STW or SOG as a Starting Point for Performance Modeling? An Empirical Study using Operational Sensor Data from 20 Oil Tankers

Camille Colle, Toqua, Gent/Belgium, camille@toqua.ai
Casimir Morobé, Toqua, Gent/Belgium, casimir@toqua.ai
Michaël Deschoolmeester, Toqua, Gent/Belgium, michael@toqua.ai

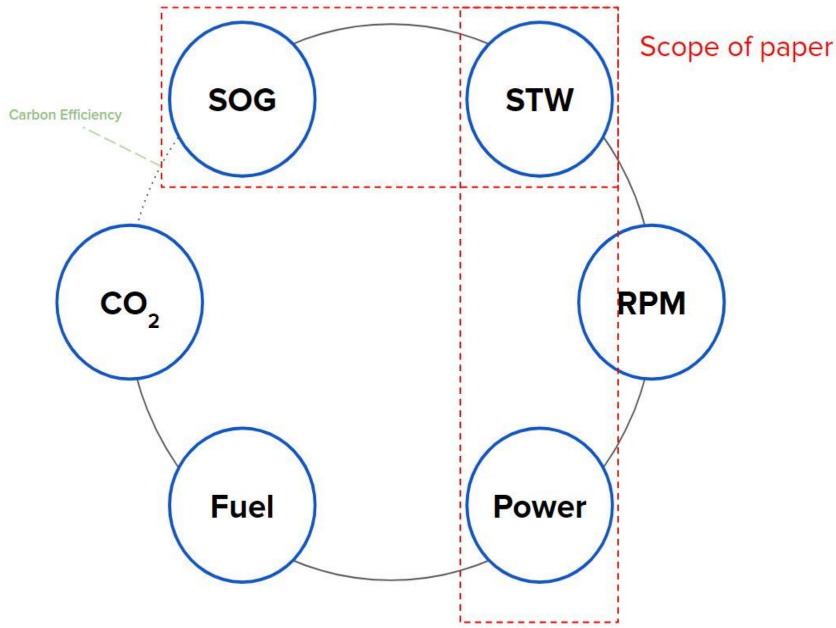
Abstract

This study details data-driven findings based on actual operational high-frequency sensor data of over 20 Crude Oil Tankers owned by Euronav, with the goal of deciding between STW (Speed-Through-Water) or SOG (Speed-Over-Ground) as a starting point for accurate ship performance modeling.

1. Introduction

Efficiency gains are the go-to answer to reach short-term decarbonization targets in shipping. Accurate speed-fuel models of vessels are a prerequisite to capture these efficiency gains. The challenge of creating accurate speed-fuel models - also called ship performance models - lies not only in accounting for all the secondary factors influencing this relationship (waves, wind, currents, draft, trim, water depth, etc.), but also in getting accurate data on the crucial variables speed and fuel.

The rise of telemetry equipment and high-frequency data collection on-board vessels has enabled many improvements for ship performance modeling, *DeKeyser et al. (2022)*. Nevertheless, with sensor data, an even more critical mindset is necessary to decide what data can be trusted. Especially when it comes to the speed of the vessel, a dilemma often ensues to choose for Speed-Over-Ground (SOG) data based on GPS-location or to choose for Speed-Through-Water (STW) data based on the speed log.



This study analyzes data from 20 oil tankers (V1-V20) to find a data-driven answer to the above dilemma. Should we use SOG or STW as a starting point for performance modeling? What options do we have and how can we maximize performance modeling accuracy? The 20 vessels are VLCC's and Suezmax's. On average we analyzed 1 year of sensor data for every ship. The data consists of measurements at 5-minute intervals.

2. Difference between SOG and STW

In theory the difference between SOG and STW should only be due to currents. As a result, the following formula is often used to convert SOG into STW:

$$\mathrm{STW} = \mathrm{SOG} + \mathrm{Current\ Speed} * \cos(\mathrm{Heading} - \mathrm{Current\ Direction})$$

The factor "current_speed * cos(heading - current_direction)" is also referred to as Current Product (CP), as it represents the vector component of the currents in the direction of the ship.

