Comparative Study of Ship Motion Prediction Models: Data-Driven Physics-Based vs Pure Machine Learning

1st Michail Mathioudakis
DeepSea Technologies
Athens, Greece
m.mathioudakis@deepsea.ai

2nd Petros Iatropoulos
DeepSea Technologies
Athens, Greece
p.iatropoulos@deepsea.ai

3rd Theodoros Stouraitis
DeepSea Technologies
Athens, Greece
t.stouraitis@deepsea.ai

4th Christos Papandreou
DeepSea Technologies
Athens, Greece
c.papandreou@deepsea.ai

5th Antonis Nikitakis
DeepSea Technologies
Athens, Greece
a.nikitakis@deepsea.ai

6th Stavros Paschalakis
DeepSea Technologies
Athens, Greece
s.paschalakis@deepsea.ai

7th Konstantinos Kyriakopoulos
DeepSea Technologies
Athens, Greece
k.kyriakopoulos@deepsea.ai

***Abstract*—Accurate ship motion prediction is essential for safety and efficiency in maritime navigation. Data-driven techniques offer potential accuracy improvements to ship model prediction, yet their utility and reliability are being explored. This study compares two data-driven approaches for vessel trajectory prediction: a data-driven optimization of physics-based model parameters and a purely machine learning approach. Both models are trained and evaluated based on a dataset comprising of 90 trajectories generated via a vessel simulator. The physics-based model is based on hydrodynamic principles while key parameters are fitted, using constrained nonlinear least squares, to tailor its prediction accuracy to the dataset. For the machine learning approach, a feed-forward neural network learns motion patterns directly from data without prior domain knowledge. A 5-fold cross-validation is employed for both models ensuring robust evaluation. The models' prediction accuracy is quantified using: (i) Euclidean Distance, (ii) Euclidean Distance with Heading Penalization, and (iii) Custom Vessel Distance Measure. The results across the dataset show that the data-driven physics-based model achieves higher prediction accuracy (40 per cent lower summarized absolute error) and consistency (35 per cent lower summarized error spread). Individual prediction scenarios are examined to highlight the trade-offs between accuracy, reliability and adaptability of the two approaches.**

***Index Terms*—Autonomous shipping, Machine Learning, Parameter optimization, Data Driven Models**

I. Introduction

In recent years, the maritime industry has seen a growing demand for accurate ship motion prediction models to enhance safety, efficiency, and decision-making in navigation. This need has become of higher importance within the context of autonomously navigating ships. To address this challenge, the research community has been exploring innovative approaches that leverage both traditional physics-based models [1]–[3] and data-driven techniques, with significant attempts from [4]–[6].

This research is part of the developments with Nabtesco corporation for the Nippon Foundation `MEGURI2040` Fully Autonomous Ship Program.

Maneuvering models trace back to [7], with major contributions to hydrodynamics [8], [9], and course stability [10], [11]. Taylor expansion was used for maneuvering models [12], [13], which were utilized for autonomous navigation [14].

The Maneuvering Modeling Group (MMG) model, first proposed by [1] and refined by [15], has been widely used for autonomous surface vehicle control [16], path following [17], course keeping [18], reinforcement learning-based control [19], and autonomous berthing [5], [20], [21]. Other models include the Hydrodynamic Derivatives model [15], the Cross-Flow Drag model [22], and the Table model [23], each offering different trade-offs between complexity, accuracy, and physical justification [2], [3], [22], [24]–[29].

Rapid innovation in data analysis techniques and measurement equipment has opened new horizons and possibilities for constructing end-to-end models based on ship data. Purely data driven, Machine Learning (ML) models can be constructed to predict the motion of the vessel. Related work in this field include artificial neural networks [4], [6], [30]–[33] and Gaussian processes [34], [35]. These models are easy to construct and can have high accuracy when the data measured on the vessel is of good accuracy and captures as large an operational spectrum as possible. However, these purely data-driven models (black box) often lack physical interpretability, leading to unexpected outcomes, especially in areas of the domain where datapoints are sparse or not seen in training.

Physics-based models incorporate established hydrodynamic principles, ensuring physical consistency and generalizability to unseen conditions, yet they often require manual tuning of parameters based on excessive model tests [3] and may not capture all real-world complexities. Hence, an alternative is a hybrid approach that combines physics-based modeling with data-driven optimisation of physics-based parameters. *i.e.* optimize parameters of physics-based model based on measured behaviour of a vessel. In the literature, [36], [37] employed a gradient-free optimization method (namely CMA-ES) to tune parameters of an MMG model [15] using sea-trial data of a real-scale ship. [38] proposed a framework that fits open parameters of an Abkowitz model [12] to data of the target ship while using a well-validated model of similar ship to regularize the fitting process.

Recent efforts propose hybrid approaches, such as [39], combining statistical methods with ML to enhance real-time prediction accuracy using AIS data. Similarly, [39] leverages recurrent neural networks to capture temporal dependencies in maritime data, and improve the predictive accuracy of the model under varying environmental conditions. Meanwhile, [40] incorporates multi-source sensor data to refine motion predictions. These contributions highlight the growing synergy between data-driven methodologies and domain-specific knowledge, aligning with the hybrid and pure ML approaches evaluated in this study.

This work introduces and compares two distinct approaches for predicting the motion of a ship: (i) a data-driven optimization of physics-based model parameters, and (ii) a pure ML approach using neural networks. The former method leverages existing domain knowledge to account for the impact of environmental factors and vessel appendages [41], while providing structure for data optimization techniques. The latter approach uses neural networks to learn patterns from data, treating the problem as a regression task. The model predicts the vessel's next state from its current state and control signals, without integrating physics-based information model into the structure or training process.

