
A Hybrid Machine Learning Model For Ship Speed Through Water: Solve And Predict

Zakarya Elmimouni
ENSAE, IP Paris
Palaiseau, France
zakarya.elmimouni@ensae.fr

Ayoub Atanane
Université du Québec à Rimouski
Montreal, Canada
ayoub.atanane@uqar.ca

Loubna Benabbou
Université du Québec à Rimouski
Montreal, Canada
Loubna_Benabbou@uqar.ca

Abstract

1 This research proposes a hybrid model for predicting ship speed through water, ad-
2 dressing challenges in estimating GHG emissions from shipping while contributing
3 to climate change mitigation. Predicting ship speed through water is a key element
4 in calculating GHG emissions. However, few models address this prediction in a
5 way that integrates both physical principles and machine learning. Our approach
6 combines physical modeling with data-driven techniques to predict real ship speed
7 through water in two key steps: "Solve" and "Predict". In the first step "Solve", a
8 differential equation is resolved to estimate speed through calm water. "Predict"
9 step uses a machine learning model that incorporates maritime and meteorological
10 conditions and historical data to improve speed predictions in real-world conditions.
11 This hybrid approach leverages both physics-based knowledge and machine learn-
12 ing models to provide a more comprehensive solution for accurately predicting
13 ship speed through water.

14 1 Introduction

15 With the growing focus on greenhouse gas emissions [3, 4] the maritime transport segment needs
16 to reduce its total annual greenhouse gas emissions by 50% compared to 2008 . In this context,
17 through-water speed predictions are becoming increasingly relevant for validating ship performance
18 under real operational conditions, as well as for optimizing and predicting ship emissions. Ship speed
19 determines fuel consumption, a major source of greenhouse gases. Therefore, to predict the amount
20 of emissions caused by a ship, prediction of ship speed is an essential factor [8].

21 In our work we are going to predict the reel through water speed V_{wr} of ships using a hybrid model
22 which is composed of a first physical part where we solve a differential equation in order to have a
23 physical parameter useful for the second part of the model where we predict V_{wr} using a data driven
24 model. The physical part of our model is constructed mainly of a Physics Informed Neural Network
25 (PINN) which allows us to solve a differential equation and to generate the speed through calm water
26 V_w which will be useful for the prediction of real speed through water .

27 [7] have demonstrated the effectiveness of PINNs for solving differential equations using collocation
28 points as inputs to the network and a loss function that integrates both the differential equation
29 formula and initial conditions.

30 In the literature, various approaches have been utilized to address issues related to speed prediction:
31 Physical models, which rely on established physical formulas, represent the first category. These
32 models, although grounded in scientific principles, have been widely criticized for their inability to
33 account for the specific characteristics of individual ships [5].

34 Statistical models offer another approach, focusing on the analysis of historical data to predict
35 ship speed. For instance [6] compared three statistical models to predict ship speed based on the
36 revolutions per minute of the main engine and prevailing weather conditions.

37 The advent of large-scale data collection from ships and tracking systems has paved the way for
38 data-driven models. These models utilize machine learning techniques to predict ship speed more
39 accurately. [9] applied a sequential random forest for short-term speed prediction, leveraging the vast
40 amounts of data available. Similarly, [10] developed a machine learning-based regression model that
41 uses ship navigation data and weather forecasts to predict speed under different weather conditions.[2]
42 explored deep learning sequence models, including LSTM, BiLSTM, and transformers, finding that
43 the transformer model delivered the best performance, albeit with a high parameter count.

44 In recent years, some hybrid models have been used. These models aim to leverage the interpretability
45 of physical models while enhancing prediction accuracy with machine learning. [1] developed
46 a hybrid model that integrates a physical model with an LSTM to predict state variables . In
47 this approach, the LSTM predicts the residuals of the physical model, which itself combines first
48 principles derived from Newton’s second law with a regression component for data-driven learning.
49 Additionally, [10] proposed a hybrid framework based on Gaussian Processes to capture the complex
50 relationships between engine power and ship speed. This method integrates machine learning with
51 domain knowledge grounded in physical principles, resulting in a robust model that minimizes the risk
52 of overfitting. Similarly, Lang et al. [5] employed a hybrid model that combines a physics-informed
53 neural network, utilizing a differential equation, with an XGBoost model to predict the speed over
54 ground based on weather conditions.

55 2 Our Proposal

56 We adapt the approach proposed by Lang et al.[5] to predict the real speed through water, and we
57 generalize it so that it is not limited to a model specific to a single vessel. In this way, we make the
58 method applicable to any ship that has historical data. Our hybrid model is mainly composed of two
59 parts: a differential equation solver and a data-driven model that enables us to predict the reel speed
60 through water V_{wr}

61 The first part of our model is a differential equation solver, it enables us to solve a differential equation
62 linking the propulsion power P , the draft T and the speed through calm water V_w , this last variable is
63 a theoretical variable representing the speed we would have had under P and T if the sea was calm.

64 Lang et al.[5] have used a general form of the differential equation linking the three parameters
65 mentioned above, which can be written as:

$$a_1 \frac{\partial P}{\partial V_w} + a_2 \frac{\partial P}{\partial T} + \mu(V_w) = 0, \quad V_w \in [V_{w_{min}}, V_{w_{max}}], \quad T \in [T_{min}, T_{max}] \quad (1)$$

66 where a_1 and a_2 are coefficients, and μ is a nonlinear function of V_w .

67 In the first part of our model, which represents a solver for our differential equation, the aim is to find
68 a solution to this equation, allowing us to link the three variables P , T and V_w . This solution then
69 enables us to generate values for the calm-water speed V_w , which will be used in the second part of
70 the model to predict V_{wr} .