3. Inaccuracies for different SOG to STW models

Put simply, we have two different approaches to calculate the STW:

1. Simple: STW = SOG
2. Current Formula: STW = SOG + Current Product (CP)

For the 20 vessels, this generates the following results, on average.
A table with all 20 ship-specific results can be found in Appendix A.
For more information on the accuracy metrics, please refer to the Blue Modeling Standard, *Deschoolmeester and Morobé (2023).*

Acc. Metric	**STW = SOG**	**STW = SOG + CP**
MAPE	6.05%	5.25%
Voyage Error	3.66%	3.73%
R^2	0.65	0.71

Average scores over 20 vessels

Unexpectedly, the second approach with the correction factor for currents barely outperforms the first very simple approach. For the voyage error, it even worsens. This means that the Current Product (CP) has very limited explaining power. This is an unexpected finding, as in theory the CP should explain all the differences between SOG and STW.

Given these findings, we might need to reframe the question. Is the approach to predict STW incorrect, or is the value we are trying to predict incorrect? Given the known flaws of speed log sensors to measure STW accurately, *Ikonomakis et al. (2021),* a likely answer could be that STW values are simply inaccurate.

The second part of this study explores the following hypothesis: if the inaccuracy is really due to inaccurate STW measurements, rather than an incorrect formula to predict STW from SOG, then this will be reflected in end-to-end SOG to Power modeling accuracy. Or in other words, it might be that the formula above predicts close to the 'true STW value' of the vessel, but that the measured STW value we validate against is simply inaccurate. If this is true, then if we would predict from SOG all the way to the Main Engine Power of the vessel, it would be more accurate to start modeling from a calculated STW instead of the measured STW. This hypothesis is tested below.

4. Impact of the STW inaccuracies on Speed-to-Power modeling

To validate the hypothesis above, we model the Main Engine Power, starting from SOG in three different ways. All three use the same modeling approach: physics-informed machine learning in

Toqua's proprietary Ship Kernels, *Collé and Morobé (2022)*. The only difference is what version of the STW is used as starting point.

1. Traditional: Measured STW - Train Power model starting from measured STW. Generate STW from SOG + CP.
2. Simple: SOG - Train Power model starting from SOG. No current corrections, so STW=SOG.
3. Current formula: Calculated STW - Train Power model starting from calculated STW = SOG + CP.

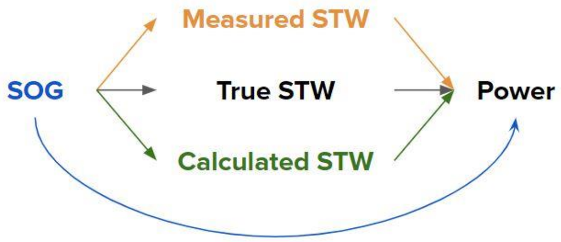
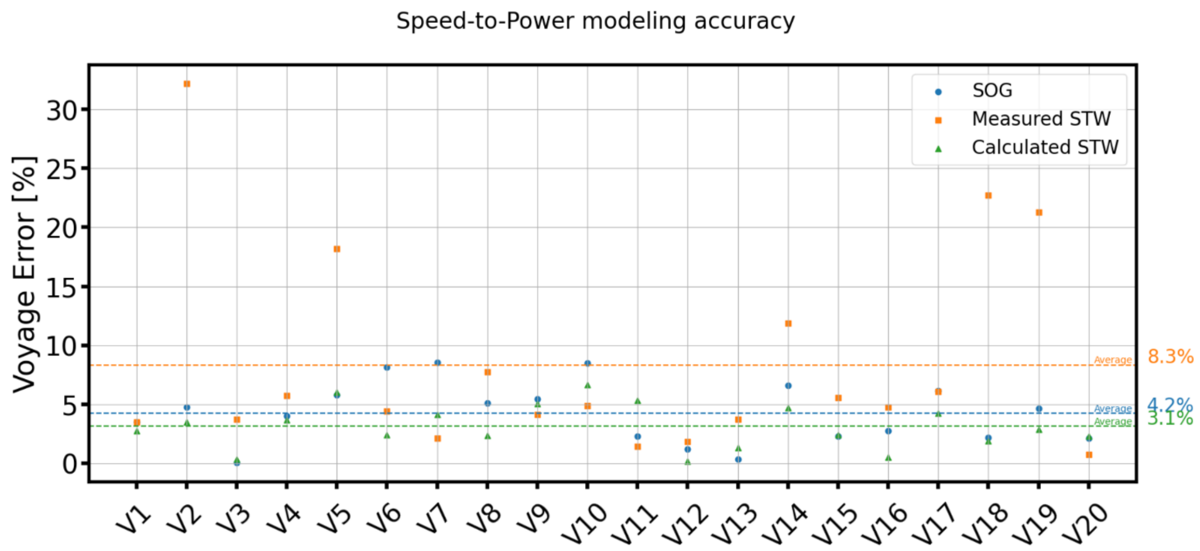