The two resulting models (corresponding to the two approaches) are compared and validated on their ability to accurately predict a set of trajectories (test set), after being trained on a different set of trajectories (train set). To avoid dataset biases, a 5-fold cross-validation is being implemented. The train and test sets are extracted from a dataset D, which for this paper has been produced using a simulator, provided by the Hiroshima University for a 130 meter containership [3]. It is assumed that the response of the simulator for the particular 130 meter containership replicates the real world behaviour of the containership.

The main contribution of this work lies in comparing a data-driven physics-based approach against a feed-forward neural network trained and tested on the same data. This provides a quantitative investigation into the accuracy and reliability trade-offs of incorporating marine engineering domain knowledge in data-driven modeling with a small-scale dataset. This investigation builds on top of our prior work [41], where a physics-based motion prediction model is being compared in terms of accuracy against real-world data, and [42], where the same physics-based motion prediction model has been being parametrized and fitted to a trajectory dataset of a container ship. Here, we compare the predictive capabilities of the data-driven physics-based approach [42] against a pure ML modeling approach.

The rest of the manuscript is organized as follows. Section II describes the characteristics and the creation process of the dataset. Section III describes the two modeling approaches of the predictive models, while section IV provides the evaluations performed in this study. Finally, in section V and section VI we discuss and conclude this work.

TABLE I: State and control features/dimensions.

Symbol	Description	Metric Units
x	position x	m
y	position y	m
ψ	heading	rad
$\dot{x}$	surge speed	m/s
$\dot{y}$	sway speed	m/s
r	yaw	rad/s
n	RPM	none
δ	rudder angle	rad
F_b	bow thruster	N
F_s	stern thruster	N
n_c	RPM command	none
δ_c	rudder angle command	rad
$F_{b,c}$	bow thruster command	N
$F_{s,c}$	stern thruster command	N

II. Dataset

A. Synthetic Dataset Generation

In our validation setup, we utilize a synthetic dataset D. The dataset consists of $N = 90$ distinct trajectories, created using the Hiroshima University simulator that serves as the ground truth for this study.

The trajectories are generated via a closed-loop setup, where a controller drives a simulated containership. The controller aims to follow reference paths that are typical maritime operational scenarios, such as approaching and departing a port. The controller's outputs are commands, such as n_C and δ_c, that are fed into the simulator to generate the trajectories of the dataset. Leveraging this approach enables us to synthesize a dataset that includes a diverse set of maneuvers under realistic maritime conditions. Such a synthetic dataset allows for a structured evaluation of the optimization process and the robustness of different fitted predictive models. A visual representation of the trajectory dataset is shown in fig. 1. To ease the reader's understanding, all trajectories have been shifted to the origin and rotated to have the same initial heading (0°).

B. Trajectory Characteristics and Dataset Statistics

Each trajectory $\mathcal{T}$ has a fixed time duration $T = 120s$, it is sampled at fixed time intervals $T_s = 1s$ leading to a $K+1$ number of knots where $K = \frac{T}{T_s} = 120$. Each knot of the trajectory consists of the ten first features described in table I, which compose the state of the vessel denoted as $\mathbf{s}_k = [x, y, \psi, \dot{x}, \dot{y}, r, n, \delta, F_b, F_s]^T$.

The dataset provides a detailed insight into the velocity characteristics of a 130-meter container ship during low-speed approaching or departing from port maneuvers. The vessel's surge speed $\dot{x}$ ranges from -0.47 m/s to 6.17 m/s, with a mean of 3.82 m/s. The 10th percentile at 3.11 m/s and the 90th percentile at 4.97 m/s highlight that the vessel operates predominantly within a narrow speed band, reflecting the

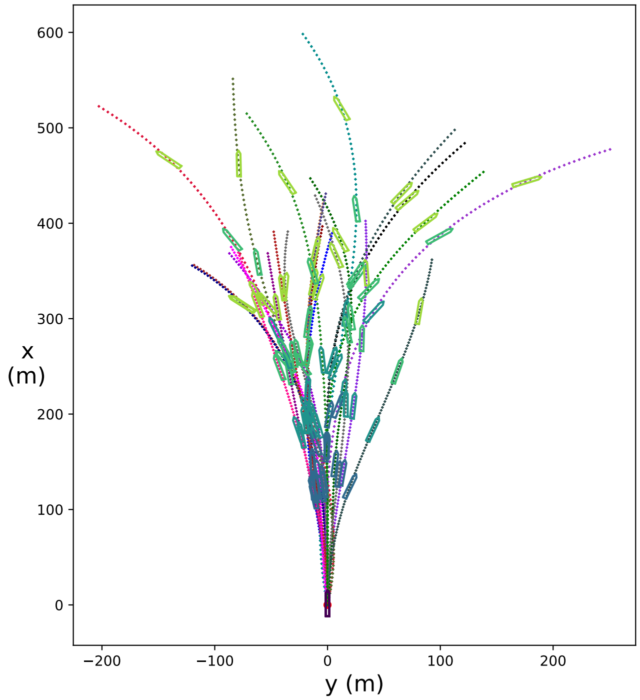


Fig. 1: The trajectories of the dataset are shifted to the origin and rotated to have zero initial heading, so they appear superimposed.

controlled nature of approach and departure phases. Note, that the maneuvers are not berthing or unberthing as we can see that bow and stern thruster commands and forces are 0 from $10th\ Perc.$ (P_{10}) to $90th\ Perc.$ (P_{90})

Sway velocity $\dot{y}$ is significantly lower, with a mean of 0.05 m/s and extreme values ranging from -1.04 m/s to 1.03 m/s. The 10th and 90th percentiles, at -0.3 m/s and 0.46 m/s, both demonstrate the limited but critical lateral motion required for precise berthing. The yaw rate r, representing the rate of heading change, remains close to zero, with a mean of 0.0 rad/s and extreme values of ± 0.02 rad/s, indicating slow, deliberate rotational adjustments during confined maneuvers.

