Efficient Representation of New Ship Dynamics Through Reuse of Existing Maneuvering Models

Motoyasu Kanazawa, Tongtong Wang,
Ryota Wada, Houxiang Zhang

***Abstract*—Reliable ship maneuvering models are essential for simulation, control, and autonomy, but constructing ship-specific models remains costly and time-consuming due to dedicated experiments and parameter identification. Meanwhile, many validated maneuvering models already exist from previous projects, yet their use is typically limited to the ships for which they were developed. This study proposes a framework for predicting the dynamics of a new ship by blending multiple trusted maneuvering models rather than identifying a new model from scratch. A neural network computes state-dependent blending weights over a dictionary of existing ship models, allowing the target dynamics to be represented as a convex combination of validated responses. The proposed method is evaluated through simulation studies and full-scale experiments using the research vessel RV Gunnerus. Results show that prediction accuracy improves rapidly when the dictionary provides sufficient coverage of the target dynamics, while additional models yield only marginal benefits. When the target dynamics lie outside the covered response space, the method converges toward the closest available model rather than producing unconstrained extrapolations. These findings suggest that validated maneuvering models can be reused as engineering assets and that new ship dynamics can be represented efficiently through blending existing models.**

***Note to Practitioners*—Developing a trustworthy maneuvering model for a new ship often requires costly experiments, expert knowledge, and extensive system identification. As a result, constructing a dedicated model for every new vessel is not always practical. This work proposes a different perspective: instead of building a new model from scratch, a new ship can be represented as a blend of previously validated ship models. The proposed method uses a neural network to combine existing maneuvering models in a state-dependent manner, enabling prediction of new ship motions while reusing accumulated modeling assets. A key practical insight is that prediction performance depends primarily on the coverage of the model library rather than on selecting a single best model. Consequently, organizations that already possess multiple validated maneuvering models can systematically leverage these assets for future projects. The method is most effective when the target ship lies within the range of behaviors represented in the library; outside this range, predictions naturally converge toward the closest available model. Overall, the framework provides a practical and interpretable approach for reusing existing ship models and reducing modeling effort for new vessels.**

***Index Terms*—ship dynamics, ship motion prediction, maneuvering model, ship representation, model reuse**

I. Introduction

Future ships are increasingly operated as large-scale intelligent systems that integrate perception, prediction, decision support, autonomy, and model-based control [1], [2]. Within such systems, maneuvering models play a critical role because they predict how a ship responds to control inputs and directly affect guidance [3], simulator-based support and training [4], digital twins, and autonomous navigation systems. Consequently, the availability of reliable maneuvering models has become an essential requirement for modern maritime operations.

Over the past decades, substantial effort has been devoted to developing maneuvering models through captive tests, towing-tank experiments, sea trials, CFD studies, and system-identification campaigns [5], [6]. As a result, a large collection of validated maneuvering models now exists for different vessel classes and operating conditions. These models represent not only prediction tools but also repositories of accumulated hydrodynamic knowledge and engineering experience.

Despite this accumulation of knowledge, maneuvering models are generally treated as ship-specific products. When a new vessel is introduced, the modeling process is often restarted, requiring additional experiments, parameter identification, or data collection. Consequently, the engineering value of previously developed maneuvering models is only partially exploited, even when those models may contain information relevant to the new vessel.

Constructing a maneuvering model for a new ship remains a challenging task. Physics-based approaches remain the standard methodology [5], [6]. When successful, the resulting models are interpretable, trustworthy, and suitable for simulation and control design. However, obtaining such models requires significant experimental effort, expert knowledge, and time-consuming calibration procedures.

Data-driven approaches have emerged as an alternative route. Instead of identifying a detailed physical parameter set, they attempt to learn the maneuvering response directly from free-running ship data using machine-learning techniques. End-to-end neural models and other regression-based approaches have been actively explored in this direction [7]–[11]. These methods are attractive because they can reduce the burden of explicit hydrodynamic identification. However, they generally provide weaker physical interpretability, and it remains difficult to guarantee that the learned model is trustworthy outside the observed data region. For ship maneuvering, this is a serious limitation because the target application domain is safety-critical.

Hybrid approaches attempt to combine the strengths of physics-based and data-driven modeling. Representative examples include trajectory correction [12], trajectory mapping [13], and response-level correction of existing maneuvering

M.Kanazawa and R.Wada are with the department of ocean technology, policy, and environment, the University of Tokyo, 277-8561 Chiba, Japan (e-mail: kanazawa@edu.k.u-tokyo.ac.jp). T.Wang and H.Zhang are with the department of ocean operations and civil engineering, faculty of engineering, Norwegian University of Science and Technology, 6009 Aalesund, Norway.

models [14]. While these approaches utilize prior physical knowledge, the final prediction is typically generated through learned corrections or mappings whose relationship to previously validated maneuvering models is not explicitly preserved. Consequently, the contribution of individual source models becomes difficult to track, and the resulting prediction is not explicitly restricted to the response space represented by existing validated maneuvering behaviors.

This observation motivates a different research question. Instead of asking how a new maneuvering model should be constructed, can a new ship be represented directly using previously validated maneuvering models?

There are several possible ways to reuse existing maneuvering knowledge. One possibility is to learn residual corrections or mappings between existing ships and a target vessel. While such approaches provide considerable flexibility, they introduce learned components whose behavior may be difficult to interpret in terms of existing maneuvering models. At the other extreme, constructing a completely new model largely discards the information already embedded in previously developed models.

The present study explores an alternative viewpoint. Rather than correcting, adapting, or replacing existing models, we investigate whether a new ship can be represented directly through combinations of previously validated maneuvering behaviors. Such a representation offers three attractive properties. First, it enables efficient reuse of existing maneuvering-model assets without requiring complete model reconstruction. Second, the resulting prediction remains explicitly interpretable because it can be traced back to known maneuvering models. Third, the prediction remains within the response space represented by validated maneuvering behaviors, avoiding unrestricted corrections generated by purely learned models.

A key challenge in such a framework is the construction of the maneuvering-model dictionary. The objective is not to identify a single source ship that most closely resembles the target vessel. Instead, the dictionary should provide a sufficiently rich collection of maneuvering behaviors from which useful information can be extracted. From this perspective, prediction performance depends less on selecting the "best" source ship and more on whether the dictionary provides adequate coverage of the target dynamics.

At present, however, it remains unclear how such coverage should be quantified and how an effective dictionary should be constructed systematically. The present study therefore adopts a simple dictionary-construction strategy and investigates a more fundamental hypothesis: given a sufficiently diverse response space, can a neural network automatically identify and exploit informative maneuvering behaviors for representing a new ship?

There are practical reasons to believe that such a hypothesis is meaningful. In ship design practice, a new vessel is rarely considered completely unrelated to existing ships. Designers routinely evaluate new designs by comparing them with previously built vessels of similar size, type, or operational purpose. Likewise, maneuvering characteristics of a new ship often resemble those of previously studied vessels while differing only in degree rather than in fundamental behavior. This suggests that a new ship may not always require an entirely new dynamic representation, but may instead be expressible through suitable combinations of existing maneuvering behaviors.

