Ocean Engineering 266 (2022) 112998

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Knowledge and data in cooperative modeling: Case studies on ship trajectory prediction

Motoyasu Kanazawa *, Tongtong Wang, Robert Skulstad, Guoyuan Li, Houxiang Zhang

The Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6009 Ålesund, Norway

ARTICLE INFO

Keywords:
Intelligent transportation systems
Ship automation
Trajectory prediction
Hybrid modeling
Model identification

ABSTRACT

A ship automation will be a key to the future maritime. In particular, ship dynamic models play an integral role. However, it is challenging to develop an accurate model readily. Recent studies proposed a physics-data cooperative model that predicts a future trajectory by compensating for position errors made by the physics-based model by using a machine learning model, which learns such a multiple-step-ahead compensation based on onboard sensor measurements. It seems to be promising to reduce effort in model development by exploiting observation data while having physics knowledge and a stable foundation in the model. However, it has been an open question "how much does the cooperative model benefit from physics knowledge and observation data?". We tackled this key question experimentally. To investigate the benefit of the physics-based model and the data amount, by changing the accuracy of the physics-based model and the size of observation dataset, simulation and full-scale experiments were conducted. Results show that the accuracy of the physics-based model and the data amount were complementary to each other to some extent. A wide range of physics-based models worked as prior knowledge, however, too inaccurate models disturbed the training.

1. Introduction

Recently, ship autonomy has gained an increasing attention from the research and industrial communities for ensuring operational safety and efficiency in the busy marine traffic, such as narrow channels and ports. They are expected to open new vistas in supporting or even substituting human onboard decision making to avoid human errors and make more efficient decisions (Norwegian Shipowners Association, 2019). Although we have seen the rapid success of the autonomous field robots and cars, that for autonomous ships remains topical. One of the reasons is the fact that an autonomous ship is a more comprehensive system composed of versatile marine robotics, automation, and sensing technologies.

In particular, it is of great importance to have a good situation awareness understanding what is happening now and will happen in the future surrounding a ship (e.g., Xiao et al. (2020) and Zhang et al. (2022)). A ship is a dynamic system with poor maneuverability, thereby, poor situation awareness may easily lead to fatal consequences, such as colliding with obstacles or stranding. In this context, for decades, researchers have been devoting their research effort to building an accurate ship dynamic model so that it predicts a future trajectory used for the early warning & prediction of the future collision risk. Fig. 1 illustrates the relationship between the ship dynamic model identification and the early warning & prediction in the versatile technologies for the ship autonomy in the future maritime transport.

Research communities have grouped approaches of ship dynamics modeling into three categories; namely, model-based, data-driven, and cooperative approaches. Model-based approaches formulate a linear regression model based on the understanding of physics. This study refers to such models as physics-based models. Model's parameters are identified by employing numerical simulations and model/full-scale experimental data (e.g., Wang et al. (2019)). The biggest advantage lies in the fact that they can be easily calibrated with a small dataset finding an optimal function parsimoniously in a parametric manner. Moreover, we can easily check the model's validity by looking into identified parameters. They have been playing a dominant role in applications by providing a stable and reliable foundation of the understanding of ship dynamics. However, the practice of parameter identification is rather sensitive revolving around a good-quality, sufficient, and balanced dataset in addition to the explicit understanding of real-world phenomena. Thereby, mostly, it ends up with poor performance although it takes prohibitive time & cost to be built with physically reasonable parameters. On the other hand, data-driven approaches offer parsimonious non-parametric models to achieve better performance by finding patterns in the dataset without depending on scientific

* Corresponding author.
E-mail address: motoyasu.kanazawa@ntnu.no (M. Kanazawa).

https://doi.org/10.1016/j.oceaneng.2022.112998
Received 30 August 2022; Received in revised form 10 October 2022; Accepted 22 October 2022
Available online 8 November 2022

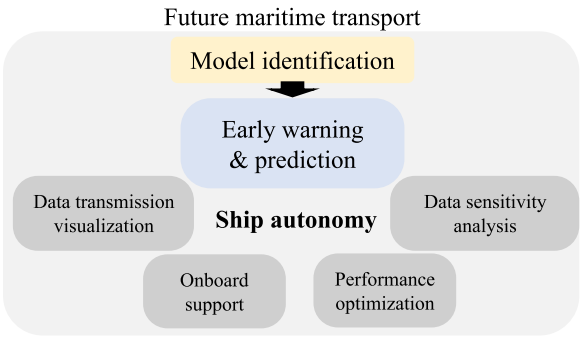

Fig. 1. A schematic overview of ship dynamic model identification and the early warning & prediction of the future collision risk in versatile technologies for ship autonomy in future maritime transport.

knowledge, although it may require more data to be calibrated than the physics-based model. Since the wake of the so-called third wave of artificial intelligence since the 1990s, data-driven approaches have been applied to challenging tasks in the maritime domain (e.g., the evaluation of ship pollutant emissions (Xiao et al., 2022) and the ship detection from videos (Chen et al., 2021)). Ship dynamics modeling is no exception to this trend (e.g., Kawan et al. (2017) and Schirmann et al. (2022)). However, pure data-driven models have no scientific interpretability. It becomes more of an issue in the application since the maritime industry is highly conservative in safety-critical systems. In addition, it is not a wise step for data-driven models to discard our domain knowledge packaged into the physics-based model.

A cooperative approach combines model-based and data-driven approaches. Recently, the maritime applications are not exceptions in a trend of making a synergy of scientific knowledge and data (e.g., Fonseca and Gaspar (2021)). Skulstad et al. (2021a) and Wang et al. (2021) presented breakthrough ideas that compensate/map trajectories made by the available physics-based model into true trajectories in a data-driven manner. Such data-driven geometrical compensation/mapping achieved a good performance while keeping the stability of the physics-based model untouched as a stable foundation of the model. Moreover, in their approaches, we can clearly distinguish the contributions of the physics-based and data-driven models on the prediction performance as they play different roles in the model, thus having good interpretability and maintainability of the model with a good performance. It seems to be promising to overcome time & cost challenges the maritime industry faces since it would lower rigorous hurdles of model-based or data-driven approaches by combining two approaches.

An open question about such cooperative approaches is "how much does the cooperative model benefit from physics knowledge and observation data?" to achieve a good performance in a physics-data cooperative way (referred to as the cooperative performance, hereinafter). In not only the academic but also the industrial views, this question is important from the two perspectives. On the one hand, the accuracy of the available physics-based model is diverse. It could be degraded, for instance, due to the poor conduct of the parameter identification, parameters identified in a compromised manner (e.g., copy & paste parameters of similar ships), low-fidelity actuator models, and being tuned to the other operations. On the other hand, available data are mostly limited. Real-world ship maneuvers are required for the data collection, however, it is money- and time-consuming. Thereby, a better understanding of the impact of these two components on the cooperative performance is of great interest to our industrial partners. To the best of the authors' knowledge, so far, this open question has not been addressed in any literature, albeit its importance in industrial applications. In this study, we validate a cooperative framework, which builds an accurate ship dynamic model by combining a compromised physics-based model and limited observation data.

To offer one solution to the open question "how much does the cooperative model benefit from physics knowledge and observation data?", this study conducted experimental investigations, which are divided into two parts. First, simulation experiments enabled us to investigate the impact of the accuracy of the physics-based model and the data amount on the cooperative performance. Second, in the full-scale experiment where only limited observation data are available, we further explore the reasonable range of the physics-based model's accuracy on the cooperative performance. The full-scale experiment was conducted by the 33.9m-length research vessel Gunnerus. The results showed that we could achieve a good performance by using the combination of the compromised physics-based model and a small dataset. The cooperative performance was equivalent to the performance of the accurate physics-based model, which takes much more time & cost to be built. Contributions of this study are summarized as follows:

- It was found that the balance of the accuracy of the physics-based model and the data amount was key to achieve a good performance of the physics-data cooperative model rather than relying on either of them. In addition, the full-scale experiment presented the validity of building a cooperative model with a compromised physics-based model and a small observation dataset. These findings make the cooperative model more promising for reducing effort dedicated for the model development by using physics knowledge and observation data.
- Although a wide range of physics-based models successfully facilitated the model identification, however, it disturbed the training if it was too inaccurate. This finding highlights the importance of technologies that develop a simplified physics-based model readily without compromising its performance drastically.

Hereinafter, this paper unfolds as follows. Section 2 illustrates related works aiming at the synergy of the scientific knowledge and data in different applications. In Section 3, we explain the cooperative ship dynamic model employed in this study. An experimental study in the simulation environment is presented in Section 4. A full-scale experiment is illustrated in Section 5. Conclusions are given in Section 7.

2. Related works

In many applications, it is seen to leverage scientific knowledge and data for better performance and reliability. In Karpatne et al. (2016), Karpatne et al. named such approaches Theory-Guided Data Science (TGDS), and grouped diverse approaches into five categories; namely, theory-guided design of data science models, theory-guided learning of data science models, theory-guided refinement of data science outputs, hybrid models of data science and theory, and augmenting theory-based models using data science. Besides TGDS, different terminologies (e.g., transfer learning (Panigrahi et al., 2021), physics-informed Machine Learning (ML) (Karniadakis et al., 2021), informed ML (Vonrueden et al., 2021), and gray-box/semi-parametric modeling (von Stosch et al., 2014)) fully/partly cover ideas of the cooperative approach in the field of ship dynamics. There is no domain-agnostic "best practice" of the cooperative approach, thereby, it is necessary to carry out a domain-specific investigation to achieve the good harmony of scientific knowledge and data.