The propulsion system operates primarily within the low to moderate RPM range, with a mean of 92.8 rpm and 90th percentile reaching 154.5 rpm, emphasizing controlled motion rather than high-speed maneuvers.

These velocity statistics collectively reflect the vessel's careful navigation strategy during port operations, where fine speed adjustments and precise heading control are crucial for safe and efficient maneuvering.

III. Vessel Motion Prediction

Trajectory prediction is based on an initial state and a sequence of control commands. Formally, given a state $\mathbf{s}_k$ and control signals $\mathbf{u}_k$ the system (vessel) proceeds to its next state $\mathbf{s}_{k+1}$ according to a function $g(\cdot)$. By iteratively applying (1) trajectories can be predicted.

$$\mathbf{s}_{k+1} = g(\mathbf{s}_k, \mathbf{u}_k),\ \ k = 0, 1, \dots K-1 \tag{1}$$

TABLE II: Statistics of the dataset.

	Min	Max	Mean	Std	P_{10}	P_{50}	P_{90}
x	-1935	1051	-362	554	-1368	-205	136
y	-2622	3594	260	1194	-1469	227	1592
ψ	-1.52	6.12	2.84	1.46	1.58	2.43	4.83
$\dot{x}$	-0.47	6.17	3.82	0.74	3.11	3.53	4.97
$\dot{y}$	-1.04	1.03	0.05	0.3	-0.3	0.02	0.46
r	-0.02	0.02	0.00	0.01	-0.01	0.00	0.01
n	-158.8	169.7	92.8	47.9	32.1	95.7	154.5
δ	-40.0	40.0	-1.0	11.3	-13.9	-0.7	11.6
F_b	-51228	63000	725	8384	0.0	0.0	0.0
F_s	-63000	62996	203	5711	0.0	0.0	0.0
n_c	-180.0	180.0	94.8	50.8	36.5	98.4	166.7
δ_c	-40.0	40.0	-1.4	12.1	-15.5	-0.7	12.0
$F_{b,c}$	-51228	63000	725	8421	0.0	0.0	0.0
$F_{s,c}$	-63000	62996	207	5734	0.0	0.0	0.0

The data-driven modeling task of predicting a ship's motion is to estimate function $g(\cdot)$ through a parametrized model $g_\theta(\cdot)$. θ is the set of the model's parameters which can be learned by utilizing data. In this work, we compare two different approaches in terms of trajectory prediction, the two models have 3 degrees of freedom (3-DoF) and are: (i) a physics-based data-driven marine model and (ii) a pure ML approach that has no prior knowledge of the underlying physics.

A. Physics-Based Model and Parameter Optimization

A physics-based motion prediction model for vessels involves simulating the ship's motion in 3 degrees of freedom (3-DoF) using nonlinear ordinary differential equations (ODEs) that represent the vessel's dynamics. Forces and moments acting on the vessel, including those from the rudder, propeller, thrusters and environmental factors like wind and currents, are incorporated into the equations of motion. The next state s_{k+1} is calculated by integrating the ODEs given the current state s_k [41]. This physics-based approach captures the ship's behavior by modeling the hydrodynamic effects and resistance forces based on the ship's geometry and operating conditions. For the equations of motion (dynamics), a 3-DoF nonlinear ODE from [43] is used. This is a Hydrodynamics Derivatives model described by (2).

$$\begin{aligned}
&(m - X_{\dot{u}})\dot{u} + (Y_{\dot{v}} - X_{vr} - m)vr + \\
&\qquad\qquad (Y_{\dot{r}} - mx_G)r^2 = \Sigma X \\
&(m - Y_{\dot{v}})\dot{v} + (mx_G - Y_{\dot{r}})\dot{r} - Y_v U v + \\
&(mu - Y_r U)r - Y_{vv}v|v| - Y_{vr}v|r| - Y_{rr}r|r|) = \Sigma Y \\
&(mx_G - N_{\dot{v}})\dot{v} - N_v U v + (mx_G u - N_r U)r + \\
&(I_z - N_{\dot{r}})\dot{r} - N_{rr}r|r| - N_{rrv}\frac{rrv}{U} - N_{vvr}\frac{vvr}{U} = \Sigma N
\end{aligned} \tag{2}$$

In (2), m is the mass of the ship, I_{zz} is the yaw moment of inertia and x_G is the longitudinal distance of ship's centre of gravity from the moving axes' origin, O_s. X, Y and N terms with surge and sway velocities u, v, equal to $\dot{x}$, $\dot{y}$, yaw rate of turn r and accelerations $\dot{u}$, $\dot{v}$, $\dot{r}$ subscripts are the maneuvering coefficients of the ship per dimension, which capture information regarding the hydrodynamic effects based on its geometric characteristics. The left-hand side (LHS) of (2) includes terms representing added mass, damping, and restoring forces. The parameters on the LHS are considered accurate, as the hydrodynamic coefficients are calculated individually for each ship based on [9], [44].

The right-hand side (RHS) of (2) aggregates all external forces and moments acting on the ship, as detailed in (3).

$$\begin{aligned}\Sigma X &= X_n + X_\delta + X_e + R(u)\\ \Sigma Y &= Y_n + Y_\delta + Y_e\\ \Sigma N &= N_n + N_\delta + N_e\end{aligned} \tag{3}$$

For each force dimension, the subscripts denote its source and $R(u) \in \mathbb{R}$ denotes the resistance of the vessel in Newton.

