? Shall we spot blast the hull in DD?

1. Introduction

Hull grooming is a proactive maintenance strategy to reduce hull resistance and maintain vessel efficiency. With in-transit grooming, the hull is cleaned while the vessel is underway. The primary motivations for this include continuous cleaning without disrupting transport schedules, eliminating reliance on third-party suppliers and port authorities, and ensuring non-damage to the paint.

Traditionally, hull efficiency has been estimated using indirect methods such as fuel efficiency analysis, video documentation of cleaning operations or inspections with remotely operated vehicles. However, these approaches have limitations, including low measurement accuracy, inconsistent data collection, insufficient sample sizes for robust statistical analysis, and high operational costs. In contrast, modern industrial processes routinely utilize high-precision measurements and large datasets to optimize performance. Given its impact on fuel consumption and environmental performance, hull roughness deserves advanced measurement methodologies.

Recent advancements in In Transit Cleaning of Hulls have introduced new possibilities to record high-resolution videos during grooming operations. A new system that measures hull roughness while simultaneously recording videos and performing grooming is now in its semi-commercial stage. It will be coupled with automated data routing and automated analysis.

2. In Transit hull cleaning and monitoring

In transit hull cleaning was introduced in 2020 and now operates at ship speeds between 10 and 17 kn. The aim is to always keep a clean hull rather than to wait for degrading hull performance. Recent real life accelerated testing indicates no damage or wear to the antifouling even after a full 5-year cycle worth of swipes. We also minimize the time taken to operate and manage the equipment to allow frequent cleaning. Later developments include:

- Increased hydraulic stability of robots leading to increased vessel speed window.
- Improved cleaning efficiency
- Data management with automated data transfer, reporting, analysis and dashboard
- Integrating video, sensor output and analysis in cloud application

The tools are carried by the ship and operated by the crews. As it is operated offshore, cleaning does not require port permits. A cleaning operation can be performed within hours, but time depends on ship size and fouling level. The system is primarily intended for slime, but further development allows removal of heavier fouling, including barnacles. Removal of heavy slime was exemplified in a study by DNV on two large container vessels that showed fuel savings of 5 and 16%, *Hollenbach (2024)*. ITCH is used on container, bulk, tank and cruise vessels from 128 m to 400 m length and with fouling release, self-polishing paints and even a vessel without antifouling paint.

A survey conducted at HullPIC 2019, *Schmode et al. (2019),* established that inadequate measurement and analysis was the biggest challenge in hull performance. The ITCH system has been independently proven with indirect analysis methods (traditional fuel efficiency derived analysis), but such analysis can be expensive and time consuming. Gaining confidence in quality analysis can take years from the first cleaning till the analysis is ready. With FuelEU efficiency penalties being introduced, *EU (2023),* and others in the pipeline, hull performance is becoming a competitive differentiator.

Qualitative evaluation of ITCH has been done by watching fouling removal videos captured during cleaning operations. At daytime offshore, the water is clear and video quality good. The fouling plumes are immediately carried away and never obstruct the view. The video is combined with areal coverage plots to evaluate efficiency. New analysis software allows fast work processes.

One example of monitoring of hull performance during transit is Air Lubrication Systems (ALS). Understanding ALS has centred around Computational Fluid Dynamics (CFD). However, as ALS enters real life, idealized conditions cannot predict performance. Trim, swell, speed, wind and other parameters affect performance. Being able to document in what conditions a hull will benefit from ALS will determine the ship owner's investment decisions.

A ship owner used the ITCH to investigate the departure of air from under the hull. The air was found to escape near the bow, indicating no air lubrication along the flat bottom at those working conditions.

A chemical tanker was cleaned by ITCH over 17 months before DD at 22 months from previous DD. The vessel was cleaned every 2 months in the period. As a test of paint wear by ITCH, the ships operator was cycling two sections of the hull with an additional number of swipes equal to 1 and 3 years of ITCH cleaning operation. This was to simulate a full 5 year docking cycle worth of cleaning. The areas cleaned were investigated by Safinah, an independent paint consultant and showed no indication surface wear due to cleaning through inspection and chip samples.

3. Reasons

3.1. Why do we need hard data in the cloud

Light, frequent cleaning improves fuel efficiency of vessels versus infrequent or absent hull cleanings. Grooming reduces fouling roughness and makes less paint damage than reactive cleaning. A concern by shipowners with hull grooming is follow-up. Most grooming requires third-party involvement, port permits, planning, cleaning specialists and ship crew involvement. Assuming grooming 4-6 times a year requires the performance engineer to keep track of the vessels and follow them up to make sure the operations are taking place. Automating the data flow is therefore a key to automating consistency and regularity of the system.

3.2. Why is roughness measurement important?

Loss of fuel efficiency is a symptom of increasing hull roughness and is the most common hull efficiency monitoring. However, measurement of roughness is hard data to be actioned by cleaning the hull. Hydraulic surface roughness is a part of fluid viscous flow equations and allows calculation of the losses. In shipping we try to derive the roughness from fuel consumption. But waves, current, wind, engine condition, water temperature, propeller, trim and loading play in. Calculations may assume that fouling and paint wear is equally distributed between boot top and bilge, but that is untrue. The division into paint roughness and hydraulic roughness is also important because it separates effects that can be affected by hull cleaning and effects needing sand blasting of the hull.

3.3. Why measure roughness in the water and during transit?

Fouling is organic matter, but is made "fluffy" by water. Once fouling gets in dry dock, the volume is lost and the hull performance information of a roughness measurement is lost. Roughness

measurement in water in port means that the fouling will not lay flat onto the hull. This is most important for fibrous fouling. In-transit hydrodynamic roughness is fuel efficiency relevant data.

Biofouling documentation will be increasingly stringent with demands on frequent monitoring for port access. To maintain revenue generation, inspection during transit generates more profits than during standstill.

3.4. Why do we need "big data"?

Traditional roughness measurements of paints are made on hulls with handheld sensors in dry dock. These measurements may be taken with tens or max hundreds of measurements. A roughness measurement of paint will tell about the paint surface, but does not tell about the paints fouling repelling properties. Good roughness estimates require tens of thousands of randomized measurements. Different parts of hulls have different working conditions, and breakdown over different parts of the hulls is needed to understand how well the paint works. The new method involves measuring the surface before and after cleaning. Traditional paint roughness measurements have a one-off and insufficient for diagnosis of a fleet.

3.5. Why do we need <u>really</u> big data?

A bad paint design may have million USD consequences in fuel cost and maintenance. A full roughness plot of a hull is still just data prone to random effects. Root cause analysis of underperformance needs a population of ships, time series and detailed breakdowns like:

- Hydraulic roughness and paint roughness separated. Hydraulic roughness determines hull efficiency while paint roughness is an expression of the paint surface.
- Roughness located
    - Draft location. Is the fouling worse on shallow or deep draft?
    - Longitudinal position. Are areas in bow cleaner than aft?
    - Berthing areas vs just flow exposed areas. Are paints resilient from berthing?
- Paint roughness before and after cleaning. Does cleaning damage paints?
- Hydraulic roughness before and after cleaning. Is cleaning frequency, cleaning intensity or brush selection adequate to keep the hull clean?
- Time series. Which paint delays fouling most?
- A fleet should be analysed with statistical significance on a fleet level, not one-off vessels.
- Manage the fleet with updated information online.

3.6. Why does it matter where the fouling is on a hull?

Selection of paint is inaccurate science with cost impact for dock works and fuel consumption. Hulls in the same trade may be painted with the same paints but experience different results. But this may be caused by biology, not the paint. One ship painted with different paints may undergo nearly the same biological conditions. Painting a ship with 4 different paints and locating and measuring the resulting roughness in different depths in a 2-month cycle is possible. One may also analyze where the roughness is. How far up on the bottom do you see fouling? This can determine if different paints can be used in different parts of the hull. Hull roughness maps can be used as a part of the biofouling documentation towards port authorities and relieve the crew from administrative burdens.

3.7. Why can paint roughness be of value be for planning paint work?

The most important hull performance decisions are the selections of paint. Some paint suppliers for fouling release coatings (FRC) claim fuel efficiency benefits with less paint roughness degradation than others. Others argue that these FRC paints are fragile. Roughness measurement on the hull can develop hard numbers on how this acts on my trade pattern and my ship type. Can big data back it up?

And can you quantify the benefits of a premium paint beyond a premium cost? If you evaluate the paints during a docking cycle, all paint selection decisions can be accelerated. Knowing the condition of the paint on before a dry dock, you can decide on blasting the hull or just spot blasting.

3.8. How is roughness measured?

The ITCH-Performance robot measures hull roughness using optical methods. By applying image processing techniques, the ITCH-P robot can measure both hydrodynamic surface roughness and hard-surface roughness. This method provides high-resolution, non-contact measurements that are effective even in underwater environments. While the hard-surface roughness measurements have been tested and validated, the hydrodynamic surface roughness measurements require further verification and validation to ensure their accuracy and reliability.

4. Case study

4.1. Single Deployment Analysis

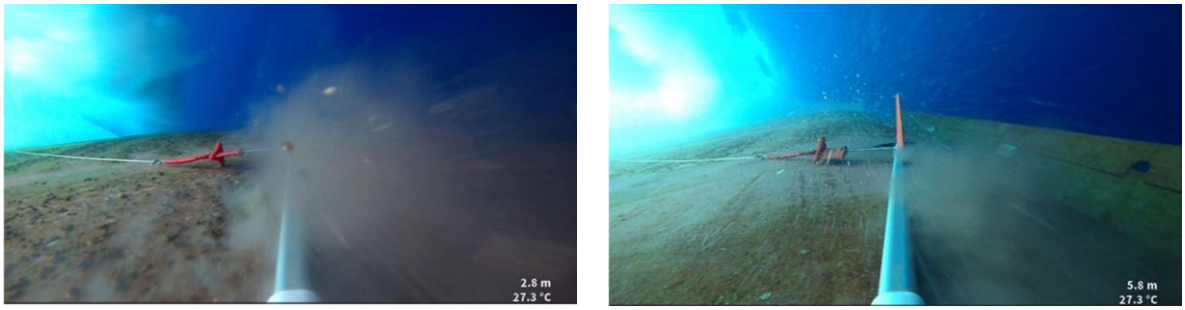


Fig.1: ITCH operates and removes fouling underwater. (Left) At boot top area capturing various fouling such as heavy slime and barnacles as seen in dislodged materials; (Right) Near bilge keel where less fouling can be observed.

In this section, we focus on a single deployment of the ITCH-Performance (ITCH-P) robot to evaluate the changes in hull roughness before and after a cleaning operation. On the starboard side of a Oil/Chemical Tanker with a length of 180 m, cleaning was conducted several times using another ITCH system prior to the deployment of ITCH-P. The ITCH-P robot was deployed after these cleaning operations to assess residual roughness. During this deployment, the ITCH-P recorded 121,079 data points.

ITCH-P does not perform data processing while underwater. Instead, video files captured during deployments are processed post-operation using a designated application once ITCH-P is connected to a laptop. This post-processing enables users to obtain detailed roughness measurements.

To evaluate the changes in hull roughness resulting from the cleaning operation, measurements from the ITCH-P deployment were analyzed. Each ITCH-P deployment included a bow-to-aft movement followed by an aft-to-bow movement. Roughness values recorded during the bow-to-aft movement represented the hull condition before cleaning, while values from the aft-to-bow movement reflected the post-cleaning condition. Depth sensor readings and the frequencies of the winch spooling the rope connected to ITCH-P were employed to estimate the robot's position throughout both movements. However, the estimation of the horizontal position is yet insufficiently accurate to allocate repeatedly to the resolution of the graphics. Specifically, data such as the spooled rope length, rope tension, and their timestamps were lacking, which constrained the accuracy of the positional estimates. The estimated path of the ITCH-P robot is presented in Fig.2.

The processed roughness data was synchronized with depth sensor readings using time stamps. Figs.3 and 4 illustrate differences in hull roughness, before and after the cleaning operation, represented by Rt_{25}, mapped to corresponding sections of the hull, each measuring 10 m in length and 1 m in depth.

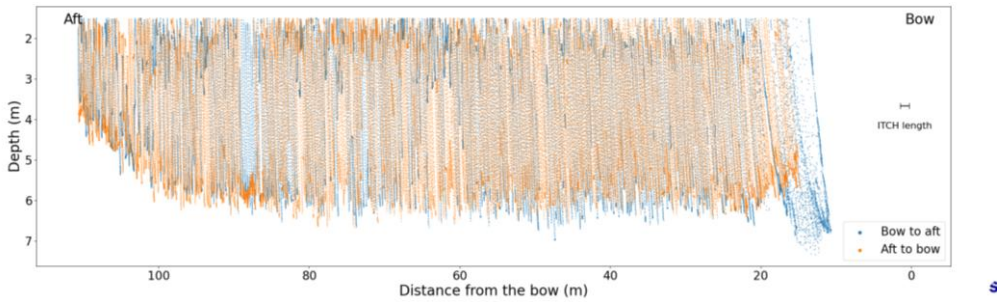


Fig.2: Paths of the ITCH-Performance robot during the cleaning operations on the starboard side. The blue path represents the bow-to-aft movement, while the orange path indicates the aft-to-bow movement during the deployment. Depth readings and winch speed were used to track the robot's position.

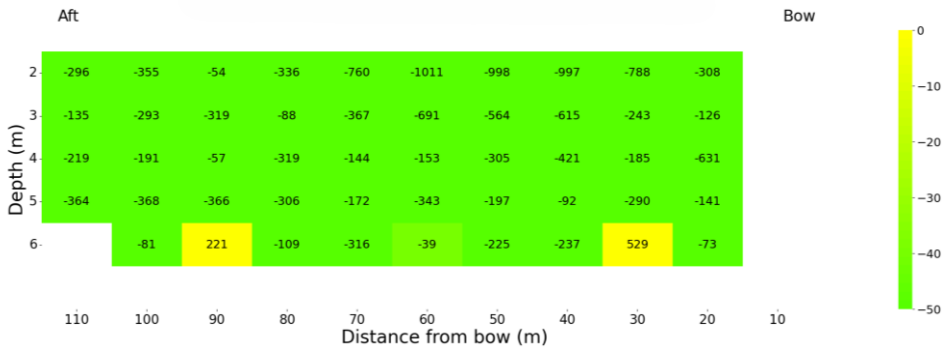


Fig.3: Difference in hydrodynamic roughness ($\mathbf{Rt_{25}}$) mapped to hull sections after the cleaning operation. The color gradient represents the reduction in roughness, indicating effective removal of transparent biofouling.

Fig.4 provides a comparative analysis of average hydrodynamic and hard-surface roughness before and after the cleaning operation, illustrating the reduction in hydrodynamic roughness achieved, while hard-surface roughness remains largely unchanged.

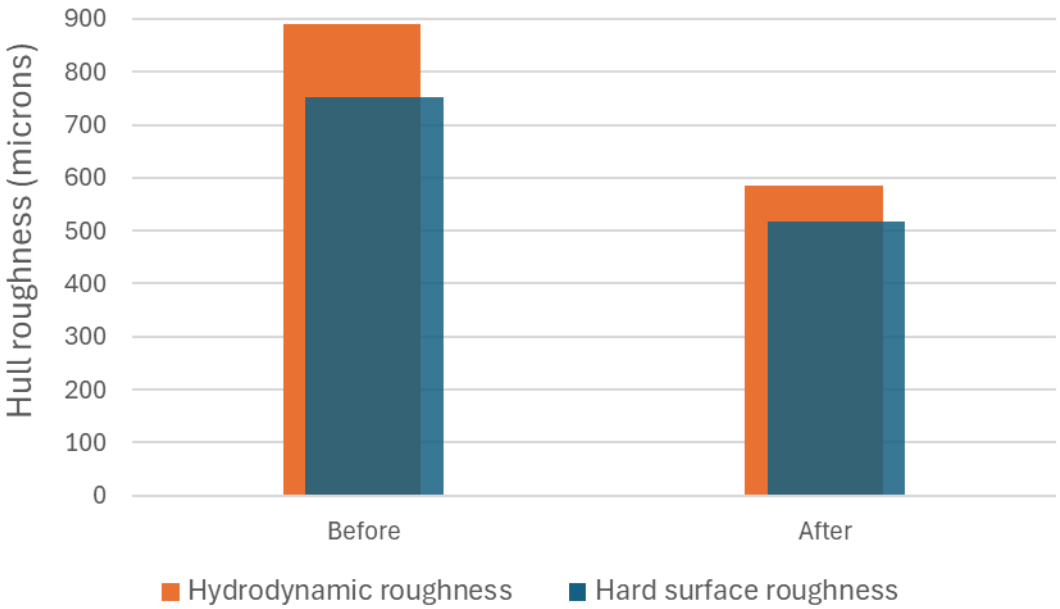


Fig.4: Comparison of average hydrodynamic and hard-surface roughness before and after the cleaning operation. The stacked bar chart shows reduction in hydrodynamic roughness, while hard-surface roughness remains largely unchanged. The hard surface roughness is reduced, signifying barnacle removal.

5. Conclusions

This paper has presented the deployment and performance of the ITCH-Performance robot in measuring hull roughness through a case study. The analysis demonstrated the system's capability to deliver reliable roughness measurements across different hull surfaces, contributing valuable insights for hull maintenance and performance optimization.

The paper proposes paint evaluation as important. Paints are much larger investments than cleanings and lasts for 5 years, but a single cleaning can destroy a paint. Paint selection will benefit of increased quantification to enter in an AI assisted evaluation scenario

The case study highlighted the effectiveness of the ITCH-Performance system in processing the collected data and extracting key surface parameters such as Rt for the paint. Additionally, the ITCH-Performance robot can measure significantly more points than traditional hull roughness measurement methods, providing more comprehensive surface characterization. The hydrodynamic roughness measurement (fouling roughness) will however never be as accurate as the hard surface.

Future work on the will include refining the measurement algorithms, particularly for hydrodynamic roughness. It will also focus on improving the position estimation of the robot during cleaning operations to make roughness comparisons more accurate and consistent. The fully commercial version, increased use, automated data management and further analysis will enable a technology that is reliable and accurate, not only for paint, but also fouling. Cloud integration with fleet management and big data statistics will also be a key in future releases.

References

EU (2023), *FuelEU Maritime Regulation*, European Commission, https://transport.ec.europa.eu/transport-modes/maritime/decarbonising-maritime-transport-fueleu-maritime/questions-and-answers-regulation-eu-20231805-use-renewable-and-low-carbon-fuels-maritime-transport_en

HOLLENBACH, U. (2024), *Performance Assessment: Assessing the performance of Shipshave ITCH system at Hapag Lloyd container vessels BARZAN and CHICAGO EXPRESS*, DNV

MIHAYLOVA, R. (2026), *ITCH Impact Evaluation*, Report, ShipShave

SCHMODE, D.; HYMPENDAHL, O.; BELLUSCI, F. (2019), *Influence of Data Sources on Hull Performance Prediction*, 4th HullPIC Conf., Gubbio, pp.162-171

A Methodology for Evaluating and Revising Roughness Allowance Formulation for Modern Hull Coatings

Hamed Vaseghnia, Jotun, Sandefjord/Norway, hamed.vaseghnia@jotun.no
Irma Yeginbayeva, Jotun, Sandefjord/Norway, irma.yeginbayeva@jotun.no
Angelika Brink, Jotun, Sandefjord/Norway, angelika.brink@jotun.no

Abstract

This paper focuses on the challenges of using traditional predictive modeling frameworks for drag performance predictions for contemporary coatings. Contemporary coatings use advanced polymers and precise application methods, producing smoother, more uniform surfaces revealing significant differences in performance than traditional antifouling coatings. Industry-standard empirical formulas such as the ITTC-1957 correlation line, Schoenherr's formula, and Townsin's roughness allowance are based on older datasets, simplified flat-plate assumptions using older coating technologies. Consequently, they often represent the performance of today's drag-reducing surfaces and modern hull geometries with lower accuracy. In this work, we outline a methodology for considering and evaluating effects of modern coatings on overall ship performance. The aim is to reflect roughness effects observed in coatings and how these qualities influence hydrodynamic behavior of ship. The approach is compared with commonly used empirical estimations, and the discussion highlights the variations that may appear under different hydrodynamic assumptions, suggesting that certain conditions may need revising of how existing empirical formulations are interpreted.

1. Introduction

Frictional resistance is a major component with up to 80% for large marine vessels of the total hydrodynamic and is therefore a crucial factor for performance assessment, *Sanz et al. (2026)*. Because hull resistance directly determines the propulsion power required to maintain a given speed, its prediction aids the management of fuel consumption and greenhouse gas emissions, *Schultz (2014)*. This motivates the need for detailed friction analysis and the development of effective friction-reduction strategies. Among the most effective strategies for reducing frictional losses is the application of advanced hull treatments and marine coatings that aim to minimize surface roughness and biofouling on the hull surface, *Zhang (2024)*. The characteristics of antifouling coatings directly and indirectly influence the overall hydrodynamic resistance by modifying the turbulent boundary layer developing along the ship hull, *Yeginbayeva (2018)*. Accordingly, analysis and prediction of the hydrodynamic effects of surface roughness have mostly relied on empirical roughness allowance formulations and friction correlation lines. Among these approaches, the scaling of similarity law proposed by *Granville (1978)* is a widely adopted method to estimate the impact of surface roughness on ship resistance and has been applied over several decades. However, despite the complex material and surface–flow interactions associated with modern antifouling systems, their hydrodynamic impact in empirical formulations is commonly represented through simplified descriptions. Firstly, such empirical formulations and upscaling approaches are commonly based on experimental datasets and theoretical assumptions that reflect older coating technologies. In contrast, modern hull coatings employ more advanced polymer chemistry, controlled rheology, and precision application techniques, resulting in smoother and more uniform surfaces with fundamentally different roughness characteristics, *Brady (2000)*. Secondly, empirical roughness allowance formulations are derived from flat-plate experiments and zero-pressure-gradient turbulent boundary layer assumptions. While these assumptions are appropriate for canonical boundary layer flows, the hydrodynamic environment around a ship hull is more complex, involving three-dimensional geometry, surface curvature, free-surface effects, and spatially varying pressure gradients. As a result, uncertainties arise when transferring roughness measurements obtained under flat-plate conditions to full-scale ship applications. A further simplifying assumption in many roughness models is the uniform and constant distribution of the roughness Reynolds number k^+, and the associated roughness function, ΔU^+, along the surface. In practice, both parameters vary spatially over a ship hull due to local changes in flow conditions, wall shear stress, and

pressure gradients, which are showing the flat-plate assumption an approximation rather than a representation of real ship flows, *Atlar (2020)*.