In this study, we assume we have a physics-based model and a new dataset of the ship maneuver. The physics-based model could be derived in the compromised manner and the new dataset is not satisfactory in terms of its amount, quality, or distribution. In such settings, the most straightforward and classic way to model ship dynamics is to re-identify parameters of the physics-based model so that it performs well in the prepared dataset by using ML algorithms (e.g., support vector machine (Wang et al., 2019; Luo and Li, 2017), a Bayesian approach (Xue et al., 2020)). This approach belongs to "theory-guided

(a) Simulation experiment (b) Full-scale experiment

Fig. 2. Snapshots of (a) simulation experiments in Section 4 and (b) full-scale experiments in Section 5. (b) was taken on November 21st, 2019 in the west coast of Norway.

design of data science models" in the TGDS's categorization. If the identification succeeds, it provides a highly interpretable model. However, it has been a challenging task since our understanding of ship dynamics mostly does not fully capture real-world phenomena. This challenge brought us to use non-parametric models in this domain. Since the maritime industry highly values the reliability of physics-based models, previous research in this field has devoted its effort to building a cooperative architecture of the physics-based and data-driven models. It corresponds to "hybrid models of data science and theory" in the TDGS's categorization. Ven et al. employed a neural network for representing damping terms of the physics-based model (van de Ven et al., 2007). In Skulstad et al. (2021b) and Kanazawa et al. (2022), neural networks were used for compensating for the single-step-ahead prediction error made by physics-based models. Mei et al. (2019) employed a random forest to map the estimation of the acceleration made by the similar ship's dynamic model into that of the targeting ship by using a dataset of the targeting ship. Those approaches directly intervene in the performance of the physics-based model by data-driven models, thus making one unified trajectory in the physics-data cooperative manner. Their cooperative approaches are efficient thanks to their simple architecture, however, their stability in the numerical iteration is hardly validated. Moreover, once trajectories are generated, it is impossible to isolate the contribution of the physics-based and data-driven models from the generated trajectory. Thereby, in practice, they can be used only when we have a relatively-accurate physics-based model due to reliability reasons.

On the other hand, Skulstad et al. (2021a) trained a neural network with onboard sensor measurements so that it compensates for the multiple-step-ahead position error made by the physics-based model. In Wang et al. (2021), Wang et al. proposed a data-driven model that maps the future position calculated by the similar ship's dynamic model into that of the targeting ship. In their approaches, the roles of the physics-based and data-driven models are clearly distinguished, thus contributing to better interpretability and maintainability of the cooperative architecture. Moreover, in their approaches, the physics-based model serves not only as prior knowledge of ship dynamics but also as a stable foundation of the prediction.

3. Cooperative ship model

As explained in Section 2, in the field of ship dynamics, previous studies have presented different types of cooperative models combining physics-based and data-driven models. In this study, we employ a geometry-based cooperative model, that makes a data-driven compensation for multiple-step-ahead position errors made by the physics-based model, based on the idea presented in Skulstad et al. (2021a) and Kanazawa et al. (2021).

3.1. Overview

In the experiments of this study, we employed a cooperative model of ship dynamics as shown in Fig. 3. The cooperative model is composed of a physics-based model highlighted in orange and a data-driven compensator highlighted in green. The physics-based model makes Ts prediction of a future trajectory based on the initial state of the ship, environmental disturbances, and commands to actuators. On the other hand, based on onboard measurement data, the data-driven compensator compensates for errors in the position made by the physics-based model, Thus, a multiple-step-ahead position prediction vector made by the physics-based model is calibrated in a data-driven manner. By adding outputs of both of them, the cooperative model makes Ts prediction of the future trajectory in a data-driven manner while having a stable and reliable model-based prediction made by the physics-based model as its foundation. Details of the physics-based model and data-driven compensator will be explained hereinafter.

3.2. Physics-based model

In the maneuvering theory of ship dynamics, the ship kinematics is expressed as:

$$\dot{\boldsymbol{\eta}} = \boldsymbol{R}(\psi)\boldsymbol{\nu} \tag{1}$$

where $\boldsymbol{\eta}$ is the vector of the ship's positions in the inertial coordinate, $\boldsymbol{R}$ is the rotation matrix between the inertial and body-fixed coordinate, ψ is the ship's heading, and $\boldsymbol{\nu}$ is the vector of the ship's velocities in the body-fixed coordinate. We define $\boldsymbol{R}$ as:

$$\boldsymbol{R}(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The ship kinetics is expressed as:

$$\boldsymbol{M}_{\boldsymbol{RB}}\dot{\boldsymbol{\nu}} + \boldsymbol{M}_{\boldsymbol{A}}\dot{\boldsymbol{\nu}}_{\boldsymbol{r}} + \boldsymbol{C}_{\boldsymbol{RB}}(\boldsymbol{\nu})\boldsymbol{\nu} + \boldsymbol{C}_{\boldsymbol{A}}(\boldsymbol{\nu}_{\boldsymbol{r}})\boldsymbol{\nu}_{\boldsymbol{r}} + \boldsymbol{D}(\boldsymbol{\nu}_{\boldsymbol{r}}) = \boldsymbol{q} \tag{3}$$

$$\boldsymbol{q} = \boldsymbol{q}_{\text{wind}} + \boldsymbol{q}_{\text{wave}} + \boldsymbol{q}_{\text{thr}} \tag{4}$$

where $\boldsymbol{M}_{\boldsymbol{RB}}$ is the rigid-body mass matrix, $\boldsymbol{M}_{\boldsymbol{A}}$ is the added-mass matrix, $\boldsymbol{C}_{\boldsymbol{RB}}(\boldsymbol{\nu})$ is the rigid-body coriolis-centripetal matrix, $\boldsymbol{C}_{\boldsymbol{A}}(\boldsymbol{\nu}_{\boldsymbol{r}})$ is the added-mass coriolis-centripetal matrix, $\boldsymbol{D}(\boldsymbol{\nu}_{\boldsymbol{r}})$ is the damping matrix, $\boldsymbol{\nu}_{\boldsymbol{r}} = \boldsymbol{\nu} - \boldsymbol{\nu}_{\boldsymbol{c}}$ is the relative velocity vector, $\boldsymbol{\nu}_{\boldsymbol{c}}$ is the current velocity vector, $\boldsymbol{q}_{\text{wind}}$ is the wind-force vector, $\boldsymbol{q}_{\text{wave}}$ is the wave-force vector, and $\boldsymbol{q}_{\text{thr}}$ is the thruster force vector. In this study, we assume the effects of the ocean current and waves on the ship motion are marginal due to the limitation that ships are not equipped with sensors measuring ocean currents and waves in real time. This assumption is acceptable under mild environmental disturbances. Thereby, this study introduces $\boldsymbol{\nu}_{\boldsymbol{r}} = \boldsymbol{\nu}$ and $\boldsymbol{q}_{\text{wave}} = \boldsymbol{0}$. The wind-force model is constructed by dedicated numerical simulation and experiments as:

$$\boldsymbol{q}_{\text{wind}} = \frac{1}{2}\rho_a V_{rw}^2 \begin{bmatrix} C_X(\gamma_{rw})A_{FW} \\ C_Y(\gamma_{rw})A_{LW} \\ C_N(\gamma_{rw})A_{LW}L_{oa} \end{bmatrix} \tag{5}$$

where C_X, C_Y, and C_Z are the wind coefficients identified for the surge, sway, and yaw directions. A_{FW} and A_{LW} are the frontal and

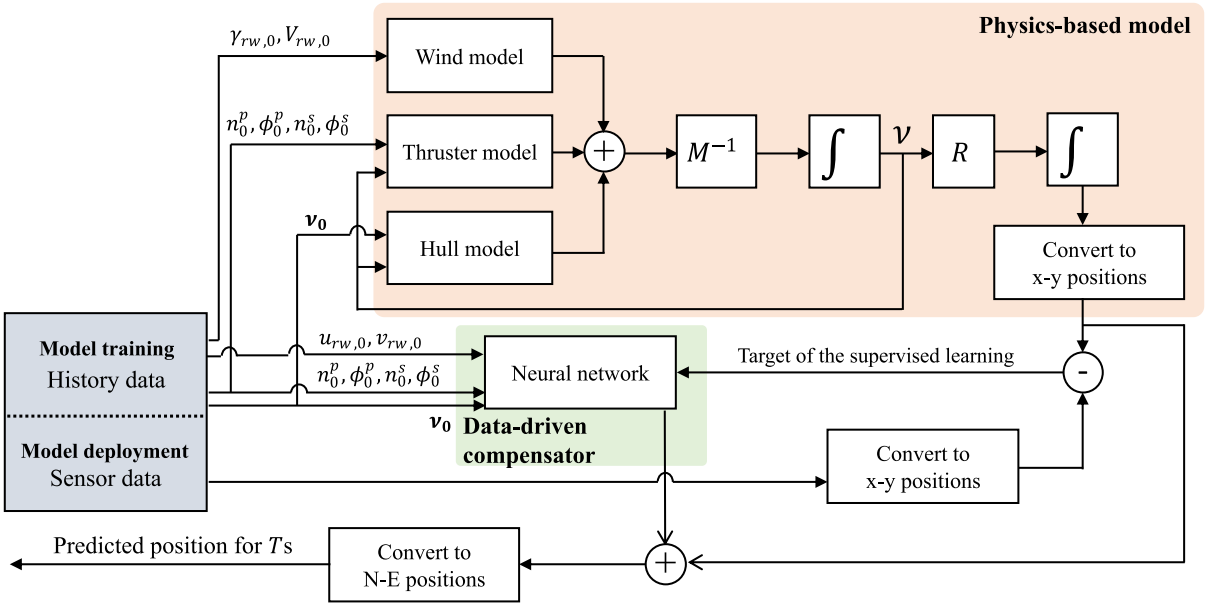

Fig. 3. An overview of the cooperative model. Sensor data (current ship's states, thruster command values, and wind information) are given to the physics-based model and data-driven compensator. The physics-based model is a 3DOF maneuvering model outputting a trajectory prediction. The data-driven compensator, of which input–output relationship is shown in (13), compensates for multiple-step-ahead position errors made by the physics-based model.