For the data-driven physics-based approach, we fit key parameters on the RHS of (2), *i.e.* parameters of (3). The selected parameters influence the thrust, resistance, and rudder forces. The propeller thrust X_n follows a propulsion model [41], while the lateral force is estimated as:

$$Y_n = p_0 X_n \ . \tag{4}$$

The moment N_n is computed geometrically from the propeller's position x_P and Y_n, with p_0 capturing course instability effects, as

$$N_n = Y_n x_P = p_0 X_n x_p \ . \tag{5}$$

Hull resistance is modeled as

$$R(u) = p_1 u + p_2 u^2 + p_3 u^3 \ , \tag{6}$$

with the 0_{th} order coefficient being set to zero, since stationary resistance is zero. The coefficients p_1, p_2, p_3 balance resistance with thrust, improving surge speed predictions and turning radius. Rudder lift and drag coefficients are modeled as:

$$c_L(a_r) = p_4 + p_5 a_r + p_6 a_r^2 + p_7 a_r^3 \ , \tag{7}$$

$$c_D(a_r) = p_8 + p_9 a_r + p_{10} a_r^2 \ . \tag{8}$$

These polynomials are chosen to match experimental trends, describe turning performance, with $p_4, p_5, \ldots, p_{10}$ being the tunable parameters. Together, these components allow the model to capture key 3-DoF maneuvers like speed, turning, and course instability. To fit the parameters to vessel-specific data, with M trajectories, we solve the following constrained nonlinear least squares (cNLS) problem (9):

$$\begin{aligned}\min_{\mathbf{p}} \quad & \frac{1}{M}\sum_{i=0}^{M}(\bar{\xi}_i - \xi_i)^T \mathbf{W} (\bar{\xi}_i - \xi_i)\\ \text{s. t.} \quad & \mathbf{b_l} \leq \mathbf{p} \leq \mathbf{b_u},\\ & g(\mathbf{p}) \leq b_j, \ j = 1, \ldots, J\end{aligned} \ , \tag{9}$$

which minimises the squared deviation between predicted ξ_i and measured trajectories $\bar{\xi}_i$. The weight matrix is $\mathbf{W} \in \mathbb{R}^{10\times 10}$ and the The optimization vector is $\mathbf{p} = [p_0, p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}]^T$, where $\mathbf{b_l}$, $\mathbf{b_u}$ are parameter bounds, and $g(\mathbf{p})$ are nonlinear functions encoding additional constraints. The problem is solved numerically using interior-point optimization to minimize trajectory error while enforcing physical constraints. For more details we point the reader to [42].

B. Machine Learning Approach

In the pure ML approach, we use a parameterized neural network g_θ to learn to predict the next state $\mathbf{s}_{k+1}$ using information from the current state $\mathbf{s}_k$ and control commands $\mathbf{u}_k$. This neural network has no prior knowledge of either the vessel characteristics or the physical laws that govern the vessel's motion. We do not use the positions directly as input and output of the model but we instead predict speeds and integrate them into position and heading. In this way, we enable translation invariance, which is essential for robust training and allows deployment to any reference coordinate frame.

The input-output partial state vector of the model consists of the seven following features $\mathbf{y} = [\dot{x}, \dot{y}, r, n, \delta, F_b, F_s]$ and input control signals $\mathbf{u} = [n_c, \delta_c, F_{b,c}, F_{s,c}]$. Using the estimated vector $\mathbf{y}$ we obtain a position vector $\mathbf{p} = [x, y, \psi]$ by performing forward Euler integration. Formally, we determine the difference on each axis (surge $\rightarrow \Delta x$, sway $\rightarrow \Delta y$, yaw $\rightarrow \Delta\psi$) by

$$\begin{bmatrix}\Delta x\\ \Delta y\\ \Delta\psi\end{bmatrix} = \begin{bmatrix}\dot{x}_{k+1}\cos(\psi_k) - \dot{y}_{k+1}\sin(\psi_k)\\ \dot{x}_{k+1}\sin(\psi_k) + \dot{y}_{k+1}\cos(\psi_k)\\ r_{k+1}\end{bmatrix} \cdot \Delta t \tag{10}$$

and we add the vector to the current position in order to obtain the next position with

$$\mathbf{p}_{k+1} = \mathbf{p}_k + \begin{bmatrix}\Delta x\\ \Delta y\\ \Delta\psi\end{bmatrix} . \tag{11}$$

This 2-step procedure can be summarized as

$$\mathbf{y}_{k+1} = g_\theta(\mathbf{y}_k, \mathbf{u}_k), \tag{12}$$

$$\mathbf{p}_{k+1} = I(\mathbf{y}_{k+1}, \mathbf{p}_k), \tag{13}$$

where $I(\cdot)$ denotes an integration operation in the form of a function. The predicted full next state $\mathbf{s}_{k+1}$ is then obtained as the concatenation of position $\mathbf{p}_{k+1}$ and model's output $\mathbf{y}_{k+1}$, formally written as

$$\mathbf{s}_{k+1} = \mathbf{p}_{k+1} \oplus \mathbf{y}_{k+1} \ . \tag{14}$$

Our ML model is a feed-forward neural network with three hidden layers with one hundred nodes each and a ReLU activation function. A Mean-Squared-Error (MSE) loss is used to fit the model with Stochastic-Gradient-Descent (SGD) optimization. Optimization is performed with Adam algorithm using an initial learning rate of 0.001 and a batch size of 32. The model is trained for a maximum of 200 epochs and it is early stopped if the validation error does not drop at least 10^{-4} after 10 epochs.