To investigate this idea, this study proposes a dictionary-based maneuvering framework in which the dynamics of a new ship are represented as a state-dependent blend of existing validated maneuvering models. A dictionary of trusted ship models is first constructed from previously developed maneuvering models. Rather than predicting accelerations directly, a neural network computes blending weights that combine the responses of the dictionary ships through a convex combination. Consequently, the prediction remains anchored to responses generated by validated models, while the contribution of each ship remains explicitly interpretable through the blending weights.

The proposed formulation differs from conventional ensemble-learning and correction-based hybrid approaches. The purpose of blending is not to construct a stronger predictor through model averaging, nor to generate unrestricted corrections to existing models. Instead, blending serves as a mechanism for representing the dynamics of a new ship through relationships with previously validated ships. In this sense, the blending weights themselves provide an interpretable description of how the target vessel relates to the available maneuvering-model assets.

The major contributions of this article are summarized as follows:

1) We introduce a model-anchored response formulation in which the learnable component generates only non-negative normalized weights over acceleration responses produced by validated maneuvering models, rather than directly generating a new response or an unrestricted residual correction. Consequently, each prediction remains explicitly attributable to the source models and constrained to their convex response space.
2) We develop a component-wise, state-dependent blending mechanism that estimates separate blending weights for surge, sway, and yaw while using the full coupled vessel state as input. This allows the relationship between the target vessel and the dictionary ships to vary across both operating conditions and motion components without neglecting motion coupling.
3) Through controlled simulation and full-scale experiments, including comparison with a direct neural network trained on the same target-ship data, we demonstrate the effectiveness of the model-anchored representation and characterize how prediction performance depends on dictionary response-space coverage.

This perspective has an important engineering implication. Maneuvering models need not be viewed solely as ship-specific products. Instead, they can also be regarded as reusable engineering assets that contribute to the representation of future vessels. The proposed framework provides a practical mechanism for leveraging such accumulated assets while reducing the effort required to model new ships.

The remainder of the paper is organized as follows. Section II describes the related works of this study. Section III

presents the proposed response-space blending framework. Section IV describes the simulation and full-scale experiments. Finally, the results are discussed and concluding remarks are provided.

II. Related Works

A. Maneuvering-Model Construction

The development of reliable maneuvering models has long been a central topic in ship dynamics and control. Existing approaches can be broadly classified into physics-based, data-driven, and hybrid methodologies. Although these approaches differ substantially in formulation, they share a common objective: the construction or refinement of a maneuvering model for a target ship.

Physics-based approaches remain the standard methodology in maneuvering modeling. Their basic philosophy is to identify the parameters of a physically interpretable dynamic model through captive tests, towing-tank experiments, sea trials, and hydrodynamic analysis. Among these frameworks, the Maneuvering Modeling Group (MMG) model is one of the most established formulations. Yasukawa et al. organized the MMG standard method into a widely used modular framework in which hull, propeller, and rudder contributions are modeled separately while preserving clear physical interpretations of the governing coefficients [5]. Related formulations based on structured hydrodynamic decomposition have also been reviewed and extended in the broader MMG literature [6]. The MMG literature has also continued to grow through practical extensions and refinements, including work on roll-coupling and shallow-water effect, showing that this family of models remains an active and evolving engineering asset [15], [16]. Because the model structure is grounded in rigid-body mechanics and hydrodynamic principles, the resulting models are suitable not only for prediction but also for analysis, simulation, and control design. However, obtaining such models typically requires substantial experimental effort, expert knowledge, and parameter-identification procedures.

To reduce these requirements, several studies have investigated identification of physics-based maneuvering models directly from operational data. Alexandersson et al. employed Kalman filtering for hydrodynamic parameter estimation under environmental disturbances [17], while Mitsuyuki et al. introduced a Bayesian MMG identification framework with uncertainty quantification [18]. More recently, An et al. employed physics-informed neural networks (PINNs) for hydrodynamic-derivative estimation by embedding maneuvering equations directly into the learning process [19]. Their framework improves data efficiency and robustness while preserving consistency with physical laws. Nevertheless, the objective remains the estimation of unknown parameters within a predefined maneuvering model for a single target vessel. Although these approaches improve practicality, they still aim to construct a dedicated model for the target ship.

Data-driven approaches take a different route by learning maneuvering behavior directly from measured input–output data without explicitly identifying hydrodynamic coefficients. Early studies explored neural-network-based identification of nonlinear ship dynamics [7], [8]. More recent work has employed recurrent neural networks, long short-term memory (LSTM) architectures, Gaussian processes, and other machine-learning techniques [9]–[11]. These approaches can reduce modeling effort and exploit the increasing availability of operational data. However, the resulting models often provide limited physical interpretability, and their behavior outside the observed operating domain may be difficult to justify.

Hybrid approaches seek to combine the strengths of physics-based and data-driven modeling. Representative examples include residual-correction approaches, trajectory-mapping approaches, and response-level correction methods. Skulstad et al. integrated a neural-network compensator into a ship dynamic model for docking-motion prediction [12]. Wang et al. introduced a data-driven mapping on top of trajectories generated by a physics-based predictor [13]. Mei et al. used machine-learning techniques to correct the maneuvering response of a reference model from another ship [14]. More recently, Zhou et al. proposed a unified physics–data-driven framework that combines maneuvering equations with neural networks by modeling unexplained dynamics as residual corrections [20]. Tian et al. developed a Hybrid Modeling and Online Update (HMOU) framework that combines reinforcement learning and recurrent neural networks to adapt maneuvering predictions online under environmental disturbances and sensor uncertainty [21]. Hatayama et al. proposed a hybrid MMG–FNN model in which a neural network augments the modular MMG structure while incorporating output constraints and out-of-distribution detection [22]. Related cooperative modeling frameworks have also been explored in recent years [23]. Although these methods differ in implementation, they share a common philosophy: improving prediction accuracy through model refinement, residual compensation, or online adaptation of a target-ship model.

B. Knowledge Transfer and Reuse of Existing Ship Models

The increasing availability of validated maneuvering models has motivated growing interest in transferring knowledge from previously studied ships to new vessels. This direction is attractive because constructing a new maneuvering model often requires substantial resources, whereas many organizations already possess collections of validated models accumulated through previous projects.

Several studies have explored this idea implicitly through hybrid modeling frameworks. For example, Mei et al. employed a maneuvering model from another ship as a reference model and learned corrections using machine-learning techniques [14]. Similarly, Wang et al. leveraged approximate physical predictors as prior knowledge for trajectory prediction [13]. More broadly, transfer-learning and model-adaptation techniques attempt to exploit similarities between previously observed systems and target systems in order to reduce modeling effort.

Recent studies have investigated model adaptation and knowledge reuse more explicitly. Kanatas et al. proposed an anchor-based adaptation framework in which a previously validated maneuvering model is retained as a baseline while

residual corrections are learned to account for environmental changes [24]. Such an approach preserves existing validated model structures and reduces retraining requirements. However, the prior model still functions primarily as an anchor for adaptation, and the final objective remains the construction of a vessel-specific predictor.

These studies establish that the reuse of previously developed ship models is itself an existing research direction. Therefore, model reuse alone is not the novelty claimed in the present study. The key distinction lies in how the learned component contributes to the final prediction. In existing correction- and adaptation-based approaches, a previous model is used as a reference, prior predictor, or anchor, while a learned residual or mapping directly modifies the predicted response.