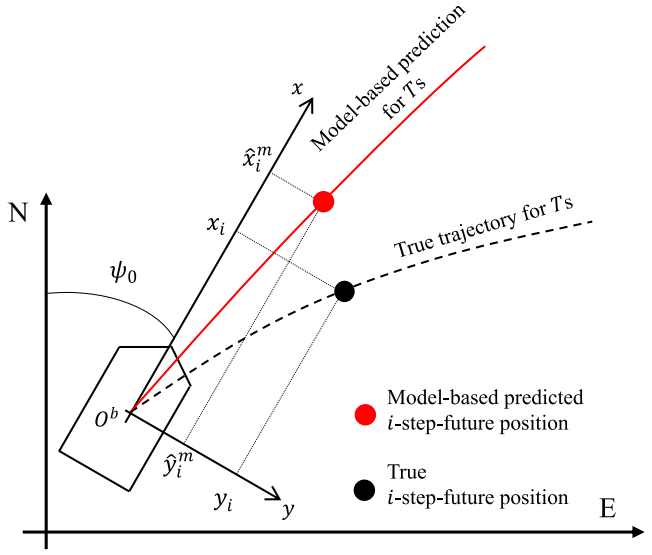

Fig. 4. A schematic relationship between the body-fixed coordinate when making a prediction and the inertial coordinate. The data-driven compensator compensates for position errors made by the physics-based model in the body-fixed coordinate when making a prediction.

lateral projected areas, respectively. L_{oa} represents the ship length. The relative wind velocity V_{rw} and direction γ_{rw} are given as:

$$V_{rw} = \sqrt{u_{rw}^2 + v_{rw}^2} \tag{6}$$

$$\gamma_{rw} = -\mathrm{atan2}(v_{rw}, u_{rw}) \tag{7}$$

where:

$$u_{rw} = u - V_w \cos(\beta_w - \psi) \tag{8}$$

$$v_{rw} = v - V_w \sin(\beta_w - \psi) \tag{9}$$

V_w and β_w are the true wind velocity and direction in the inertial coordinate. The thruster-force vector $\boldsymbol{q}_{\mathrm{thr}}$ is calculated based on thruster commands and $\boldsymbol{\nu}$ by the mathematical model f_{thr} provided by manufacturers of thrusters. In this study, the ship is equipped with two azimuth thrusters and one bow thruster. The bow thruster was turned off in the experiments. Namely:

$$\boldsymbol{q}_{\mathrm{thr}} = f_{thr}(\boldsymbol{\nu}, n^p, \delta^p, n^s, \delta^s) \tag{10}$$

where n^p and n^s represent thruster revolutions of the port- and starboard-side azimuth thrusters. δ^p and δ^s are thruster angles of the port- and starboard-side thrusters. Hence, by using wind, thruster, and hull models, total forces and moment acting on the hull are calculated. By multiplying $\boldsymbol{M}^{-1} = (\boldsymbol{M_{RB}} + \boldsymbol{M_A})^{-1}$, the acceleration vector is estimated. We numerically integrate the estimated acceleration vector into the velocity vector in the body-fixed coordinate. The numerical integration of this velocity vector over the prediction horizon yields the model-based predicted trajectory $[\hat{N}_1^m, \ldots, \hat{N}_i^m, \ldots, \hat{N}_{n_T}^m, \hat{E}_1^m, \ldots, \hat{E}_i^m, \ldots, \hat{E}_{n_T}^m]$ in the future, where $\hat{N}_i^m$ and $\hat{E}_i^m$ represent model-based predicted ship's north and east positions at ith step future, respectively. n_T denotes the number of time steps of the prediction horizon. In this study, the Euler method is employed for the numerical integration with 1s time step; namely, $n_T = T$.

As shown in Fig. 4, this study expresses trajectories in the $x-y$ coordinate of which origin is located at the center of gravity of the ship when making a prediction. The positive directions of the x and y axes are the longitudinal and lateral directions of the ship. Thereby, future positions in the $x-y$ and $N-E$ coordinates are interconvertible as:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\psi_0 & \sin\psi_0 \\ -\sin\psi_0 & \cos\psi_0 \end{pmatrix} \begin{pmatrix} N - N_0 \\ E - E_0 \end{pmatrix} \tag{11}$$

where N_0, E_0, and ψ_0 represent the north, east positions and heading when making a prediction. Hence, the model-based predicted trajectory $[\hat{N}_1^m, \ldots, \hat{N}_{n_T}^m, \hat{E}_1^m, \ldots, \hat{E}_{n_T}^m]$ in the $N-E$ coordinate is converted to $[\hat{x}_1^m, \ldots, \hat{x}_{n_T}^m, \hat{y}_1^m, \ldots, \hat{y}_{n_T}^m]$ in the $x-y$ coordinate to provide target vectors of the data-driven compensator.

3.3. Data-driven compensator

The data-driven compensator makes prediction of $[\Delta\hat{x}_1^m, \ldots, \Delta\hat{x}_{n_T}^m, \Delta\hat{y}_1^m, \ldots, \Delta\hat{y}_{n_T}^m]$ where $\Delta\hat{x}_i^m = x_i - \hat{x}_i^m$ and $\Delta\hat{y}_i^m = y_i - \hat{y}_i^m$, respectively, based on onboard sensor measurements. x_i and y_i are the true position at the i-step future. This target vector is given not in the model deployment but in the model training.

3.3.1. Input features

Input features of any ML models must be carefully selected. Otherwise, ML models would suffer from missing important information. In such settings, ML models fail to be efficiently trained. This study selects input features based on the theory of ship dynamics that future trajectories are determined by inertial and hydrodynamic parameters of the ship, the initial state of the ship, commands to thrusters, and environmental disturbances. As we develop data-driven compensators for the specific loading condition of the specific ship, the inertial and hydrodynamic parameters are assumed to be constant. Wave and ocean current data are not included in input features as they are mostly not measured in real time. We assume commands to thrusters, the true wind velocity, and the true wind direction are kept unchanged over the prediction horizon. These assumptions yield the formulation of the data-driven compensator f_N:

$$[\Delta\hat{x}_1^m, \dots, \Delta\hat{x}_{n_T}^m, \Delta\hat{y}_1^m, \dots, \Delta\hat{y}_{n_T}^m] = f_N(\boldsymbol{\nu_0}, n_0^p, \delta_0^p, n_0^s, \delta_0^s, u_{rw,0}, v_{rw,0}) \tag{12}$$

The suffix 0 represents the values when making a prediction. The input vector is z-score normalized with the statistic values in the training dataset. The same values are applied to the normalization in the validation and test datasets. In the experiments of this study, two azimuth thrusters are manipulated with the same commands. Thereby, (12) is reduced to:

$$[\Delta\hat{x}_1^m, \dots, \Delta\hat{x}_{n_T}^m, \Delta\hat{y}_1^m, \dots, \Delta\hat{y}_{n_T}^m] = f_N(\boldsymbol{\nu_0}, n_0, \delta_0, u_{rw,0}, v_{rw,0}) \tag{13}$$

where $n_0 = n_0^p = n_0^s$ and $\delta_0 = \delta_0^p = \delta_0^s$.

3.3.2. Model training

This study employs a MultiLayer Perceptron (MLP), a fully-connected feedforward neural network, which is one of the most classic architectures of neural networks. It consists of an input layer, hidden layer(s), and an output layer. tanh and linear functions are used for the hidden layer(s) and the output layer, respectively. Weights and biases are updated in the manner of the backpropagation by using Adam (Kingma and Ba, 2015) optimizer so that it minimizes the mean squared error between MLP's output and target vectors. During training, we separate some maneuvers from a training–validation dataset and keep them as a validation dataset. The validation loss is monitored to avoid overfitting the training dataset. If the validation loss does not improve over 200 epochs, the training is automatically terminated, and the best model is used for the prediction (early stopping). In this study, we build an MLP in the Pytorch (Paszke et al., 2019) framework in Python.

3.3.3. Hyperparameter tuning

Hyperparameters are parameters to be fixed in advance to determine ML model's architecture and training setting. A hyperparameter tuning is important to achieve a good performance of ML models. In this study, the number of hidden layers $\in [1, 3]$, the number of units in hidden layer(s) $\in [10, 500]$, the drop-out rate in the input layer $\in [0.0, 1.0]$, the drop-out rate in hidden layer(s) $\in [0.0, 1.0]$, and the learning rate of the optimizer $\in [10^{-5}, 10^{-1}]$ are optimized. The hyperparameter tuning is an optimization problem finding the best set of hyperparameters that performs the best in the validation dataset. In this study, such an optimum set is searched by using the Tree-structured Parzen Estimator (TPE) optimizer in the optuna (Akiba et al., 2019) framework. The TPE optimizer is one of the Bayesian optimization methods. It has been widely used for the hyperparameter tuning with a great performance and small computational time. Thereby, in the optuna, the TPE is selected as a default algorithm. In this section, details of the TPE algorithm are not revisited as it is not the focus of this study. For further information, original articles (Bergstra et al., 2013, 2011) for the TPE can be referred. The number of trials for the parameter search is 50 as further drastic improvement of the validation loss was not found with the larger number of trials than 50. The learning rate is searched in the log domain. After 50 trials of the hyperparameter search, a set of hyperparameters with the best performance in the validation dataset was selected as a set of optimum hyperparameters. Hyperparameter tuning was conducted independently for having different physics-based models and dataset.

3.3.4. Model deployment

In the model deployment, the data-driven compensator makes prediction $[\Delta\hat{x}_1^m, \dots, \Delta\hat{x}_{n_T}^m, \Delta\hat{y}_1^m, \dots, \Delta\hat{y}_{n_T}^m]$ based on input vectors provided by onboard sensors. By adding the model-based predicted position vector $[\hat{x}_1^m, \dots, \hat{x}_{n_T}^m, \hat{y}_1^m, \dots, \hat{y}_{n_T}^m]$, the cooperative prediction yields $[\hat{x}_1, \dots, \hat{x}_{n_T}, \hat{y}_1, \dots, \hat{y}_{n_T}]$ where $\hat{x}_i = \Delta\hat{x}_i^m + \hat{x}_i^m$ and $\hat{y}_i = \Delta\hat{y}_i^m + \hat{y}_i^m$. It is re-converted to the position vector in the $N-E$ coordinate by (11).

3.4. Evaluation metrics

The accuracy of the physics-based model A is evaluated with the Root Mean Squared Error (RMSE) of the geometrical similarity between true and predicted trajectories in the test dataset.

$$A = \frac{1}{S}\sum_{k=1}^{S}\sqrt{(N_k - \hat{N}_k^m)^2 + (E_k - \hat{E}_k^m)^2} \tag{14}$$

N_k and E_k represent the true north and east positions of the kth sample. $\hat{N}_k^m$, and $\hat{E}_k^m$ are the north and east positions of the kth sample predicted by the physics-based model. S is the number of data samples. N_D represents the number of maneuvers in the dataset used for the training. The larger the N_D is, the larger the dataset is. Hence, cooperative models are characterized by the combination (A, N_D) in this study. It should be noted that the ship dynamics is highly complex and nonlinear, thereby, a single metric A does not fully represent the characteristic of the physics-based model.