71 In this part, which serves as a solver for the differential equation, we use a PINN to solve the differ-
 72 ential equation based on experimental data and collocation points. PINNs are proving particularly
 73 a powerful performance for solving this type of equation. A PINN is a neural network whose loss
 74 function integrates a term corresponding to the loss on the initial conditions, as well as a term
 75 quantifying the residual of the differential equation.

76 PINN training requires experimental data on P , T and V_w . However, a problem arises here: V_w
 77 represents the speed through water under calm sea conditions, a theoretical information that is difficult
 78 to obtain when sailing at sea. This would require experiments test in towing tanks to recreate calm
 79 water conditions, which is costly and often impractical.

80 Our strategy, therefore, is to exploit the historical data available and select the moments corresponding
 81 to calm sea conditions. In addition to the ship's operational data, our database contains maritime and
 82 meteorological information on sea (waves,current and wind conditions) at the time of data collection.
 83 Based on this meteorological data, we identify the periods in our database that correspond to the calm
 84 sea hypothesis, and use this data as experimental data to train our PINN.

85 The coefficients a_1 , a_2 and μ of the differential equation are optimized to obtain the most appropriate
 86 parameters for our data. We compare three methods for optimizing these coefficients in order to
 87 choose the best solution to adopt:

- 88 1. Optimization during PINN training
- 89 2. Genetic algorithms
- 90 3. Variable Neighborhood Search (VNS)

91 As the form of the differential equation indicates, solving this equation allows us to obtain information
 92 about P , but not about V_w . Therefore, we will use our trained PINN to train a neural network capable
 93 of predicting V_w based on T and P . We compare two approaches to building this network:

94 1. **First approach:** This involves generating the training data for this network from the
 95 previously constructed PINN. Data generation from the PINN proceeds as follows: we
 96 define a list of collocation points for V_w and T , which are then used as inputs in the trained
 97 PINN. The PINN generates a list P , corresponding to the solutions of the differential
 98 equation for the input data. These solutions are then used to form triplets (P_i, T_i, V_{w_i}) ,
 99 which serve as training data for the neural network designed to predict V_w .

100 The loss function for this neural network is written as :

$$\text{MSE}_{NN} = \frac{1}{N} \sum_{i=1}^N \left| \hat{V}_{w(i)} - V_{w(i)} \right|^2$$

101 2. **Second approach :** This involves using a neural network which takes the variables P and T
 102 as inputs and provides the variable \hat{V}_w as an output. PINN is integrated into this approach
 103 via the loss function, which is written as :

$$\text{MSE}_{NN} = \frac{1}{N} \sum_{i=1}^N \left| \text{PINN} \left(\hat{V}_{w(i)}, T_{(i)} \right) - P_{(i)} \right|^2$$

104 In this approach, the physical information represented in the differential equation is trans-
 105 mitted to the speed via the trained PINN (which acts as a function) integrated into the loss
 106 function. This procedure makes it possible to estimate the speed in calm water from the
 107 draft T and the propulsion power P .

108 The second part is a data-driven model that enables us to predict the real speed through water V_{wr} .
 109 This part of the model allows us to add the meteorological and current conditions corresponding to
 110 the calm water speed V_w estimated previously in order to predict the real speed through water V_{wr} .
 111 In this part of the model, we utilize various machine learning algorithms (XGBoost, Random Forest
 112 Regressor...), compare their accuracy, and determine which one performs best.

113 3 Acknowledgements

114 This work was conducted as part of an internship at MILA - Quebec Artificial Intelligence Institute
115 and The Research Chair of AI for Supply Chains at University of Quebec at Rimouski (UQAR). We
116 gratefully acknowledge the financial support provided by MILA and the Chair which made this
117 research possible.

118 References

- 119 [1] S.Staab A.Baier, Z. Boukhers. Hybrid physics and deep learning model for interpretable vehicle
120 state prediction. *arXiv preprint arXiv:2103.06727*, 2021.
- 121 [2] Sara El Mekkaoui, Loubna Benabbou, Stéphane Caron, and Abdelaziz Berrado. Deep learning-
122 based ship speed prediction for intelligent maritime traffic management. *Journal of Marine
123 Science and Engineering*, 11(1), 2023.
- 124 [3] IMO. Amendments to the annex of the protocol of 1997 to amend the international convention
125 for the prevention of pollution from ships, 1973, as modified by the protocol of 1978 relating
126 thereto.
- 127 [4] ITCC. 2014a the seakeeping committee technical report.
- 128 [5] Xiao Lang, Da Wu, and Wengang Mao. Physics-informed machine learning models for ship
129 speed prediction. *Expert Systems with Applications*, 238:121877, 2024.
- 130 [6] Wengang Mao, Igor Rychlik, Jonas Wallin, and Gaute Storhaug. Statistical models for the speed
131 prediction of a container ship. *Ocean Engineering*, 126:152–162, 2016.
- 132 [7] Perdikaris P. Karniadakis G. E. Raissi, M. Physics-informed neural networks: A deep learning
133 framework for solving forward and inverse problems involving nonlinear partial differential
134 equations. *soner, o., akyuz., Journal of Computational Physics*, 378, 686–707., 2019.
- 135 [8] Nguyen Khoi Tran and Jasmine Siu Lee Lam. Effects of container ship speed on co2 emission,
136 cargo lead time and supply chain costs. *Research in Transportation Business Management*,
137 43:100723, 2022.
- 138 [9] Jun Wang, Yuhan Guo, and Yiyang Wang. A sequential random forest for short-term vessel
139 speed prediction. *Ocean Engineering*, 248:110691, 2022.
- 140 [10] Byunghyun Yoo and Jinwhan Kim. Probabilistic modeling of ship powering performance using
141 full-scale operational data. *Applied Ocean Research*, 82:1–9, 2019.