In the proposed framework, multiple validated maneuvering models directly generate candidate acceleration responses, and the learned component produces only nonnegative normalized weights for combining these responses. No learned residual acceleration or unrestricted correction is added. Consequently, every predicted response is explicitly attributable to the dictionary models and remains within their convex response space. The proposed method therefore treats existing models not merely as starting points for adaptation, but as the exclusive response-generating elements of the target-vessel representation.

C. Model Blending and Representation

The concept of combining multiple models has been extensively studied across engineering and scientific disciplines. Ensemble learning, stacking, and model averaging are widely used to improve predictive performance and robustness when no single model is consistently superior. Recent studies in automation science have employed ensemble-based approaches to improve generalization and uncertainty quantification [25], while stacking methods introduce meta-models that learn how to combine multiple predictors [26].

Outside machine learning, model blending also appears in numerical weather prediction, hydrology, and control engineering. Multi-model weather ensembles improve forecast reliability by aggregating predictions from multiple physical models [27]. Similar approaches have been adopted in hydrology to address structural uncertainty in rainfall–runoff relationships [28]. In control engineering, gain scheduling and LPV frameworks combine locally valid models to represent nonlinear systems over broader operating regions [29]. Blending strategies have also been explored in hybrid physics–machine-learning frameworks that combine physical simulations and data-driven models [30], [31].

A conceptually related direction is the use of multiple-model frameworks. Wang et al. proposed an interactive multiple-model (IMM) framework for autonomous docking trajectory prediction, where physics-based and data-driven predictors are combined through adaptive Bayesian weighting [32]. The framework improves prediction accuracy and interpretability by dynamically selecting the most reliable predictor under changing operational conditions. However, the constituent models are treated as competing predictors, and the objective remains prediction rather than vessel representation.

Most model-combination methods treat individual models as interchangeable predictors and seek improved predictive performance through averaging, weighting, aggregation, adaptation, or selection. In such frameworks, blending serves primarily as a mechanism for enhancing prediction accuracy.

The proposed framework shares mathematical similarities with ensemble-learning and model-combination approaches in that multiple models are combined through state-dependent weights. However, the objective is fundamentally different. The purpose of blending is not to construct a stronger predictor through model averaging, nor to generate corrections to an existing model. Instead, blending serves as a representation mechanism through which a new ship is expressed in terms of previously validated maneuvering behaviors.

III. Methodology

In this section, we introduce a dictionary-based framework for representing the maneuvering dynamics of a new ship through combinations of previously validated ship models. Rather than constructing a dedicated model for the target ship, the proposed method expresses the target dynamics as a state-dependent blend of existing maneuvering behaviors.

A. Overview

1) Problem Formulation: The objective of this study is not to construct a new maneuvering model for a target ship, but to represent the target ship using a collection of previously validated ship models.

We assume that an organization possesses multiple trusted maneuvering models developed through previous projects. These models constitute a library of accumulated engineering knowledge. The central question addressed in this study is whether a new vessel can be represented through suitable combinations of such existing models. A schematic overview of the proposed approach is drawn in Fig. 1. We assume the availability of the following three assets:

- **Dictionary ships:** A set of K reliable physics-based maneuvering models constructed in previous projects of the model builder. Each model contains trusted hydrodynamic coefficients.
- **Target ship data:** Time-series measurements obtained from onboard sensors, including velocity $\nu = [u, v, r]^T$ and actuator command feedback.
- **Actuator model:** A mathematical model of the target ship's actuators, which converts prescribed actuator commands into forces and moments. In this study, the actuator commands are predefined according to standard maneuvering tests and are therefore available throughout the prediction horizon. The actuator model describes the actuator dynamics and is simulated together with the vessel dynamics during recursive prediction. Consistent with conventional maneuvering simulation practice, state-dependent actuator responses are evaluated using the current simulated vessel state. No future measured vessel states or state-feedback control actions are required.

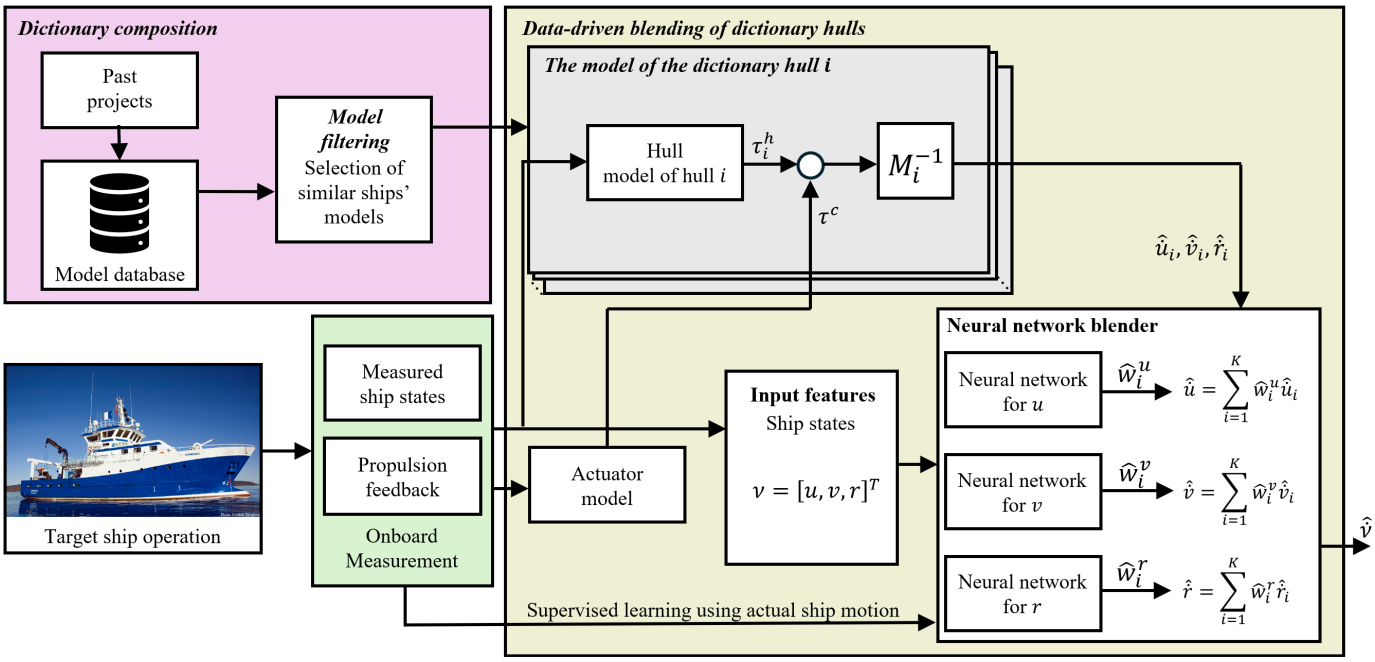


Fig. 1: A schematic overview of the proposed method. Given the current ship state and actuator inputs, multiple dictionary ship models generate candidate acceleration responses. Neural networks then computes state-dependent blending weights, which are used to combine these responses into a final prediction through a convex combination.