Diagram: speed→power accuracy as a proxy to uncover the closest estimate of 'true STW'

It is expected that the closer the approach gets to the 'true STW value', the more accurate the end-to-end SOG→Power model will be. The 'true STW value' is unknown, so the end-to-end speed-to-power accuracy serves as a proxy for which approach is most accurate to predict the 'true STW value'.



For the 20 vessels and these 3 different starting points, this generates the following results, on average. A table with all 20 ship-specific results can be found in Appendix B.

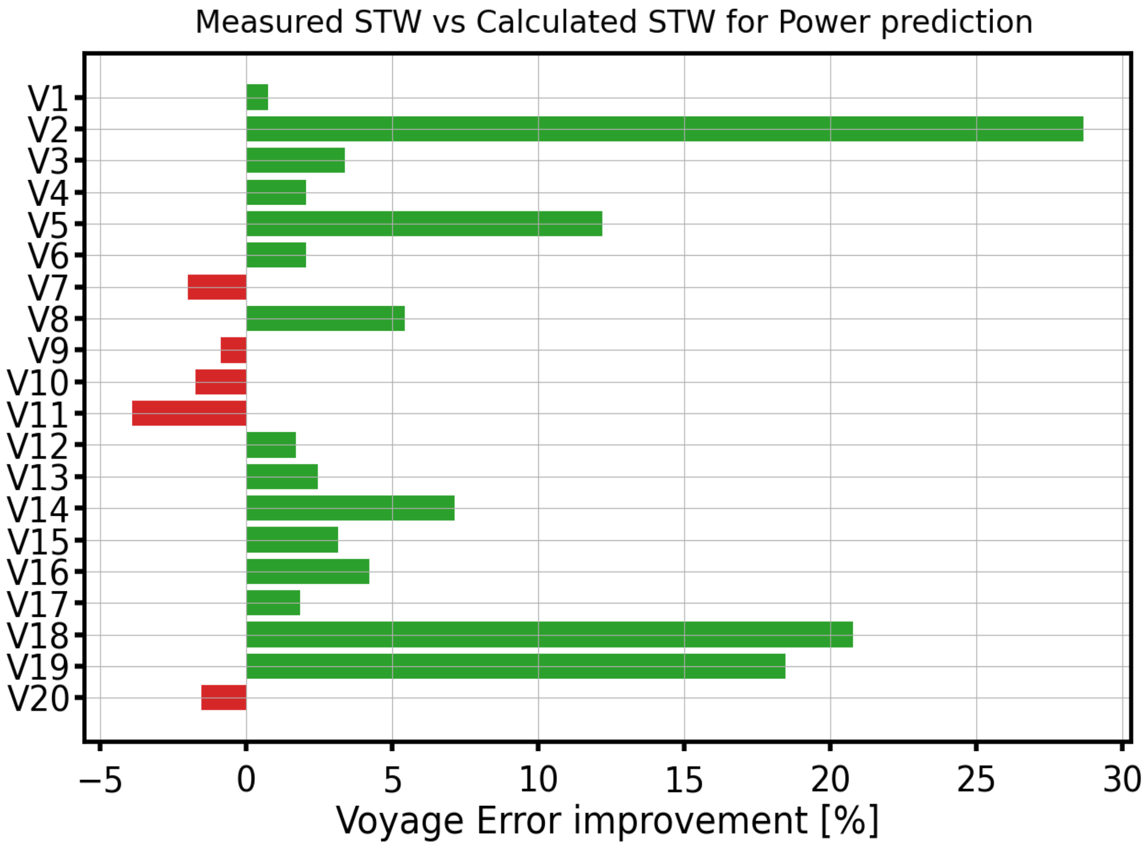
Acc. Metric	**Measured STW**	**SOG**	**Calculated STW**
MAPE	15.43%	14.15%	11.46%
Voyage Error	8.33%	4.22%	3.12%
R^2	0.49	0.54	0.67