Nevertheless, the flat-plate-based similarity scaling laws are widely used in industry for extrapolating laboratory-scale roughness measurements to full-scale ship performance. In particular, the Granville upscaling procedure is consistent with the methodology adopted within the RightShip performance assessment framework. In this procedure, the full-scale frictional resistance increment derived from roughness measurements is incorporated into ship speed–power predictions to quantify efficiency degradation associated with hull condition. The resulting performance estimates contribute to the vessel's efficiency rating, which is used in commercial and environmental evaluations and may influence decisions and acceptance criteria.

On the other hand, alongside with empirical approaches, significant efforts have been done for modelling surface roughness effects using Computational Fluid Dynamics (CFD), *Schultz (2014), Martic (2018), Song (2019)*. A key advantage of CFD based approaches is their ability to overcome the limitations of traditional flat-plate roughness models. In CFD simulations, three-dimensional flow effects can be explicitly resolved, and scale effects can be eliminated when simulations are performed at full scale, *Atlar (2020)*. Moreover, the local friction velocity is computed dynamically at each computational cell, which allows the roughness Reynolds number and the associated roughness function to vary spatially and temporally throughout the flow field. Consequently, recent studies have employed CFD to investigate the influence of surface roughness on ship resistance, propeller performance, and self-propulsion characteristics. These studies have demonstrated that hull roughness can affect not only frictional resistance but also other resistance components, including viscous pressure resistance and wave-making resistance, *Murphy (2019)*. Given that ship hydrodynamic behaviour depends strongly on vessel type, scale, and operating conditions, the impact of surface roughness is also expected to vary across different ship forms. Recent comparative studies have already reported distinct roughness effects between different ship types, which highlight the variations in form factors and wake characteristics, *Demirel (2020), Van (2011), Min (2010)*. In this context, a systematic investigation of roughness effects across different ship geometries, scales, and speeds is necessary.

The present work proposes a methodology for revising empirical formulations of hull resistance calculations combined with modern hull coatings effects. The approach combines Jotun's in-house flow-cell roughness measurements with CFD simulations to assess the influence of coating roughness on ship-scale frictional resistance. Hybrid roughness function representations are introduced to capture non-linear roughness behaviour, and their impact on full-scale friction predictions is compared against traditional empirical formulations, including ITTC-1957 and Granville-type extrapolation methods. By systematically comparing empirical roughness allowance formulations with CFD-based predictions informed by modern coating data, this study aims to clarify the limitations of classical models and to provide a practical pathway for improved resistance prediction of coated hulls.

2. Background

2.1. Roughness function formulation based on flow-cell experiments

Surface roughness enhances near-wall turbulence and leads to an increase in skin-friction drag. Its effect is commonly represented as a downward shift of the mean velocity profile in the turbulent boundary layer, quantified by the roughness function, ΔU^+ , *Granville (1987).* For rough-wall conditions, the non-dimensional mean velocity profile can be expressed as:

$$U^+ = \frac{1}{\kappa}\ln y^+ + B - \Delta U^+, \tag{1}$$

$U^+ = U/U_\tau$, $y^+ = yU_\tau/\upsilon$ and $U_\tau = \sqrt{\tau_w/\rho}$ is the friction velocity. τ_w denotes the wall shear stress, ρ the fluid density, υ kinematic viscosity, κ the von Kármán constant, and B the smooth wall logarithmic

intercept. The roughness function is parameterized as a function of the roughness Reynolds number,

$$k^+ = \frac{kU_\tau}{\upsilon}, \tag{2}$$

k denotes the characteristic roughness height. In the present study, the roughness functions are obtained from Jotun's in-house flow-cell experiments conducted under fully developed turbulent channel flow conditions, *Yeginbayeva (2022)*. The flow-cell facility enables the quantification of surface induced drag by measuring the streamwise pressure gradients over coated and smooth reference panels. For a given flow rate, the wall shear stress is determined from a momentum balance for fully developed channel flow,

$$\tau_w = -\frac{H}{2}\frac{dp}{dx}, \tag{3}$$

H is the channel height and dp/dx is the measured pressure gradient between the pressure taps located along the test section. The corresponding skin-friction coefficient is obtained as:

$$C_f = \frac{\tau_w}{\frac{1}{2}\rho U_m^2}, \tag{4}$$

U_m denotes the bulk mean velocity.

From experimentally determined wall shear stress, the friction velocity is evaluated for each surface and flow condition. The roughness function, ΔU^+, is then extracted by comparing the rough-wall response against the smooth-wall reference under identical flow conditions, *Schultz (2007).*

2.2. Assumptions for integrating flow cell derived roughness functions into CFD

The integration of the roughness functions obtained from flow-cell experiments into the ship scale CFD simulations relies on several physical assumptions in classical turbulent boundary layer theory. First, the approach assumes inner–outer layer scale separation, whereby the near-wall turbulence dynamics are governed primarily by local wall shear stress and viscosity. In accordance with *Townsend's (1956)* similarity hypothesis, the inner layer structure is assumed to be only weakly influenced by outer-layer conditions, provided that the roughness height remains small compared to the boundary layer thickness. Under this assumption, the roughness effect can be represented through a shift in the logarithmic mean velocity profile. Second, the local boundary layer along the hull is assumed to be approximately similar to a turbulent boundary layer under zero-pressure-gradient conditions when expressed in inner scaling variables. While this approximation is not strictly valid in regions of strong adverse pressure gradient, such as near the bow and stern, it is considered acceptable over most of the ship region where pressure gradients are low to moderate. Third, the roughness function, expressed as a downward shift of the velocity profile, ΔU^+, is assumed to depend primarily on the roughness Reynolds number, $k^+ = kU_\tau/\upsilon$, and not explicitly on the global Reynolds number, provided that the flow remains fully turbulent. This assumption follows classical rough-wall similarity arguments for hydraulically rough and transitionally rough regimes. Finally, the roughness function derived under controlled flow-cell conditions is assumed to be independent, to first order of free-surface effects, wave-induced motions, and bubble interactions. In this framework, such large-scale flow features are considered to influence primarily the outer flow, while the near-wall roughness-induced velocity deficit is governed by local inner scaling.

2.3. Resistance components

The total resistance of a ship can be decomposed into two primary contributions: frictional resistance R_F, and residuary resistance, R_R, such that:

$$R_T = R_F + R_R, \tag{5}$$

The frictional resistance comes from viscous shear stresses acting along the wetted hull surface. The residuary resistance is mainly pressure-related and consists of two components: the viscous pressure resistance (form drag), R_{VP}, and the wave-making resistance, R_W. Thus,

$$R_T = R_F + R_{VP} + R_W, \tag{6}$$

The viscous pressure resistance, also referred to as form drag, is commonly assumed to be proportional to the frictional resistance. Introducing the form factor, k_f, this relationship can be written as:

$$R_{VP} = k_f R_F, \tag{7}$$

Substituting into the total resistance expression gives:

$$R_T = (1 + k_f)R_F + R_W, \tag{8}$$

For non-dimensional analysis, the resistance components are normalized using the dynamic pressure, $0.5\rho V^2$, and the wetted surface of the hull, S. The corresponding resistance coefficients are defined as:

$$C_T = (1 + k_f)C_F + C_W, \tag{9}$$

C_T, C_F, and C_W denote the total, frictional, and residuary resistance coefficients, respectively

3. Numerical modeling

3.1. Governing equations and physical modeling

The flow is modelled by the incompressible Reynolds-averaged Navier-Stokes (RANS) equations, *Fureby (1998)*. The simulations employ the k-ω Shear Stress Transport (SST) turbulence model. The simulations were performed using the navalHydroPack framework, *Jasak (2018)*, which is developed for ship hydrodynamics applications and extends the OpenFOAM finite-volume environment. The solver employs a segregated pressure-velocity coupling strategy based on the PIMPLE algorithm.

The air-water interface is captured using a level-set method, *Osher (1994)*. The interface is reconstructed implicitly from the zero level-set contour, allowing accurate resolution of wave elevation and free-surface deformation while maintaining numerical stability.

Roughness effects at the hull are incorporated via calibrated rough-wall functions derived from experimental flow-cell measurements, ensuring consistency with shear-dominated boundary-layer behavior under equilibrium conditions explained in section 3.5.

3.2. Geometry and boundary conditions

To investigate the influence of antifouling coatings and their associated surface roughness under the assumption of a newly applied coating following dry-docking, CFD simulations were performed on three benchmark ship hulls: a container Ship (KCS), a tanker (KVLCC2), and a bulk carrier (JBC), Fig.1, Table I. These vessels were selected to represent different ship classes and assessment of roughness effects across distinct hydrodynamic characteristics and form factors. The simulations were performed in a fixed reference frame, where the ship remains stationary in space and the incoming flow represents the vessel advancing at constant speed in calm water. A uniform velocity boundary condition was prescribed at the inlet, corresponding to the target ship speed. At the outlet, a fixed static pressure condition was imposed, allowing the flow to exit the domain. The top boundary of the domain represents the open atmosphere and was prescribed with a constant atmospheric pressure condition. A symmetry-plane condition was applied along the ship centre plane, exploiting geometric symmetry to reduce computational cost. The bottom boundary was modelled as a no-slip wall. The water depth was selected sufficiently large to ensure negligible seabed influence on the hull boundary layer and wave system,

representing deep-water conditions. The hull surface was treated as a no-slip wall with wall-function treatment for turbulence quantities.

Although the computational mesh remains fixed in space, rigid-body motion of the hull was permitted in heave and pitch degrees of freedom. This allows the vessel to dynamically adjust its vertical position and trim in response to hydrodynamic forces and moments, while surge, sway, roll, and yaw motions were constrained.

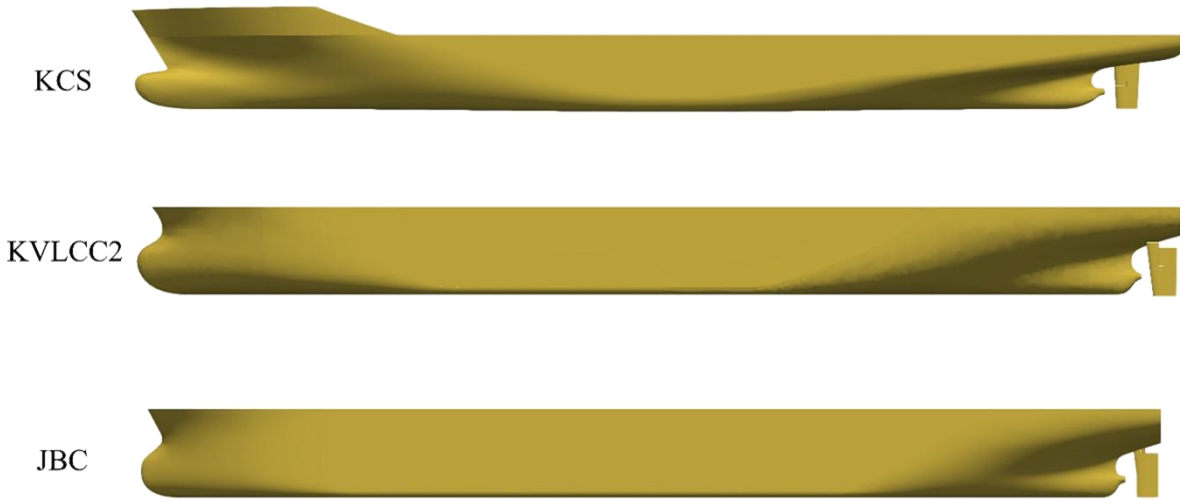


Fig.1: Geometry of KCS, KVLCC2 and JBC hulls with rudder

Table I: Geometrical characteristics of KCS, KVLCC2 and JBC at full scale

Particulars	KCS	**JBC**	KVLCC2
Length between the perpendiculars LPP	230 m	280 m	320 m
Length of waterline LWL	232.5 m	285 m	325.5 m
Beam at waterline BWL	32.2 m	45 m	58 m
Design draught T	10.8 m	16.5 m	20.8 m
Wetted surface area S	9539 m^2	19633 m^2	27194 m^2
Block coefficient C_b	0.651	0.8580	0.8098
Design speed	12.3456 m/s	14.5 m/s	15.5 m/s
Froude number at the design speed	0.26	0.142	0.142
Speed range used in simulations	8.28-12.35 m/s	6.918-7.966 m/s	6.948-7.956 m/s

3.3. Mesh generation and near wall treatment

The computational mesh was generated using cfMesh+, employing a Cartesian-dominant discretization with polyhedral transition cells and prism boundary layers along the hull surface. The meshing strategy was designed to simultaneously resolve (i) free-surface wave physics and (ii) turbulent boundary-layer behavior compatible with a wall-function approach as shown in Fig.2. The horizontal refinement strategy follows the characteristic Kelvin wake pattern generated by a displacement ship advancing in calm

water. The refinement regions were therefore aligned with Kelvin waves angle to ensure adequate resolution of the associated wave field and pressure gradients, Fig.3. To accurately capture free-surface elevation and wave steepness, anisotropic refinement was applied in the vertical direction around the mean waterline. The refinement band extends both above and below the interface to ensure sufficient resolution of the multiphase transport of the air–water interface. The vertical grading provides adequate resolution of pressure variations contributing to wave-making resistance, while maintaining numerical robustness through controlled mesh transitions. For all three investigated operating conditions, the mean y^+values over the wetted surface are between 150 and 250, with local minimum and maximum values ranging approximately between 30 and 300. These values confirm that the solution remains within the validity range of the applied rough-wall wall-function formulation across the hull.

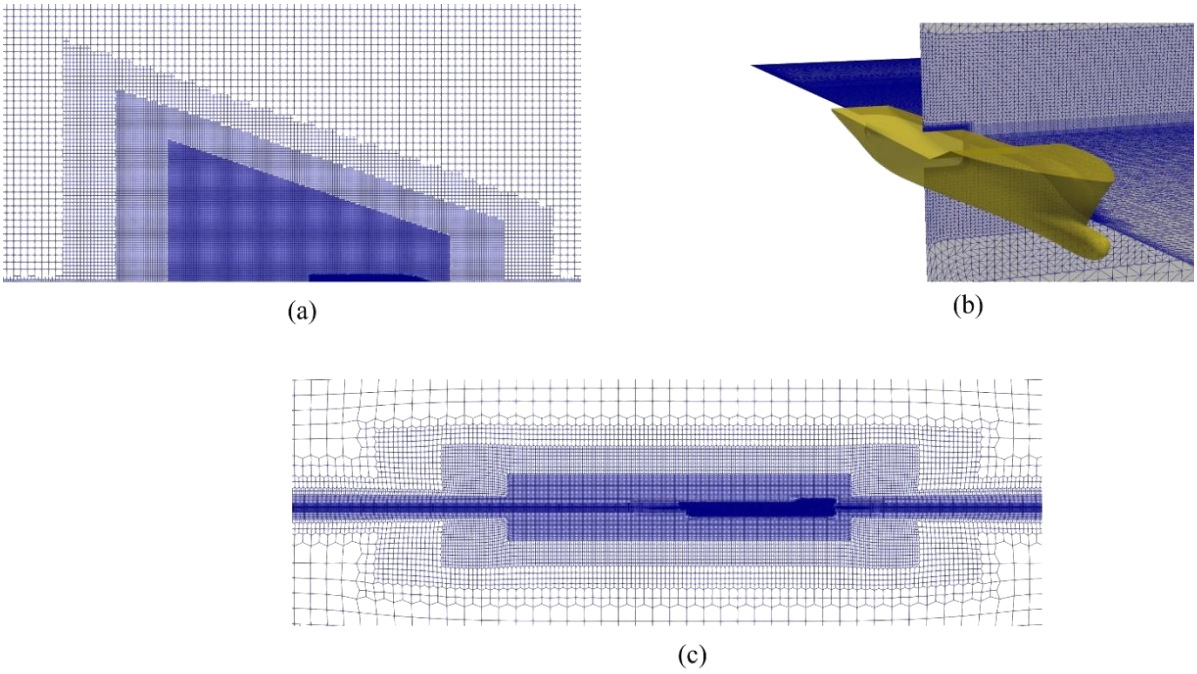


Fig.2: Computational mesh used in the simulations: (a) tapered refinement regions aligned with the Kelvin wake direction, (b) global mesh topology around the hull, and (c) anisotropic vertical refinement near the free surface for wave resolution.

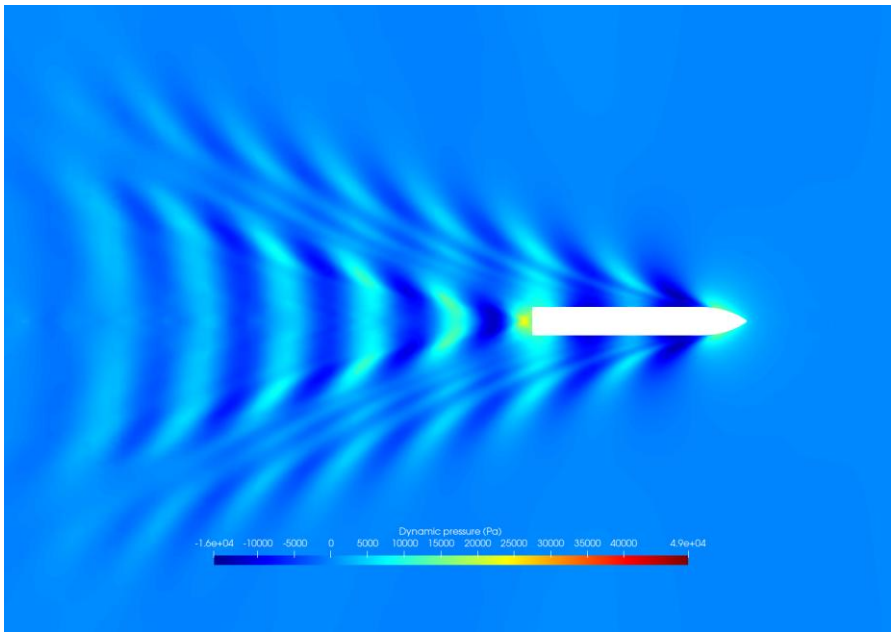


Fig.3: Kelvin wave patterns obtained from the current simulations for the KCS in design speed.

3.4. Verification and validation

To quantify the discretization error and assess the numerical uncertainty associated with the computational mesh, a grid convergence study was conducted. Three systematically refined mesh resolutions – coarse, medium, and fine were generated for the full-scale KCS model under smooth hull conditions. The simulations were performed under identical physical and numerical settings with the speed of 12.35 m/s. The spatial uncertainty was evaluated using the Grid Convergence Index (GCI) method following the procedure recommended by *Celik (2008)*. This approach provides a standardized estimation of discretization error based on Richardson extrapolation and apparent order of convergence. Table II summarizes mesh resolutions and corresponding total resistance values obtained from the simulations.

TableII: Spatial convergence study of full-scale KCS with smooth reference

	No. Cells (Millions)	$\boldsymbol{R_T}$ $(\boldsymbol{N})$
Coarse	~ 2.5	1.732×10^6
Medium	~ 6.5	1.670×10^6
Fine	~ 17	1.649×10^6
Spatial uncertainty		0.675%

Based on the GCI analysis, the estimated spatial discretization uncertainty on the fine grid was 0.672%. The monotonic convergence behavior observed across the three mesh levels indicates that the solution lies within the asymptotic range of grid convergence. The fine grid was therefore considered sufficiently resolved for subsequent simulations, as the estimated discretization uncertainty is below 1%.

To validate the numerical setup employed in the present study, the computed total resistance coefficient $C_T$of the KCS hull at $Fn = 0.26$ was compared against published CFD benchmarks and extrapolated experimental fluid dynamics (EFD) data available in the literature, Table III.

Table III: $\boldsymbol{C_T}$ values of KCS obtained from present CFD and other sources $\boldsymbol{Fn = 0.26}$

	$\boldsymbol{C_T}$	Relative difference %
Present CFD	0.002134	
CFD of *Dogrul et al. (2020)*	0.002113	0.98893
CFD of *Farkas et al. (2020*)	0.002081	2.51483
CFD of *Demirel et al. (2017)*	0.002097	1.74757
Extrapolation based on EFD of *Kim et al. (2001)*	0.002084	2.37079

The present CFD prediction shows agreement with both numerical and experimental references. The relative differences remain below 2.6% for all comparisons, with deviations below 1% relative to the CFD results of *Dogrul (2020)*. The close correspondence with independent CFD studies and experimental extrapolation confirms the reliability and consistency of the present numerical setup for predicting the hydrodynamic performance of the hull.

