How to consistently select the right ship performance model in a fleet with mixed data availability

Camille Colle, Toqua, Gent/Belgium, camille@toqua.ai
Casimir Morobé, Toqua, Gent/Belgium, casimir@toqua.ai
Lukas Demey, Toqua, Gent/Belgium, lukas@toqua.ai

Abstract

Modeling ship performance can be done in many different ways. The spectrum includes purely theoretical formulas, purely data-driven models, and everything in between. With different data available for different vessels, how does one make the right choice for a whole fleet? This paper proposes a framework to select the best model in a consistent way, over a whole fleet where certain vessels may or may not have sea trial data, model tests, noon reports, sensor data, etc.

1. Introduction

Efficiency gains are the go-to answer to reach short-term decarbonization targets in shipping. To capture these efficiency gains, accurate speed-fuel models of vessels are a prerequisite. The challenge of creating accurate speed-fuel models - also called ship performance models - holds many layers of complexity from a theoretical point of view: different speeds, different drafts, different weather conditions, changing hull performance, changing engine performance, etc. However, in recent years the rise of sensor data and data-driven modeling has shown great promise to overcome these theoretical challenges, *DeKeyser et al. (2022)*.

Unfortunately, today, the potential of applying data-driven technologies such as machine learning to ship performance modeling remains largely untapped in the maritime industry. Not due to theoretical reasons, but for practical reasons. There's too much heterogeneity in the data across a fleet for a single type of data-driven model to provide consistently great results. As a result, simpler, traditional approaches are used to ensure consistency. This leaves the potential of big data and machine learning on the table.

This paper explores a practical framework to capture the full modeling potential across a data-heterogenous fleet, to always deliver the best model possible given the available data.

2. Heterogeneity in ship performance data across a fleet

There is an endless list of causes for heterogeneity in performance data. This paper initially focuses on a single cause for heterogeneity: different data types (public data, design data, noon report data, sensor data). After tackling heterogeneity due to different data types, section 6. explores three additional sources of heterogeneity and how orchestrations can overcome them. Many other sources of heterogeneity remain undiscussed within this paper, as it would lead us too far.

Heterogeneity due to different data types

This paper identifies 4 fundamental 'types' of data that can be used to model vessel performance:
1. Sensor Data (SD): High-frequency data collected onboard using sensors.
2. Noon Reports (NR): Daily manual reports.
3. Design Data (DD): Seatrial curves, shop tests, etc.

4. Public Data (PD): Anything that can be publicly retrieved based on IMO number such as vessel type, DWT, LOA, etc.

This paper assumes a fleet of 10 vessels with mixed data types according to the randomly selected distribution below. For some vessels only a single source of data is available, for others there can be multiple sources of data. The goal is to represent a realistic amount of heterogeneity as can occur operationally in the industry today. Public data is left out of scope.

Vessel ID	**Design Data Available**	**Noon Report Data Available**	**Sensor Data Available**
V1	X	X	
V2		X	
V3	X	X	X
V4		X	X
V5	X		
V6	X	X	X
V7		X	X
V8		X	X
V9		X	
V10			X
Coverage	4/10	8/10	6/10

3. Different model options

Different data types require different modeling techniques. Design Data (DD) is typically combined with traditional formula-based and filter-based approaches (ISO15016, DNV VTI). Noon Reports (NR), due to their operational nature, can be valuable for assessing different conditions and tracking performance changes over time. Yet, extreme caution is required when using NR data for data-driven techniques given the data is infrequent and error-prone, *Collé and Morobé (2022)*. Sensor Data (SD) is suitable for data-driven techniques such as machine learning, but always requires extreme caution to safeguard data quality.

This paper applies the following techniques to the following scenarios:
1. Design Data: Seatrial data and Main Engine Shop Test data are combined using a variation of ISO15016 that allows for the modeling of different operational conditions.
2. Noon Reports: A combination of physics-based and data-driven methods.
3. Sensor Data: A proprietary version of physics-informed machine learning.

4. Validating model accuracy

To guarantee an objective and consistent way of evaluating model accuracy over different approaches, the 'Blue Modeling Standard' is applied, *Deschoolmeester and Morobé (2023)*. The most important details are summarized below.