IV. Evaluation Setup and Results

The goal of this evaluation is to compare the performance of a data-driven physics-based motion prediction model and a ML prediction model against ground truth trajectories generated by a ship simulator. The dataset is divided into five folds for 5-fold cross-validation, resulting in five sets, each containing 72 training trajectories and 18 test trajectories. Both models are trained on the same training datasets of each of the five folds and they evaluated on the corresponding test sets, allowing for a fair and systematic comparison. This cross-validation scheme enables a comprehensive analysis of how well both models generalize to unseen data while minimizing any potential bias introduced by the specific dataset split.

A. Evaluation measures

The performance of the models is determined by comparing their predicted trajectories to the ground truth test trajectories using three evaluation measures. The first is given in (15) and it is the Euclidean distance (ED) between the (x, y) coordinates of the predicted and ground truth trajectories, providing a direct measurement of spatial accuracy. This measure does not capture the vessel's deviation in heading, which is critical in ship motion prediction. To address this, the second measure extends the Euclidean distance by incorporating a heading penalization term (EDH), given in (16). This accounts for differences in vessel orientation $|\Delta\psi|$ and thus provides a more complete evaluation of the trajectory prediction accuracy. The term $\frac{L}{2} * |\Delta\psi|$ represents the arc length a vessel would traverse when rotating around its longitudinal center to align its heading. The third measure, referred to as cVDM (Custom Vessel Distance Measure), was developed specifically for evaluating ship trajectories [41]. Unlike the first two (measured in meters), cVDM quantifies trajectory error as a percentage, offering insight into the relative deviation. Together, these three measures capture spatial accuracy and orientation consistency. This enables us to systematically assess and compare the performance of the two models.

$$ED = \sqrt{(\bar{x}_i - x_i)^2 + (\bar{y}_i - y_i)^2} \tag{15}$$

$$EDH = ED \; + \; \frac{L}{2} * |\Delta\psi| \tag{16}$$

$$cVDM = \frac{100}{K} \sum_{k=1}^{K} \left(\frac{|\bar{x}_k - x_k|}{\bar{L}} + \frac{|\bar{y}_k - y_k|}{\bar{L}} + \frac{|\bar{\psi}_k - \psi_k|}{\pi} + \frac{|\bar{u}_k - u_k|}{\bar{U}} + \frac{|\bar{v}_k - v_k|}{\bar{U}} + \frac{|\bar{r}_k - r_k|}{r_{max}} \right) \tag{17}$$

In (15), (16), (17), we denote the ground truth value of each dimension with a bar, *e.g.* $\bar{x}$ denotes the value of x value in the ground truth trajectory. $\bar{L}$ is the total length of the ground truth trajectory and $\bar{U}$ is the average total speed of the ground truth trajectory. r_{max} is set to $\frac{pi}{100}$ radians per second (equivalent to 1.8 degrees per second). Further, we use the subscript i to denote the knot points within the trajectory. This specifies the number of timesteps required to predict the state of the vessel, *e.g.* x_{10} indicates the value of x for the vessel after 10 timesteps. The total number of knot points is denoted with n.

For each trajectory in the test set of each fold, we calculated the average value of the aforementioned metrics in (15), (16), (17). For each of those averaged metrics we calculated the following statistical values across folds:

TABLE III: Summarized statistics over 5 folds

Metric	μ	m	σ	IQR
ED (DD)	13.52	10.82	10.31	13.10
ED (ML)	20.84	17.10	12.94	17.87
EDH (DD)	18.79	15.59	14.52	17.87
EDH(ML)	28.49	22.62	18.37	22.86
cVDM (DD)	14.23	13.27	8.24	8.83
cVDM (ML)	28.95	22.99	20.45	20.10

- μ - mean value,
- m - median value,
- σ - standard deviation,
- IQR - interquartile range.

For every statistical value we calculate the mean value across 5 folds and we report these result in table III. We also provide a boxplot (see fig. 2) of the mean value (μ) of each metric to illustrate the performance of each model.

B. Observations on quantitative evaluation

The experimental results indicate that the data-driven physics-based model (DD) achieves lower error metrics compared to the pure ML model. This is reflected in the mean and median values across all evaluation measures. We can observe that DD μ & m values are around 40% lower than ML. This suggests that, on average, the physics-based approach provides more accurate trajectory predictions. For a 130-meter containership, the median position error μ of 13.52 meters for ED and 18.79 meters for EDH, for the DD model means that, on average, the predicted vessel position is off by about 10-15% of the vessel's length,while μ of 20.84 meters for ED and 28.49 meters for EDH, for the ML model represent a 16-22% of the vessel's length, a significant increase. It is important to note that these errors (especially for ML) tend to grow larger towards the later steps of each trajectory. This is because errors compound over time, especially for purely data-driven models that lack physical constraints. In practical terms, this means that for longer maneuvers, the ML model's predictions become progressively less reliable, while the DD model maintains better course-stability over time because its physics-based foundation resists drift that violates physics principles.

More importantly, a crucial observation in this comparison is the consistency of the models. The standard deviation (σ) and interquartile range (IQR) are significantly lower, around 35%, for the DD model, indicating that its predictions exhibit less variance. Consistency is particularly important in setups where such motion prediction models are used with decision making algorithms to enable informed planning and control.