It is important to note that the actuator configuration of the target ship may differ from those of the dictionary ships. Therefore, actuator inputs are first converted into generalized forces and moments using the target ship's actuator model, ensuring compatibility with all dictionary hull models.

2) Dictionary Composition: The purpose of dictionary construction is not to identify a single ship that most closely resembles the target vessel. Instead, the objective is to provide a collection of maneuvering behaviors that collectively span the expected dynamic range of the target ship.

From this viewpoint, the quality of a dictionary is determined primarily by its coverage of relevant maneuvering behaviors rather than by the similarity of any individual ship. Ideally, the dictionary should contain sufficient behavioral diversity to represent the range of dynamics that may be encountered in the target vessel.

This perspective differs from conventional source-model selection approaches, where prediction performance is assumed to depend strongly on identifying the single most appropriate source ship. In the proposed framework, the neural network determines which maneuvering behaviors are informative for the target vessel under different operating conditions. The role of the dictionary is therefore not to provide the best source ship, but to provide a sufficiently rich set of maneuvering behaviors from which useful information can be extracted in a data-driven manner.

At the same time, effective coverage must be achieved within a physically relevant region of the maneuvering space. Ship models with substantially different principal characteristics may exhibit fundamentally different dynamic behaviors and may therefore contribute little to the representation of the target vessel. Including such models may increase the nominal size of the dictionary without meaningfully expanding the relevant response space.

Accordingly, the dictionary is not constructed arbitrarily but is formed through a filtering process based on principal particulars. In this study, we restrict the dictionary to ships with comparable scales by selecting models whose length between perpendiculars (L_{pp}) lies within a specified range. This filtering step is intended to exclude vessels with fundamentally different maneuvering characteristics while retaining a diverse set of transferable maneuvering behaviors.

This viewpoint is particularly relevant in practical shipbuilding environments, where validated maneuvering models from previous projects are often available for a wide variety of vessel types. Rather than treating these models solely as ship-specific products, the proposed framework views them as reusable maneuvering-model assets that collectively define a response space for representing future vessels.

3) Dictionary Models: Given the current ship state ν and actuator commands, each dictionary ship model is used to compute a candidate acceleration. Specifically, the i-th dictionary model computes hull forces including hydrodynamic forces and moments $\tau_i^h(\nu)$ based on its hydrodynamic coefficients. The actuator model provides the corresponding actuator-induced forces and moments τ^c. The total force and moment applied to the i-th model are then given by

$$\tau_i = \tau_i^h + \tau^c. \tag{1}$$

Using the mass matrix M_i of the i-th dictionary ship, the resulting acceleration is estimated as

$$\hat{\dot{\nu}}_i = M_i^{-1}\tau_i, \tag{2}$$

where $\hat{\nu}_i = [\hat{u}_i, \hat{v}_i, \hat{r}_i]^T$, $\hat{u}_i$ is the estimated acceleration of the ship using the i-th dictionary ship in the surge direction, $\hat{v}_i$ is that for the sway direction, and $\hat{r}_i$ is that for the yaw direction. This represents the hypothetical acceleration that would occur if the same forces and moments were applied to the i-th dictionary ship. Each dictionary model provides a candidate maneuvering response corresponding to a previously validated ship.

Individually, these responses do not necessarily represent the target vessel. Collectively, however, they define a response space composed of maneuvering behaviors observed in existing ships. The target dynamics are subsequently represented as a combination of these responses.

For making prediction, the future actuator inputs are assumed to be known and are provided from the measured rudder-angle and propeller-command records. Consequently, the prediction problem considered in this study corresponds to open-loop trajectory prediction rather than closed-loop simulation with an onboard controller.

4) State-Dependent Blending via Neural Networks: To represent the target ship within the dictionary response space, we introduce a state-dependent blending mechanism. The role of the neural network is not to generate a new maneuvering response. Instead, it determines how existing dictionary responses should be combined under the current operating condition. For each motion component (u, v, and r), a neural network maps the current state ν to a set of blending weights. For example, in the surge direction, the network outputs

$$\hat{\mathbf{w}}^u = \left[\hat{w}_1^u, \ldots, \hat{w}_K^u\right], \tag{3}$$

which satisfies the convex constraints

$$\hat{w}_i^u \geq 0, \quad \sum_{i=1}^{K} \hat{w}_i^u = 1. \tag{4}$$

The blended acceleration is then computed as

$$\hat{\dot{u}} = \sum_{i=1}^{K} \hat{w}_i^u \hat{\dot{u}}_i. \tag{5}$$

The same formulation is applied to the sway and yaw directions:

$$\hat{\dot{v}} = \sum_{i=1}^{K} \hat{w}_i^v \hat{\dot{v}}_i, \quad \hat{\dot{r}} = \sum_{i=1}^{K} \hat{w}_i^r \hat{\dot{r}}_i. \tag{6}$$

The final estimated acceleration vector is $\hat{\dot{\nu}} = [\hat{\dot{u}}, \hat{\dot{v}}, \hat{\dot{r}}]^T$. The convex constraint ensures that every prediction can be interpreted as a weighted combination of previously validated ship responses. Consequently, the blending weights provide an explicit description of how the target vessel relates to the ships contained in the dictionary. The estimated acceleration $\hat{\dot{\nu}}$ can be integrated over time to predict the future motion of the ship. Given an initial state, the trajectory is obtained by numerical integration over a prediction horizon.

The blending weights are estimated independently for surge, sway, and yaw motions. This design choice is intentional. The objective of the proposed framework is not to identify a single latent ship that explains all motion components through a common weight vector. Instead, the goal is to represent the target dynamics within the response space spanned by the dictionary models. In practice, similarity between ships is often direction- dependent. A target vessel may exhibit surge behavior similar to one dictionary ship while its sway and yaw responses are better represented by others. Allowing component-specific blending weights provides additional flexibility to capture such relationships. Importantly, this does not imply that the coupling between surge, sway, and yaw motions is ignored. Each blending network receives the full ship state vector $\nu = [u, v, r]^T$ as input, allowing the estimated weights to depend on the coupled motion state.

B. Physics-Based Model of Dictionary ships

In this section, we describe the physics-based model employed for the dictionary ships. The model is derived based on Newton's second law and represents the ship dynamics as a balance between inertial forces and external forces and moments.

In this study, we adopt the MMG model [5], which is one of the most widely used frameworks for describing ship maneuvering dynamics. The MMG model is a modular modeling approach that decomposes the total hydrodynamic forces into contributions from the hull and actuators, enabling flexible and interpretable modeling of ship motion.