What data is considered the ground truth?
Sensor data with Speed-Over-Ground above 5 knots is used to validate model accuracy. Operational sensor data of good quality is available for all 10 vessels. Following the scenarios listed in Table 1, sensor data is frequently not used to train the model. However, it is always used to validate model accuracy, to ensure consistent and representative results.
What model validation technique is used?
A fit-for-purpose k-fold cross-validation technique is applied, preventing leakage and guaranteeing independent and identically distributed random variables over the folds.
What relationship is modeled?
Main Engine Fuel consumption is modeled using SOG as input. Secondary variables such as draft and weather conditions are also used.
What time horizon is used for the accuracy?
Daily. So the predicted daily consumption is compared to the actual.
What accuracy metric is used?
Mean Absolute Percentage Error (MAPE) is applied at daily intervals. This combination is also referred to as 'MADPE' (Mean Absolute Daily Percentage Error). (See 'Blue Modeling Standard' for more details on the accuracy metrics.)

5. Consistently selecting the best option: results

With a system in place to continuously assess model accuracy against the latest operational data, it's possible to compare different modeling approaches for a single vessel, and then select the most accurate option. The below table does this for 10 vessels using different data types. If multiple options are available, the 'Orchestration' ensures the best model is selected and made operational.

Mean Absolute Daily Percentage Error (MADPE) per scenario				
	Design Data-based	**Noon Report-based**	**Sensor Data-based**	**Orchestration**
V1	18%	9%		9%
V2		8%		8%
V3	13%	7%	4%	4%
V4		17%	6%	6%

V5	16%			16%
V6	26%	14%	9%	9%
V7		11%	5%	5%
V8		14%	4%	4%
V9		13%		13%
V10			5%	5%
Coverage	**4/10**	**8/10**	**6/10**	**10/10**
Avg. MADPE	**18.3%**	**11.6%**	**5.5%**	**7.9%**

It can be observed that 'orchestration' over different data types, for this specific case, has two major benefits. First of all, the coverage (=vessels that can be modeled) of the fleet increases to 10/10. Approaches based on only a single data type, would have to leave some vessels unserved. Secondly, the average accuracy also improves considerably. If we were to only use NR-based models because it has the largest coverage, the average daily error would be 11.6%. Because Orchestration enables to benefit from sensor data - when available - the error drops to 7.9% on average, and even to 5.5% on average for sensor-based vessels, while still guaranteeing a 10/10 coverage.

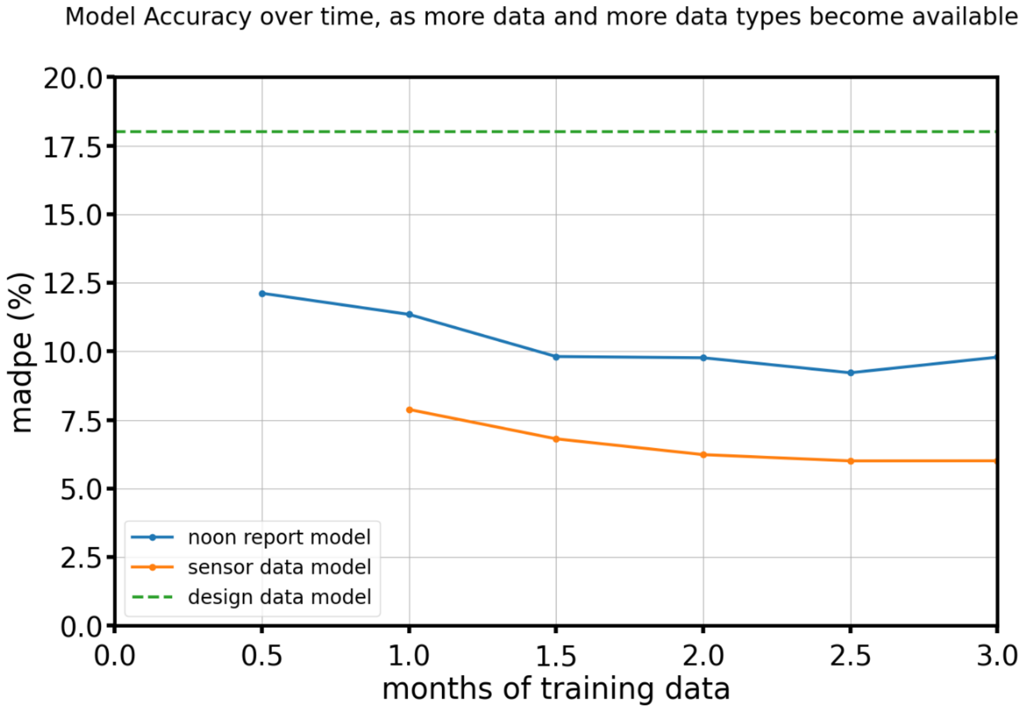
6. Other forms of orchestration to solve for heterogeneity