3.5. Numerical implementation of flow-cell based roughness model

The roughness function derived from flow-cell measurements is implemented into the RANS framework through a custom wall-function boundary condition developed within OpenFoam. The objective is to transfer the experimentally observed velocity shift directly into full-scale hull resistance simulations while maintaining compatibility with standard turbulence closures. The implementation modifies the near-wall eddy viscosity through an adjusted wall law. For each wall face, the local friction velocity is estimated from the turbulence kinetic energy as

$$u_\tau = C_\mu^{1/4}\sqrt{ke}, \tag{10}$$

ke is the turbulence kinetic energy and $C_\mu = 0.09$ is a turbulence model constant. The corresponding non-dimensional wall distance is then evaluated:

$$K_s^+ = \frac{u_\tau K_{eff}}{\nu}, \tag{11}$$

ν denotes the kinematic viscosity, and K_{eff} is prescribed locally at the wall faces, allowing spatial variation of coating conditions along the hull surface. The roughness function $\Delta U^+(K_{eff}^+)$, obtained from flow-cell and represented through fitted correlations, is evaluated locally. Rather than modifying the velocity profile explicitly, the roughness effect is introduced through a shift of the near wall law. This can be achieved by adjusting the smooth-wall log-law constant:

$$E' = \frac{E}{\exp\left(\kappa \Delta U^+\right)}, \tag{12}$$

which incorporates the experimentally measured velocity deficit into the wall formulation. The turbulent eddy viscosity at the wall is then computed as:

$$\nu_t = \nu\left(\frac{\kappa y^+}{\ln\left(E' y^+\right)} - 1\right), \tag{13}$$

thus, preserving the classical structure while embedding the roughness-induced shift.

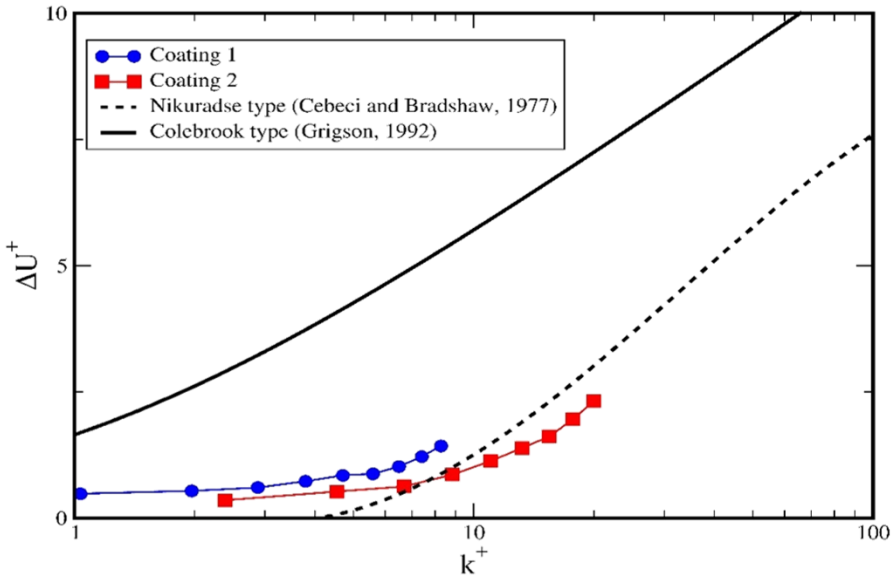


Fig 4: Roughness function $\boldsymbol{\Delta U^+}$ as function of roughness Reynolds number $\boldsymbol{k^+}$ obtained from flow-cell measurements for investigated coatings 1 and 2 ($\boldsymbol{k_{eff} = 17.45\mu m}$ and $\boldsymbol{41.98\,\mu m}$). Experimental data compared with classical Nikuradse-type and Colebrook-type correlations.

This formulation ensures that the experimentally calibrated coating behaviour is directly reflected in the predicted hull frictional resistance. At the same time, the approach remains fully consistent with standard RANS practice and robust for large-scale ship simulations, where local variations in surface condition may occur along the hull. Accordingly, this study investigates the effects of hull coatings on the ship hulls by considering two different coating conditions. Specifically, $k_{eff} = 17.45\mu m$ and Coating 2 with $k_{eff} = 41.98\mu m$ are examined. The corresponding roughness functions of these coatings as a function of the roughness Reynolds number are presented in Fig.4.

4. Results and discussion

To estimate full-scale performance curves and the influence of coating roughness on ship resistance, empirical similarity-based approaches such as Granville's method are commonly employed. In this framework, roughness effects are first quantified from canonical flat-plate or boundary-layer measurements through the roughness function ΔU^+. The smooth-wall friction line is then shifted accordingly, and the resulting full-scale friction coefficient is determined using similarity arguments. The hull is treated as an equivalent flat plate of representative length, and the corrected friction coefficient is subsequently incorporated into the total resistance formulation. This procedure inherently assumes that roughness behavior can be described by a representative roughness Reynolds number, evaluated under zero-pressure-gradient conditions. The correction is therefore applied globally, implying spatially uniform roughness behavior along the hull.

CFD simulations, however, can show that the roughness Reynolds number is not uniform over the wetted surface. Since the local friction velocity u_τ is obtained directly from the computed wall shear stress, k^+varies along the hull in response to boundary-layer development and hull curvature. Higher values are observed in shear-dominated midship regions and near the stern shoulder, whereas lower values appear in regions of flow acceleration and pressure-gradient influence. Consequently, the roughness function ΔU^+varies spatially, leading to non-uniform modification of the near-wall velocity profile and frictional resistance. The deviations observed between CFD-based predictions and empirical upscaling can therefore be attributed to the assumption of spatially uniform $k^+$and the neglect of three-dimensional and hull design effects inherent to real ship flows.

Fig 5 presents the distribution density function (DDF) of the roughness Reynolds number k^+along the full-scale KCS hull for the investigated speeds using the (a) Coating 1 and (b) Coating 2. The distributions demonstrate that $k^+$is not represented by a single characteristic value but instead exhibits a spatial spread over the wetted surface.

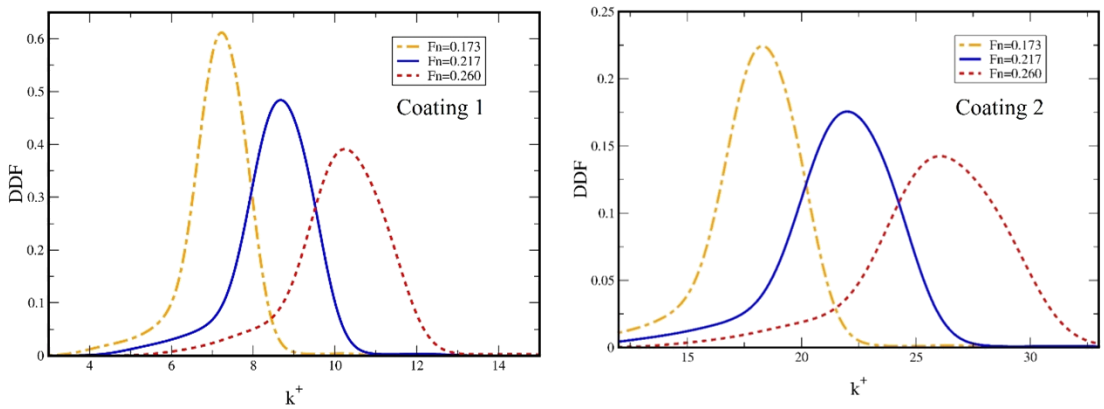


Fig 5: Distribution density function of the roughness Reynolds number $\boldsymbol{k}^+$along the full-scale KCS hull for the investigated speeds using the (a) Coating 1 and (b) Coating 2.

At the lowest speed ($Fr = 0.173$), in both cases the distribution is relatively narrow with slightly higher peak, indicating that a large portion of the hull operates within a limited k^+range. This suggests a more uniform shear environment, where the roughness effect is more concentrated around a characteristic value of k^+.

As the speed increases ($Fr = 0.217$ and $Fr = 0.260$), the distribution shifts toward higher k^+values and becomes slightly broader. The peak flattens and the tail extends toward larger k^+, indicating increased spatial variability of wall shear stress along the hull. Consequently, the roughness function ΔU^+, which depends on k^+, varies more over the surface.

The variation of the k^+distribution with speed might highlights a key limitation of Granville-type upscaling. While the similarity-law approach relies on a single representative k^+derived under flat-plate conditions, the CFD results demonstrate that real ship flows exhibit a spectrum of roughness Reynolds numbers whose spread increases with operating speed. This heterogeneity implies that the roughness correction cannot be strictly represented by a single global shift of the friction line.

As shown in the DDF plots, the distribution of k^+along the hull behaves slightly different for the two examined coatings as the speed increases. For Coating 1, the distribution remains relatively narrower across the Froude numbers investigated. Although the mean k^+shifts to higher values with increasing speed, the overall spread does not increase significantly. This indicates that the spatial variation of wall shear stress along the hull remains relatively uniform for smoother coating with lower effective roughness height.

For the Coating 2, the k^+distribution not only shifts toward higher values with increasing speed but also becomes slightly broader. This widening of the distribution reflects increased spatial heterogeneity of friction velocity along the hull. As the ship speed increases, pressure-gradient effects and stern-flow development become more pronounced, leading to stronger local variations in wall shear stress. It can be assumed that with a rougher surface, these local differences are amplified because k^+scales directly with friction velocity.

In simple terms, the smoother coating might respond more uniformly to changes in speed, while the rougher coating might magnify local flow differences along the hull. This might explain why the Coating 2 distribution becomes wider with increasing speed, whereas the Coating 1 distribution remains slightly narrower.

Table IV quantifies how the roughness Reynolds number distribution shifts and widens with increasing speed for both coatings. For both Coating 1 and Coating 2, the mean k^+increases monotonically with Froude number.

Table IV: Mean and standard deviation of the roughness Reynolds number k^+over the wetted hull surface for Coating 1 and Coating 2 at different Froude numbers

	Coating 1		Coating 2	
Froude (Fr)	Mean k^+	Std deviation	Mean k^+	Std deviation
0.173	7.10	0.84	17.93	2.27
0.217	8.52	1.02	21.58	2.77
0.260	10.14	1.22	25.81	3.31

More importantly, the standard deviation shows how k^+varies along the hull. For Coating 1 maintains as mentioned before a relatively compact distribution: the standard deviation increases only from 0.84 to 1.22 as Fr increases. In contrast, Coating 2 exhibits a broader distribution at all speeds, with standard deviation increasing from 2.27 to 3.31. This indicates that Coating 2 leads to a slightly wider range of local k^+values along the hull, i.e., stronger spatial differences in near-wall shear conditions, so different regions of the hull operate in more distinct roughness regimes, especially as speed increases. This effect can be seen in Figs.6 and 7 which represent the spatial distribution of wall shear stress magnitude on

the wetted surface of the KCS hull for Coating 2 and Coating 1 respectively, at three Froude numbers. For both coatings, the wall shear stress increases with increasing Froude number, as expected due to the increase in ship speed and friction velocity. At $Fn = 0.173$, the shear distribution remains relatively uniform along the hull, with moderate values concentrated along the midship region and lower levels near the stern wake. As the Froude number increases to 0.217 and 0.260, the overall magnitude of wall shear stress increases. However, differences can be observed between Coating 2 and Coating 1.

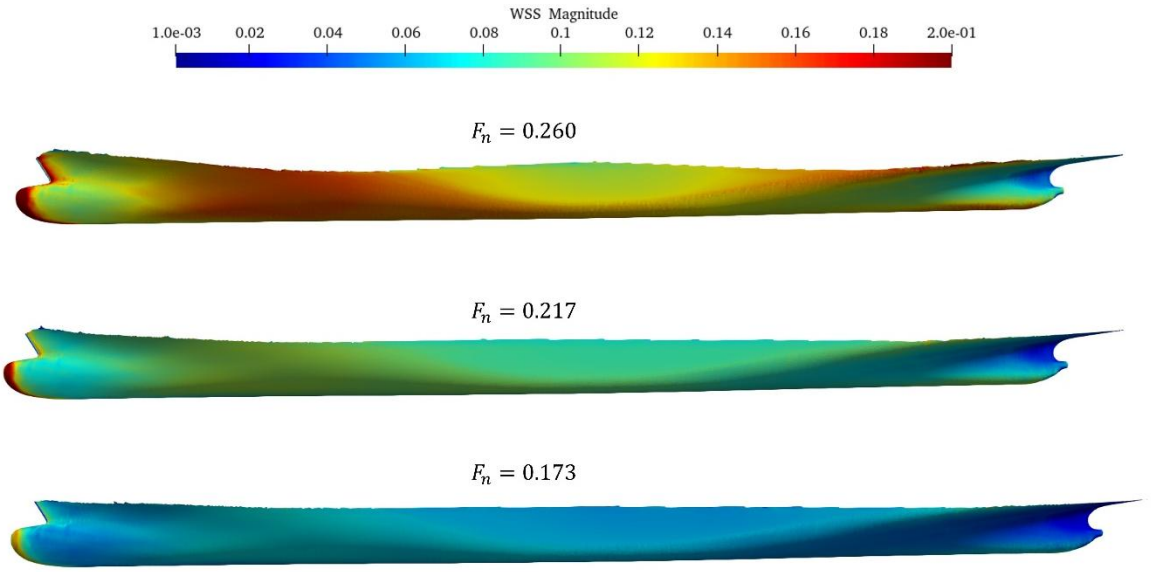


Fig.6: Distribution of wall shear stress magnitude on the wetted surface area for the KCS along three different ***Fn*** numbers under effects of Coating 2

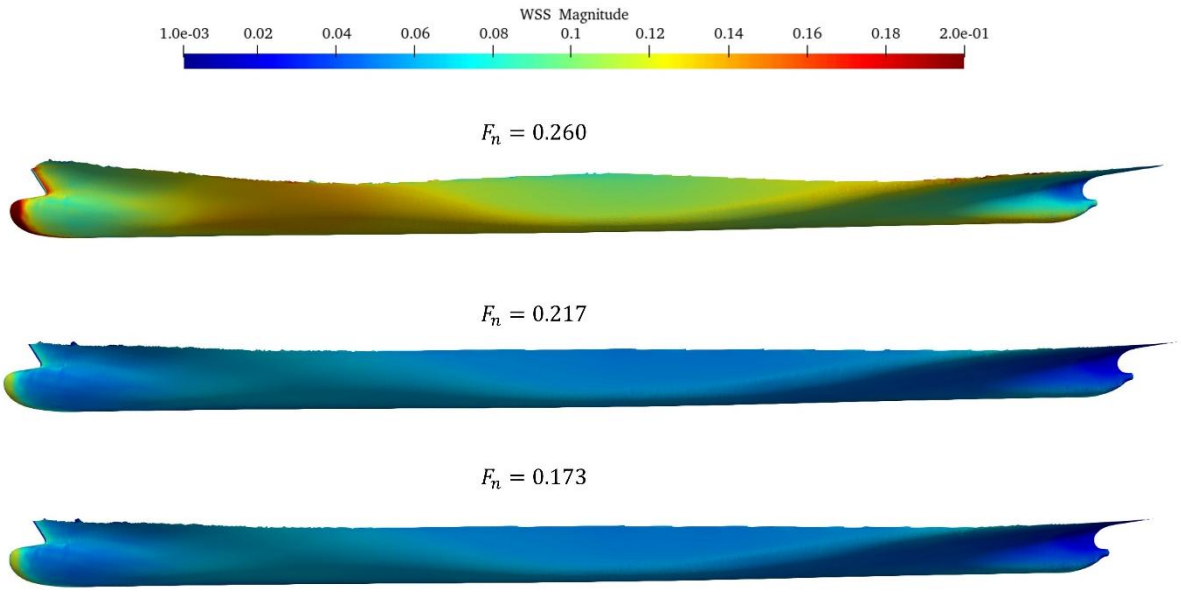


Fig.7: Distribution of wall shear stress magnitude on the wetted surface area for the KCS along three different Fn numbers under effect of Coating 1

For the Coating 2, higher wall shear magnitudes are can be seen across the hull at all speeds. At Fn = 0.260, pronounced high-shear regions develop along the midship and stern shoulder, and the spatial gradients of wall shear stress become stronger. The stern region shows intensified shear patterns, indicating increased sensitivity to pressure-gradient effects. The high-shear zones occupy a larger portion of the wetted surface compared to Coating 1 case. Overall, the wall shear stress contours confirm that coating roughness influences not only the magnitude of frictional forces but also their spatial distribution along the hull.

Fig.8 presents a comparison of the C_f values predicted by the Granville similarity-law extrapolation against full-scale CFD results for three benchmark hull forms of KCS, JBC, and KVLCC2, Combined with Coating 1 and Coating 2. In all cases, the Granville method underpredicts the full-scale C_f values obtained from CFD, which resolves the 3D hull geometry, pressure gradients, and boundary-layer development that increase the area-averaged wall shear stress. The difference is most visible at lower Reynolds numbers, where frictional resistance dominates and boundary layer is thicker. The percentage increase ranges from approximately 2.3–6.5% for Coating 1 and 8.0–11.5% for Coating 2, with the largest values observed on the fuller hull forms (JBC and KVLCC2). This trend confirms the hypothesis of *Oliveira et al. (2018)* that hull-form effects on resistance penalties are amplified on hulls with higher block coefficient. The amplification is consistently larger for Coating 2 than for Coating 1, as expected from the increased roughness regime.

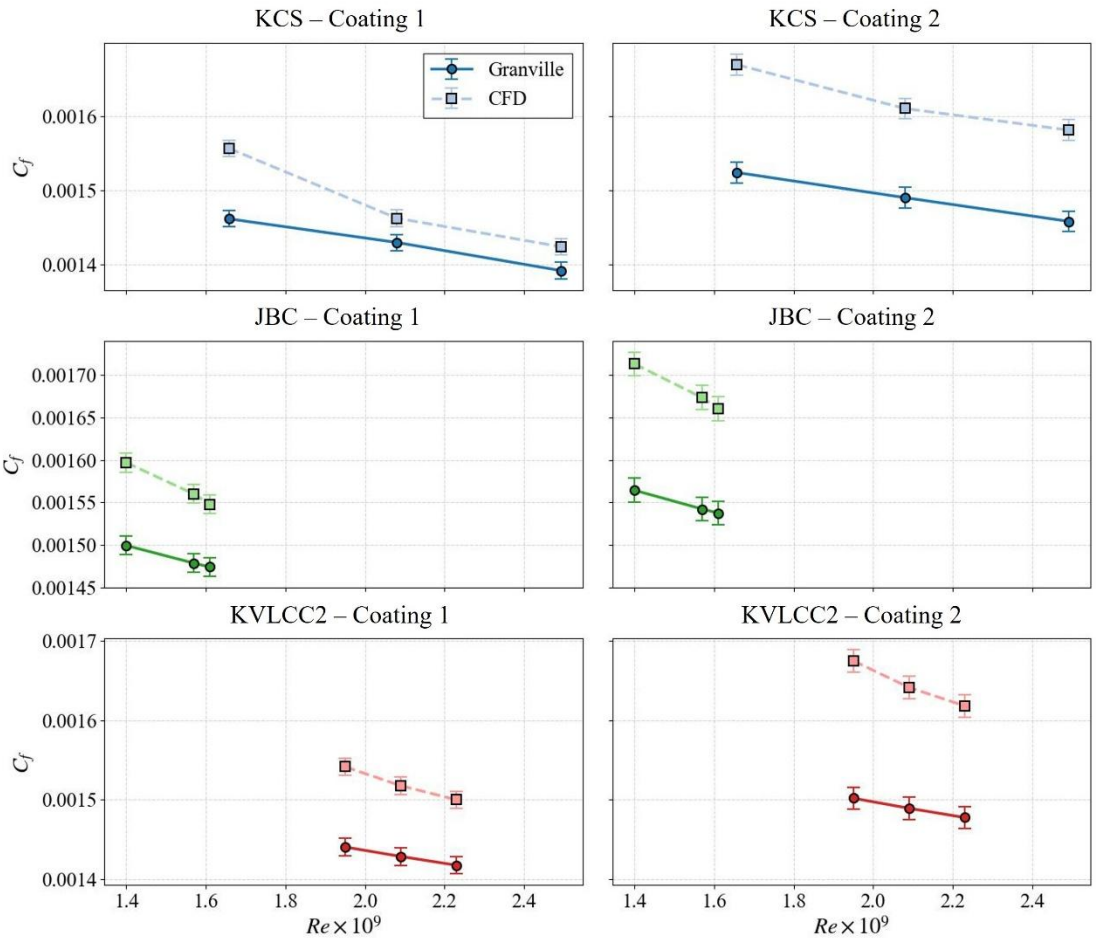


Fig.8: Friction coefficient C_f obtained from CFD and values upscaled using the Granville method for Coating 1, and Coating 2 surfaces over increasing Reynolds number

These comparisons highlight the need for hull-form-aware corrections when applying laboratory-derived roughness functions to full-scale predictions. This behavior difference is quantified in Fig.9 where the percentage difference is presented explicitly, highlighting the growing divergence with decreasing speed. Direct comparisons in Figs.8 and 9 show that the Granville similarity-law method consistently underpredicts the full-scale frictional resistance coefficient and it is confirming that hull-form effects amplify roughness penalties.

Given that block coefficient, roughness height, Reynolds number, and coating type all influence the magnitude of the correction, a data-driven approach is then considered that can capture these interactions without relying on strong functional assumptions.

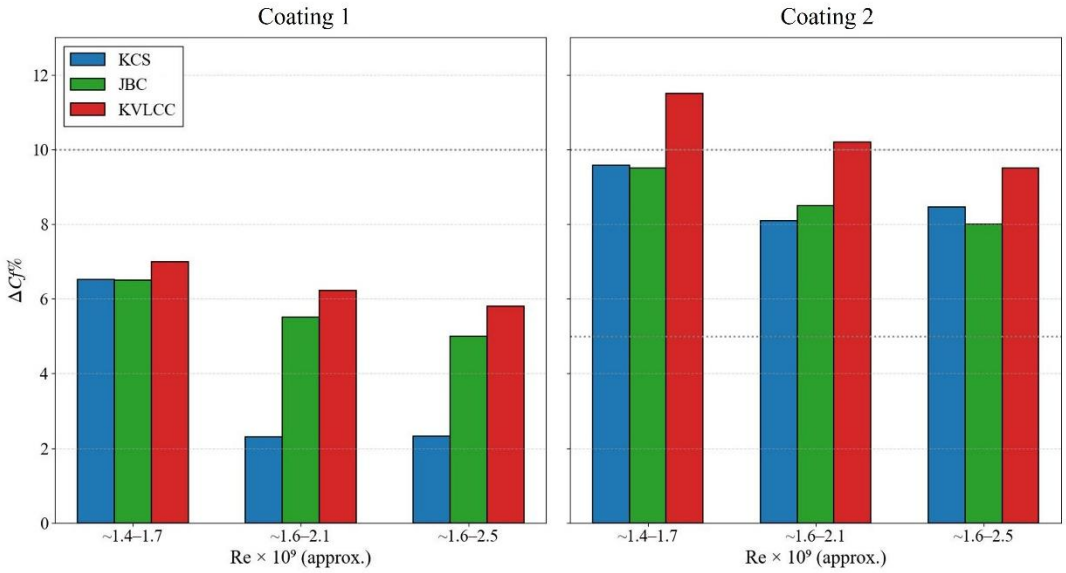


Fig.9: Percentage difference between CFD predictions and Granville-upscaled C_f values for Coating 1, and Coating 2 at three Reynolds numbers and hull forms.