The three-degrees-of-freedom (3-DOF) dynamic model for each dictionary ship is given by

$$M\dot{\nu}+C_{RB}(\nu)\nu+C_A(\nu_r)\nu_r+D(\nu_r) = \tau_c+\tau_{wind}+\tau_{wave} \tag{7}$$

where $\nu = [u, v, r]^T$ denotes the body-fixed velocity vector consisting of surge, sway, and yaw rate, and ν_r represents the velocity relative to the surrounding current. Here, $M \in \mathbb{R}^{3\times3}$ is the mass matrix including added mass effects. The matrices $C_{RB}(\nu)$ and $C_A(\nu_r)$ denote the Coriolis and centripetal matrices associated with rigid-body motion and added mass, respectively. The term $D(\nu_r)$ represents the hydrodynamic damping, which includes both linear and nonlinear components. The control input $\tau_c \in \mathbb{R}^3$ represents the forces and moments generated by the propulsion and steering system, such as propellers, rudders, or thrusters. Hydrodynamic terms in (7) are typically modeled using a set of hydrodynamic derivatives, which are identified through captive model tests or other experimental procedures. In this study, we assume that such hydrodynamic coefficients have been reliably identified for each dictionary ship in prior projects. This assumption is reasonable in practical settings such as shipyards, where multiple ships of similar classes are designed and tested. The terms τ_{wind} and τ_{wave} represent external disturbances due to wind and waves, respectively.

In this study, we introduce the following simplifying assumptions for trajectory prediction. First, the effect of ocean currents is neglected, and thus we assume $\nu_r = \nu$. Second, environmental disturbances due to wind and waves are not considered, i.e.,

$$\tau_{wind} = 0, \quad \tau_{wave} = 0. \tag{8}$$

These assumptions are introduced to isolate the central question investigated in this study: whether the dynamics

of a target vessel can be represented through combinations of previously validated maneuvering models. If strong environmental disturbances are included, prediction errors may arise from both environmental effects and limitations of the representation framework, making it difficult to evaluate the proposed concept itself.

It should also be noted that maneuvering models are frequently applied in relatively sheltered environments, such as harbor maneuvering, docking operations, pilot support, and low-speed ship handling, where environmental disturbances are often moderate compared with offshore operations. In such situations, accurate prediction of the vessel's intrinsic maneuvering behavior remains a primary requirement.

The full-scale experiments were therefore conducted under relatively calm open-water conditions, where environmental disturbances were limited but not completely absent. Consequently, the experimental results should be interpreted as a validation of the proposed ship-representation framework under conditions where maneuvering dynamics remain the dominant factor in the prediction problem. Extending the framework to explicitly incorporate wind, waves, and current effects remains an important direction for future research.

Under these assumptions, each dictionary ship provides a mapping from the applied control forces and moments to the resulting acceleration. As described in the overview, by applying the actuator-generated forces of the target ship to each dictionary model, we can obtain candidate acceleration estimates for the target ship.

C. State-Dependent Ship Representation

To obtain a compact representation of the target ship within the dictionary response space, we employ neural networks as state-dependent weighting functions. Among many data-driven models, we select neural networks since it is expected to serve as a strong approximator of state-dependent function. The blending is performed independently for each motion component, namely surge, sway, and yaw. For each direction, a gating network maps the current ship state $\nu = [u, v, r]^T$ to a set of blending weights over the K dictionary models. Each gating network is implemented as a fully connected feedforward neural network with two hidden layers, each consisting of 64 hidden units, and hyperbolic tangent (Tanh) activation functions. The output layer produces K values, which are normalized using a softmax function so that the weights satisfy (4). This ensures that the blended prediction remains a convex combination of the dictionary models, preserving physical consistency. Prior to being fed into the network, the input state is z-normalized using the mean and standard deviation computed from the training dataset. The blended acceleration is computed component-wise as (5) and (6).

a) Training Data and Preprocessing: The dataset consists of time-series measurements obtained from free-running trials of the target ship. In this study, zigzag and/or turning maneuver data are used for training, as they sufficiently excite the ship dynamics in all three degrees of freedom. The dataset is divided into training and test sets. The training set is used for optimizing the neural network parameters, while the test set is used exclusively for evaluation and is not used during training. Each data sample consists of the ship state ν, actuator commands, and the corresponding ground-truth acceleration. The ground-truth acceleration is calculated by using the difference between the velocity at the subsequent time step and that at the current time step. A subset of the training dataset is held out as a validation dataset, and early stopping is applied by monitoring the validation loss and terminating training when no sufficient improvement is observed for a predefined number of epochs. The model parameters corresponding to the best validation loss observed up to that point are retained.

b) Loss Function and Optimization: Neural networks are trained in a supervised manner using measured acceleration data. For each motion component, the loss is defined as a normalized mean squared error:

$$\mathcal{L}_d = \mathbb{E}\left[\left(\frac{\hat{\dot{d}} - \dot{d}}{s_d}\right)^2\right], \quad d \in \{u, v, r\}, \tag{9}$$

where s_d denotes the standard deviation of the target acceleration computed from the training dataset. In practice, this expectation is approximated by the empirical average in the training dataset. Neural networks are optimized using mini-batch stochastic gradient descent. The three gating networks are trained independently using separate optimizers. For each direction, only the corresponding loss component is back-propagated through the associated network. We employ the AdamW optimizer [33] with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-5}. Training is performed over multiple epochs.

D. Interpretation of Blending Weights

Unlike conventional ensemble learning, the blending weights are not introduced solely to improve predictive performance. Instead, they provide a representation of how the target ship relates to previously validated ships contained in the dictionary. A dominant weight indicates that the target vessel behaves similarly to a particular dictionary ship under the corresponding operating condition. Distributed weights indicate that the target behavior lies between multiple existing ships. Therefore, the blending weights may be interpreted not only as prediction parameters but also as a compact description of the relationship between the target vessel and accumulated maneuvering-model assets.

IV. Experiments

This section evaluates the proposed method in two stages. First, a simulation study is conducted to provide controlled validation and to systematically examine the effects of dictionary composition and size. Second, a full-scale experiment is performed to assess the applicability of the framework to measured ship-motion data. The two-stage design is intentional. Although full-scale experiments are essential for practical validation, they inevitably include environmental disturbances and sensor noise, which make it difficult to isolate the influence of dictionary composition. In contrast, the simulation environment allows for a clean and controlled analysis of

how dictionary size and composition affect the response-space coverage and prediction performance. The subsequent full-scale experiment complements this analysis by verifying that the observed properties remain relevant under real-world conditions.

A. Experimental Baselines

The experimental evaluation includes two complementary types of baselines. First, the proposed framework is compared with the individual dictionary ship models. This comparison evaluates whether state-dependent blending provides a more effective representation of the target vessel than selecting any single existing maneuvering model.

Second, a small direct neural-network model is introduced as a data-driven baseline. The purpose of this comparison is to determine whether the observed prediction improvement can be obtained simply by learning the target dynamics directly from the same target-ship data, without using the dictionary models.

The direct neural network and the proposed framework differ not only in architecture but also in the form of representation they provide. The direct network generates acceleration responses without explicit attribution to existing maneuvering models. In contrast, the proposed network generates only convex blending weights over acceleration responses produced by validated dictionary models. The comparison is therefore used to evaluate predictive accuracy, while the difference in model-level traceability and response-space constraints is discussed separately.