The errors made by the cooperative model H of (A, N_D) is evaluated with the RMSE of the geometrical similarity between the true and predicted trajectories made by the cooperative model:

$$H = \frac{1}{S}\sum_{k=1}^{S}\sqrt{(N_k - \hat{N}_k)^2 + (E_k - \hat{E}_k)^2} \tag{15}$$

where $\hat{N}_k$ and $\hat{E}_k$ are the north and east positions of the kth sample, predicted by the cooperative model. It should be noted that A is not used for selecting a good physics-based model for better cooperative performance since it is defined in the test dataset. A is only employed for presenting the relationship between H and (A, N_D) in the test dataset to develop a better understanding of the contribution of the physics-based model and data to the cooperative performance. If $N_D = 0$, we substitute $\hat{N}_k^m$ and $\hat{E}_k^m$ for $\hat{N}_k$ and $\hat{E}_k$ in graphs since the cooperative model without using any data is regarded as the physics-based model. If the pure data-driven model is used, $\hat{N}_k$ and $\hat{E}_k$ are calculated without the help of the physics-based model. In the application, large prediction errors become an issue since they could negatively affect our decision making ending up with fatal consequences. Thereby, we define the 90% of percentile of $\sqrt{(N_k - \hat{N}_k)^2 + (E_k - \hat{E}_k)^2}$ as H_{90} to examine the occurrence of large prediction errors.

4. Simulation experiment

Before performing full-scale experiments in Section 5, we examine the contribution of the physics-based model and data to the cooperative performance in the simulation environment. In the simulation environment, we can build physics-based models with different accuracy in a flexible manner since the ground-truth model is known. In addition, we can efficiently investigate the impact of data amount on the cooperative performance as we can generate virtual maneuvers as much as we need. Since trajectory predictions are used for the purpose of the early

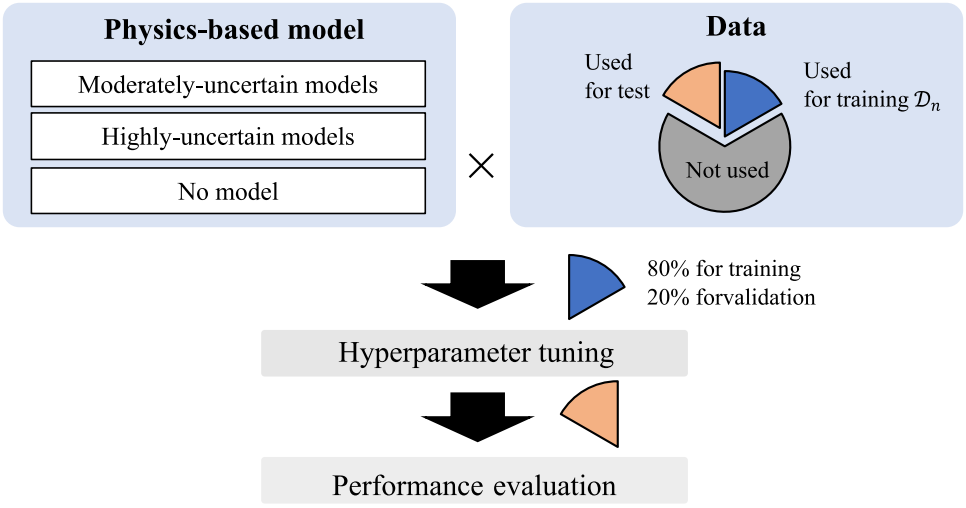

Fig. 5. Experimental setting.

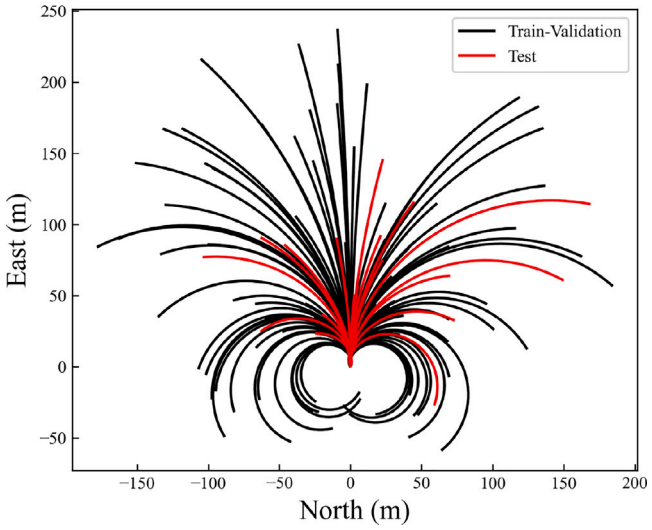

Fig. 6. Trajectories of maneuvers employed in the simulation experiment. Black trajectories show maneuvers grouped into the training–validation dataset and red trajectories show maneuvers grouped into the test dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

warning of the collision risk, T, which is longer than the time that the ship can take evasive actions, is preferable for the evaluation of ship dynamic models. In the simulation experiment, the cooperative model was trained and evaluated for making $T = 30$s trajectory prediction.

4.1. Overview

The overview of simulation experiments in this study is illustrated in Fig. 5. 18 different physics-based models were used in the cooperative models. A dataset with 120 maneuvers was prepared in this study. They are explained in detail in the following subsections. 20 maneuvers in the dataset were randomly selected for the test dataset and kept untouched during training and validation of the cooperative models. By selecting N_D maneuvers from the remaining 100 maneuvers, we built the training and validation sub dataset D_{N_D} with the different number of maneuvers. Please note that $D_{N_D=a} \subset D_{N_D=b}$ if $a < b$. In this study, ten sub datasets from $D_{N_D=10}$ to $D_{N_D=100}$ were prepared. We trained the cooperative models with different physics-based models and sub datasets; thus examining the impact of the accuracy of the physics-based model and dataset on the cooperative performance. For the different combinations of the physics-based models and sub datasets, hyperparameter tuning was conducted independently. In D_{N_D}, 80% of maneuvers were used for the training and the remaining 20% were used for the validation. The performance of the trained cooperative model was examined by using 20 maneuvers in the test dataset. The test dataset was always identical regardless of which sub dataset was used during training.

4.2. Dataset

A simulation dataset was generated by using a six Degrees of Freedom (DoF) seakeeping and maneuvering model of the R/V Gunnerus, which is a 28.9m-length Norwegian University of Science and Technology (NTNU)'s research ship. It is a high-fidelity ship dynamic model provided in the Open Simulation Platform project, which is a joint project with Kongsberg Maritime, DNV, SINTEF, and NTNU. It is composed of a hull model (Ross, 2008; Hassani et al., 2015) and thruster models running on the simulation platform Vico (Hatledal et al., 2021). Two azimuth thrusters were manipulated simultaneously. 120 unique turning maneuvers were generated by randomly selecting thruster revolution $n \in [50, 200]$ Revolution Per Minute (RPM), and thruster angle $\delta \in [-50, 50]°$. The ship's motion was disturbed by the constant wind and irregular waves in the simulation. The true wind direction $\beta_w \in [0, 360)°$, the true wind speed $V_w \in [0, 6]$ m/s, and the global wave direction $\in [0, 360)°$ are randomly chosen for each maneuver. The wave spectrum was JONSWAP spectrum (Hasselmann et al., 1973) with 1.0 m significant wave height and 5.0s significant wave period. The time step of the simulation environment was 0.05s. 50 s time series were saved in 1 Hz with 30 s future trajectory at each time step for each maneuver. Future trajectories were used only for training and evaluation purposes. Generated trajectories are shown in Fig. 6. A snapshot of a simulation experiment is shown in Fig. 2(a). Minimum, mean, and maximum input values of the data-driven compensator in datasets are shown in Table 1.

4.3. Physics-based models

In this experiment, cooperative models were trained with different physics-based models. They were developed by shifting parameters of the ground-truth model used in the simulation environment. This procedure introduced the model's uncertainty that we had in reality due to poorly identified parameters. We randomly produced 18 physics-based models with parameter uncertainty to examine the impact of the

Table 1
Minimum, mean, and maximum input values in the sub datasets for the training and validation and the test dataset in the simulation experiment.

		$\mathcal{D}_{10}$	$\mathcal{D}_{20}$	$\mathcal{D}_{30}$	$\mathcal{D}_{40}$	$\mathcal{D}_{50}$	$\mathcal{D}_{60}$	$\mathcal{D}_{70}$	$\mathcal{D}_{80}$	$\mathcal{D}_{90}$	$\mathcal{D}_{100}$	Test
u_0 (m/s)	min	0.3	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
	mean	2.2	2.3	2.2	2.2	2.2	2.2	2.3	2.4	2.4	2.4	1.9
	max	5.4	5.4	5.4	5.4	5.8	5.8	5.8	6.0	6.0	6.0	5.3
v_0 (m/s)	min	−0.9	−0.9	−0.9	−1.1	−1.1	−1.1	−1.1	−1.1	−1.1	−1.1	−0.8
	mean	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	max	1.0	1.0	1.0	1.0	1.1	1.1	1.1	1.1	1.1	1.1	0.8
r_0 (°/s)	min	−8.2	−8.2	−8.2	−8.2	−8.2	−8.2	−8.2	−8.2	−8.2	−8.2	−3.9
	mean	−0.5	−0.3	−0.4	−0.1	−0.1	−0.1	0.0	0.1	−0.1	0.0	0.2
	max	6.4	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	7.8	5.2
n_0 (RPM)	min	62.8	62.8	55.9	50.4	50.4	50.4	50.4	50.4	50.4	50.4	50.0
	mean	133.4	133.3	126.9	126.2	122.7	123.6	125.2	128.9	130.1	129.4	103.9
	max	191.7	192.4	192.4	192.4	199.6	199.6	199.6	199.6	199.6	199.6	189.1
δ_0 (°)	min	−37.3	−45.8	−45.8	−47.4	−47.4	−47.4	−47.4	−48.7	−48.7	−48.7	−38.7
	mean	1.7	0.8	1.9	−0.4	−0.7	−2.3	−2.6	−3.6	−2.1	−2.7	0.7
	max	48.3	48.3	48.3	48.3	48.3	48.3	48.3	48.3	49.4	49.4	41.4
$u_{rw;0}$ (m/s)	min	−4.0	−4.0	−4.0	−4.0	−4.2	−4.2	−4.4	−4.4	−4.4	−4.4	−3.0
	mean	2.9	3.1	3.1	2.7	2.5	2.5	2.6	2.7	2.8	2.8	2.5
	max	7.5	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.4	10.4	7.4
$v_{rw;0}$ (m/s)	min	−5.1	−5.4	−6.0	−6.0	−6.0	−6.0	−6.0	−6.0	−6.0	−6.0	−6.4
	mean	0.3	0.1	−0.1	−0.1	0.0	0.0	−0.2	−0.1	0.1	0.0	−0.4
	max	5.9	5.9	5.9	5.9	5.9	5.9	5.9	5.9	6.3	6.3	5.8

accuracy of the physics-based model on the cooperative performance. 18 models were grouped into moderately- and highly-uncertain models. It should be noted that none of them were identical to the ground-truth model used in the simulation. They produced prediction errors at different levels due to different reasons.