A Random Forest regression model was therefore developed to predict the correction multiplier M based on the CFD results defined as:

$$C_{f,\text{corrected}} = C_{f,\text{Granville}} \times M \tag{14}$$

Random Forest was selected because it can model non-linear relationships and feature interactions in small, heterogeneous datasets, while being robust to noise and requiring minimal hyperparameter tuning. The model was trained on all 18 benchmark points (3 hulls × 3 Reynolds numbers × 2 coatings), using five input features: block coefficient C_B, roughness height k_s, Reynolds number and binary coating indicator I (0 = Coating 1, and 1 = Coating 2). The target variable was the excess multiplier $M - 1$, where $M = C_{f,\text{CFD}}/C_{f,\text{Granville}}$. The final model (200 trees, maximum depth 5) achieved $R^2 = 0.942$ on the full dataset and reproduced the observed ΔC_f%with mean absolute error < 1.2%. Although the small number of data points limit statistical power and may introduce some overfitting risk, the approach illustrates a methodology for improving roughness upscaling accuracy when traditional analytical corrections are insufficient. The corrected Granville predictions (green lines in Fig.8) are closer to the CFD results than the original upscaled values across all cases, validating the inclusion of hull geometry, roughness, and flow parameters. Consequently, a simplified linear correction of the Random Forest results can be proposed as:

$$M \approx 1 + 0.10(C_B - 0.65) + 0.035\left(\frac{k_s}{50}\right) + 0.008\log_{10}\left(\frac{10^9}{Re}\right) + 0.015I \tag{15}$$

This methodology indicates the benchmark penalties within ±2–3% and can be directly implemented in existing Granville workflows using only standard ship parameters and CFD generated data.

Fig.10 shows that applying the hull-form correction brings the predictions closer to the CFD results than the original Granville extrapolation across all hulls, coatings (1 and 2), and Reynolds numbers. Overall, the corrected Granvile represented by black triangle symbol demonstrate that the Random Forest model successfully captures the hull-form, roughness, and Re-dependent amplification missing in flat-plate Granville upscaling. The correction reduces the mean absolute error to CFD from ~5–10% (uncorrected) to ~1–2% (corrected) across the dataset, which validates its effectiveness for more accurate full-scale predictions.

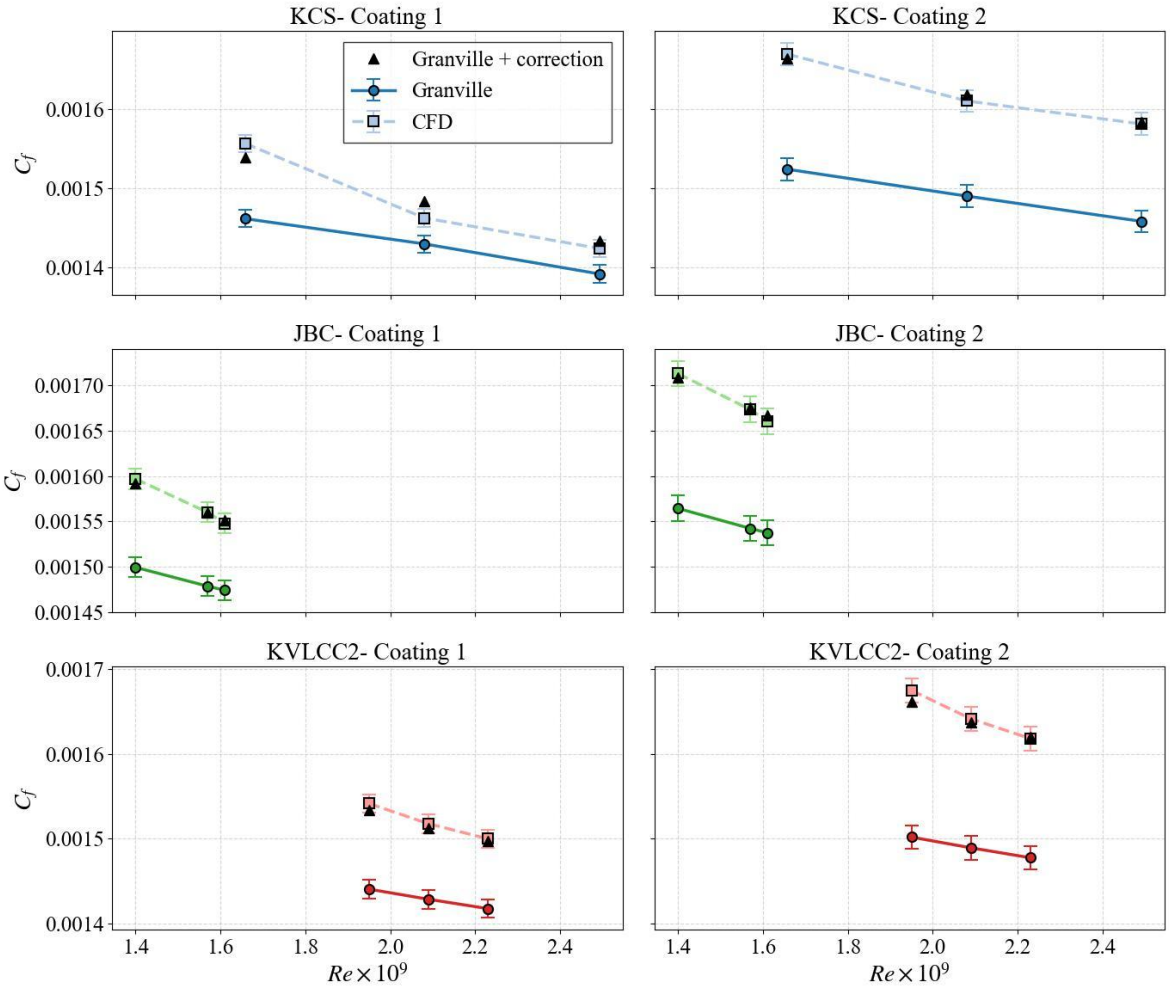


Fig.10: CFD, Granville and CFD-based corrected Granville frictional resistance coefficient versus Reynolds number for benchmark hulls under effects Coating 1, and Coating 2.

5. Conclusions

This study presented a methodology for evaluating and revising empirical roughness formulations for modern hull coatings by integrating flow-cell measurements with full-scale CFD simulations. The objective was to assess the validity of classical flat-plate-based upscaling approaches, particularly the Granville similarity-law method, when applied to contemporary coatings and realistic three-dimensional ship geometries.

The results demonstrate that the Granville method consistently underpredicts the full-scale frictional resistance coefficient when compared with CFD simulations that explicitly resolve hull geometry, pressure gradients, and spatial boundary-layer development. The deviation ranges from approximately 2–6% for a hydrodynamically smoother coating namely Coating 1 and 8–12% for rougher coating namely Coating 2, depending on Reynolds number and hull form. The amplification of roughness penalties is found to be larger for fuller hull forms such as JBC and KVLCC2 compared to the slenderer KCS, confirming that hull geometry and associated pressure-gradient effects influence roughness scaling. CFD results show that the roughness of Reynolds number $k^+$is not spatially uniform along the hull surface. Instead, it exhibits a distribution that broadens with increasing speed, particularly for rougher coatings. This behavior highlights a fundamental limitation of similarity-law extrapolation: the assumption of a single representative k^+value is insufficient to describe real ship boundary layers, where local friction varies due to curvature, pressure gradients, and stern flow development. To address the limitations of flat plate assumptions, a data-driven correction methodology was introduced. A

Random Forest regression model was trained using block coefficient, roughness height, Reynolds number, and coating type as input features. The model achieved $R^2 = 0.942$ and reduced the mean absolute error between Granville predictions and CFD results from approximately 5–10% to 1–2%. A simplified explicit correction formula was proposed, for practical implementation within existing Granville workflows without requiring CFD for each case. Overall, the study demonstrates that roughness penalties are hull-form dependent, fuller ships exhibit amplified frictional deviations, rougher coatings amplify spatial shear variability, Granville-type global upscaling is insufficient for modern coatings without geometry-aware correction and finally data-driven correction approaches provide a practical and robust pathway for improved full-scale resistance prediction. The proposed methodology connects experimental roughness characterization and full-scale hydrodynamic prediction, offering a systematic framework for revising empirical resistance formulations. Future work should expand the dataset to additional hull types and operating conditions, incorporate self-propulsion effects, and evaluate long-term coating behavior to further generalize the correction model.

Acknowledgement

We would like to express their gratitude to the technicians in the Fouling Protection Department at Jotun for their expert paint application and valuable assistance during experimental testing.

References

ATLAR, S.S.F. (2020), *Fouling effect on the resistance of different ship types*, Ocean Eng. 216

BRADY, R.F. (2000), *Mechanical factors favoring release from fouling release coatings*, Biofouling 15, pp.73-81

CELIK, I.B. (2008), *Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications*. J. Fluids Eng. 130

DEMIREL, A.D. (2020), *Scale effect on ship resistance components and form factor*, Ocean Eng. 209

DEMIREL, Y.K. (2017), *Predicting the effect of biofouling on ship resistance using CFD*, Applied Ocean Research 62, pp.100-118

DOGRUL, A.; SONG., S.; DEMIREL, Y.K. (2020), *Scale effect on ship resistance components and form factor*, Ocean Eng. 209

FARKAS, A.; DEGIULI, N.; MARTIC, I. (2020), *An investigation into the effect of hard fouling on the ship resistance using CFD*, Applied Ocean Research 100

FUREBY, H.G. (1998), *A Tensorial Approach to Computational Continuum Mechanics using Object Oriented Techniques*, Computers in Physics, pp.620-631

GRANVILLE, P.S. (1978), *Similarity-Law Characterization Methods for Arbitrary Hydrodynamic Roughnesses*, David Taylor Naval Ship Research and Development Center, Bethesda

GRANVILLE, P.S. (1987), *Three indirect methods for the drag characterization of arbitrarily rough surfaces on flat plates,* J. Ship Research 31, pp.70-77

JASAK, V.V. (2018), *The Naval Hydro Pack: Current Status and Challenges*, 13th OpenFOAM Workshop, Shanghai

KIM, W.J.; VAN, S.H.; KIM, D.H. (2001), *Measurement of flows around modern commercial ship models*, Experiments in Fluids 31, pp.567-578

MARTIĆ, A.F. (2018), *Towards the prediction of the effect of biofilm on the ship resistance using CFD*, Ocean Eng. 167, pp.169-186

MIN, K.S.; a.-H. (2010), *Study on the form factor and full-scale ship resistance prediction method*, J. Marine Science and Technology 15, pp.108-118

MURPHY, A.C. (2019), *Ship performance monitoring dedicated to biofouling analysis: Development on a small size research catamaran*, Applied Ocean Research 89, pp.224-236

OSHER, M. S. (1994), *A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow*, J. Computational Physics 114, pp.146-159

SANZ, D.S.; GARCIA, S.; TRUEBA-CASTANEDA, L.; TRUEBA, A. (2026), *CFD analysis of the influence of biofouling growth on ship resistance under variable trim and draft conditions*, Ocean Eng. 343

SCHULTZ, M.P. (2007), *Effects of coating roughness and biofouling on ship resistance and powering*, Biofouling 23, pp.331-341

SCHULTZ, Y.K. (2014), *A CFD model for the frictional resistance prediction of antifouling coatings*, Ocean Eng.89, pp.21-31

SONG, S.; DEMIREL, Y.K.; ATLAR, M. (2019), *An investigation into the effect of biofouling on full-scale propeller performance using CFD*, Int. Conf. Offshore Mechanics and Arctic Eng. (OMAE)

TOWNSEND, A.A. (1956), *The structure of turbulent shear flow*, Cambridge University Press

VAN, S.; AHN, H.; LEE, Y.Y.; KIM, C.; HWANG, S.; KIM, J.; KIM, K.S.; PARK, I.R. (2011), *Resistance characteristics and form factor evaluation for geosim models of KVLCC2 and KCS*, Advanced Model Measurement Technology for EU Maritime Industry, pp.282-293

YEGINBAYEVA, I. (2018), *An experimental investigation into the surface and hydrodynamic characteristics of marine coatings with mimicked hull roughness ranges*, Biofouling 34, pp.1001-1019

YEGINBAYEVA, I. (2022), *New measurement facility: Enhanced skin-friction measurement over large-scale plates in a channel flow*, HullPIC Conf., Tullamore

ZHANG, X.S. (2024), *Advanced strategies for marine antifouling based on nanomaterial-enhanced functional PDMS coatings*, Nano Materials Science 6, pp.375-395

Real Vessel Data Challenge: First Results, Food for Thought & Discussion

Vasileios Tsarsitalidis, ErmaTech, Athens/Greece, V.Tsarsitalidis@ermafirst.com
Dmitry Ponkratov, Siemens, London/UK, Dmitry.Ponkratov@siemens.com
Nikolaos Tsoulakos, Laskaridis Shipping, Athens/Greece, Tsoulakos@Laskaridis.com

Abstract

This paper presents the outcomes of the Real Vessel Data Challenge, a benchmarking workshop designed to compare and critically assess different methodologies applied to the same real-world dataset from a 76000 dwt bulk carrier. The challenge focused on generating speed–power baseline models, estimating performance degradation and fouling effects, reconstructing missing data, and assessing data quality and uncertainty, while remaining broadly aligned with ISO 19030. Beyond presenting comparative outcomes, the paper aims to stimulate an open technical discussion on the limits of objectivity in long-term performance analysis, the trade-off between robustness and sensitivity, and the need for clearer guidance on the application of standards to non-ideal datasets.

.

1. Introduction

The Real Vessel Data Challenge was conceived as a benchmarking exercise built around a single anonymized real-vessel dataset, with the specific aim of comparing how different analysts interpret the same operational evidence. The challenge was intentionally positioned at the intersection of ship-performance monitoring, fouling assessment, data quality control, and standard interpretation. Rather than seeking a single winning model, it was designed to expose areas of agreement and disagreement, reveal the sensitivity of results to modeling choices, and stimulate a more objective discussion around the practical use of ISO 19030 on non-ideal datasets.

The participant pool covered a diverse set of affiliations and backgrounds, including contributions from Albis Marine Performance, Chugoku Marine Paints, Nippon Kaiji Kyokai, NMRI, SRC, Lloyd's Register, MARIN, CMCC Foundation, Enamor; NTNU, SINTEF Ocean, NTU, DNV, We4Sea, NAPA, GCMD, Mærsk Mc-Kinney Møller Center for Zero Carbon Shipping, Seaspan Corporation, Foreship, RINA Group, JoRes project, and Perception (Vessel Performance Expert). The associated questionnaire responses also showed a mix of academic/research institutions, technology vendors, consultancies, startups, nonprofit bodies and mixed consortia, with declared tools ranging from Excel, MATLAB and R to Python-based regression, Bayesian, machine-learning, deep-learning and hybrid physics-informed workflows. These declarations are not used to de-anonymize the comparative discussion below, but they help explain why the submissions differed not only in numerical outputs, but also in problem framing and data-handling philosophy.

The dataset and reporting framework were accompanied by a code of conduct, confidentiality provisions, and a structured questionnaire. This provided a second layer of information beyond the technical reports: it captured declared tools, methods, organizational profiles, and baseline assumptions, which are used here to enrich the anonymized comparison. The central objective of this paper is therefore twofold: first, to compare the exported speed-power curves and the methodological choices behind them; second, to extract lessons for future benchmarking exercises and for the continued interpretation and possible refinement of ISO 19030.

2. Challenge Scope, Submission Set and Harmonization Procedure

2.1. Submission set and quantitative summary

The full participant set comprised 13 teams and 14 documented entries, because one team submitted two distinct machine-learning models (there was the option for multiple entries per team, but only one exercised this). The quantitative comparison presented here was limited to the exported speed-power

curves contained in the submitted CSV files. Since the exports were not fully standardized, an organizer-side harmonization step was required before any common plotting was possible.

Table I: Submission and harmonization summary used in the quantitative comparison

Participating teams	13
Documented entries discussed in the paper	14
Entries represented by exported speed-power curves	14
CSV files processed	49
Retained speed-power points	3624

2.2. Harmonization of submitted curve data

The common-curve comparison was based on all exported files containing explicit speed-power relationships, whether submitted as single consolidated datasets or as one file per draft, condition, or temporal variant. Organizer-side harmonization normalized the speed coordinate to knots and the power coordinate to kW, preserved participant labels, and introduced a condition-family layer used only for comparison plots.

Condition naming was one of the first issues revealed by the exercise. Some participants used explicit clean/current labels; others used fouled, baseline, newbuilt, December 2024, or month-specific labels. The resulting families were therefore grouped as clean-exact, current-like, baseline-like, monthly-variant, and unlabeled special cases. This was not done to erase semantic differences, but to make those differences visible in a manageable visual structure.

For the mean comparison line in each common plot, pointwise outlier removal was first applied and a smooth exploratory mean curve was then constructed. The resulting mean curves are therefore useful anchors for discussion, but not authoritative physical references. In addition, descriptive deviation scores relative to the smoothed family mean were calculated to test whether simple spread metrics can help summarize agreement and disagreement. These scores are used here descriptively only and are not interpreted as rankings.

2.3. Updated scope

In order to produce a commonly acceptable ranking, some “ground truth” would have to be established. Such a ground truth, could be a high fidelity CFD baseline model, but this has not been agreed upon with the participants and the scoring of the challenge (and the consequential declaration of the winner) will be concluded after open deliberation with all the participants. Thus, this paper will be limited to the anonymized comparison of the submissions among themselves, and the “competitive” part of the challenge will be shown in the presentation, after proper discussion with all participants.

3. Comparative Methodological Review

A key outcome of the challenge is that the entries differed not only in algorithm, but in problem definition. Some entries approached the task as a largely ISO 19030-compliant speed-power and degradation assessment. Others treated it first as a sensor-diagnostics problem. A third group framed it as a physics-based baseline reconstruction exercise, while another group treated it primarily as a statistical or machine-learning estimation task. The questionnaire responses reinforce this distinction by showing materially different declared tools and modeling traditions behind the entries. This distinction is essential because it explains why disagreement between submissions should not immediately be read as disagreement between models alone. In several cases, the participants were solving subtly different inverse problems with the same data. The condensed problem-definition table therefore includes both the report-based analytical interpretation and a short questionnaire-derived profile for each anonymized entry.

Table II: Condensed problem-definition summary for all entries discussed in the paper

	Questionnaire profile	**Effective question being solved**	**Main trust anchor**	**Main limiting factor**	**ISO-related implication**
Entry 1	Consult./eng.; Matlab; regressions + hybrid	Can ISO-like processing recover comparable outputs?	Standards workflow	Weak coverage after aggressive filtering	Needs clearer guidance for non-ideal datasets
Entry 2	Acad./research; Python; regressions + physics + Bayesian	Can monthly shifts be detected robustly?	Bayesian consistency bands	Attribution ambiguity between fouling and instrumentation	Detected shift should be distinguished from confirmed fouling
Entry 3	Tech vendor; Excel/SQL; ML + physics + hybrid	Can normalized trends support blind prediction?	Availability checks + hybrid model	Incomplete blind dataset	Define handling of blind/incomplete continuation periods
Entry 4	Acad./research + tech vendor; Excel/OCTARVIA; regressions + physics	How do actual-sea curves compare to clean baselines?	Towing-tank + sea-trial baseline	Limited sensor critique in report	Bridge still-water and actual-sea benchmarking more explicitly
Entry 5	Acad./research; Python; regressions	Can HF data be statistically condensed?	Hourly aggregation + OLS transparency	Smoothing may suppress fouling	Clarify acceptable temporal aggregation
Entry 6	Tech vendor; Python; DL + physics + hybrid	Can sensor diagnostics precede fouling inference?	Shaft-power/RPM envelopes + SFOC	Recalibration + state imbalance	Add explicit sensor-health stage
Entry 7	Non-profit; Python; regressions + physics + hybrid	Can corrected calm-water curves reveal fouling?	Cross-sensor engineering checks	Post-July consistency failure	Require torque/power/rpm cross-validation
Entry 8	Tech vendor; R; regressions + hybrid	Can gradual fouling be modeled as time drift?	Physics baseline + regression	Limited case-specific audit	Hybrid formulations may deserve more explicit recognition
Entry 9	Small team; Python; PINN + physics + hybrid	Can physical feasibility improve extrapolation?	Physical decomposition in AI	Too little detail for validation	Useful philosophy, limited evidence here
Entry 10	Acad./research consortium; Python/Excel; regressions + physics + hybrid	What remains possible under weak metadata?	Engineering standards + checks	Missing metadata and sensor trust	Specify minimum metadata/sensor verification requirements
Entry 11	Tech vendor/startup; Python/Excel; regressions + physics	Can fuel benchmarking work without full HF engine detail?	Digital twin + noon fuel	Noon timing uncertainty	Consider a fuel-centric pathway
Entry 12	Acad./research + consultancy; Python; regressions + ML	Can fouling proxies be learned from HF data?	Large preprocessed HF set	Missing GPS/draft/context	State more clearly how missing context limits objectivity
Entry	Acad./research +	Same with	Proxy + SHAP	Location blindness	Allow hybrid data-

13	consultancy; Python; regressions + ML	nonlinear flexibility	interpretability	+ collinearity	driven methods with stricter assumption reporting
Entry 14	Tech vendor/consultancy; Python/Tableau; regressions + physics + hybrid	Can heavy corrective preprocessing recover usable baselines?	Corrected variables + model-test guidance	Raw variables too compromised	Better address corrective preprocessing of incomplete inputs

4. Quantitative Comparison of Exported Curves

The main task of the challenge was to provide baseline curves for the clean and current condition of the vessel, given the real time data of a whole year along with the sea trials curves and main particulars of the vessel.