1) Direct Neural-Network Baseline: The direct neural-network baseline receives the current vessel state and four actuator-related control inputs,

$$\boldsymbol{x} = \begin{bmatrix} u & v & r & c_1 & c_2 & c_3 & c_4 \end{bmatrix}^{\mathrm{T}}, \tag{10}$$

and directly predicts the acceleration vector,

$$\hat{\boldsymbol{\nu}} = \begin{bmatrix} \hat{\dot{u}} & \hat{\dot{v}} & \hat{\dot{r}} \end{bmatrix}^{\mathrm{T}}. \tag{11}$$

The network consists of a fully connected input layer, one hidden layer with eight units and a hyperbolic tangent activation function, and a three-dimensional output layer. It was trained using the same target-ship training and test datasets as the proposed method. The loss function was the normalized mean squared error of the three acceleration components. Training was performed for 60 epochs using the AdamW optimizer with a learning rate of 8×10^{-4}, a weight decay of 1×10^{-4}, and a batch size of 128.

B. Simulation experiment

1) Dictionary composition: In the simulation experiments, MMG models of multiple ships are used to generate target data and to construct dictionary ship models. First, an original dataset consisting of 60 synthetically generated ship models is prepared based on the reference hydrodynamic values reported in [5] and [6]. These models are generated by introducing both small perturbations and larger systematic variations to the reference parameters. From this original dataset, a subset of ships is selected through a filtering process based on the principal particulars. Specifically, ships with L_{pp} in the range of 10–35 m are selected, resulting in a set of 13 ships. These ships are used either to generate target data for training or to serve as dictionary ship models in the simulation experiments. Fig. 2 shows the responses of the selected ships to a set of test commands used in the simulation experiments. As illustrated, the dynamics of these ships exhibit a certain spread, forming both overlapping and distinct patterns in their maneuvering behavior.

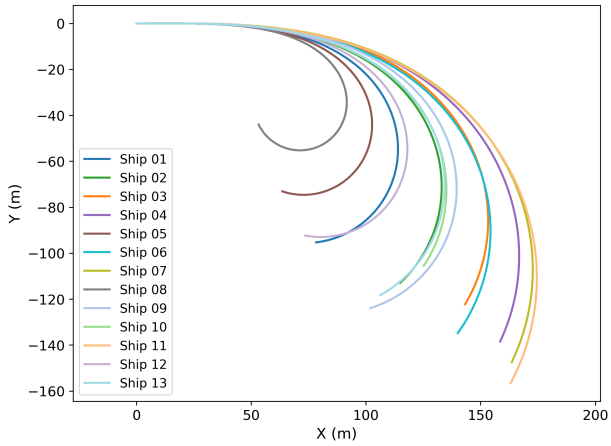


Fig. 2: A comparison of an example turning operation conducted by 13 ship models.

2) Experimental setting: The objective of the simulation experiments is to provide insights into key research questions associated with the proposed method, particularly regarding how the coverage of the response space by the dictionary ships affects prediction performance. Here, the notion of "coverage" does not refer to strict inclusion in a mathematical sense, but rather to the extent to which the target ship dynamics can be approximated by combinations of the response patterns available in the dictionary over the relevant operating conditions.

To this end, three representative test cases are considered:

- **Inside-of-Space Case**: In this case, the target ship dynamics are well covered by the response patterns of the dictionary ships. Specifically, Ship 1, Ship 2, and Ship 4 are used as dictionary ships, and Ship 3 is used as the prediction target. Although the target dynamics are not strictly contained within those of the dictionary ships, their responses can be effectively approximated by combining the available models over the considered operating range.
- **Outside-of-Space Case**: In this case, the target ship exhibits dynamics that are not sufficiently represented by the available dictionary ships. Specifically, Ship 1 and Ship 2 are used as dictionary ships, and Ship 3 is used as the prediction target. Under the same actuation inputs, Ship 3 tends to exhibit slower rotational motion compared to the dictionary ships, representing a dynamic characteristic that is difficult to reproduce through their combination over the considered operating range.
- **Dictionary-Growth Case**: This case examines how pre-

diction performance changes as the number of dictionary ships increases, including situations where ships with highly different dynamics are added. In this setup, Ship 3 is used as the prediction target, and the dictionary is progressively expanded to analyze how additional models affect response-space coverage and prediction accuracy.

In this experiment, both the training dataset and the evaluation dataset are generated using the model of the target ship to be predicted. Each operation has a duration of 300 s with a time step of 0.5 s. The training dataset consists of 16 turning operations, while the test dataset includes 7 turning operations.

The turning operations are standardized maneuvering tests defined by the International Towing Tank Conference (ITTC) for evaluating maneuvering models. For each operation, representative thruster angles and RPM values are randomly assigned, while the execution procedure follows the standard ITTC guidelines [34]. In generating the dataset, it is assumed that the ship is equipped with two azimuth thrusters, which operate under identical command inputs. To address the research questions, the evaluation focuses on trajectory prediction performance, which is of primary importance from a practical application standpoint. In the simulation experiments, the deviation of the predicted trajectories from the ground truth in the turning operations is assessed using four performance indices defined by the ITTC, namely Steady Turning Radius, Maximum Advance, Maximum Transfer, and Speed Loss in Steady Turn.

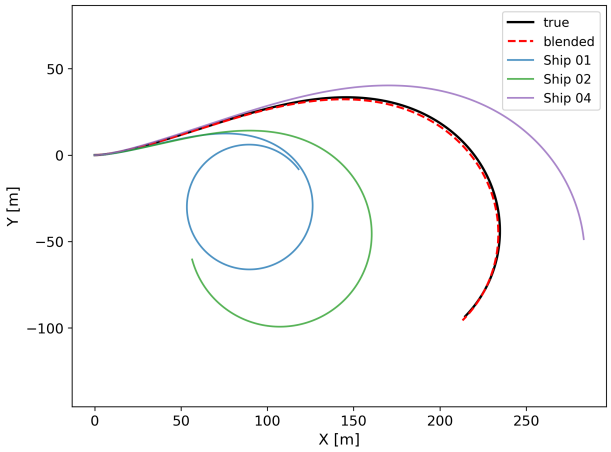


Fig. 3: Trajectory predictions in the inside-of-space case.

3) Simulation Results:

TABLE I: Prediction errors in turning performance metrics. Values represent mean absolute error (MAE) with respect to the ground truth generated by the Ship 3 model. STR: steady turning radius, MT: maximum transfer, MA: maximum advance, SLS: speed loss in steady turn.