4.3.1. Moderately-uncertain models

We prepared ten physics-based models by randomly shifting parameters of the ground-truth model in $\boldsymbol{D}(\boldsymbol{\nu}_r)$. They were grouped into moderately-uncertain models in this paper. The ground-truth model has 32 hydrodynamic derivatives $\theta_1 - \theta_{32}$ in $\boldsymbol{D}(\boldsymbol{\nu}_r)$ (such as X^L_{uu}, see Ross et al. (2015) for details). A set of disturbed parameters $\theta'_{i,j}$ the ith hydrodynamic derivative of the jth moderately-uncertain model was introduced as:

$$\theta'_{i,j} = \Delta_{i,j}\theta_i \tag{16}$$

$0.4 < \Delta_{i,j} < 1.6$ was randomly selected for the ith hydrodynamic derivative of the jth moderately-uncertain model. Although they made prediction errors due to poorly identified parameters, predicted trajectories they made could represent the basic characteristics of the true dynamics of the targeting ship. The mass, inertia moment, added-mass coefficients, and thruster models were kept unchanged from the ground-truth model. This situation could occur if we copy and paste hydrodynamic parameters of similar ships, we have a physics-based model adjusted to the different operational conditions, and so on.

4.3.2. Highly-uncertain models

We prepared another eight physics-based models by randomly shifting the mass, inertia moment, added-mass coefficients, and propeller diameter of the thruster model up to 40%, in the same procedure as (16), in addition to the parameter shift introduced in the moderately-uncertain models. The trajectories they predicted had very different characteristics from the true trajectories since the basic parameters of the model were shifted. This situation could occur if we copy and paste parameters of very different ships or actuator models have large uncertainty.

4.3.3. No model

If no model was assigned to the physics-based model, a pure data-driven model was built in the experiment. It was trained in the same manner as the cooperative model, however, the target vector was not the residual vector $[\Delta\hat{x}^m_1, \dots, \Delta\hat{x}^m_{n_T}, \Delta\hat{y}^m_1, \dots, \Delta\hat{y}^m_{n_T}]$ but the future position vector $[\hat{x}_1, \dots, \hat{x}_{n_T}, \hat{y}_1, \dots, \hat{y}_{n_T}]$ without the help of any model-based guides.

4.4. Results

Snapshots in Fig. 7 show predictions with different physics-based models with different sub datasets at one of the example prediction time instance in the test dataset. Black dotted lines show the 30 s true trajectories, which are the same in the three subfigures. In Fig. 7(a), predicted trajectories made by pure data-driven models trained with $\mathcal{D}_{10}$, $\mathcal{D}_{30}$, and $\mathcal{D}_{100}$ are shown. Since they were not supported by any prior knowledge of ship dynamics, it is seen that they needed a large dataset to make predictions accurately. Models trained with $\mathcal{D}_{10}$ and $\mathcal{D}_{30}$ ended up making discontinuous trajectories with less similarity to the true trajectory. In Fig. 7(b), predicted trajectories made by cooperative models with one of the moderately uncertain physics-based models with $\mathcal{D}_{10}$, $\mathcal{D}_{30}$, and $\mathcal{D}_{100}$ are shown. Although physics-based models made prediction error in (b), it was rather small and captured the basic geometry of the true trajectory. It is seen that data-driven compensators compensated for such errors well only by using a small dataset $\mathcal{D}_{10}$. As the pure data-driven model with $\mathcal{D}_{10}$ failed at making an accurate prediction in (a), it implies that the physics-based model successfully supported the cooperative performance. In Fig. 7(c), predicted trajectories made by cooperative models with one of the highly-uncertain physics-based models with $\mathcal{D}_{10}$, $\mathcal{D}_{30}$, and $\mathcal{D}_{100}$ are shown. In (c), a trajectory predicted by the physics-based model notably diverges from the true trajectory. The poor performance of the physics-based model was induced by its parameters with higher uncertainty. We see it deteriorated the cooperative performance with $\mathcal{D}_{10}$ and $\mathcal{D}_{30}$ significantly while the performances in (b) were very good with the same datasets. Moreover, although the large prediction error was mitigated by having a large sub dataset $\mathcal{D}_{100}$, the cooperative performance with $\mathcal{D}_{100}$ in (b) outperforms that in (c).

An overview of results is illustrated in Fig. 8. As addressed in Section 3.4, this figure shows the relationship between the errors made by the cooperative model H as a height of bars and its (A, N_D) as a position of bars on the bottom plane, where A denotes the errors made by the physics-based model and N_D denotes the number of maneuvers in the training dataset. For instance, the height of the bar located at the small A and large N_D on the bottom plane shows the errors made by the cooperative model with a combination of such an accurate physics-based model and a large dataset for the training. Bars at $N_D = 0$ shows the original performance of the physics-based model without using any data for the training. We see that most cooperative models with highly-uncertain physics-based models made larger errors than cooperative

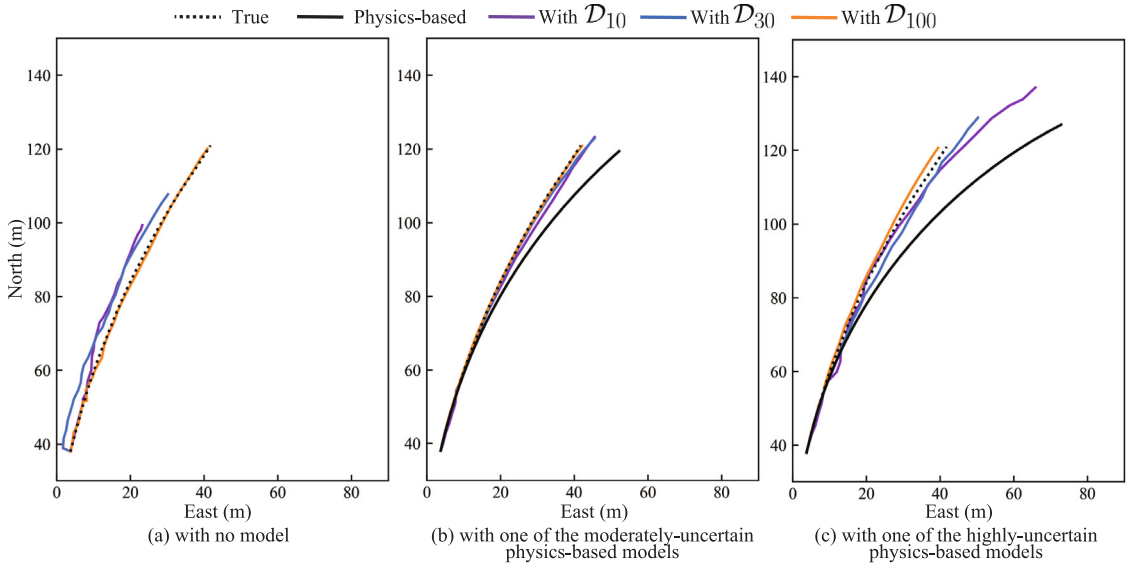

Fig. 7. Snapshots of trajectory predictions made by (a) the pure data-driven model, (b) the cooperative model with one of the moderately-uncertain physics-based models, and (c) the cooperative model with one of the highly uncertain physics-based models with sub dataset D_{10}, D_{30}, and D_{100}.

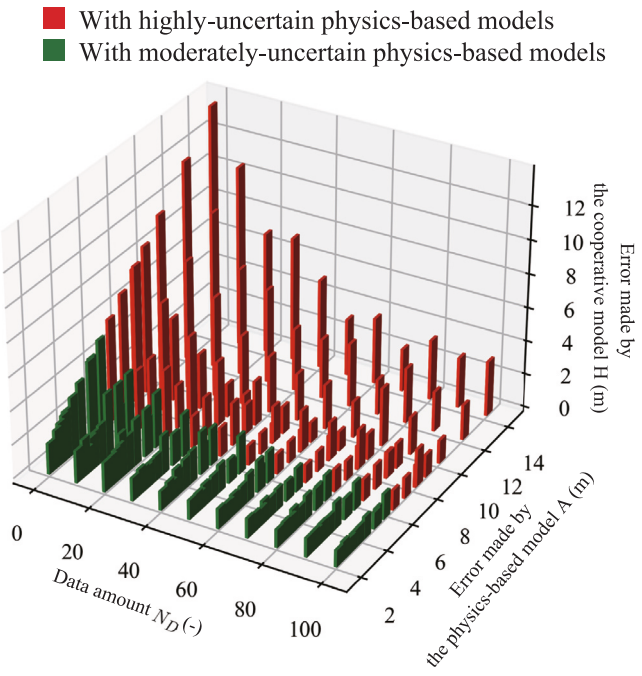

Fig. 8. The effect of the errors of the physics-based model A and the data amount N_D on the errors made by the cooperative model H in the test dataset.

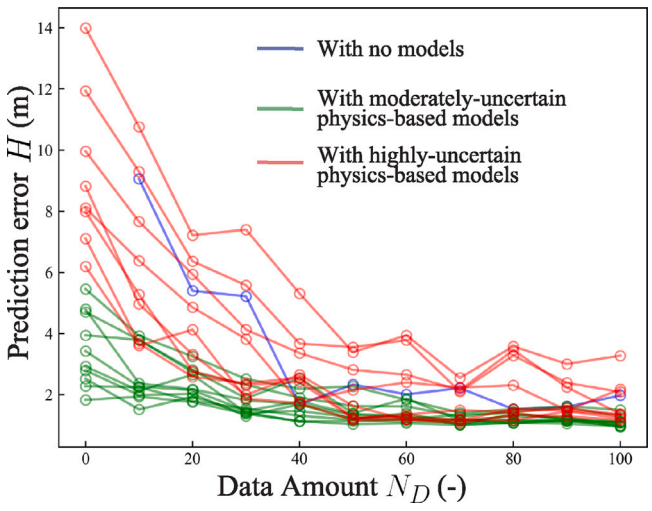

Fig. 9. A projected 2D graph of the effect of the accuracy of the physics-based model A and the data amount N_D on the errors made by the cooperative model H in the test dataset.

models with moderately-uncertain physics-based models with the same amount of data for training. In addition, a trend was seen that the higher cooperative performance was achieved with a larger dataset and a more accurate physics-based model. A good cooperative performance was achieved by either having an accurate physics-based model or having a large dataset; thereby, they are complementary to each other to some extent.