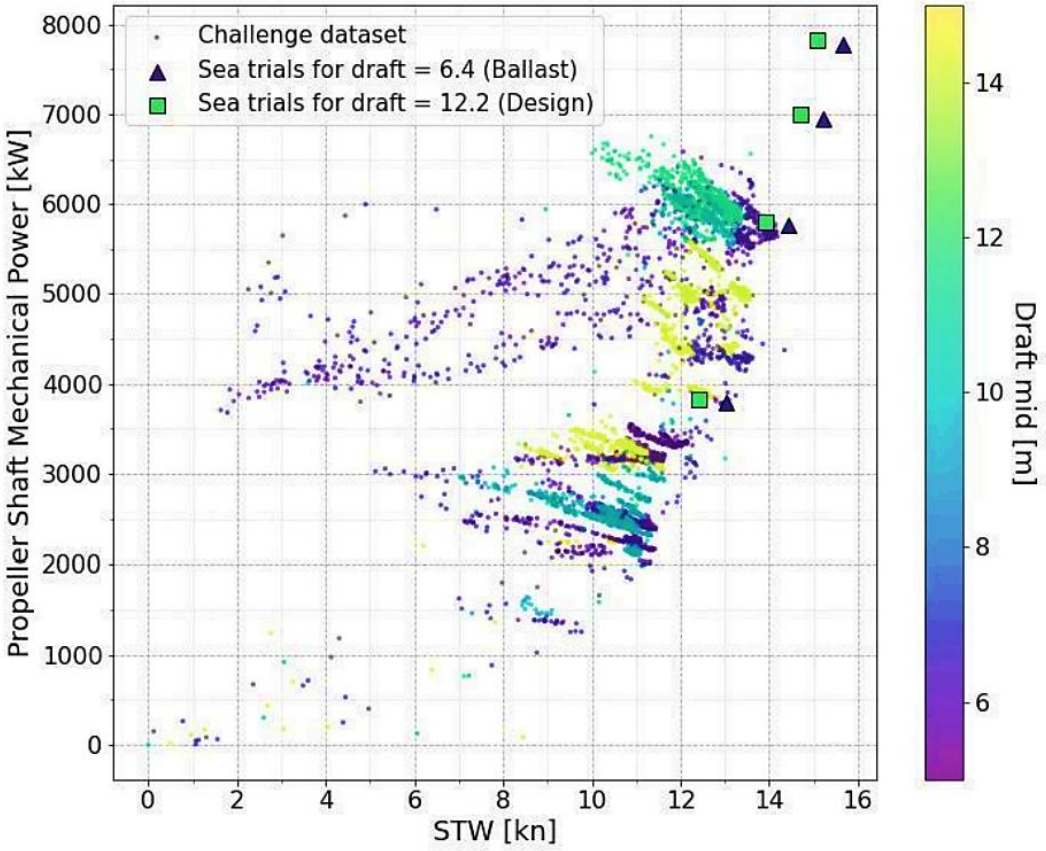


Fig.1: All speed vs power data of the dataset given to participants (shown to indicate the complexity)

The harmonized curve comparison reinforces the methodological findings. Once the exported curves are grouped by practical condition families, the plots show both where the entries converge and where they diverge due to semantics, baseline philosophy, or possible data-treatment differences. The intention of the figures is therefore diagnostic rather than competitive.

The common plots were built to reveal issues rather than suppress them. Each figure overlays anonymized submissions, grouped by draft where possible, and adds an exploratory smoothed family mean after pointwise outlier removal (dashed line). This allows families with reasonable overlap to be visually compared while still preserving outliers and special branches that may themselves be informative.

Table III: Exploratory quantitative summary of the harmonized curve families. The deviation metrics are descriptive only and are not used as rankings.

	Entries	Files	Points	Drafts represented	Median MAPE to mean	Median RMSE to mean
clean t	8	38	1,679	Ballast, Heavy ballast, 10 m, 11 m, Design, Scantling	5.5%	293 kW
current	7	8	802	Ballast, Heavy ballast, 10 m, 11 m, Design, Scantling	8.8%	470 kW
monthly variant	2	2	245	Ballast, Scantling	4.7%	154 kW

4.1. Speed VS Power Curves

Figs.2 to 6 show the collected curves for the clean and current condition (columns) and for the different drafts (lines). The clean-condition family shows the best overlap in the better-populated draft bands, but even there the spread is not negligible. This indicates that the challenge did not produce a single clean-baseline consensus. Some entries were close in both curvature and absolute power level, while others remained systematically higher or lower, reflecting different assumptions on what constituted clean reference behavior. The current condition family is more heterogeneous still, because it pools labels such as current, fouled, and December 2024. The common plot makes clear that current was not defined uniformly across participants. In several drafts the spread is wider than in the clean family, which is consistent with the report-based finding that sensor-health interpretation and baseline semantics were major differentiators.

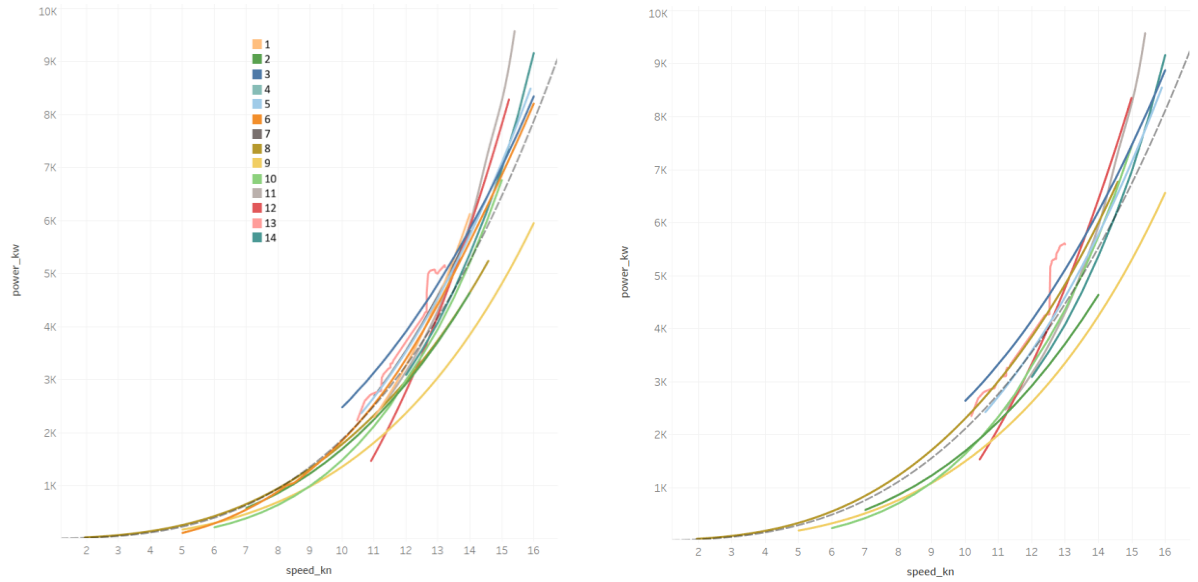


Fig.2: Curves for clean hull (left) and current condition (right) at ballast draft

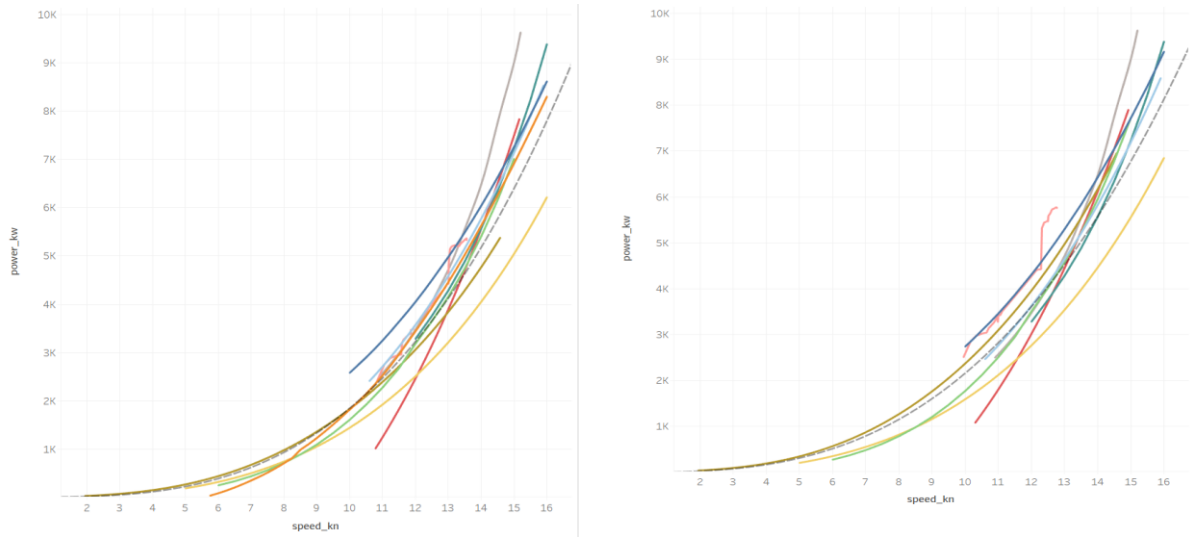


Fig.3: As Fig.2, at heavy ballast draft

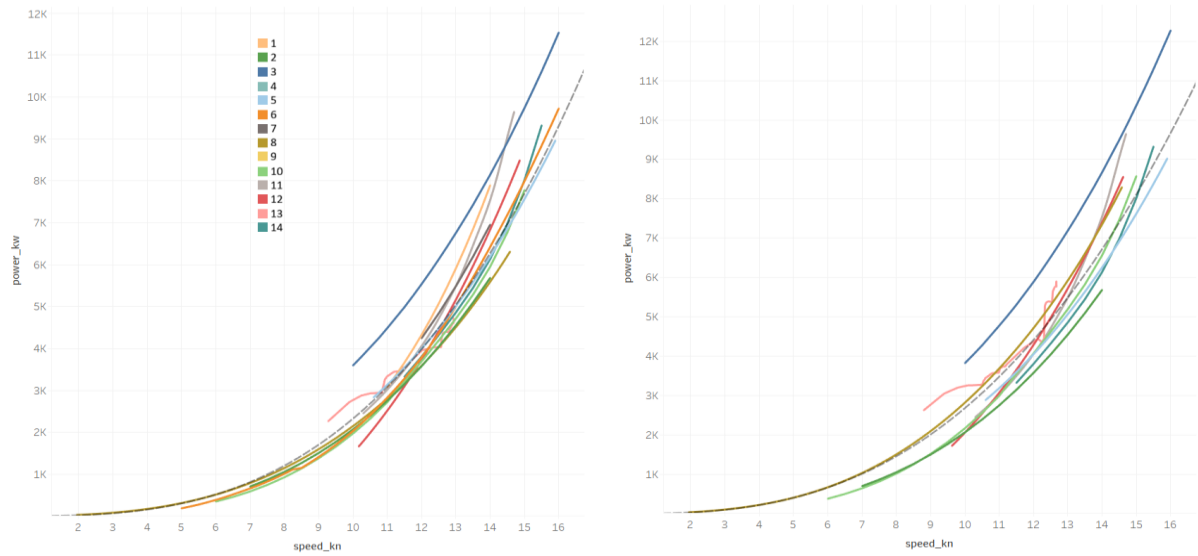


Fig.4: As Fig.2, at design draft

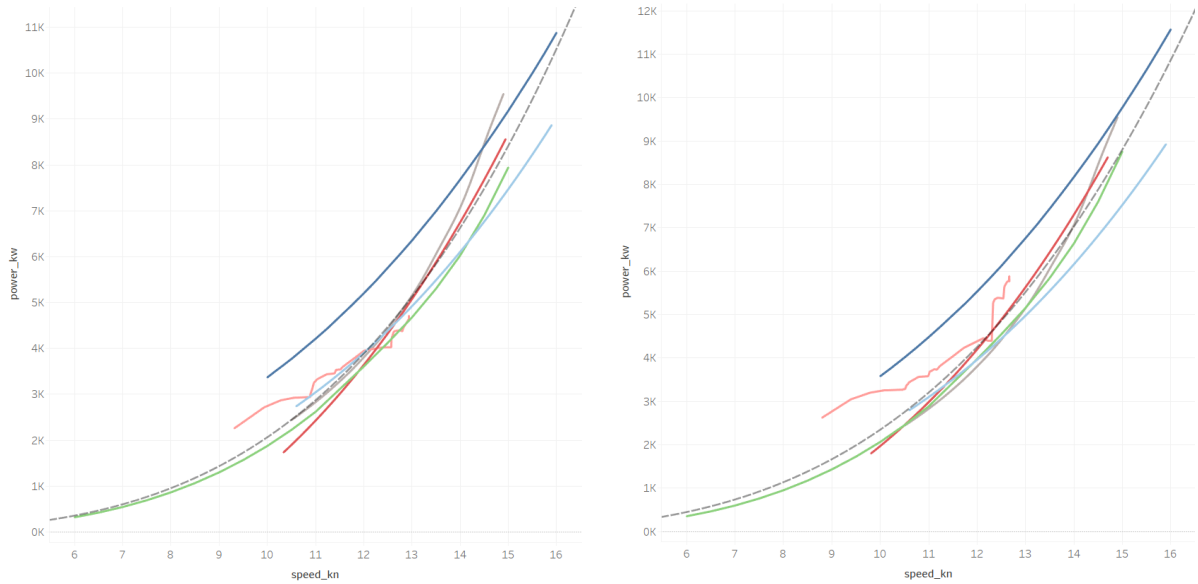


Fig.5: As Fig.2, at 11 m draft

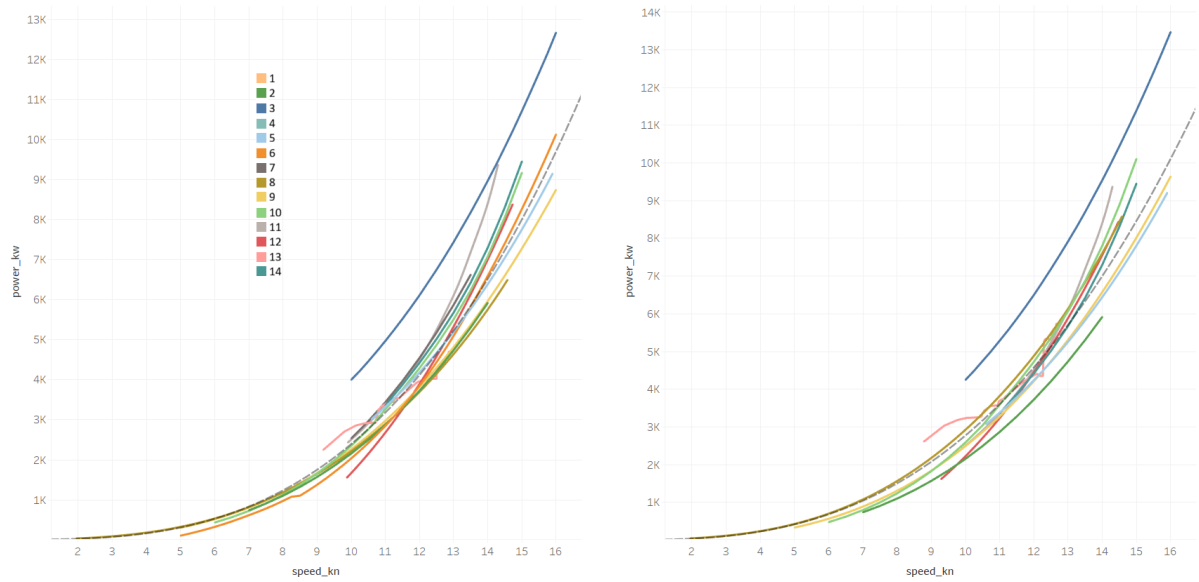


Fig.6: As Fig.2, at scantling draft

5. Discussion

The curve comparison and the report review point in the same direction: the largest differences between entries are not simply algorithmic. They arise much earlier, in how participants define the problem, choose or reconstruct baselines, decide which channels are trustworthy, and determine whether a shift in the data is operational, hydrodynamic, or instrumental.

A second major result is that data trustworthiness often prevails over model sophistication. Several entries independently pointed to structural breaks in torque- or propulsion-related signals around mid-2024, although they framed these breaks differently: recalibration, maintenance, or related sensor inconsistency. This convergence is important because it suggests that part of the challenge became an exercise in diagnosing instrumentation consistency before interpreting fouling trends.

The quantitative curve plots also show that direct averaging can be misleading when baseline semantics differ. The exploratory family mean is useful as a visual anchor, but it should not be interpreted as a physical reference line whenever the family itself mixes different clean or current philosophies. In that sense, the mean curve is helpful precisely because it highlights where averaging ceases to be conceptually safe. These observations also explain why the challenge should not be interpreted as a ranking exercise. The descriptive deviation scores are useful for testing spread metrics, but they cannot decide whether a curve is good or bad without reference to a common baseline philosophy and a common view of data validity. The challenge was therefore more valuable as a structured comparison of problem framings than as a competition for a single best line.

6. Implications for ISO 19030

The reports, the exported curve files, and the harmonized comparison together suggest several practical improvements or clarifications relevant to ISO 19030 and to the broader practice of long-term performance assessment on operational data. The points below combine the report review with the issues surfaced by the quantitative curve harmonization.

1. Separate performance-change detection from fouling attribution. A detected shift should not automatically be labeled fouling unless maintenance and instrumentation effects are excluded.

2. Add an explicit sensor-health stage. Torque-meter recalibration, unit inconsistencies, and cross-sensor propulsion checks should appear before fouling assessment, not only as informal pre-checks.
3. Provide clearer guidance for non-ideal datasets. Missing draft, missing GPS, uncertain noon timing, corrupted propulsion channels, and incomplete blind periods were all material in this challenge.
4. Clarify acceptable temporal aggregation. The challenge included 10-minute, 15-minute, 20-minute, hourly, and daily approaches, each with different implications for noise rejection and trend meaning.
5. Standardize weather and wave-treatment practice more clearly. Weather-source choice, wave filtering, and wave correction materially influenced the results.
6. Tighten minimum reporting requirements. Baseline definition, resampling logic, weather source, sensor corrections, uncertainty method, and synthetic inputs should always be disclosed.
7. Clarify acceptable clean-baseline concepts. Sea-trial, towing-tank, digital-twin, statistical, and post-cleaning baselines are not interchangeable and should not be treated as such.
8. Develop a common operational uncertainty language. The reports quantified very different uncertainties using very different tools and assumptions.
9. Be cautious about hull/propeller separation. Simultaneous cleaning and correlated proxies can make this distinction unreliable.
10. Consider a fuel-centric benchmarking pathway. Some operationally meaningful methods are not naturally shaft-power-centric and should not be excluded from comparison frameworks.

7. Conclusions

The first Real Vessel Data Challenge succeeded in demonstrating that common access to a single real-vessel dataset does not guarantee common interpretation. The participant reports, the questionnaire responses, and the harmonized curve comparison together show that variability arises from baseline philosophy, data-quality diagnosis, temporal treatment, environmental correction, and the willingness to question the instrumentation before trusting the trend. The report review and the curve comparison both indicate that the most important source of disagreement is often upstream of the final model. Sensor trust, missing context, incomplete metadata, and semantic differences in what counts as clean or current can dominate the apparent outcome. This is not a weakness of the challenge; it is one of its most useful results.

Future iterations of the challenge should preserve this openness while tightening the numerical export schema and minimum reporting requirements. More rigorous harmonization rules, richer metadata, and clearer distinction between change detection and attribution would make future comparisons sharper without suppressing the methodological diversity that made this first exercise valuable.

Acknowledgement

We gratefully acknowledge the participants of the Real Vessel Data Challenge, Laskaridis Shipping, the donor side for enabling the anonymized release of the dataset, and the broader HullPIC community for providing the forum in which these results can be discussed openly.