	Mean Absolute Error (MAE)			
Model	STR [m]	MT [m]	MA [m]	SLS [m/s]
blended	0.114	0.243	0.345	0.000
Ship 01	33.102	66.700	62.424	0.116
Ship 02	17.282	33.662	38.101	0.081
Ship 04	23.232	11.907	35.561	0.021

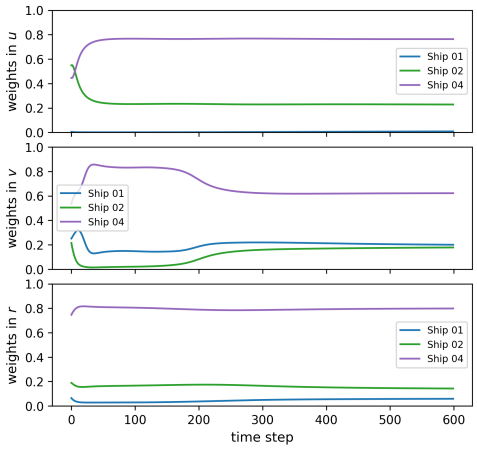


Fig. 4: Time histories of the blending weights in the inside-of-space case.

a) Inside-of-space case: Fig. 3 shows the turning trajectory for one representative case in the test dataset, along with the corresponding predictions. In the inside-of-space case, Ship 1, Ship 2, and Ship 4 are used as dictionary ships. However, each of these individual models predicts trajectories that deviate from the true trajectory of Ship 3, which is shown as the black solid line. In contrast, the blended prediction closely follows the true trajectory, demonstrating that high prediction accuracy can be achieved through the blending of dictionary ships. Fig. 4 presents the time evolution of the blending weights for this operation. In all directions (u, v, and r), Ship 4 has the dominant contribution, while the other dictionary ships also contribute effectively. Furthermore, as intended in the design of the proposed method, the weights are computed in a state-dependent manner. Specifically, the contribution of Ship 4 is relatively small in the early stage of the turning motion and increases in the later stage. These results indicate that the proposed data-driven blending approach can be interpreted as a weighted superposition of dictionary ships, thereby maintaining a high level of interpretability. Tab. I shows the average performance over the seven test operations in terms of the four ITTC evaluation metrics. It can be observed that the proposed method outperforms all individual dictionary ships across all evaluation criteria.

b) Outside-of-space case: Figure 5 shows the turning trajectory for one representative case in the test dataset, along with the corresponding predictions. In this case, Ship 1 and Ship 2 were used as dictionary ships, while Ship 3 was used as the prediction target. Compared with Ship 1 and Ship 2, the target Ship 3 exhibits slower forward motion and more gradual turning under the same actuation inputs. Such characteristics cannot be adequately represented by a convex combination of Ship 1 and Ship 2 alone. This observation is consistent with the inside-of-space case, where Ship 4 played a significant role in reconstructing the dynamics of Ship 3, indicating that additional response patterns are required to span the target dynamics. As a result, the blended prediction tends to remain close to the dictionary ship whose dynamics are most similar to the target, namely Ship 2. Rather than attempting to

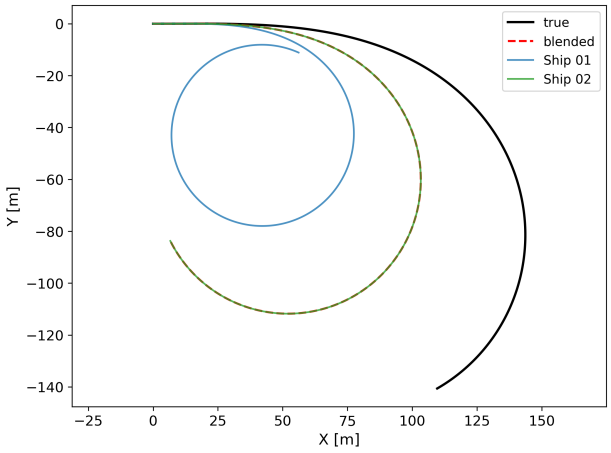


Fig. 5: Trajectory predictions in the outside-of-space case.

extrapolate beyond the available response space, the proposed method effectively collapses to the nearest feasible representation within the dictionary.

Importantly, this behavior should be interpreted as a desirable property of the framework. By restricting the prediction to the response space defined by trusted models, the method avoids generating physically implausible or unverified dynamics. In this sense, the fallback toward the nearest dictionary ship provides a controlled and reliable prediction, even when the target dynamics lie outside the span of the available models.

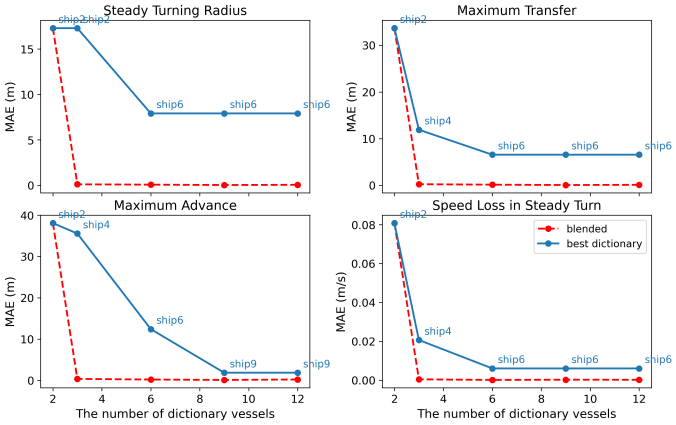


Fig. 6: The relationship between the number of the dictionary ships and the four metrics of Mean Absolute Error (MAE).

c) Dictionary-growth case: Figure 6 shows the variation of the four evaluation metrics as the number of dictionary ships increases. In all experiments, Ship 3 is used as the target ship, and the dictionary ships are selected from the remaining ships. The blue solid line represents the performance of the best individual dictionary ship in each experiment, with the corresponding ship number indicated alongside the plot. The red line represents the performance of the proposed blended method. A consistent trend can be observed across all four metrics. When the number of dictionary ships is two, this corresponds to the outside-of-space case. In this situation, the performance of the blended method is nearly identical to that of the best dictionary ship. However, by adding just one more dictionary ship, this limitation is effectively resolved. While the performance improvement of the best individual dictionary ship remains limited, the blended method shows a significant improvement in prediction accuracy. Furthermore, increasing the number of dictionary ships beyond three does not lead to noticeable performance gains in this experiment. This suggests that, in this experiment, three ships are sufficient to span the required response space, effectively serving as a set of basis elements. Adding more ships provides little additional benefit. From another perspective, increasing the number of dictionary ships may introduce models with similar dynamics or models that differ significantly from the target ship. Nevertheless, the results indicate that the proposed method can effectively suppress the influence of such redundant or irrelevant ships through the neural network training process, thereby maintaining high prediction performance.

C. Full-Scale Experiment

To validate the effectiveness of the proposed method using full-scale maneuvering data, a field experiment was conducted using the research ship *R/V Gunnerus*. The *R/V Gunnerus* is an experimental ship owned by the Norwegian University of Science and Technology (NTNU). Its principal particulars are summarized in Table II. Fig. 7 shows snapshots of the zigzag maneuver of the *R/V Gunnerus*.