This section explores three other sources of heterogeneity present in performance data & performance modeling: changes over time, different modeling approaches, and data quality issues. It also suggests how orchestration can overcome these challenges.

6.1 Updates over time

The above exercise for different data types is an oversimplification, as it doesn't account for time. Over time, different data types become available, and for operational data sources such as NR and SD more and more data continuously becomes available. These changes over time in available data types and available data duration, will continuously alter what modeling approach is the most accurate one. As a result, the orchestration exercise above, should be repeated frequently, to ensure the best possible model is always available.

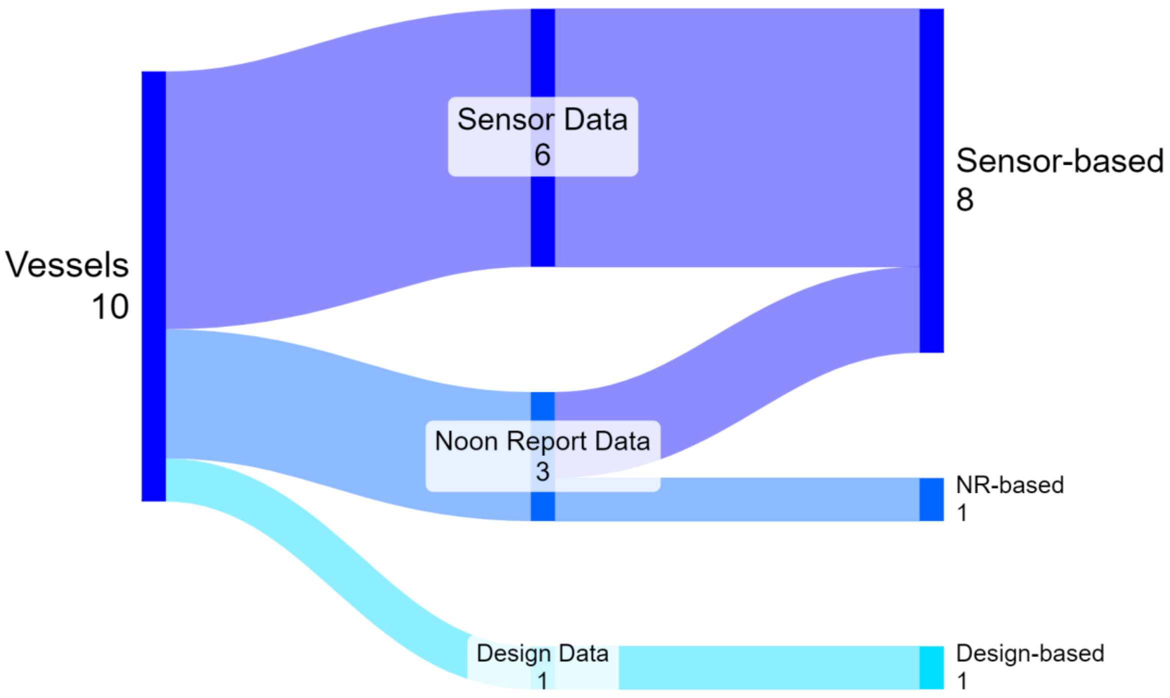
The below graph does exactly that for a vessel that initially only has Design Data, then gets access to Noon Reports after 2 weeks, and gets access to Sensor Data after 1 month. Every time a new data source becomes available, a more accurate model is deployed and used operationally.



6.2 Different modeling approaches within the same data type

There is no single model that is always the best choice - even within a certain data type, the best modeling approach might differ depending on many factors. For example, sometimes it can be beneficial to use data over multiple vessels to improve modeling accuracy. Below we explore a case where the 'Augmented Approach' is applied, *Collé and Morobé (2022)*. This approach takes sensor-based learnings from similar vessels in the fleet, and transfers those modeling insights to vessels with only Noon Reports. This enables the creation of a model that is much more accurate than just a NR-based model, as it also incorporates the sensor-based insights from similar vessels in the fleet.