References

DNV (2023), *Recommended practice for verification and use of ship performance models*, DNV-RP-675, DNV, Hovik

ISO (2016), *ISO 19030 - Measurement of changes in hull and propeller performance*, Int. Standard Org., Geneva

ISO (2025), *ISO 15016 - Guidelines for speed and power trials*, Int. Standard Org., Geneva

Appendix A: Master comparison of submitted entries

Table A1: Master comparison of all entries in anonymized form, enriched with questionnaire-derived profile information

	Questionnaire profile	**Problem framing**	**Method class**	**Main data-issue diagnosis**	**Data-quality stance**	**Practical takeaway**
Entry 1	Consult./eng.; Matlab; regressions + hybrid	ISO-style recovery of clean/current/fouling outputs	Standards-led engineering	No single dominant instrumentation fault isolated; strict filtering/correction central	Formal, staged, standards-driven	ISO usable, but practical deviations were needed
Entry 2	Acad./research; Python; regressions + physics + Bayesian	Monthly performance-shift inference with uncertainty	Bayesian operational modelling	July shift may reflect recalibration or maintenance, not only fouling	Cautious, uncertainty-aware	Shift detected, cause not uniquely attributable
Entry 3	Tech vendor; Excel/SQL; ML + physics + hybrid	Reference-normalized fouling diagnosis and blind prediction	Hybrid empirical + ML	Blind data incomplete; draft and heading assumptions required	Pragmatic, availability-focused	October peak underperformance; later partial recovery
Entry 4	Acad./research + tech vendor; Excel/OCTARVIA; regressions + physics	Voyage-wise comparison to clean physical baselines	Physics/software operational-sea	Main challenge is voyage grouping and fit quality	Objective-fit, voyage-based	Actual-sea comparison feasible when clean baseline exists
Entry 5	Acad./research; Python; regressions	Statistical recovery of vessel-performance curve	Interpretable OLS	HF data too sparse/noisy for direct use	Conservative, statistical	Hourly smoothing robust, but fouling term weak
Entry 6	Tech vendor; Python; DL + physics + hybrid	Sensor diagnosis first, fouling second	Grey-box hybrid	Torquemeter inconsistency/recalibration after July	Highly diagnostic	Sensor-health correction is essential
Entry 7	Non-profit; Python; regressions + physics + hybrid	Calm-water curve after environmental correction	Engineering correction + regression	20–25% KQ step change after July	Strong engineering validation	Uncorrected split-year comparison not meaningful
Entry 8	Tech vendor; R; regressions + hybrid	Fouling as time drift on physics baseline	Physics + empirical hybrid	No specific anomaly emphasized	Structural/model-based	Useful hybrid framework; more method note than audit
Entry 9	Small team; Python; PINN + physics + hybrid	Physically feasible ML extrapolation	Hybrid hydrodynamics + ML	Too little case-specific detail	Method-focused	Insufficient detail for full comparison
Entry 10	Acad./research consortium;	Engineering procedure	Physics/engineering	Possible turbocharger or	Skeptical, realism-	Benchmark exposed

	Python/Excel; regressions + physics + hybrid	under incomplete metadata	procedures	mass-flow-meter issue; missing metadata	driven	sensor-trust and metadata limits
Entry 11	Tech vendor/startup; Python/Excel; regressions + physics	Fuel-performance benchmarking vs digital twin	Digital twin / fuel-centric	Noon timing uncertain; blind period narrow	Aggregation- and availability-centered	Operationally relevant but only partly comparable to power-centric methods
Entry 12	Acad./research + consultancy; Python; regressions + ML	Learn performance using fouling roughness proxy	ML with engineered proxy	Missing draft/GPS; corrupted propulsion channels	Explicit issue-register style	Promising, but context limitations remain strong
Entry 13	Acad./research + consultancy; Python; regressions + ML	Same proxy problem with nonlinear learning	ML / nonlinear	Same limits, plus hull/prop roughness collinearity	Issue-register + SHAP	Interpretable ML possible, but context still limiting
Entry 14	Tech vendor/consultancy; Python/Tableau; regressions + physics + hybrid	Parallel data-driven and model-test baselines	Regression + model-test hybrid	Major corrections needed for STW, power, torque, and swell descriptors	Correction-led	Preprocessing burden is itself a key result

Appendix B: Problem-definition summary

Table B1: Problem-definition table used in the main discussion

	Questionnaire profile	**Effective question being solved**	**Main trust anchor**	**Main limiting factor**	**ISO-related implication**
Entry 1	Consult./eng.; Matlab; regressions + hybrid	Can ISO-like processing recover comparable outputs?	Standards workflow	Weak coverage after aggressive filtering	Needs clearer guidance for non-ideal datasets
Entry 2	Acad./research; Python; regressions + physics + Bayesian	Can monthly shifts be detected robustly?	Bayesian consistency bands	Attribution ambiguity between fouling and instrumentation	Detected shift should be distinguished from confirmed fouling
Entry 3	Tech vendor; Excel/SQL; ML + physics + hybrid	Can normalized trends support blind prediction?	Availability checks + hybrid model	Incomplete blind dataset	Define handling of blind/incomplete continuation periods
Entry 4	Acad./research + tech vendor; Excel/OCTARVIA; regressions + physics	How do actual-sea curves compare to clean baselines?	Towing-tank + sea-trial baseline	Limited sensor critique in report	Bridge still-water and actual-sea benchmarking more explicitly
Entry 5	Acad./research; Python; regressions	Can HF data be statistically condensed?	Hourly aggregation + OLS transparency	Smoothing may suppress fouling	Clarify acceptable temporal aggregation
Entry 6	Tech vendor; Python; DL + physics + hybrid	Can sensor diagnostics precede fouling inference?	Shaft-power/RPM envelopes + SFOC	Recalibration + state imbalance	Add explicit sensor-health stage
Entry 7	Non-profit; Python; regressions + physics + hybrid	Can corrected calm-water curves reveal fouling?	Cross-sensor engineering checks	Post-July consistency failure	Require torque/power/rpm cross-validation
Entry 8	Tech vendor; R; regressions + hybrid	Can gradual fouling be modeled as time drift?	Physics baseline + regression	Limited case-specific audit	Hybrid formulations may deserve more explicit recognition
Entry 9	Small team; Python; PINN + physics + hybrid	Can physical feasibility improve extrapolation?	Physical decomposition in AI	Too little detail for validation	Useful philosophy, limited evidence here
Entry 10	Acad./research consortium; Python/Excel; regressions + physics + hybrid	What remains possible under weak metadata?	Engineering standards + checks	Missing metadata and sensor trust	Specify minimum metadata/sensor verification requirements
Entry 11	Tech vendor/startup; Python/Excel; regressions + physics	Can fuel benchmarking work without full HF engine detail?	Digital twin + noon fuel	Noon timing uncertainty	Consider a fuel-centric pathway
Entry 12	Acad./research +	Can fouling	Large	Missing	State more

	consultancy; Python; regressions + ML	proxies be learned from HF data?	preprocessed HF set	GPS/draft/context	clearly how missing context limits objectivity
Entry 13	Acad./research + consultancy; Python; regressions + ML	Same with nonlinear flexibility	Proxy + SHAP interpretability	Location blindness + collinearity	Allow hybrid data-driven methods with stricter assumption reporting
Entry 14	Tech vendor/consultancy; Python/Tableau; regressions + physics + hybrid	Can heavy corrective preprocessing recover usable baselines?	Corrected variables + model-test guidance	Raw variables too compromised	Better address corrective preprocessing of incomplete inputs

Appendix C: ISO 19030 improvement points emerging from the challenge

1. Separate performance-change detection from fouling attribution. A detected shift should not automatically be labeled fouling unless maintenance and instrumentation effects are excluded.
2. Add an explicit sensor-health stage. Torque-meter recalibration, unit inconsistencies, and cross-sensor propulsion checks should appear before fouling assessment, not only as informal pre-checks.
3. Provide clearer guidance for non-ideal datasets. Missing draft, missing GPS, uncertain noon timing, corrupted propulsion channels, and incomplete blind periods were all material in this challenge.
4. Clarify acceptable temporal aggregation. The challenge included 10-minute, 15-minute, 20-minute, hourly, and daily approaches, each with different implications for noise rejection and trend meaning.
5. Standardize weather and wave-treatment practice more clearly. Weather-source choice, wave filtering, and wave correction materially influenced the results.
6. Tighten minimum reporting requirements. Baseline definition, resampling logic, weather source, sensor corrections, uncertainty method, and synthetic inputs should always be disclosed.
7. Clarify acceptable clean-baseline concepts. Sea-trial, towing-tank, digital-twin, statistical, and post-cleaning baselines are not interchangeable and should not be treated as such.
8. Develop a common operational uncertainty language. The reports quantified very different uncertainties using very different tools and assumptions.
9. Be cautious about hull/propeller separation. Simultaneous cleaning and correlated proxies can make this distinction unreliable.
10. Consider a fuel-centric benchmarking pathway. Some operationally meaningful methods are not naturally shaft-power-centric and should not be excluded from comparison frameworks.

Optimizing Fuel Utilization by Direct Emission Monitoring

Jakob Broens, Green Instruments, Broenderslev/Denmark, jb@greeninstruments.com
Casper Jensen, Green Instruments, Broenderslev/Denmark, cnj@greeninstruments.com

Abstract

Using real-world case examples, this paper demonstrates how emissions measurements—particularly for methane—can be used to optimize fuel utilization, reduce methane slip, and improve overall engine performance. The case study compares Method C (fuel flow and default emission factors) against Method D (direct emission monitoring) for all engines onboard a large LNG carrier for almost 6 months. The resulting emission profiling included CO_2, CH_4 and CO_{2eq}, which show measurable reductions in methane slip percentage and improved fuel utilization achieved through targeted operational adjustments that do not require costly upgrades.

1. Introduction

Direct emissions monitoring is increasingly discussed as a compliance and reporting tool within maritime decarbonization. However, its potential value extends well beyond regulatory use. This white paper explores how high-frequency, engine-level emissions data can be transformed into actionable insights to improve fuel utilization and support operational decision-making and behavioral change within the marine industry.

2. The Problem

Default emission factors do not directly account for the engine load dependent nature of CH_4 slip. If succeeding in operating the engines at optimal load only the fuel consumption reduction is taken into consideration with the default emission factors. By relying on direct emission monitoring the achieved reduction related to the resulting CO_{2eq} can also be accounted for. Instead of relying on default engine performance curves the data provided by a direct emission monitoring method can be used to determine engine specific optimal load point as well as the load points to be avoided.

3. Functional Description

A Continuous Emission Monitoring System named G7200 CEMS is installed onboard, Fig.1, analyzing the GHG emissions from exhaust gas samples extracted from the stacks of up to four engines per system. Data from up to eight combustion units are transferred to an onboard physical database and calculation module named GreenView.

Based on the carbon balance method defined in the NOx technical Code. The Greenview generates data reports and totalized mass emissions based on received data related to engine, fuel and emissions. The result is comparable for mass emission data related to each individual engine making engine-to-engine, vessel-to-vessel and voyage-to-voyage comparisons possible. The data can be used for onboard optimization, regulatory reporting, as well as strategic decision making.

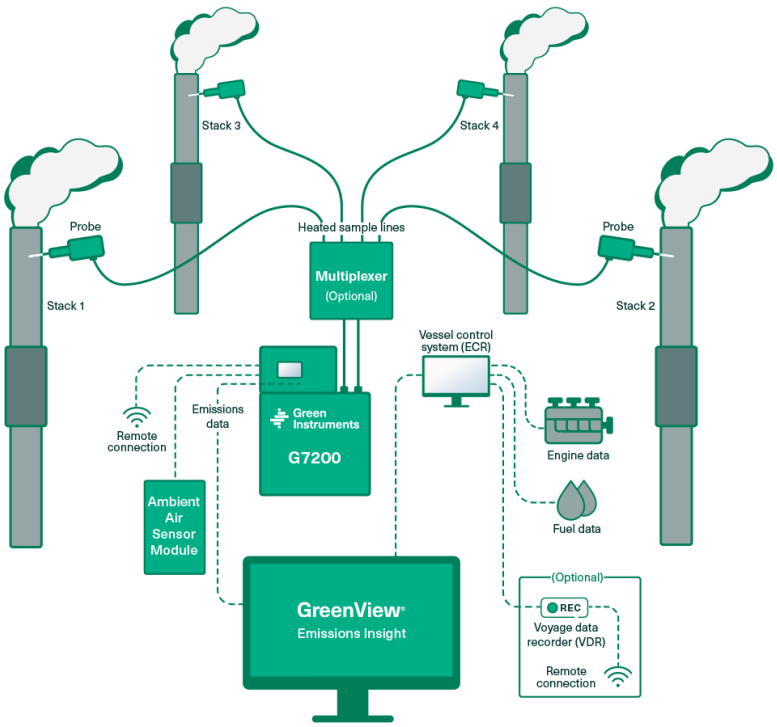


Fig.1: Schematic Overview of the Mass Emission Architecture

4. Case Studies

Using real-world case examples covering a period of almost 6 months, this paper demonstrates how combustion and emissions measurements—particularly for methane—can be used to optimize fuel utilization, reduce methane slip, and improve overall engine performance. The case study is based on a 95000 DWT LNG tanker equipped with two dual-fuel two-stroke slow speed main engines and four dual-fuel four-stroke medium speed auxiliary engines.

Table I: Default Emission Factors for Engine Types

Engine Alias	Engine Type	Rated power [MW]	$C_{f\,CO2}$ [$gCO_2/gFuel$]*	C_{slip} [%]*
ME 1	two-stroke Otto cycle	11.9	2.75	1.7
ME 2				
DG 1	four-stroke Otto cycle	4.3		3.1
DG 2				
DG 3		2.9		
DG 4				

*Default emission factors relevant to EU ETS and Fuel EU Maritime.

The fuel consumption related to each engine is based on Coriolis mass flow meters. The engines are primarily running boil off gas (BOG) from the cargo tanks; however, since the exact fuel composition is unknown, the fuel composition is defined as pure methane throughout the dataset. For practical reasons and to keep comparison simple, the modest consumption of pilot fuel of less than 10 kg/h MGO per engine is disregarded.

4.1. Case 1: Comparing Method C and D on Macro Level

In this case, we will zoom in on a macro level to compare Method C utilizing default emission factors with Method D direct emission measurement.

During the voyage, the majority of the fuel is being consumed by the main engines. Fig.2 shows a typical emission profile for Main Engine 2.

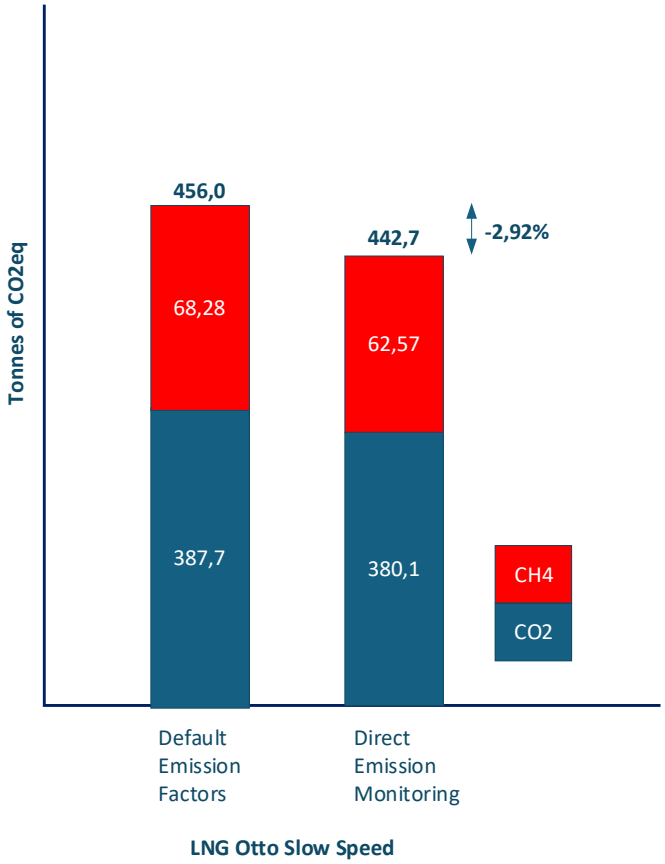


Fig.2: Comparing Methods C & D for LNG Otto Slow Speed Engine

The correlation between the two methods is evaluated by directly comparing the CO_2 emissions, which have a deviation of 1.96%. A GWP of 28 is used to convert CH_4 emissions into CO_2 equivalences (CO_{2eq}). When combining the fuel consumption with the default emission factors, the resulting emission profile is 456.0-ton CO_{2eq} of which 15.0% of the CO_{2eq} derived from CH_4. During the same time period, the emissions were reported by means of the onboard direct emission monitoring system, which resulted in a total emittance of 442.7-ton CO_{2eq} of which 14.1% of the CO_{2eq} derived from CH_4. Choosing a direct emission monitoring method, the total CO_{2eq} is measured to be 2.9%. The resulting CH_4 slip% was 1.56% against 1.7% as default. The average engine load was 40.2%.

At port stay, the majority of the fuel is consumed by the auxiliary engines. Fig.3 shows a typical emission profile of D/G 4.

The correlation between the two methods is evaluated by directly comparing the CO_2 emissions, which have a deviation of 0.32%. A GWP of 28 is used to convert CH_4 emissions into CO_2 equivalences. When combining the fuel consumption with the default emission factors, the resulting emission profile is 69.45-ton CO_{2eq} of which 24.5% of the CO_{2eq} derived from CH_4. During the same time, the emissions were reported by means of the onboard direct emission monitoring system, which resulted in a total emittance of 59.11-ton CO_{2eq} of which only 11.7% of the CO_{2eq} derived from CH_4. Choosing a direct emission monitoring method, the total CO_{2eq} is measured to be 14.87%. The resulting CH_4 slip% was 1.25% against 3.1% as default. The average engine load was 49.8%.

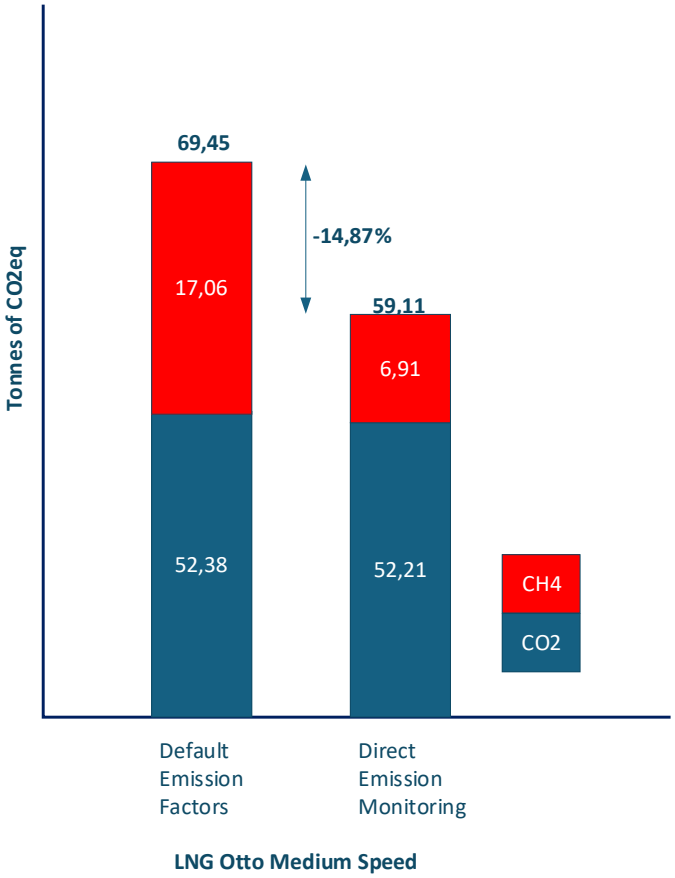


Fig.3: Comparing Methods C & D for LNG Otto Medium Speed Engine

On a macro level, this comparative case study illustrates the emission profiles of a typical two-stroke and four-stroke engine and shows the proportions on how CO_2 and CH_4 form the resulting $CO2_{eq}$. It becomes clear that CH_4 accounts for a significant part of the CO_{2eq} and that operational changes striving towards minimizing the CH_4 slip thus hold the potential of significant CO_{2eq} reductions often accompanied by reduction in fuel consumption as well.

4.2. Case 2: Comparing Method C and D over a 3.5 Month Period

In this case study, we will step out of the macro level to create an overview of emissions covering a 3.5-month period with engines operating during voyage as well as during port stay. Focus will be on evaluation of engine overall performance monitored as CH_4 slip. As shown in Table II, the average measured CH_4 slip and resulting CO_{2eq} throughout the 3.5 months period is measured to be below the Method C alternative for 5 out of 6 engines. The measured data also reveals that individual engine performance varies throughout the period.

Table II: Default and measured CH_4 slip and impact on CO_{2eq}

Engine Alias	Engine Type	Rated power [MW]	$C_{f\,CO2}$ [gCO_2/gFuel]*	C_{slip} [%]*	Actual C_{slip} [%]	Average C_{slip} [%]	Resulting CO_{2eq} Diff. [%]
ME 1	two-stroke Otto cycle	11.9	2.75	1.7	2.00	1.73	+2.04
ME 2					1.46		-1.91
DG 1	four-stroke Otto cycle	4.3		3.1	2.19	2.02%	-6.52
DG 2					2.06		-7.45
DG 3		2.9			2.07		-7.36
DG 4					1.75		-9.65

*Default emission factors relevant to EU ETS and Fuel EU Maritime.

Case 3: Comparing Emissions on an Engine-to-engine Level

Based on the data in Table II, it is possible to compare actual CH_4 slip for similar engine types on an engine-to-engine level. The comparison shows that ME 1 stands out compared to ME 2, while also DG 4 stands out to DG 3. The cause of the differences could be linked to variation of numerous parameters such as engine loads during the period and also variance in BOG composition. In order to explain the differences, let's take a closer look comparing ME1 with ME2 at a given period in which the engines are running in parallel indicating that the engine type, the engine loads and the fuel composition are comparable.

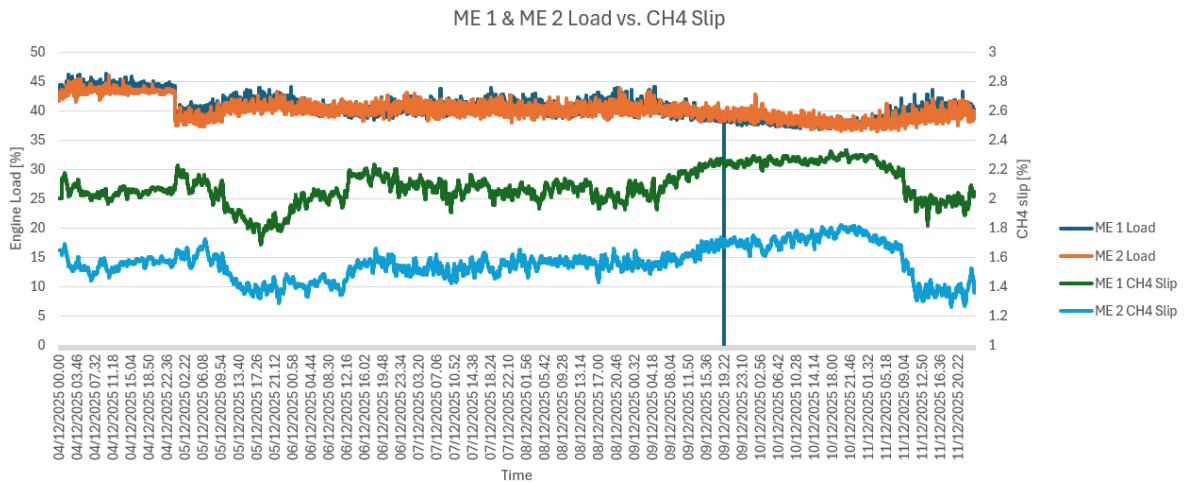


Fig.4: Comparing measured CH_4 slip for identical engines

As shown in Fig.4, the engine loads are comparable while the CH_4 slip shows significant differences. During this period, the average CH_4 slip is 2.08% for ME 1 and 1.56% for ME 2. This engine-to-engine comparison is not possible when relying on Method C or Engine Load Monitoring (ELM) methods, which demonstrates the insights only achievable by direct emission monitoring. In the given case, the CO_{2eq} related to ME 2 will be lower than for ME 1.