Fig. 9 is a projected graph of Fig. 8 on the $N_D - H$ plane for better visibility of absolute values of the cooperative performance. In Fig. 9, a trend is seen that the higher the accuracy of the physics-based model was, the higher the cooperative performance was, especially when the dataset was small. At the same time, cooperative models with a wide range of physics-based models, including some highly-uncertain models with relatively better performance, were found to outperform the pure data-driven model. It implies the possibility of the cooperative framework of building an accurate model with a compromised physics-based model and a small dataset. However, it does not mean any physics-based models are acceptable as a foundation of cooperative models. It is clearly seen that some cooperative models with highly-uncertain physics-based models ended up with poorer performance than the pure data-driven models. In such cases, physics-based models seem to not introduce prior knowledge of ship dynamics but introduce disturbances in the training. Thereby, the negative impact of having such physics-based models with the poor performance on the cooperative performance remained even if we had a large dataset. In particular, the performance of the cooperative model with the most inaccurate physics-based model fluctuated much depending on the sub dataset used for the training. It is seemingly caused by its high data dependency with a physics-based model introducing disturbance to the training.

Fig. 10 shows the relationship between N_D and H_{90} of models. Its trend is similar to that in Fig. 9. That means findings in Fig. 10 are

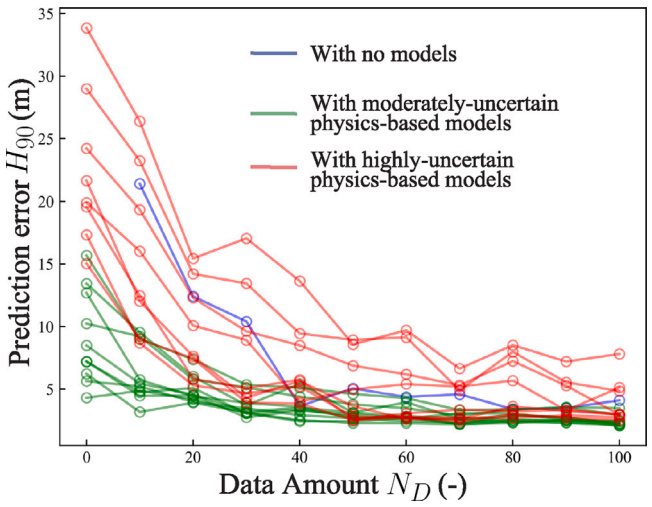

Fig. 10. A projected 2D graph of the effect of the accuracy of the physics-based model A and the data amount N_D on the 90% percentile cooperative performance H_{90} in the test dataset.

also applied to the strategy of how we reduce the occurrence of large prediction errors by using the physics-based model and data in this experiment.

5. Full-scale experiment

Hereinafter, we further explore the reasonable range of the physics-based model's accuracy on the cooperative performance in the real-life project by employing a small dataset of a full-scale experiment in the open sea. This full-scale experiment validates that we can build an accurate ship dynamic model in the practical project by combining a compromised physics-based model and a small dataset rather than relying on either of them. A snapshot of the experiment is shown in Fig. 2(b).

5.1. Overview

We made $T = 15$s trajectory predictions in the full-scale experiment since thruster commands changed drastically over the prediction horizon longer than $T = 15$s in a full-scale zigzag maneuvers. In addition, in the full-scale experiment, having too much uncertainty from unexpectable environmental disturbances in the longer prediction horizon makes a fair comparison between dynamic models challenging. Due to this limitation, errors over the prediction horizon from 15 s to 30 s have been widely used as a metrics of the accuracy of ship dynamic models in the full-scale experiment (see Skulstad et al. (2021a) and Wang et al. (2021)). The zigzag maneuver is one of the maneuvers that the International Towing Tank Conference (the ITTC) recommends as a full-scale maneuvering trials procedure. During the zigzag maneuver, ship's heading swings from side to side. The detailed definition of the zigzag maneuver can be referred to The International Towing Tank Conference (2002). In the full-scale experiment, we investigated the cooperative performances with different physics-based models and a small dataset to examine the framework building an accurate model with a compromised physics-based model and a small dataset.

5.2. Dataset

We conducted full-scale experiments in the open sea on November 21st, 2019 in Trondheim, Norway. The 33.9m-length R/V Gunnerus was employed. Under the mild weather condition, we conducted 10°/10°, 15°/15°, 20°/20°, 25°/25°, and 30°/30° zigzag maneuvers with high ($n \approx$ 145RPM) and low ($n \approx$ 125RPM) surge velocities. Each maneuver was saved in 1 Hz and cut into 85 s time series with 15 s future positions at each time step. The number of sampled maneuvers was 16. A 20°/20° zigzag maneuver with the high surge velocity was kept untouched for the test dataset. This maneuver was not included in the other maneuvers in the training–validation dataset. During the full-scale experiment, onboard sensors provided the following measurements:

- Positions: North and East positions in the NED coordinate in addition to the heading.
- Velocities: The surge, sway, and yaw velocities.
- Commands: Thruster revolution and angle of the port- and starboard-azimuth thrusters.
- Wind: The true wind direction and velocity.

During maneuvers, same commands were given to the two azimuth thrusters and the bow thruster was turned off. Except for the maneuver in the test dataset, 15 maneuvers were used for the training. The three-fold cross validation was conducted by using 15 maneuvers in the training and validation dataset. Minimum, mean, and maximum input values of the data-driven compensator in datasets are shown in Table 2.

5.3. Physics-based models

In the full-scale experiment, we employed two physics-based models; namely, accurate and inaccurate physics-based models to examine the impact of having different physics-based models on the cooperative performance with a real-life small dataset.

5.3.1. Accurate physics-based model

Before the full-scale experiment, the R/V Gunnerus was elongated from 28.9 m to 33.9 m. However, the corresponding ship dynamic model has not been fully developed. In this study, we employ a ship dynamic model of the 28.9 m R/V Gunnerus before the elongation as a physics-based model since it well captures the dynamic behavior of the elongated R/V Gunnerus as well. It is referred to as the accurate physics-based model in this experiment. It represents an optimistic assumption that an accurate physics-based model is available in the project.

5.3.2. Inaccurate physics-based model

In an inaccurate physics-based model, we shifted dominant damping coefficients in addition to removing higher-order damping coefficients in $\boldsymbol{D}(\boldsymbol{\nu}_r)$. The inaccurate physics-based model represents a pessimistic assumption that the available physics-based model performs poorly due to different reasons.

5.3.3. No model

If no physics-based model was given, as we did in the simulation experiment, a pure data-driven model was built without the help of the inaccurate physics-based model.

5.4. Results

In Fig. 11, snapshots of 15 s trajectory predictions of the maneuver in the test dataset at $t = 20$s in (a), $t = 40$s in (b), and $t = 60$s in (c) are shown. Time histories of the thruster revolution and angle of this maneuver are shown in Fig. 12. In Fig. 11(a), although the cooperative model with the accurate physics-based model made smaller errors compared to the other models, all models deviated from the true trajectory notably. In the full-scale experiment, as the dataset used for training was limited, it was seen that cooperative models did not always make a good prediction seemingly due to the lack of experience during training. On the other hand, in (b) and (c), cooperative models notably reduced prediction errors made by the corresponding physics-based models and significantly outperformed the pure data-driven model. In

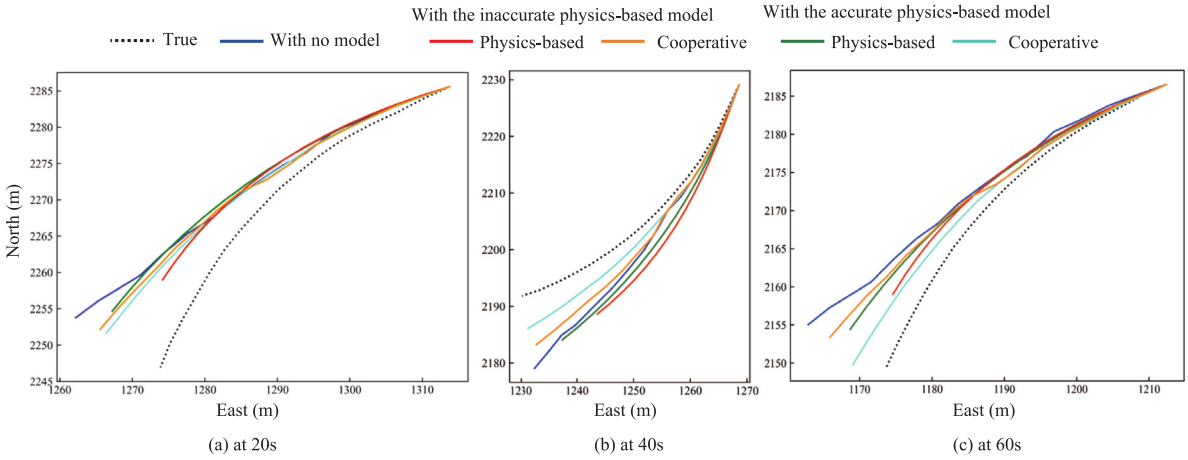

Fig. 11. Snapshots of trajectory predictions at (a) 20 s, (b) 40 s, and (c) 60 s of the maneuver in the test dataset.

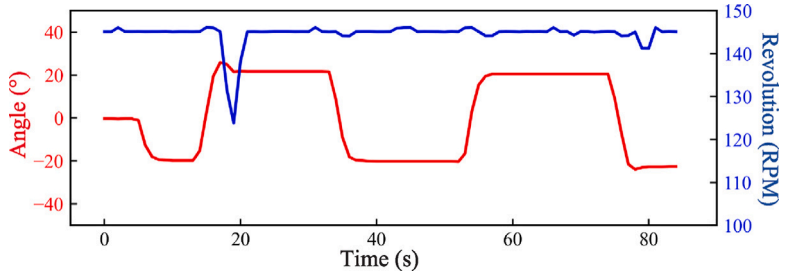

Fig. 12. A time series of thruster angle δ and revolution n in the example maneuvers in the test dataset shown in Fig. 11.