Average scores over 20 vessels

Using SOG instead of measured STW reduces voyage accuracy from 8.3% to 4.2%. Using calculated STW instead of measured STW, reduces voyage inaccuracy from 8.3% to 3.1%.

This confirms the earlier hypothesis. Training a model from the measured STW leads to large inaccuracies. Much larger than if you would simply take SOG or calculated STW as input. However, most scores for the measured STW scenario are not that much worse than the other scenarios. It is just that some vessels (V2, V5, V14, V18, V19) have exceptionally large voyage errors for the measured STW scenario (12%-32%); as a result the measured STW scenario drastically underperforms on average. This is caused by inaccurate speed logs, which are drastically more erroneous on some ships than on others. This confirms the hypothesis that measured STW values are often less accurate than calculating STW starting from SOG. This was proven indirectly by using end-to-end speed-to-power modeling accuracy as a proxy.

In some cases (V7, V9, V10, V11, V20) the measured STW scenario outperforms the other scenarios. This indicates that in some cases the STW does capture meaningful information that goes beyond what SOG and correction factors can account for. We believe these cases have highly accurate and well-calibrated speed logs. But they are the exception, not the rule.



5. Conclusion

After analyzing operational sensor data for 20 oil tankers, to analyze if SOG or STW is the best starting point for accurate performance modeling, we find that measured STW values are unreliable. Using them leads to large average inaccuracies for performance modeling (8.3% voyage error). Instead, using the more reliable SOG already reduces the voyage error to 4.2%. If we then go a step further, and not simply use SOG, but apply correction factors for currents to derive a calculated STW, the voyage inaccuracy further reduces to 3.1%. By using end-to-end modeling as a proxy, these numbers indirectly confirm the hypothesis that STW sensors are unreliable. The most robust estimation of the true STW value is found via a formula based on SOG and currents, rather than measurement devices.

References

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

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

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

IKONOMAKIS, A.; NIELSEN, U.D.; HOLST, K.K.; DIETZ, J.; GALEAZZI, R. (2021), *How Good Is the STW Sensor? An Account from a Larger Shipping Company*, JMSE 9/5, p.465, https://www.mdpi.com/2077-1312/9/5/465

Appendices

A) SOG-STW table for all 20 vessels

i. MAPE

Vessel	SOG=STW MAPE	Current formula MAPE
V1	4.33%	4.02%
V2	10.17%	9.81%
V3	6.86%	5.99%
V4	5.68%	5.09%
V5	6.77%	5.37%
V6	5.53%	4.11%
V7	5.74%	3.97%
V8	5.01%	3.80%
V9	3.52%	3.23%
V10	3.84%	3.55%
V11	6.66%	5.47%
V12	3.96%	3.09%
V13	4.07%	3.59%
V14	10.23%	9.38%
V15	5.15%	4.14%
V16	5.06%	4.41%
V17	5.36%	4.61%
V18	8.88%	8.68%
V19	10.91%	9.65%
V20	3.18%	2.97%
Average	6.05%	5.25%

ii. Voyage Error

Vessel	SOG=STW Voyage Error	Current formula Voyage Error
V1	1.59%	2.37%
V2	9.93%	10.61%
V3	3.38%	3.90%
V4	3.44%	3.81%