5. Evaluation and Interpretation of Results

To compare emissions related to the default emission factors and the direct emissions monitoring, the total CO_{2eq} must be compared. For both methods, the total CO_{2eq} is based on emissions of CO_2 and CH_4 without including emissions of N_2O. The emission of CH_4 is converted into CO_{2eq} based on CH_4 GWP_{100} factor 28. For both methods, the pilot fuel is not included and is less than 10 kg/h MGO for each engine.

When an engine is running in gas mode, the fuel composition is assumed to be 100% CH_4, however due to the engines primarily running on BOG the composition varies. Installing a fuel gas analyzer will allow for more exact fuel data input to our calculations. This will also further strengthen the carbon balance method utilized by the GreenView module, i.e., by accounting for the N_2 content of the BOG in real time, which directly adjusts the reported mass emissions.

6. Conclusion

By utilizing our proposed method based on direct emissions monitoring, it has been demonstrated that there is a good correlation between Method C and Method D with differences between 0.32% to 1.96% when comparing CO_2 emissions. The correlation is good even on macro levels during voyage and port stay. A more detailed emission profile is possible when using a direct emission method

compared with static emission factors, since direct method accounts for the load-dependent nature of the CH_4 slip%. For this case study covering multiple engine load scenarios related to main engines and auxiliary engines during voyage and port stay, it has been demonstrated that the direct emission monitoring in general documents reduced emissions compared with Method C.

As a tool to monitor engine performance, the study shows that engine-to-engine comparison of the actual measurement of CH_4 slip can be used to identify differences in performance related to each individual engine even when being of identical type. This difference questions not only the feasibility of static emission factors, but also the feasibility of the engine load monitoring method (ELM).

From an operational point of view the detailed level of insights can be used to optimize individual engine performance and fuel utilization, for instance by identifying the most and least favorable engine load points. Since emission of CH_4 greatly impacts the overall CO_{2eq} striving for optimal load points has the potential to realize substantial reduction in both fuel consumption and emission footprint. Furthermore, the engine specific emission insights

From a reporting perspective the actual emission profile can be directly converted into total CO_{2eq} making it suitable for transparent regulatory emission reporting.

Looking forward the adoption of post combustion GHG abatement technologies will break the link between fuel based default emission factors. At this point the direct emission monitoring method also serves as a fuel agnostic tool to document the performance and investment feasibility of engine upgrades and emission abatement systems like CH_4 catalysts and OCCS.

Acknowledgement

Special thanks to the team representing the shipping company for trusting our solution and having the will to implement direct emission monitoring on this vessel. Thank you for your cooperation and open-mindedness during this joint effort.

References

EU (2015), *Regulation (EU) 2015/757 on the Monitoring, Reporting, and Verification of CO_2 emissions from ships*, European Commission

EU (2020), *Commission Delegated Regulation (EU) 2020/1044 on Global Warming Potentials for greenhouse gases,* European Commission

EU (2025), *Guidelines for Reporting and Verification of Actual Methane Slip Emission Factors – Interim Procedure for Engine Load Monitoring (ELM)*, *Annex I: Interim Procedure for Engine Load Monitoring (ELM) to establish weighting factors that reflect the actual operation of a marine diesel engine and calculation of emission values*, Directorate General for Mobility and Transport, European Commission

IMO (2008), *NOx Technical Code 2008*, *Appendix 6: Guidelines for Engine Test Cycles and Compliance*, Int. Mar. Org., London

Ship Performance Prediction with a Hybrid Model for the Estimation of Ship's Added Resistance in Waves

Théo Descamps, Bureau Veritas Solutions Marine & Offshore, Nantes/France,
theo.descamps@bvsolutions-m-o.com

Abstract

For ship performance prediction, the added resistance in waves is a significant part of the total ship's resistance. In naval industry, estimating accurately this added resistance remains a challenge. Existing numerical methods are from empirical formulations to costly CFD simulations. They mostly rely on added resistance spectrum calculation and linear theory. However, this linear theory is known to be inaccurate with steep waves or short crested sea states. In such cases, CFD simulations of ships in the targeted sea states are possible but, because of the high cost and complexity, this solution is not suitable for industrial use. This paper presents several standard models for the calculation of added resistance in waves in the context of voyage data analyses and ship performance prediction. From these models, a hybrid one is presented and reviewed for specific applications. This hybrid intends to improve the accuracy of semi-empirical and potential flow models, using CFD simulation but keeping a reasonable computational cost.

1. Introduction

When studying ship performances either to investigate new designs or to estimate the gains achievable through retrofit, a variety of loads must be estimated as precisely as possible. This seek of accuracy often meets the reality of costs and a trade-off must be made between cost and calculation accuracy.

Following ITTC procedures, the first load to estimate is the calm water resistance. As this quantity largely dominates the other sources of resistance in normal navigation conditions, combination of accurate methods is used from CFD calculation to model test in basin. Along with these tests and simulations, the second most important aspect to evaluate accurately is the self-propulsion efficiency. Once again CFD simulations and model test are used.

At this step, only the performance in ideal conditions is estimated. However, this ideal case is not often encountered (not to say never) and naturally the environmental load comes into account when the performance prediction is at stake. Even for sea trials, the estimated calm water performances cannot be evaluated without minimal corrections due to wind and waves when the sea states are not negligeable as shown in *ITTC (2024)*.

In this paper, we focus on the wave resistance estimations in the context of global ship performance estimations. The objective is to explore the existing methods for such estimations, discuss their benefits regarding their cost and their accuracy, build a hybrid model, and evaluate it through statistical analysis on voyage data. For confidentiality reasons, the names and exact dimensions of the ships considered are not given, and results are only provided in terms of percentage of a reference.

2. Wave added resistance models

2.1 Semi empirical model, the SNNM procedure

To predict the added resistance in waves, the first chosen model is the easiest to apply and the requiring the smallest amount of information about the ship geometry. From few geometrical characteristics and the targeted operational conditions, the semi-empirical models provide fast estimations of this wave-added-resistance transfer function. Those models are used to get rough estimations in some studies or during first steps of a ship design process.

In ITTC procedures *ITTC (2024),* several semi-empirical models are proposed for all the ship resistance sources. From the three semi-empirical methods proposed in ITTC, only the semi-empirical SHOPERA-NTUA-NTU-MARIC (SNNM) wave-added-resistance prediction method is presently considered.

SNNM model was proposed in *Liu and Papanikolaou (2016)* and added to the *ITTC (2024)* as a valuable choice for estimating the added resistance in waves from arbitrary heading. A validation of the method is provided in *Wang et. al. (2021)*.

From the geometrical parameters, the SNNM model estimates and a given wave heading α, a transfer function denoted R_{wave} as a function of the wave angular frequency ω. Applying the linear theory approach, for a given seastate with a wave spectrum S_p characterized by its significant wave height H_s and it mean period T_m, the mean added resistance due to this seastate is,

$$R_{AW}(\alpha, H_s, T_m) = 2\int_0^\infty R_{wave}(\alpha, \omega) S_p(\omega) d\omega \quad (1)$$

2.2. Potential flow model

For more accurate predictions of added resistance in waves, the potential flow theory is often chosen as a compromise between accuracy and computational cost. With such an approach, the geometries of the ship are better considered than with semi-empirical method, but the computation time remains significantly smaller that CFD simulations presented in section 2.3.

Various potential-flow methods were developed for decades. In this study, two potential-flow solvers are used. The first one is Hydrostar, developed by Bureau Veritas Marine & Offshore. A description of the theoretical background is given in *Chen (2004)*. The second one is SEACAL developed by the Maritime Research Institute Netherlands (MARIN). Both solvers rely on the boundary element method theory but with different resolution methods.

For a given ship geometry below the free surface, the two solvers give a transfer function of the added resistance that can be used in Eq.(1).

2.3. RANSE-CFD model

2.3.1. Presentation of the method

The most accurate method used in the presented model is the CFD using based on Reynolds Averaged Navier-Stokes Equations (RANSE). This model provides accurate naval simulation in calm water and waves and has been extensively used for decades for ship performance predictions.

Using CFD, the added resistance in waves can be evaluated with two approaches. The first is the direct long-term simulation in irregular waves. It allows the simulation of a specific sea state and provides a direct calculation of the added resistance. However, extrapolating, results from one sea state is almost unfeasible and no added-resistance transfer function can be conveniently defined. Moreover, reaching an accurate estimation requires long-term simulations and consequently high computational costs.

The other approach relies on the added-wave-resistance transfer function already presented above. This approach is largely used in literature as few simulations are needed to build the wave-added-resistance transfer function from which the mean added resistance of any seastate can be estimated. For all the reasons exposed, this second approach is the chosen one in the present study. Aside from the better accuracy of RANSE-CFD regarding to potential-flow method, the error induced by the transfer-function approach remains. A comparison of the added resistance in waves calculated with the direct long-time simulation and with the transfer function approach is given in *Tierno (2024)*.

2.3.2. Application on the studied ships

In the present work, the objective is to reduce the CFD simulation to a minimal number. The solver used in this paper is ISIS-CFD developed by the LHEEA and distributed by CADENCE®. The simulations are conducted with free heave and pitch motions in regular head waves.

The simulations are run on grids about 5 to 8 million of cells in half domain. To ensure the quality of the simulated waves, for a given wavelength and wave height denoted λ and H, the grids discretization verifies: $dx < \lambda/100$ and $dz < H/10$ with dx and dz the dimensions along x and z axes of cells close to the free surface.

3. Hybrid method

3.1. Presentation of the method

The objective of the hybrid method is to limit the computational cost but keeping a good accuracy by combining all the presented methods. The present version of the hybrid method is a low-cost one with only three CFD simulations in head waves. The three CFD simulations are done at three selected wave periods corresponding to frequency at the peak of the response, high frequency and an intermediate one. The frequencies of interest can be estimated using potential flow codes. Because of the high accuracy of the model, these three CFD results are used as reference results for the following of the method.

Fig.1 gives an overview of the hybrid model. It is decomposed into four steps “combination”, “calibration”, regression” and “extrapolation”.

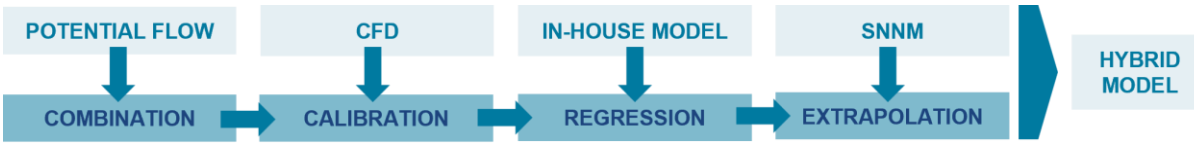


Fig.1: Overview of the hybrid model building

The first step of the method is the calculation wave resistance transfer function from potential flow at various headings and few speed and drafts. Targeting a minimal computational cost, only one speed and draft were simulated with potential-flow code. Due to limitations of the accuracy of the potential-flow code with waves from behind and forward speed, only waves from 0°(head sea) to 90° (beam sea) are simulated. To improve the model accuracy, the simulations are conducted with the two presented potential flow codes and results are combined. It is known that the ship speed and to a lesser extent the draft have a significant impact on the added resistance in waves. By only computing values for one speed and one draft, it is assumed the SNNM model is sufficiently accurate to extrapolate to other speed and draft.

The second step consists in calibrating the combined potential-flow results using CFD results. The calibration factors are determined at 0° of wave heading and applied on potential-flow solutions for other headings. Fig.2 shows the calibration step applied on the computed transfer functions.

The third step consists in simplifying the calibrated transfer functions with least-square regression approach. The resulting transfer function is characterized by only few parameters.

The last step is to use the SNNM model presented in section 2.1. to extrapolate the transfer function to all the desired draft, ship speed and wave heading. Fig.3 shows the transformation of the computed transfer functions following the two last steps.

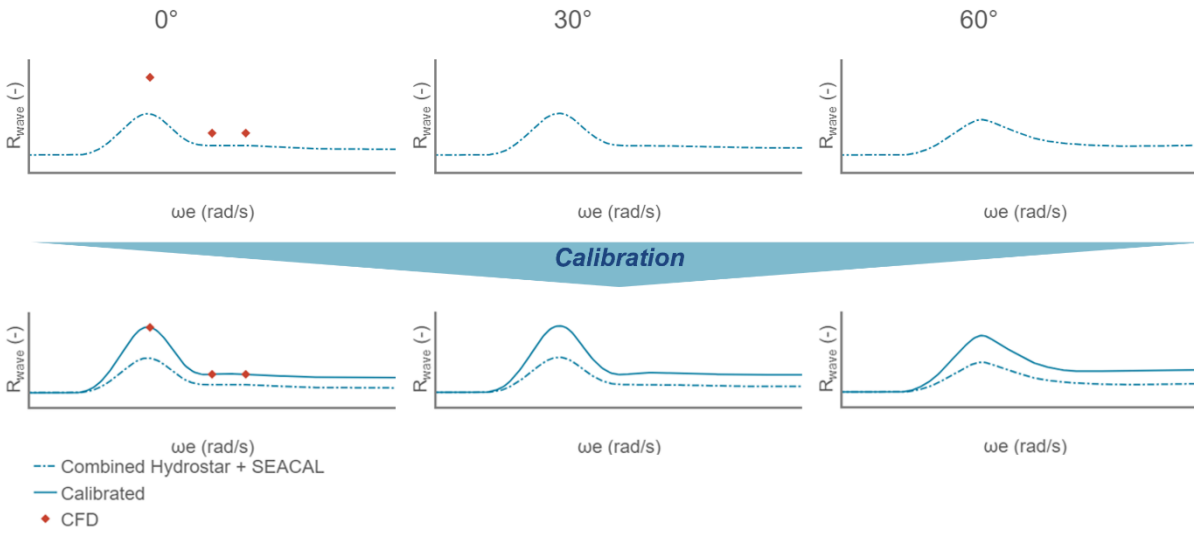


Fig.2: Calibration phase of the hybrid model

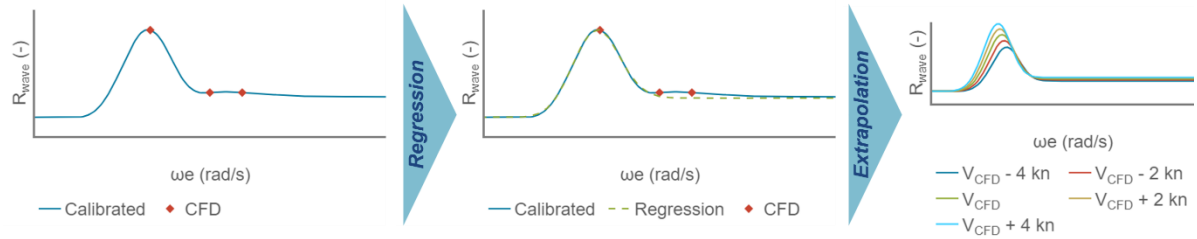


Fig.3: Regression and extrapolation phases of the hybrid model

4. Evaluation of the hybrid model accuracy through statistical analysis

To evaluate the presented hybrid model for wave-added-resistance calculation, it is included into a larger model along with calm water resistance and wind added resistance estimations. The propeller characteristics are also considered.

The model is then applied on voyage data from a ship series to predict the shaft power. The available voyage data are high frequency sampled every 10 minutes. From the model's point of view, all the sister ships of the series should have the same performances so, all the data from all the sister ships are grouped. The total duration of the available data represents 3815 sea going days.

4.1. Voyage data filtering

To statistically evaluate the contribution of calm water, waves and wind into the total ship resistance. The first step is to filter the data according to the quantity available, the physical validity of it and the periods when the ship is navigating in open sea. The filters used can be grouped as follows:

- Presence: The timestamp is valid if the data required by the models are present.
- Physical validity: The timestamp is valid if the data represents a realistic physical quantity ($RPM < RPM_{lim}$, $Speed < Speed_{lim}$, etc…)
- Navigation phase: The timestamp is valid if the ship velocity and RPM are characteristic of a navigation phase and if quantities variations are small. This filter aims avoids manoeuvring phases.

After applying the filters, the total remaining data represents a duration of 1128 days. It corresponds to 29.6% of the unfiltered data.

To keep the ship anonymous, the reference shaft power P_{ref} is defined as the mean shaft power over all the filtered data.

The data repartition is controlled. Fig.4 shows this data repartition regarding to the peak periods (T_p) and wave headings in the filtered dataset. All the wave headings are well represented. Considering the peak period, almost no waves in the dataset have a peak period below 4s and above 18s. The most represented values are around T_p = 11 s.

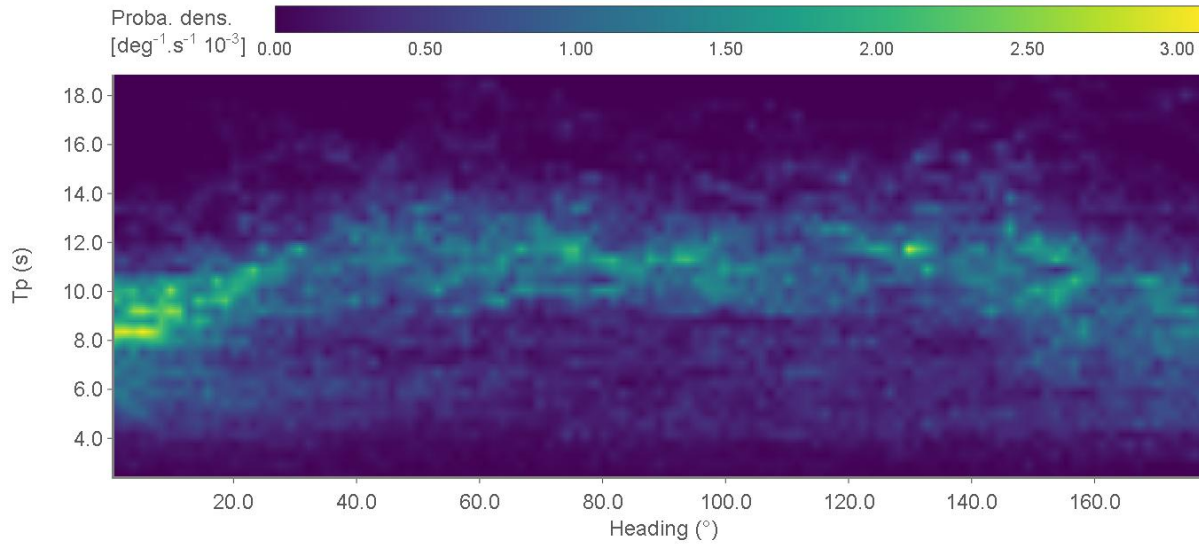


Fig.4: Probability density of combination of peak periods and wave headings in filtered dataset

4.2. Statistical model presentation

To evaluate the model accuracy a statistical approach is used. First a statistical model is built intending to model the effect of various parameters on the shaft power. On one side, the model is fitted to the measured shaft power from the filtered dataset. The resulting statistical model is denoted D-STAT in the following. On the other side, the model is fitted on the shaft power predicted by the hybrid model on the same dataset. The resulting statistical model is denoted H-STAT in the following. In the following, the two fitted statistical models are compared one to each other and to the direct hybrid method.

4.2.1. Statistical model equations

In this study, the statistical model is built with regression based on the least-square method. The total power law is as follows:

$$Pw = Pw_{ClamWater} + Pw_{Wind} + Pw_{Wave} \quad (2)$$

With $Pw_{ClamWater}$, Pw_{Wind} and Pw_{Wave}, the power induced by the calm water resistance, wind-added-resistance and wave-added-resistance, respectively. The wind effort is often correlated with the wave in voyage data, and its impact on the power is about the same order. The risk of transferring a part of the wind-added-resistance into the wave-added-resistance is to keep in mind.

In the present the focus is set on the model of added power due to waves Pw_{Wave}. The inputs of this wave-added-power model are:

- the ship speed V,
- the water density ρ,
- the significant wave height for swell $H_{s,swell}$,

- the significant wind wave height $H_{s,ww}$,
- the mean swell period $T_{m,swell}$,
- the mean wind wave period $T_{m,ww}$,
- the swell direction α_{swell},
- the wind wave direction α_{ww}.

The statistical model is based on some standard equations modeling the contribution of each phenomenon to ship resistance.

Large bias is induced by the choice of the model itself. This bias must be kept in mind when comparing other results with the ones from the statistical model. The model strongly constrained the results and can either force some power variation where it should not or, on the other side, make some behaviour impossible to catch. More complex model (eventually requiring deep neural network) can be chosen to get a better fit, but it would not necessarily ease the comparison with the other methods presented in the following. The presented model was chosen as simple as possible, with physically identifiable parameters, but still succeeding to estimate the added power due to wave.