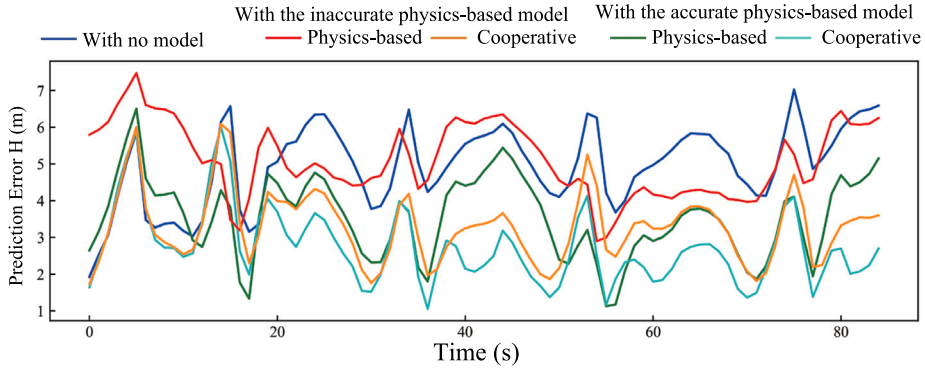

Fig. 13. Time series of the average prediction error H in the 15 s prediction horizon at each prediction instance of the maneuver in the test dataset.

(b), the cooperative model with the inaccurate physics-based model performed better than the accurate physics-based model. In (c), its performance was comparable to the accurate physics-based model. Hence, although the cooperative model with the inaccurate physics-based model did not outperform that with the accurate physics-based model, the contribution of having such a compromised physics-based model was clearly discerned. This finding corresponds to the results presented in the simulation experiment.

Time histories of the prediction error H for pure data-driven, pure physics-based, and cooperative models are shown in Fig. 13. The prediction performance fluctuated as time advanced due to having epistemic and aleatoric uncertainties. Thereby, it is seen that models did not make accurate predictions at some time steps, as shown in Fig. 11(a), however, the overall prediction performance of the cooperative model outperformed the pure data-driven and corresponding physics-based models. The performance of the cooperative model with the inaccurate physics-based model was mostly comparable with that of the accurate physics-based model.

Fig. 14 shows the average prediction error H in the 15 s prediction horizon at each prediction instance made by the pure data-driven, pure physics-based, and cooperative models. The prediction error becomes larger in the distant horizon as we have much uncertainty in the distant future. The pure data-driven model made larger prediction errors than the other models over the prediction horizon. It highlights the benefit of having the physics-based model in terms of prediction performance. Although the inaccurate physics-based model did not perform well in the experiment over the prediction horizon, its prediction error was well compensated for by using the small dataset. Its overall performance is comparable to that of the accurate physics-based model, which requires much time & effort to be developed. In addition, in Fig. 14, the discrepancy of the prediction error H made by the accurate and inaccurate physics-based model was found to be large. On the other hand, by using the small dataset, the discrepancy

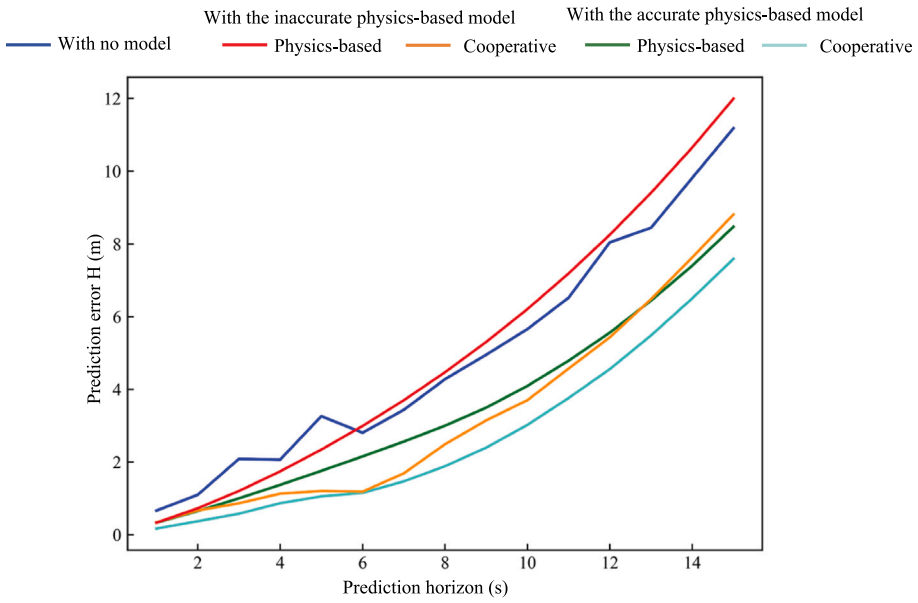

Fig. 14. The prediction error H over the prediction horizon made by pure data-driven, pure physics-based, and cooperative models.

Table 2
Maximum and Minimum input values in the training and validation dataset and the test dataset in the full-scale experiment.

		Training–validation	Test
u_0 (m/s)	min	2.4	3.5
	mean	4.1	3.8
	max	5.0	4.7
v_0 (m/s)	min	−0.6	−0.3
	mean	0.2	0.2
	max	0.8	0.7
r_0 (°/s)	min	−3.9	−3.9
	mean	−0.1	−0.1
	max	4.0	3.8
n_0 (RPM)	min	115.2	123.8
	mean	136.5	144.6
	max	145.1	146.1
δ_0 (°)	min	−29.7	−23.9
	mean	0.8	1.2
	max	29.8	25.9
$u_{rw;0}$ (m/s)	min	−3.3	−0.8
	mean	0.6	0.8
	max	7.5	2.8
$v_{rw;0}$ (m/s)	min	−5.8	−4.3
	mean	−0.8	−2.9
	max	6.0	−0.2

between corresponding cooperative performances becomes smaller. It implies the robustness of the cooperative model to the poor accuracy of the physics-based model.

6. Discussion

In this section, key findings in the simulation and full-scale experiments related to the open question "how do a physics-based model and data cooperate with each other in the cooperative model?" are summarized. In addition, this section clearly presents our suggestions for industrial applications based on the key findings. In the end, the limitations of this work and future works are discussed.

6.1. Key findings

In the simulation and full-scale experiments, **cooperative models were found to work better with a wide range of physics-based models than the pure data-driven models**. It indicates the cooperative model effectively introduced prior knowledge packaged in a wide range of physics-based models into the training in addition to acting as a stable foundation for making a prediction. The cooperative performance was notably improved by having larger datasets, thereby, the cooperative model was found to be robust to the poor accuracy of the physics-based model. On the other hand, the contribution of the physics-based model was critical when the dataset was small. If large datasets are available, the low accuracy of the physics-based model can be compensated to some extent.

In the cooperative models, **the most important perspective was to balance the physics-based model and data rather than solely relying on either of them.** In the experiments, in some cases, it was seen that the combination of a moderately-accurate physics-based model and a relatively-small dataset outperformed pure data-driven models with larger datasets or more highly-accurate physics-based models.

Although cooperative models were found to be robust to the poor accuracy of the physics-based model by being trained in a data-driven manner, **we have no such thing as a general ship model** that serves as a fundamental physics-based model of any ships. Simulation experiments in this study showed that a significantly inaccurate physics-based model may disturb the training of cooperative models instead of facilitating it with prior knowledge of ship dynamics. Its negative impact was found to remain even with a large dataset and it induced the instability of the performance due to them heavily relying on data with a disturbing foundation. Such models are not acceptable as a foundation and we should pay much attention to avoid using them.

Unlike the simulation experiment in Section 4, the full-scale experiment in Section 5 represents a real-life problem setting including the impact of real-world environmental disturbances and dataset. Thereby, the result in Section 5 validated the practical application of the framework of building an accurate model with a compromised physics-based model and a small dataset.

6.2. Suggestions for industrial applications

Findings in this study bring us to some suggestions for industrial applications. First, **we should devote more effort to developing methodologies that more easily and efficiently find highly-accurate physics-based models from the database.** It is expected for shipyards to accumulate results of numerical simulations, model/full-scale experiments, and hopefully identified parameters of the physics-based model of similar ships in their database. We could find a physics-based model of similar ships in the database in a more time- and cost-effective manner compared to the data collection; thereby, it would be a first priority in the model development. Having an accurate physics-based model makes the cooperative performance better, especially when a dataset is small. However, only a few studies (e.g., Mei et al. (2019)) have focused on the importance of this practice.

Second, **a key technology in the future would be how to build a simplified physics-based model easily**. Probably, we would not be able to find a physics-based model of similar ships from the database, especially when the project is carried out by small stakeholders without diverse experiences. In such settings, it is important to develop a physics-based model of which performance is in the acceptable range. With such a compromised model, we can improve the performance by collecting a limited amount of data as we showed in Section 5. It needs to be mentioned that we have criteria to be eligible for being a foundation model, although a wide range of physics-based models was found to be helpful. If the performance is too poor, it ruins the cooperative performance instead of helping. The co-simulation technology (e.g., Hatledal et al. (2021)) would be useful to build a simplified physics-based model only by assembling sub models. Such technology has not been fully applied to the maritime industry, however, it has great potential to provide a physics-based model with acceptable performance easily and readily.

6.3. Limitations

This study conducted experimental investigations by using different case studies in the simulation and full-scale experiments. Thereby, it is plausible to say that the findings in this study would provide basic insights for the practical applications of cooperative models. However, in future work, we must check how general our findings are in the theoretical, experimental, and practical manners involving more case studies.

This study employed a data-driven model without a physics-based model as a baseline. It does not represent a definite limitation of the data-driven approach itself. By using state-of-the-art ML architectures, pure data-driven models might perform better than we presented in this paper. However, in this field, the maritime industry does not apply such complex ML models without a physics foundation in their industrial applications due to the lack of the model's interpretability. In addition, such models would require great effort in their tuning and training. Thereby, good performance with a simple ML architecture could be seen as a practical benefit of having a physics-based model.

In the future, our research effort should be paid to developing methodologies how to balance our effort dedicated to having a physics-based model and collecting data. This study presented its basic understandings, however, it has not been revealed how to balance two methodologies in industrial practices and how it reduces the time & cost dedicated to the model development.

Cooperative models offer the development of the ship dynamic model in a timely and easy fashion for some applications including the situation awareness and onboard decision support. It can be also used as an initial-stage model of the project. However, it does not intend to substitute conventional ship dynamic models in all applications. For example, they would face a challenge in ensuring stability when they are fully implemented in the control system.