For the fleet of 10 vessels explored in this paper, 2 out of the 3 NR-based models can benefit from this different modeling approach. Meaning that this type of orchestration improves accuracy for those 2 out of 3 vessels, by leveraging the most suited modeling approach within that data type. As a result, even though there is only sensor data available for 6/10 vessels, eventually 8/10 vessels benefit from that sensor data. This allows the error to drop by 2% and 7% for those respective vessels, a meaningful improvement.

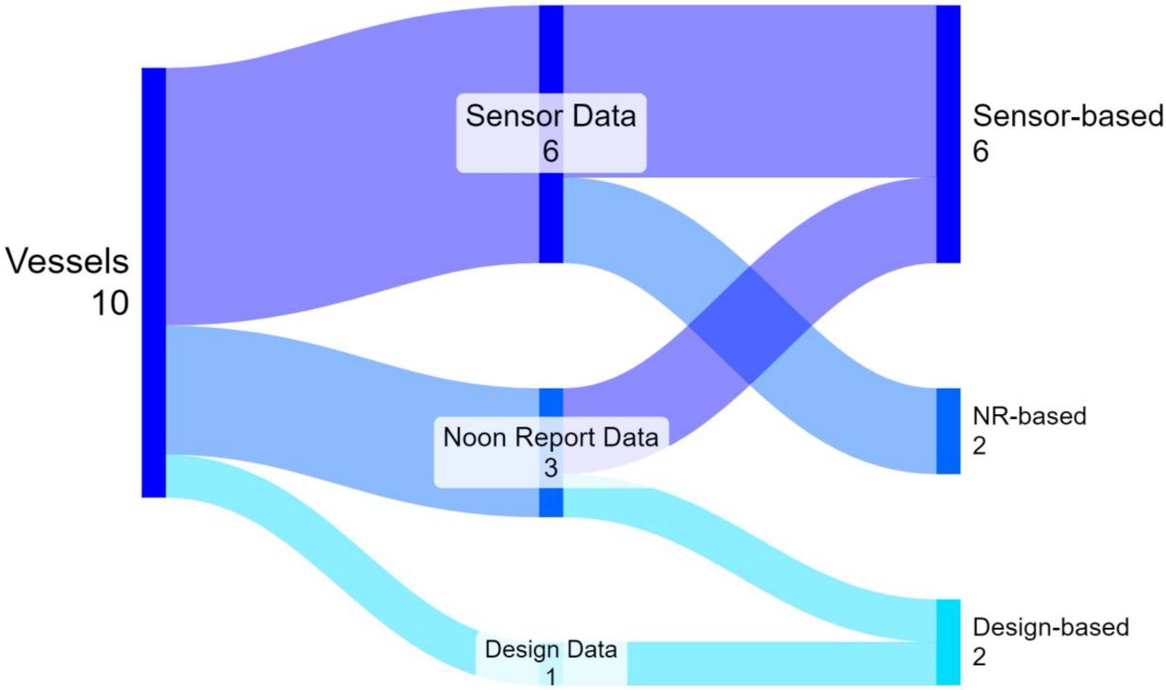


Mean Absolute Daily Percentage Error (MADPE) per scenario				
	Design Data-based	**Noon Report-based**	**Sensor Data-based**	**Orchestration**
V1	18%	9%		9%
V2		8%	*6%	6%
V3	13%	7%	4%	4%
V4		17%	6%	6%
V5	16%			16%
V6	26%	14%	9%	9%
V7		11%	5%	5%
V8		14%	4%	4%
V9		13%	*6%	6%
V10			5%	5%
Coverage	**4/10**	**8/10**	**6/10**	**10/10**
Avg. MADPE	**18.3%**	**11.6%**	**5.6%**	**7.0%**

6.3 Data Quality Issues

So far the table has always assumed the available data is free from data quality issues. But in practice, NR-data and Sensor Data are often plagued by data quality issues throughout time. If these are not flagged and resolved, this can have a very negative impact on model accuracy, *Colle et. al (2023)*. Below we assume a scenario where one vessel experiences unreliable noon report data, and another two experience unreliable sensor data.