C. Observations on qualitative evaluation

In order to visually demonstrate the differences between the predictions of the two models, five sample trajectories were chosen (see figs. 3 and 4). For each trajectory, the predicted trajectories obtained by the models are shown and compared against the ground truth trajectories. As it can be seen from the plots, the DD model is better in the average case staying

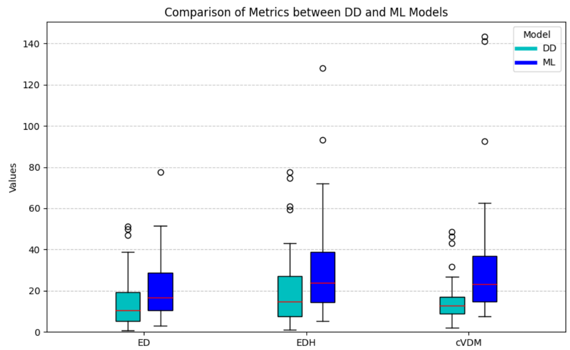


Fig. 2: The mean value distribution of each metric and each model presented in a boxplot.

closer to the real path than the ML model which has a similar behavior but in increased distance (see fig. 3). There are cases (see fig. 4a) where the ML model outperforms the DD model and predicts almost perfectly the groundtruth trajectory. Yet, in cases where the ML model drifts away, its performance rapidly deteriorates (see fig. 4b). This lack of robustness and rapid error accumulation is the main reason that the DD model exhibits better performance, which is reflected via the quantitative comparison.

V. Discussion

The physics-based model benefits from physical interpretability, making it robust to changes in operating conditions and ensuring adherence to known vessel dynamics. Its structure enables domain experts to inspect, validate, and refine parameters, leading to reliable long-term predictions. However, it requires extensive calibration and may struggle with capturing highly nonlinear effects not explicitly modeled.

In contrast, the ML model excels in flexibility and adaptability, learning complex patterns from data without requiring explicit equations. This makes it well-suited for environments where physical modeling is incomplete or difficult to formally define. Additionally, building an ML model is often easier in terms of required domain related expertise—given sufficient vessel data, an ML engineer can develop a working model, whereas a physics-based approach demands collaboration and extensive development efforts. However, the fundamental limitation of ML is its reliance on large, high-quality datasets—without sufficient data, it cannot generalize or function effectively. In such cases, the physics-based approach remains an attractive solution, as it provides structure and predictability even in data-scarce environments.

VI. Conclusion and Future Work

A. Summary of Key Findings

This study presented a comparison between two approaches for ship motion prediction: (i) a physics-based model with data-driven parameter optimization and (ii) a pure machine learning model. Both models were trained and evaluated using

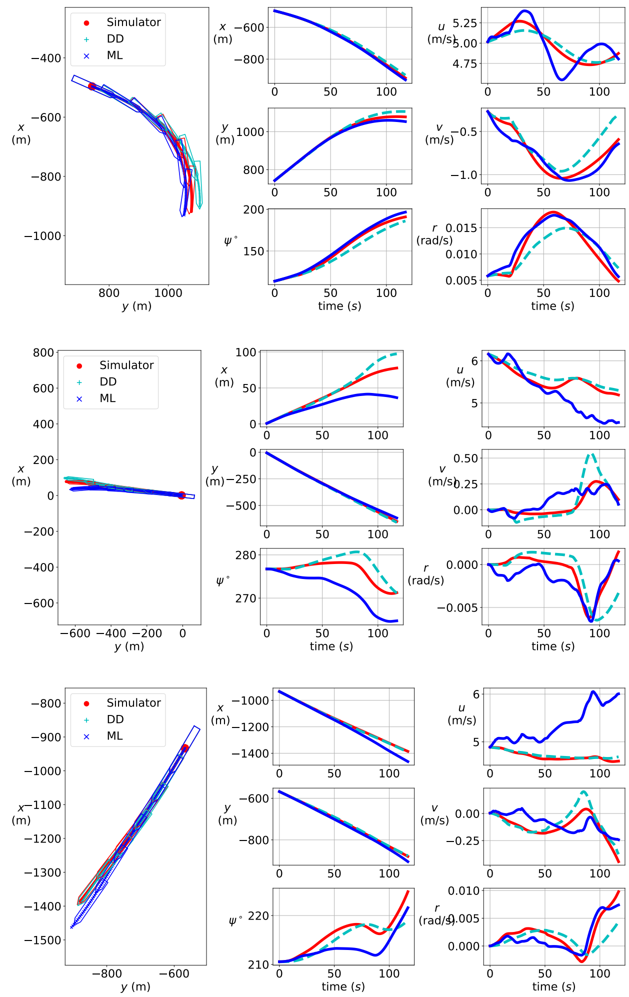
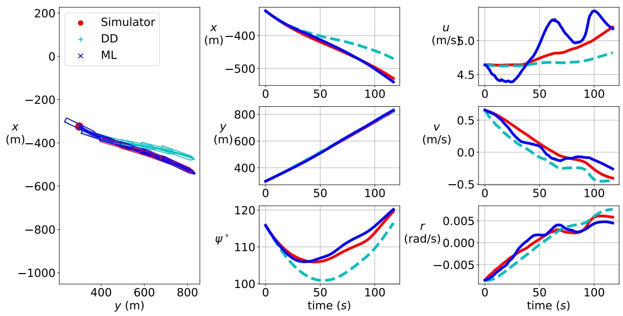


Fig. 3: Sample trajectories where the prediction error of the DD model is lower than the ML but ML-based predictions still follows closely the ground truth trajectories.

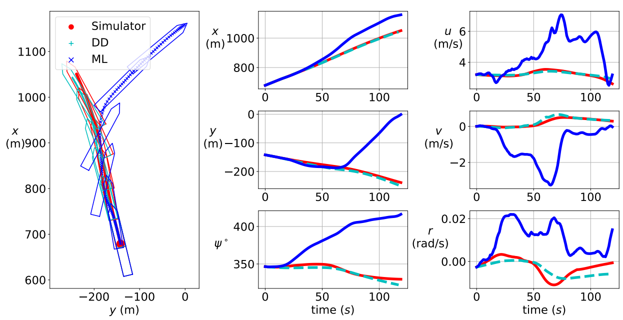
a synthetic dataset of 90 trajectories generated for a 130-meter container ship, focusing on low-speed port maneuvers. The dataset provided detailed insights into vessel dynamics, including surge, sway, and yaw velocities, under realistic operational conditions.