V5	3.44%	4.05%
V6	1.41%	1.21%
V7	1.11%	0.68%
V8	1.99%	2.28%
V9	1.06%	0.89%
V10	0.97%	0.93%
V11	4.23%	3.47%
V12	1.70%	1.16%
V13	2.37%	1.65%
V14	9.64%	9.92%
V15	2.38%	2.45%
V16	1.78%	2.71%
V17	3.49%	2.82%
V18	7.98%	8.94%
V19	9.56%	9.59%
V20	1.72%	1.23%
Average	3.66%	3.73%

iii. R^2

Vessel	SOG=STW R^2	Current formula R^2
V1	0.76	0.80
V2	-0.21	-0.11
V3	0.48	0.62
V4	0.65	0.73
V5	0.60	0.77
V6	0.76	0.88
V7	0.78	0.90
V8	0.77	0.88
V9	0.82	0.88
V10	0.77	0.83
V11	0.55	0.71
V12	0.68	0.83
V13	0.81	0.86
V14	0.34	0.45
V15	0.66	0.78
V16	0.66	0.77
V17	0.73	0.81
V18	0.45	0.49
V19	-0.65	-0.21
V20	0.83	0.83
Average	0.65	0.71

B) SOG-Power table for all 20 vessels

i. MAPE

Vessel	Measured STW MAPE	SOG MAPE	Calculated STW MAPE
V1	10.39%	12.05%	10.84%
V2	32.23%	12.52%	9.80%
V3	12.97%	13.22%	9.57%
V4	12.69%	13.37%	10.54%
V5	21.97%	18.36%	15.98%
V6	12.30%	19.32%	11.50%

V7	12.42%	18.81%	12.71%
V8	12.94%	18.64%	11.77%
V9	12.01%	14.11%	12.96%
V10	12.60%	14.27%	12.91%
V11	11.27%	11.29%	10.16%
V12	8.76%	10.08%	8.74%
V13	11.19%	11.32%	9.99%
V14	22.26%	15.44%	13.14%
V15	14.12%	14.35%	11.39%
V16	13.56%	13.35%	12.13%
V17	14.50%	14.53%	12.20%
V18	26.61%	15.96%	12.60%
V19	23.77%	10.58%	9.52%
V20	10.02%	11.41%	10.70%
Average	15.43%	14.15%	11.46%

ii. Voyage Error

Vessel	Measured STW Voyage Error	SOG Voyage Error	Calculated STW Voyage Error
V1	3.48%	3.48%	2.74%
V2	32.19%	4.76%	3.51%
V3	3.71%	0.05%	0.32%
V4	5.74%	4.03%	3.70%
V5	18.20%	5.78%	5.99%
V6	4.44%	8.16%	2.40%
V7	2.12%	8.54%	4.12%
V8	7.76%	5.11%	2.32%
V9	4.15%	5.42%	5.02%
V10	4.90%	8.47%	6.63%
V11	1.41%	2.27%	5.33%
V12	1.83%	1.18%	0.14%
V13	3.75%	0.37%	1.31%
V14	11.85%	6.59%	4.72%
V15	5.54%	2.28%	2.39%
V16	4.74%	2.75%	0.53%
V17	6.06%	6.16%	4.22%
V18	22.70%	2.19%	1.91%
V19	21.32%	4.64%	2.85%
V20	0.77%	2.11%	2.31%
Average	8.33%	4.22%	3.12%

iii. R^2

Vessel	Measured STW R^2	SOG R^2	Calculated STW R^2
V1	0.29	0.31	0.37
V2	-0.19	0.82	0.87
V3	0.70	0.75	0.82
V4	0.75	0.67	0.80
V5	0.16	0.57	0.63
V6	0.79	0.58	0.84
V7	0.76	0.59	0.78
V8	0.75	0.61	0.83
V9	0.46	0.21	0.42
V10	0.63	0.57	0.65
V11	-0.01	0.13	0.37
V12	0.40	0.29	0.54
V13	0.86	0.86	0.89
V14	0.14	0.34	0.52
V15	0.72	0.70	0.81
V16	0.76	0.78	0.83
V17	0.29	0.24	0.51
V18	0.38	0.69	0.81
V19	-0.41	0.72	0.75
V20	0.39	0.29	0.36
Average	0.49	0.54	0.67