Once the issues are detected, the best alternative modeling options are selected. For V1 with NR data quality issues, a Design Data-based model is selected instead. For V3 and V7 with sensor data issues, an NR-based model is selected instead.



If the data quality issues had remained undetected, it would have increased inaccuracy considerably for those specific vessels. For example, V3 would suddenly have an error of 20%. The average fleet error would have increased to 12.1%. After detecting the issues and redirecting to the best alternative modeling method with reliable data, the average error was reduced to 8.8%. For example for V3 specifically, the inaccuracy drops from 20% to 7%.

Mean Absolute Daily Percentage Error (MADPE) per scenario				
	Design Data-based	**Noon Report-based**	**Sensor Data-based**	**Orchestration**
V1	18%	**31%		~~31%~~ 18%
V2		8%	*6%	6%
V3	13%	7%	**20%	~~20%~~ 7%
V4		17%	6%	6%
V5	16%			16%
V6	26%	14%	9%	9%

V7		11%	**18%	~~18%~~ 11%
V8		14%	4%	4%
V9		13%	*6%	6%
V10			5%	5%
Coverage	**4/10**	**8/10**	**6/10**	**10/10**
Avg. MADPE	**18.3%**	**14.4%**	**9.3%**	**~~12.1%~~ 8.8%**

7. Results

In the below table we compare the effect of all the different types of orchestration. The orchestration of different data types, has a big effect and reduces the average daily error from 11.6% to 7.9% over the mixed fleet assessed in this paper. The second type of orchestration enables the leveraging of different modeling types within a single data type and reduces inaccuracy by 0.9% on average across the fleet. The third type of orchestration, handling data quality issues, reduces the inaccuracy by 3.3%. In total, an average fleetwide improvement of ~8% is realized through orchestration. It's important to stress this paper considers only a very limited amount of very simple orchestration processes. There is much more potential in more numerous and more advanced processes to tackle heterogeneity.

Average Fleet Modeling Accuracy (MADPE)			
	Orchestration v1 (Data Type)	**Orchestration v2 (Model Type)**	**Orchestration v3 (Data Quality)**
Before	11.6%	7.9%	12.1%
After	7.9%	7.0%	8.8%
Improvement	**3.7%**	**0.9%**	**3.3%**

8. Conclusion

This paper explores the potential of orchestration to tackle the heterogeneity present across performance data and modeling approaches within the domain of ship performance modeling. Even though only a few sources of heterogeneity are addressed within this paper, and fairly simple orchestration solutions are proposed, the benefits are clear. To capture the full potential of the data across a fleet, one must look beyond a single modeling approach and data type, and develop a holistic fleet-wide approach that's able to address and overcome the different sources of heterogeneity. Otherwise, the potential of sensor-derived big data and machine learning to decarbonize the shipping industry will remain a theoretical construct.

References

DEKEYSER, S.; MOROBÉ, C.; MITTENDORF, M. (2022), *Towards Improved Prediction of Ship Performance: A Comparative Analysis on In-service Ship Monitoring Data for Modeling the Speed-Power Relation*, https://arxiv.org/pdf/2212.13061.pdf

ISO (2015), *ISO 15016:2015 - Ships and marine technology — Guidelines for the assessment of speed* *and power performance by analysis of speed trial data*, Int. Standard Org., Geneva

COLLE, C.; MOROBÉ, C. (2022), *Saving AI from its own hype: Getting real about the benefits and challenges of machine learning for ship performance modelling aimed at operational optimizations*, 7th HullPIC Conf., Tullamore, https://toqua.ai/uploads/default/HullPic2022_Toqua_Proceedings.pdf

DESCHOOLMEESTER, M.; MOROBÉ, C. (2023), *Blue Modeling Standard: a Benchmarking Methodology for Ship Performance Models*, Whitepaper, Toqua, https://toqua.ai/whitepapers/how-to-compare-different-ship-performance-models

COLLE, C.; DESCHOOLMEESTER M.; MOROBÉ, C. (2023), *STW or SOG as a Starting Point for Performance Modeling? An Empirical Study using Operational Sensor Data from 20 Oil Tankers*, 8th HullPIC Conf., Pontignano, https://toqua.ai/whitepapers/sog-or-stw-as-a-starting-point-for-performance-modeling