The results demonstrated that the physics-based model outperformed the pure ML approach across all evaluation metrics. The physics-based model achieved, on average, 40% lower error in Euclidean Distance (ED), in Euclidean Distance with Heading Penalization (EDH), and in Custom Vessel Distance Measure (cVDM). Moreover, it exhibited greater consistency, with approximately 35% lower standard deviation and interquartile range across the five cross-validation folds. This enhanced reliability is particularly valuable for integration into planning and control setups, where stable and consistent trajectory predictions are critical.

While the pure ML model showed adaptability in learning motion patterns from data, its higher variability and occasional



(a) ML model prediction performs better than the DD model. Minor divergence of DD model from the ground truth trajectory.



(b) DD model prediction outperforms ML model prediction which rapidly drifts away from the ground truth trajectory due to accumulating error.

Fig. 4: Sample trajectories where the models predictions differ more than the average case.

divergence from realistic vessel behavior highlighted the limitations of data-driven approaches when physical constraints are not explicitly considered. These findings emphasize the advantage of combining domain knowledge with data-driven techniques to achieve robust and interpretable ship motion prediction.

B. Future Work

Future work entails comparisons between different data-driven approaches using data from an actual ship, including environmental factors and complex interaction between the ship, the sea and the surroundings. In terms of machine learning models, in future comparisons we aim to consider models that operate on time sequences e.g. LSTM networks. Using models that consider a larger window of past states will likely benefit the predictive accuracy and reduce the pace of error accumulation, thus robustifying the ML models' performance.

References

[1] A. Ogawa, T. Koyama, and K. Kijima, "Mmg report-i, on the mathematical model of ship manoeuvring," *Bull Soc Naval Archit Jpn*, vol. 575, no. 22-28, 1977.

[2] K. J. Spyrou, "Asymmetric surging of ships in following seas and its repercussions for safety," *Nonlinear Dynamics*, vol. 43, pp. 149–172, 2006.

[3] R. Okuda, H. Yasukawa, T. Yamashita, and A. Matsuda, "Maneuvering simulations at large drift angles of a ship with a flapped rudder," *Applied Ocean Research*, 2023.

[4] G. Rajesh and S. Bhattacharyya, "System identification for nonlinear maneuvering of large tankers using artificial neural network," *Applied Ocean Research*, 2008.

[5] A. Maki, N. Sakamoto, Y. Akimoto, H. Nishikawa, and N. Umeda, "Application of optimal control theory based on the evolution strategy (cma-es) to automatic berthing," *Journal of Marine Science and Technology*, vol. 25, pp. 221–233, 2020.

[6] J. Woo, J. Park, C. Yu, and N. Kim, "Dynamic model identification of unmanned surface vehicles using deep learning network," *Applied Ocean Research*, vol. 78, 2018.

[7] K. S. M. Davidson and L. I. Schiff, "Turning and course keeping qualities," *Stevens Institute of Technology*, 1946.

[8] S. Motora, "On the measurement of added mass and added moment of inertia for ship motions," *Journal of Zosen Kiokai*, vol. 1959, no. 105, pp. 83–92, 1959.

[9] S. Inoue, M. Hirano, and K. Kijima, "Hydrodynamic derivatives on ship manoeuvring," *International Shipbuilding Progress*, vol. 28, no. 321, pp. 112–125, 1981.

[10] S. Motora, "Course stability of ships," *Journal of Zosen Kiokai*, vol. 1955, no. 77, pp. 69–90, 1955.

[11] K. Nomoto, T. Taguchi, K. Honda, and S. Hirano, "On the steering qualities of ships," *International Shipbuilding Progress*, vol. 4, no. 35, pp. 354–370, 1957.

[12] M. A. Abkowitz, "Ship hydrodynamics-steering and manoeuvrability," *Hydro-and Aerodynamics Laboratory, Hydrodynamics Section, Lyngby, Denmark, Report No. Hy-5, Lectures*, 1964.

[13] K. J. Åström and C. G. Källström, "Identification of ship steering dynamics," *Automatica*, vol. 12, no. 1, pp. 9–22, 1976.

[14] N. Norrbin, "Theory and observations on the use of a mathematical model for ship manoeuvring in deep and confined waters," Tech. Rep., 1971.

[15] H. Yasukawa and Y. Yoshimura, "Introduction of mmg standard method for ship maneuvering predictions," *Journal of Marine Science and Technology*, vol. 20, pp. 37–52, 2015.

[16] W. Wang, L. A. Mateos, S. Park, P. Leoni, B. Gheneti, F. Duarte, C. Ratti, and D. Rus, "Design, modeling, and nonlinear model predictive tracking control of a novel autonomous surface vehicle," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6189–6196.

[17] R. Li, T. Li, R. Bu, Q. Zheng, and C. L. P. Chen, "Active disturbance rejection with sliding mode control based course and path following for underactuated ships," *Mathematical Problems in Engineering*, vol. 2013, no. 1, p. 743716, 2013.

[18] Q. Zhang, X. Zhang, and N. Im, "Ship nonlinear-feedback course keeping algorithm based on mmg model driven by bipolar sigmoid function for berthing," *International Journal of Naval Architecture and Ocean Engineering*, vol. 9, no. 5, pp. 525–536, 2017.

[19] Y. Zheng, J. Tao, Q. Sun, H. Sun, Z. Chen, M. Sun, and G. Xie, "Soft actor–critic based active disturbance rejection path following control for unmanned surface vessel under wind and wave disturbances," *Ocean Engineering*, vol. 247, p. 110631, 2022.