This study focused on challenges in building a ship dynamic model. Such a model is useful for a short-term prediction, estimating maneuverability, and building a simulator for training purposes. Dynamic models do not have information about the surrounding geography, traffic, and future environmental disturbances. Thereby, it should be noted that it is not suitable for making a long-term trajectory prediction such as 30 min.

7. Conclusion

In the era of ship automation in the future, precise ship dynamic models play an integral role in making the early warning of the future collision risks. In practice, it has been a great challenge how we develop such a model while minimizing the time & cost dedicated to the model development. A cooperative ship dynamic model, which employs a data-driven model for compensating for the position error made by the physics-based model, was presented in recent studies. It develops a ship dynamic model in a non-parametric manner exploiting data while having an interpretable and stable foundation of the physics-based model. Although it seems to be a promising direction to combine two approaches to overcome the time & cost challenges in industrial practices, it has been an open question "how much does the cooperative model benefit from physics knowledge and observation data?". This study conducted simulation and full-scale experiments to offer one solution through case studies and explore a safe zone of cooperative models in the physics-based model's accuracy and the data amount dimensions. In the simulation experiments, the performances of the cooperative models with different physics-based models and different datasets were examined. In addition, in the full-scale experiment, the impact of having different physics-based models on the cooperative performance with a real-life small dataset was investigated. Findings in the experiments showed that the balance of the accuracy of the physics-based model and the data amount is key to achieve a good performance of the cooperative model rather than relying on either of them. Although a wide range of physics-based models successfully facilitated the model identification, however, it disturbed the training if it was too inaccurate. In the full-scale experiment, a framework of building an accurate model with a compromised physics-based model and a small dataset was validated. Hence, for reducing the time & cost challenges in the cooperative framework, it would be pivotal to find an accurate physics-based model from database or build a simplified physics-based model with acceptable performance efficiently.

CRediT authorship contribution statement

Motoyasu Kanazawa: Conceptualization, Methodology, Investigation, Writing – original draft. **Tongtong Wang:** Conceptualization, Writing – review & editing. **Robert Skulstad:** Conceptualization, Supervision. **Guoyuan Li:** Writing – review & editing, Supervision. **Houxiang Zhang:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by a grant from the Research Council of Norway, IKTPLUSS project No. 309323 "Remote Control Center for Autonomous Ship Support" in Norway.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, Anchorage, AK, USA, pp. 2623–2631. http://dx.doi.org/10.1145/3292500.3330701.

Bergstra, J., Bardenet, R., Bengio, Y., Balazs, K., 2011. Algorithms for hyper-parameter optimization. In: Advances in Neural Information Processing Systems (NIPS 2011). pp. 1–9.

Bergstra, J., Yamis, D., Cox, D., 2013. Making a science of model search: hyper-parameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the 30th International Conference on Machine Learning. Vol. 28, Atlanta, Georgia, USA, http://dx.doi.org/10.1080/01459740.2015.1058375.

Chen, X., Ling, J., Wang, S., Yang, Y., Luo, L., Yan, Y., 2021. Ship detection from coastal surveillance videos via an ensemble canny-gaussian-morphology framework. J. Navig. 74, 1252–1266. http://dx.doi.org/10.1017/S0373463321000540.

Fonseca, A., Gaspar, H.M., 2021. Challenges when creating a cohesive digital twin ship: a data modelling perspective. 68, (2), pp. 70–83. http://dx.doi.org/10.1080/09377255.2020.1815140.

Hassani, V., Fathi, D., Ross, A., Sprenger, F., Selvik, Berg, T.E., 2015. Time domain simulation model for research vessel Gunnerus. In: Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE. Vol. 7, St.John's, Newfoundland, Canada, http://dx.doi.org/10.1115/OMAE201541786.

Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Eake, K., Euring, J.A., Gicnapp, A., Hasselmann, D.E., Kruseman, P., Meerburg, A., Mullen, P., Olbers, D.J., Richren, K., Sell, W., Walden, H., 1973. Measurements of wind-wave growth and swell decay during the joint North Sea wave project (JONSWAP). Deutsche Hydrogr. Zeitschrift 12 (Deutches Hydrographisches Institut).

Hatledal, L.I., Chu, Y., Styve, A., Zhang, H., 2021. Vico: An entity-component-system based co-simulation framework. Simul. Model. Pract. Theory 108 (September 2020), 102243. http://dx.doi.org/10.1016/j.simpat.2020.102243.

Kanazawa, M., Skulstad, R., Li, G., Hatledal, L.I., Zhang, H., 2021. A multiple-output hybrid ship trajectory predictor with consideration for future command assumption. IEEE Sens. J. 1–13. http://dx.doi.org/10.1109/JSEN.2021.3119069.

Kanazawa, M., Skulstad, R., Wang, T., Li, G., Hatledal, L.I., Zhang, H., 2022. A physics-data co-operative ship dynamic model for a docking operation. IEEE Sens. J. 22 (11), 11173–11183. http://dx.doi.org/10.1109/JSEN.2022.3171036.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3 (6), 422–440. http://dx.doi.org/10.1038/s42254-021-00314-5.

Karpatne, A., Atluri, G., Faghmous, J., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2016. Theory-guided data science: A new paradigm for scientific discovery from data. http://dx.doi.org/10.1109/TKDE.2017.2720168.

Kawan, B., Wang, H., Li, G., Chhantyal, K., 2017. Data-driven modeling of ship motion prediction based on support vector regression. In: Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58) Reykjavik, Iceland, September 25th – 27th, 2017. Vol. 138, pp. 350–354. http://dx.doi.org/10.3384/ecp17138350.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. pp. 1–15.

Luo, W., Li, X., 2017. Measures to diminish the parameter drift in the modeling of ship manoeuvring using system identification. Appl. Ocean Res. 67, 9–20. http://dx.doi.org/10.1016/j.apor.2017.06.008.

Mei, B., Sun, L., Shi, G., 2019. White-black-box hybrid model identification based on RM-RF for ship maneuvering. IEEE Access 7, 57691–57705. http://dx.doi.org/10.1109/ACCESS.2019.2914120.

Norwegian Shipowners Association, 2019. Maritime outlook report. Technical Report, p. 56, URL https://maritimpolitikk.no/en/2021.

Panigrahi, S., Nanda, A., Swarnkar, T., 2021. A survey on transfer learning. Smart Innov. Syst. Technol. 194, 781–789. http://dx.doi.org/10.1007/978-981-15-5971-6_83.

Paszke, A., Gross, S., Bradbury, J., Lin, Z., Devito, Z., Massa, F., Steiner, B., Killeen, T., Yang, E., 2019. Pytorch : An imperative style, high-performance deep learning library. (NeurIPS).

Ross, A., 2008. Nonlinear manoeuvring models for ships: A lagrangian approach. In: Department of Engineering Cybernetics, Faculty of Information, Technology, Mathematics, and Electrical Engineering. (Ph.D. thesis). p. 181.

Ross, A., Hassani, V., Selvik, Fathi, D., 2015. Identification of nonlinear manoeuvring models for marine vessels using planar motion mechanism tests. In: Proceedings of the ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering. St.John's, Newfoundland, Canada.

Schirmann, M.L., Collette, M.D., Gose, J.W., 2022. Data-driven models for vessel motion prediction and the benefits of physics-based information. Appl. Ocean Res. 120, http://dx.doi.org/10.1016/j.apor.2021.102916.

Skulstad, R., Li, G., Fossen, T.I., Vik, B., Zhang, H., 2021a. A hybrid approach to motion prediction for ship docking - integration of a neural network model into the ship dynamic model. IEEE Trans. Instrum. Meas. 70, http://dx.doi.org/10.1109/TIM.2020.3018568.

Skulstad, R., Li, G., Fossen, T.I., Wang, T., Zhang, H., 2021b. A co-operative hybrid model for ship motion prediction. Model. Identif. Control: Norwegian Res. Bull. 42 (1), 17–26. http://dx.doi.org/10.4173/mic.2021.1.2.

The International Towing Tank Conference, 2002. Ittc-recommended procedures: full scale measurements manoeuvrability full scale manoeuvring trials procedure. Technical Report.

van de Ven, P.W., Johansen, T.A., Sørensen, A.J., Flanagan, C., Toal, D., 2007. Neural network augmented identification of underwater vehicle models. Control Eng. Pract. 15 (6), 715–725. http://dx.doi.org/10.1016/j.conengprac.2005.11.004.

von Stosch, M., Oliveira, R., Peres, J., Feyo de Azevedo, S., 2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Comput. Chem. Eng. 60, 86–101. http://dx.doi.org/10.1016/j.compchemeng.2013.08.008.

Vonrueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Walczak, M., Pfrommer, J., Pick, A., Ramamurthy, R., Garcke, J., Bauckhage, C., Schuecker, J., 2021. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans. Knowl. Data Eng. 1–19. http://dx.doi.org/10.1109/TKDE.2021.3079836.

Wang, T., Li, G., Hatledal, L.I., Skulstad, R., Aesoy, V., Zhang, H., 2021. Incorporating approximate dynamics into data-driven calibrator: A representative model for ship maneuvering prediction. IEEE Trans. Ind. Inf. 1. http://dx.doi.org/10.1109/tii.2021.3088404.

Wang, Z., Zou, Z., Soares, C.G., 2019. Identification of ship manoeuvring motion based on nu-support vector machine. Ocean Eng. 183 (January), 270–281. http://dx.doi.org/10.1016/j.oceaneng.2019.04.085.

Xiao, Z., Fu, X., Zhang, L., Goh, R.S.M., 2020. Traffic Pattern Mining and Forecasting Technologies in Maritime Traffic Service Networks: A Comprehensive Survey. http://dx.doi.org/10.1109/TITS.2019.2908191.

Xiao, G., Wang, T., Chen, X., Zhou, L., 2022. Evaluation of ship pollutant emissions in the ports of los angeles and long beach. J. Mar. Sci. Eng. 10, 1206. http://dx.doi.org/10.3390/jmse10091206.

Xue, Y., Liu, Y., Ji, C., Xue, G., 2020. Hydrodynamic parameter identification for ship manoeuvring mathematical models using a Bayesian approach. Ocean Eng. 195 (November 2019), 106612. http://dx.doi.org/10.1016/j.oceaneng.2019.106612.

Zhang, X., Fu, X., Xiao, Z., Xu, H., Qin, Z., 2022. Vessel trajectory prediction in maritime transportation: current approaches and beyond. IEEE Trans. Intell. Transp. Syst. http://dx.doi.org/10.1109/TITS.2022.3192574.