Another aspect of the model precision regarding to the data is the data uncertainties itself. In this paper the quantification of the data uncertainties is not explored but some insight can be provided. The most significant part of the total ship resistance is the calm water resistance. The calm-water resistance prediction is strongly correlated with the ship speed. The ship speed is here an input value, but it includes uncertainties from speed loc sensors. The second source of major uncertainties are the wind and wave hindcast.

The equation of the chosen statistical model for the wave's contribution is as follows:

$$Pw_{Wave} = Pw_{ref} \cdot \left(\frac{V}{V_{ref}} - \frac{p_1}{2}\right) \cdot \frac{\rho}{\rho_{ref}} \cdot \left(\frac{H_s}{H_{s,ref}}\right)^{2p_2} \cdot \left[g(\alpha) e^{\left(-2 \frac{\left(\frac{T_m}{T_{ref}} - T_{m0}(\alpha)\right)^2}{p_9} \right)} + f_a(\alpha) \right] \quad (3)$$

with,

$$T_{m0}(\alpha) = p_{10} + \frac{p_{11}}{4} \left| \alpha - \frac{\pi}{2} \right| \quad \text{Eq.4}$$

$$g(\alpha) = \begin{cases} f_b(\alpha), & if\ T_p \le T_{p0}(\alpha) \\ f_b(\alpha) - f_a(\alpha), & if\ T_p > T_{p0}(\alpha) \end{cases} \quad \text{Eq.5}$$

$$f_a(\alpha) = a_2\alpha^2 + a_1\alpha + a_0 \quad \text{Eq.6}$$

$$where \begin{cases} f_a(0) = p_3 \\ f_a\left(\frac{\pi}{2}\right) = p_4 \\ f_a(\pi) = \frac{-p_5}{2} \end{cases}$$

$$f_b(\alpha) = b_2\alpha^2 + b_1\alpha + b_0 \quad \text{Eq.7}$$

$$\begin{cases} f_b(0) = 4p_6 \\ f_b\left(\frac{\pi}{2}\right) = 2p_7 \\ f_b(\pi) = \frac{p_8}{2} \end{cases}$$

For simplicity, the wave characteristics are only denoted by the wave heading α the significant wave heigh H_s and the mean period T_m. When the model is fitted Eq.(3) is applied on both the wind wave and swell keeping the same parameters p_i. In all the above equations, some factors (2,4, 1/2,1/4) are applied. These factors are used to scale the parameters p_i and keep the fitted values close to 1. V_{ref}, $H_{s,ref}$ and T_{ref} represent the chosen reference speed, wave height and wave mean period, respectively.

4.2.2. Statistical model applications

The statistical model is first applied on the entire filtered dataset using the measured shaft power on one side and the calculated one from hybrid model on the other side. In the following , the resulting two statistical models are denoted D-STAT Global and H-STAT Global. The D-STAT model can be seen as the targeted values as it represents the measure data when H-STAT model represents a validator both the hybrid model and the statistical model itself. If the fitted parameters from the H-STAT model are in line with the implemented hybrid model, it means that the statistical model is accurate enough to catch the underlaying wave model. On the other hand, if the fitted H-STAT model is far from the D-STAT it is a strong indicator of the bad quality of the hybrid model.

The evaluated standard deviations of the error are about 8.7% for the model fitted on the measured shaft power and 2.5% for the one fitted on hybrid-model results. Getting a smaller standard deviation with the model fitted on a calculated shaft power was expected but the remaining standard deviation reveals that the chosen statistical model cannot perfectly fit the hybrid model. Besides, the difference between the two standard deviations is representative of the combination of the uncertainties inside the data itself and badly modeled phenomenon in both the statistical approach and the hybrid model.

To control the consistency of the statistical model and evaluate the sensitivity of each parameter, the dataset is divided into 50 random sub-sets, and the least-square method is applied on each sub-set.

Fig.5 shows views of the evaluated probability density of the values of each p_i parameters from the statistical models D-STAT (from data) and H-STAT (from hybrid model). The probability density is identified by colored zones. The values from the statistical models fitted on the entire dataset D-STAT global and H-STAT global are identified by vertical dotted lines.

For some parameters, the distributions are significantly narrower than for others showing the lower sensitivity of some parameters compared to the other in Eq.3. From these first results, D-STAT and H-STAT are in line for some fitted parameters (p_1, p_2, p_5, p_6, p_{10}) but far apart on others (p_3, p_4, p_7, p_8, p_{11}). We can also notice that the probability distribution is globally narrower with H-STAT than with D-STAT. This is representative of the expected higher uncertainty in the statistical model based on the voyage data than in the statistical model based on computed shaft power from equations of the hybrid model.

For each parameter, the distance between the expected value from the hybrid model and the evaluated value from D-STAT and H-STAT can be discussed. For instance, from the probability distribution of p_2, both the D-STAT and H-STAT tend to a value close to 1 with p_2 = 1.0 for H-STAT global and p_2 =1.1 for G-STAT global. Regarding Eq.3, it means that H-STAT predicts a power of Hs about 2.0 and D-STAT a power of Hs about 2.2. Such results show that the statistical model succeeds in catching the expected value of 2 (that is the implemented one) fitting the results from the hybrid model. It also shows that the identified dependency between the added resistance in waves and the significant wave heigh in the voyage data is close to 2.

For other parameters the most interesting thing to notice is that the most discrepancies between H-STAT and D-STAT are encountered in the function's f_a from Eq.6 and f_v from Eq.7. Those functions and the associated parameters (p_3, p_4, p_5, p_6, p_7, p_8) characterize the impact of the wave heading on the amplitude of the added power due to waves. It reveals either that the build hybrid model tends does not model well the true data or that the statistical model is inefficient for the estimation of this wave-heading dependency.

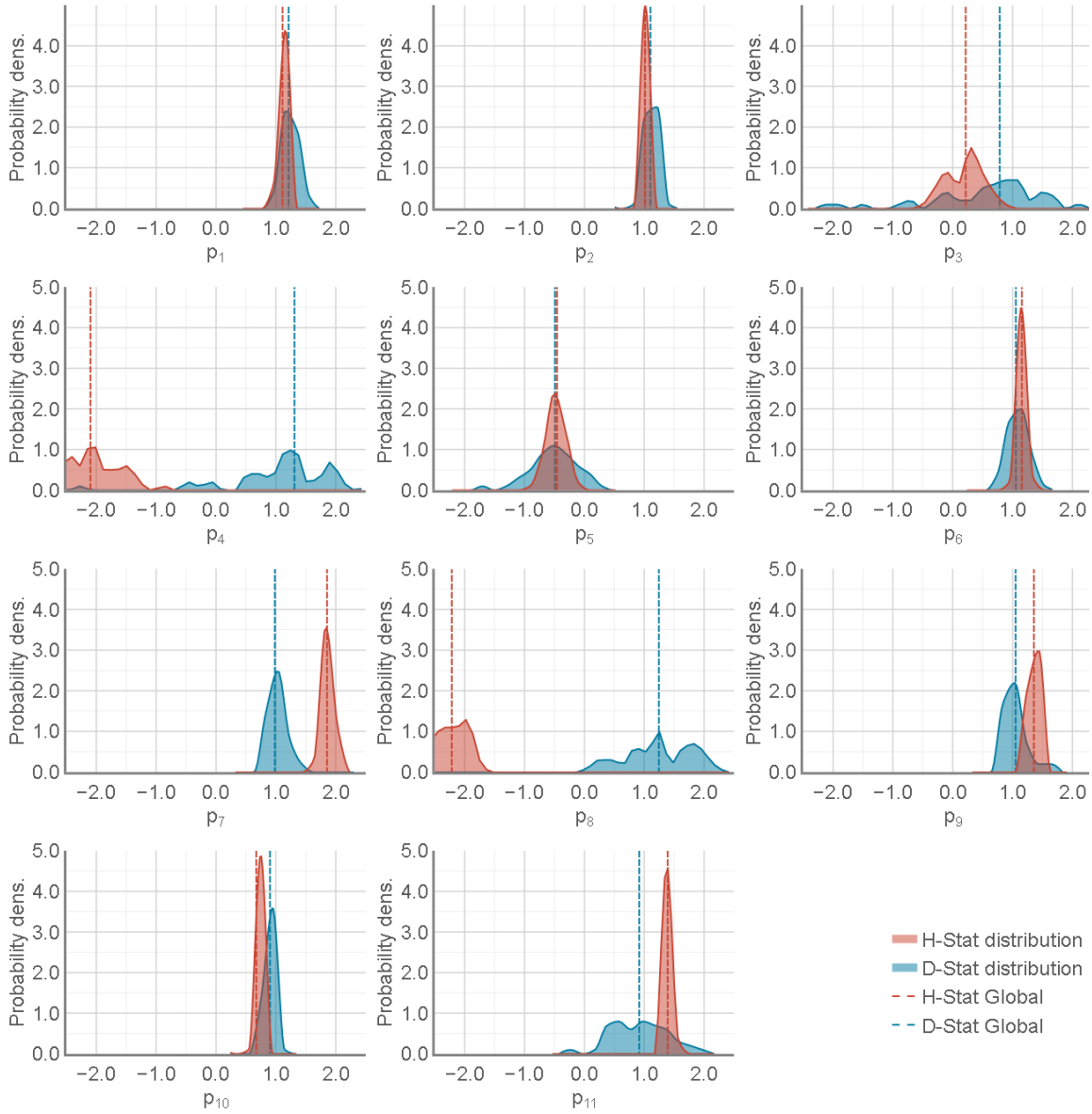


Fig.5: Probability distribution of fitted wave-model parameters on the shaft power from the dataset and the one calculated with the hybrid model.

To explore more the effects of the identified discrepancies in the fitted parameters, Fig.6 shows views of reconstructed wave added power as a function of the wave peak periods for several wave headings.

The blue zone represents a confidence interval defined as the mean value get from the 50 statistical models D-STAT fitted on the dataset plus or minus the standard deviation of these 50 values. The red zone is the same but considering the statistical models H-STAT fitted on the hybrid-model result. The solid blue and red lines represent the calculated values with the global models D-STAT Global and H-STAT Global. The green line is the direct result from the hybrid model.

The figure shows that the statistical model is only roughly accurate to fit the expected hybrid model results. This is due to the simplicity of the statistical model and to the data distribution. However, the confidence interval is significantly smaller for the H-STAT model than with the D-STAT one. This is directly correlated to the observation made on Fig.5.

The identified discrepancies between D-STAT and H-STAT regarding the wave heading are confirmed here. For the presented headings, there is a relatively good catch of the hybrid model prediction by the H-STAT model generally between $T_p = 10$ s and $T_p = 18$ s. Putting aside 0° of wave heading, the statistical model fitted on the voyage data is far from the one fitted on the hybrid model results and for all the results, D-STAT is lower than H-STAT. It tends to show that the hybrid model overestimates the added power due to wave resistance in this T_p range with a more significant over estimation between heading from 30° to 120°.

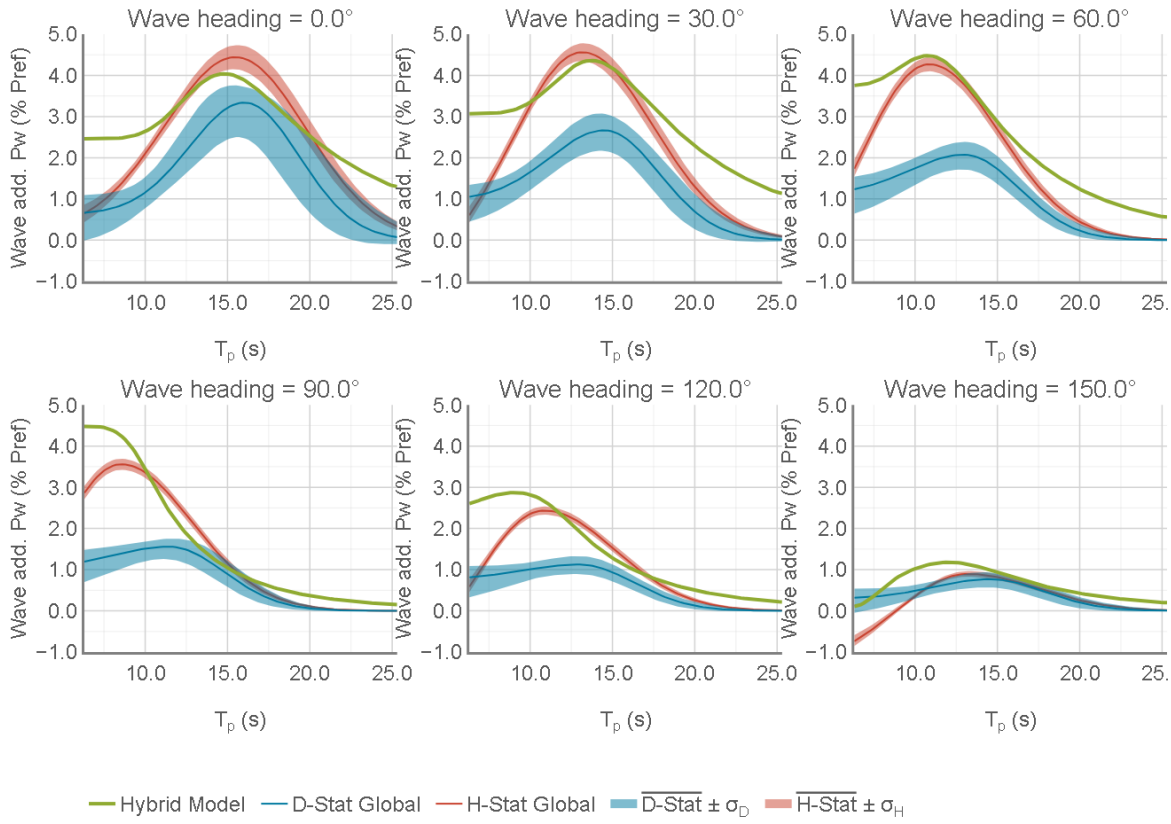


Fig.6: Calculated added power in waves as a function of the peak period

To confirm this analysis, Fig.7 shows the views of reconstructed wave added power as a function of the wave heading for several peak periods.

As in Fig.5, the blue and red zones and represent the confidence interval defined as the mean value plus or minus the standard deviation for the D-STAT and H-STAT models. The solid blue and red lines represent the calculated values with the global models D-STAT Global and H-STAT Global. The green line is the direct result from the hybrid model.

The good catch of hybrid model results by the statistical model for T_p around 10s to 18s is clearly noticeable on this figure as well as the significant overestimation of power by the hybrid model compared to the voyage data. This overestimation is particularly identifiable for $T_p = 10$ s and for wave heading around 90°.

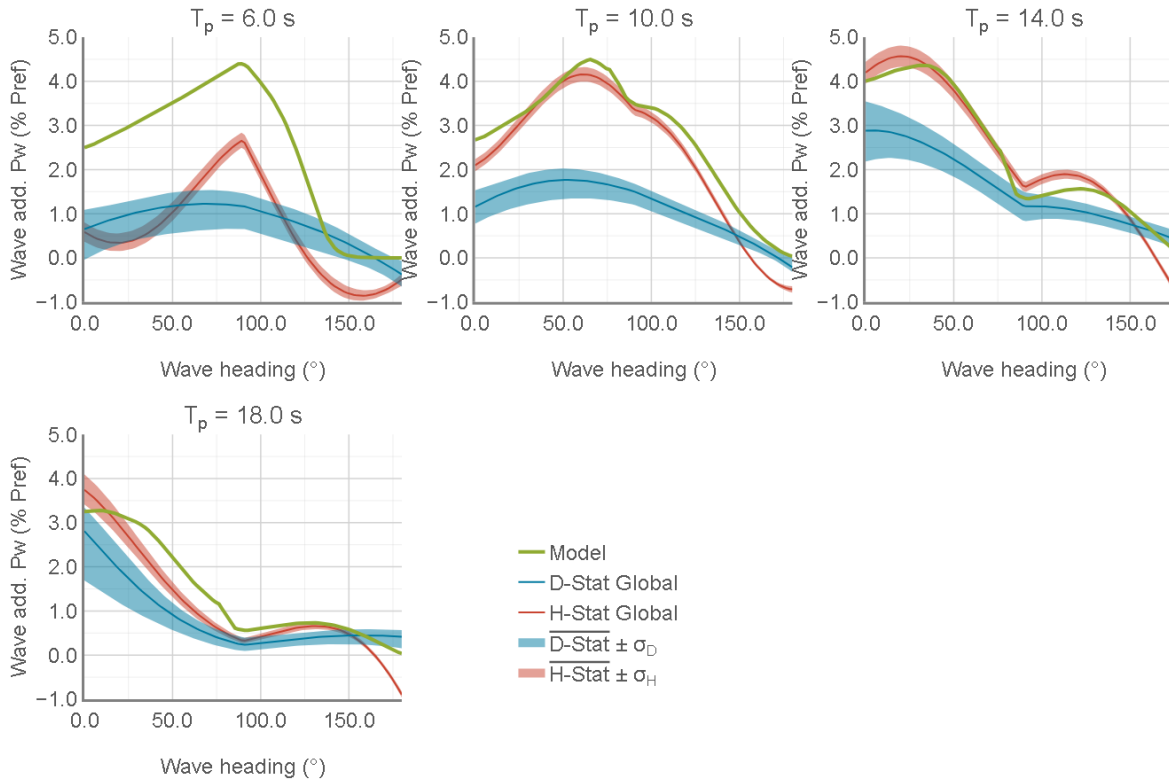


Fig.7: Calculated added power in waves as a function of the wave heading

5. Conclusion

In this paper, a wave added-resistance hybrid model has been presented. The objective was to find a compromise between computational time and accuracy. To evaluate the model a statistical approach was developed and applied to a ship series.

It has been shown that the statistical analysis was accurate enough to draw conclusions on the validity of the tested hybrid model. Several parameters defined in the hybrid model were validated and some other were clearly identified as badly defined.

The main observation is that the hybrid model tends to overestimate the added power and the closer the wave heading is to 90°, the larger is the error between the hybrid model and the statistical one. This is all the most noticeable with a wave peak period close to 10 s. For small peak period, no clear conclusion can be drawn as the statistical model do not catch well the expected values.

A reason that might explain this overestimation, is the impact of the motion (pitch and roll) on the wave resistance and its bad capture by the potential-flow results calibrated with CFD results in head waves. Limiting the number of CFD computation to only three is m too restrictive and simulation in oblique and beam seas would be needed for future investigation on this ship series.

Finally, statistical procedure exposed in this paper will help us to improve the presented hybrid model by progressively adding accurate computations where the needs are identified.

Acknowledgement

Thanks to Stephane and Eliott for their help on all the project and more particularly on the voyage data collection and its filtering.

References

CHEN, X.B. (2004), *Hydrodynamics in offshore and naval applications*, 6th Int. Conf. Hydrodynamics, Perth

ITTC (2024), *Preparation, Conduct and Analysis of Speed/Power Trials,* ITTC Recommended Procedure 7.5-04-01-01.1, Revision 08, Int. Towing Tank Conf.

LIU, S; PAPANIKOLAOU, A. (2016), *Fast approach to the estimation of the added resistance of ships in head waves,* Ocean Eng. 112, pp.211-225

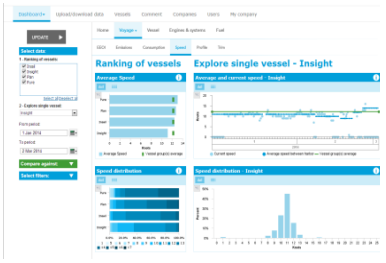
WANG, J.; BIELICKI, S.; KLUWE, F.; ORIHARA, H.; XIN, G.; KUME, K.; OH, S.; LIU, S.; FENG, P. (2021), *Validation study on a new semi-empirical method for the prediction of added resistance in waves of arbitrary heading in analyzing ship speed trial results,* Ocean Eng. 240

TIERNO, M.; DESCAMPS, T.; DUCROZET, G.; BOUSCASSE, B. (2024), *Evaluation of the added resistance of a containership in irregular sea state through RANSE computations in regular and irregular waves,* 35th Symp. Naval Hydrodynamics, Nantes

Index by Authors

12th Hull Performance & Insight Conference (HullPIC)

Tutzing/Germany, 19-21.4.2027




Topics: ISO 19030 and beyond / sensor technology / human factors in reporting / information fusion / big data / sea trials / uncertainty analysis / hydrodynamic models / business models / CII management

Organiser: Volker Bertram (VB conferences)
Morten Sten Johansen (Jotun)

Advisory Committee:

Gerry Docherty	Ardmore Shipping	**Erik Hagestuen**	Danelec	**Casimir Morobé**	Toqua
Falko Fritz	Albis MP	**Søren V. Hansen**	VesOPS	**Richard Marioth**	Idealship
Inno Gatin	Cloud Towing Tank	**Martin Köpke**	Hapag-Lloyd	**Ivana Melillo**	GNV
Wojciech Gorski	Enamor	**Manolis Levantis**	Jotun	**Geir Axel Oftedahl**	UNeFAi
Ditte Gundermann	Hempel	**Carsten Manniche**	Navigator Gas	**Beom-Jin Park**	KRISO
				Sofia Werner	RISE SSPA

Venue: The conference will be held in the Certosa di Pontignano.

Format: Papers to the above topics are invited and will be selected by a selection committee.
The proceedings will be made freely available to the general public.

Deadlines:

anytime	Optional “early warning” of interest to submit paper / participate
20.11.2026	First round of abstract selection (1/3 of available slots)
04.12.2026	next 1/3 selection
20.12.2026	Final round of abstract selection (remaining slots)
28.02.2027	Payment due for authors
28.02.2027	Final papers due

Admin. Fees: **900 €** – early registration (by 6.1.2027)
1000 € – late registration

Fees are subject to VAT

Sponsors: Jotun, Tutech Innovation, COACH Solutions, Enamor, Hempel, Hoppe Marine, Idealship, Miros, Veinland, Vessel Performance Solutions, *more tbd*

Media Partners: Hansa, The Royal Institution of Naval Architects

Information: volker@vb-conferences.com

Operators: tbd