[20] Y. Miyauchi, R. Sawada, Y. Akimoto, N. Umeda, and A. Maki, "Optimization on planning of trajectory and control of autonomous berthing and unberthing for the realistic port geometry," *Ocean Engineering*, vol. 245, p. 110390, 2022.

[21] R. Suyama, Y. Miyauchi, and A. Maki, "Ship trajectory planning method for reproducing human operation at ports," *Ocean Engineering*, vol. 266, p. 112763, 2022.

[22] Y. Yoshimura and Y. Masumoto, "Hydrodynamic force database with medium high speed merchant ships including fishing vessels and investigation into a manoeuvring prediction method," *Journal of the Japan Society of Naval Architects and Ocean Engineers*, vol. 14, pp. 63–73, 2012.

[23] S. Sutulo and C. Soares, "Development of a core mathematical model for arbitrary manoeuvres of a shuttle tanker," *Applied Ocean Research*, vol. 51, pp. 293–308, 2015.

[24] K. Kose, H. Hinata, Y. Hashizume, and E. Futagawa, "On a mathematical model of maneuvering motions of ships in low speeds," *Journal of the*

Society of Naval Architects of Japan, vol. 1984, no. 155, pp. 132–138, 1984.

[25] J. Takashina, "Ship maneuvering motion due to tugboats and its mathematical model," *Journal of the Society of Naval Architects of Japan*, vol. 1986, no. 160, pp. 93–102, 1986.

[26] K. Kijima, "On the practical prediction method for ship manoeuvring characteristics," *Trans West Jpn Soc Nav Archit*, vol. 105, pp. 21–31, 2002.

[27] Y. Yoshimura, "Mathematical model for the manoeuvring ship motion in shallow water (2nd report)-mathematical model at slow forward speed," *Journal of Kansai Society of Naval Architects*, vol. 210, no. 210, pp. 77–84, 1988.

[28] Y. Yoshimura, I. Nakao, and A. Ishibashi, "Unified mathematical model for ocean and harbour manoeuvring." International Conference on Marine Simulation and Ship Maneuverability, 2009.

[29] G. Delefortrie, K. Eloot, E. Lataire, W. Van Hoydonck, and M. Vantorre, "Captive model tests based 6 dof shallow water manoeuvring model," in *4th MASHCON*. Bundesanstalt f'ur Wasserbau, 2016, pp. 273–286.

[30] L. Moreira and C. Soares, "Dynamic model of manoeuvrability using recursive neural networks," *Ocean Engineering*, vol. 30, no. 13, pp. 1669–1697, 2003.

[31] X. Zhang and Z. Zou, "Black-box modeling of ship manoeuvring motion based on feed-forward neural network with chebyshev orthogonal basis function," *Journal of Marine Science and Technology*, vol. 18, pp. 42–49, 2013.

[32] D. Oskin, A. Dyda, and V. Markin, "Neural network identification of marine ship dynamics," *IFAC Proceedings Volumes*, vol. 46, no. 33, pp. 191–196, 2013.

[33] W. Luo and Z. Zhang, "Modeling of ship maneuvering motion using neural networks," *Journal of Marine Science and Application*, vol. 15, pp. 426–432, 2016.

[34] W. Ariza Ramirez, Z. Leong, H. Nguyen, and S. Jayasinghe, "Non-parametric dynamic system identification of ships using multi-output gaussian processes," *Ocean Engineering*, vol. 166, pp. 26–36, 2018.

[35] Y. Xue, Y. Liu, C. Ji, G. Xue, and S. Huang, "System identification of ship dynamic model based on gaussian process regression with input noise," *Ocean Engineering*, vol. 216, p. 107862, 2020.

[36] R. Suyama, R. Matsushita, R. Kakuta, K. Wakita, and A. Maki, "Parameter fine-tuning method for mmg model using real-scale ship data," *Ocean Engineering*, vol. 298, 2024.

[37] Y. Miyauchi, A. Maki, N. Umeda, D. Rachman, and Y. Akimoto, "System parameter exploration of ship maneuvering model for automatic docking/berthing using cma-es," *Journal of Marine Science and Technology*, vol. 27, 06 2022.

[38] M. Kanazawa, T. Wang, Y. Ichinose, R. Skulstad, G. Li, and H. Zhang, "Bridging similar ships' dynamics for safeguarding the system identification of maneuvering models," *Ocean Engineering*, vol. 280, p. 114874, 2023.

[39] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, 2018.

[40] E. Chondrodima, N. Pelekis, A. Pikrakis, and Y. Theodoridis, "An efficient lstm neural network-based framework for vessel location forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 4872–4888, 2023.

[41] M. Mathioudakis, C. Papandreou, T. Stouraitis, V. Margari, A. Nikitakis, S. Paschalakis, K. Kyriakopoulos, and K. J. Spyrou, "Towards real-world validation of a physics-based ship motion prediction model," 2025. [Online]. Available: https://arxiv.org/abs/2501.13804

[42] C. Papandreou, M. Mathioudakis, T. Stouraitis, P. Iatropoulos, A. Nikitakis, S. Paschalakis, and K. Kyriakopoulos, "Interpretable data-driven ship dynamics model: Enhancing physics-based motion prediction with parameter optimization," 2025. [Online]. Available: https://arxiv.org/abs/2502.18696

[43] K. J. Spyrou, "Dynamic instability in quartering seas: the behavior of a ship during broaching," *Journal of Ship Research*, vol. 40, no. 01, pp. 46–59, 1996.

[44] D. Clarke, P. Gedling, and G. T. Hine, "The application of manoeuvring criteria in hull design using linear theory," *International Shipbuilding Progress*, 1982.