TABLE II: Principal particulars of R/V Gunnerus

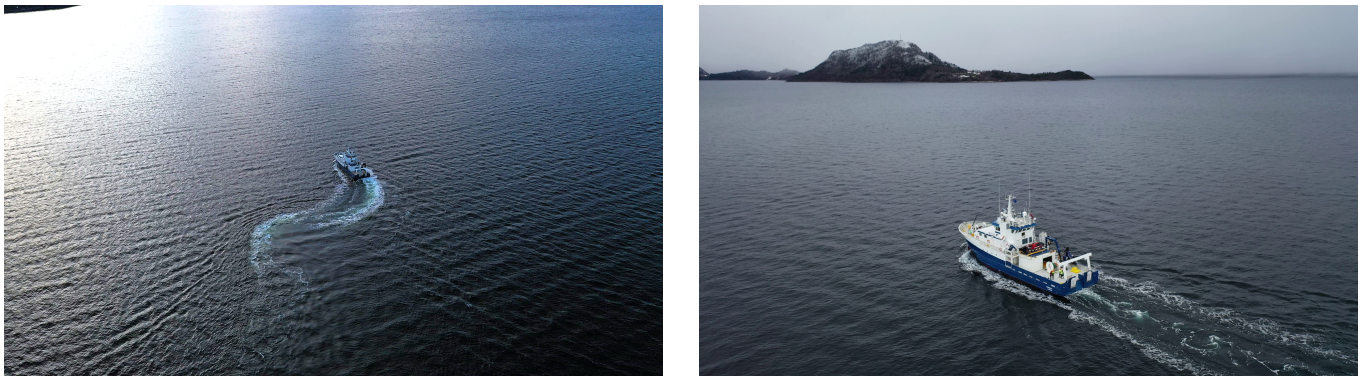
Item	Value
Length overall	33.90 m
Breadth	9.60 m
Draught	2.60 m

During the experiment, the *R/V Gunnerus* performed both zigzag and turning maneuvers under various rudder angles and thruster RPM settings. Throughout the experiment, the ship's motion was recorded using various onboard channels. Among these, the following monitoring data were utilized in this full-scale experiment:

- **Position**: Position records in the NED (North-East-Down) frame and heading angle.
- **Velocity**: Linear surge and sway velocities, as well as yaw rate.
- **Command inputs**: RPM and azimuth angles of the port and starboard thrusters. Although the *R/V Gunnerus* is also equipped with a tunnel thruster, it was not used during the experiment.

These monitoring data were used as input for both the dictionary models and the neural network to predict ship motion. For performance evaluation, a portion of the zigzag maneuver data obtained under rudder angles and thruster RPM conditions different from those of training dataset was extracted and used as a test dataset. The remaining data were used to train the neural network.

The target vessel used in the full-scale experiment, *R/V Gunnerus*, has a length around 30m, which falls within the range of the dictionary ship group constructed in the simulation experiments. An important finding from the simulation experiments



(a) Bird-eye view (b) Side view

Fig. 7: R/V Gunnerus conducting a zigzag maneuver in the open sea.

is that increasing the number of dictionary ships enhances the representational capability of the proposed method. However, in the cases examined, the marginal benefit of adding more dictionary ships gradually diminishes. Moreover, as the number of dictionary ships increases, the visualization and interpretability of the results become more challenging. Therefore, in the full-scale experiment, we employ Ship 1, Ship 2, and Ship 4, which achieved satisfactory performance in the simulation study.

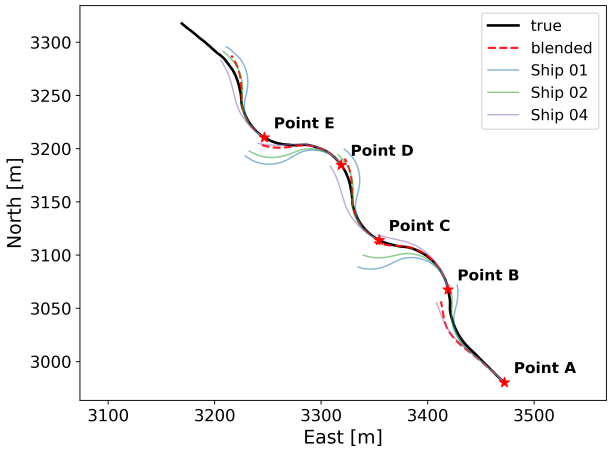


Fig. 8: Trajectory predictions in the full-scale experiment.

Figure 8 illustrates the actual ship trajectory in the test dataset, along with the predicted trajectories obtained from the dictionary models and the blended prediction, evaluated at multiple prediction points. The ship moves from the lower right to the upper left in the figure. Five prediction starting points are indicated by red stars at 30-second intervals. From each of these points, future trajectories for 30s are predicted using each method and plotted accordingly. For prediction starting at point A and D, the blended prediction closely resembles that of dictionary ship 4. In contrast, at points C and E, the blended prediction is more similar to that of dictionary ship 2. At point B, the proposed method produces a trajectory that is closer to the true trajectory than any individual dictionary ship. These results demonstrate that the proposed method effectively combines the strengths of multiple dictionary ships in a state-dependent manner, leading to improved prediction accuracy.

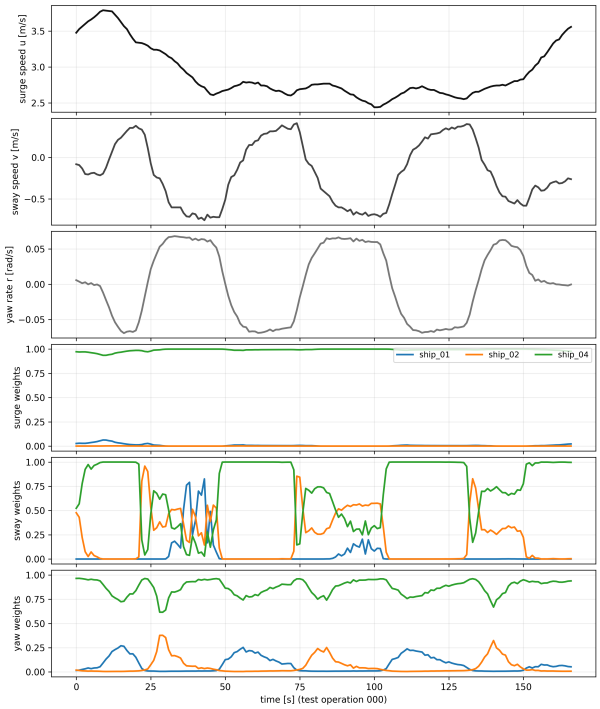


Fig. 9: A time history of the dictionary weights in the full-scale experiment.

Figure 9 further examines the relation between the learned weights and the measured ship state, because the proposed model is a state-dependent 3-DOF blender rather than a single global interpolation model. The upper panels show the time histories of the input state variables, namely the surge velocity u, sway velocity v, and yaw rate r, while the lower panels show the corresponding dictionary weights learned separately for the surge, sway, and yaw acceleration components.

A clear component-dependent structure can be observed. For the surge component, Ship 04 remains dominant throughout almost the entire maneuver. This suggests that, in the longitudinal response, the target vessel is represented mainly by the surge dynamics of a single nearby dictionary model,

and only minor corrections from the other ships are required.

The sway component shows a more state-dependent behavior. When the yaw rate is negative, the weight on Ship 04 becomes dominant in a relatively stable manner. By contrast, when the yaw rate becomes positive, the dominant contribution is shared mainly by Ships 02 and 04, while Ship 01 also retains a non-negligible contribution. Thus, no single dictionary vessel explains the sway response over the entire maneuver. Instead, the model switches and blends among several existing ship responses according to the instantaneous motion state, especially the sign and evolution of the yaw rate.

For the yaw component, Ship 04 is again the main contributor over much of the operation, but the secondary weights exhibit a consistent trend. The contribution of Ship 01 increases when the yaw rate is decreasing, whereas the contribution of Ship 02 increases when the yaw rate is increasing. This indicates that the yaw response is not represented by an arbitrary black-box correction, but by a state-dependent redistribution of weights among physically defined dictionary vessels.

These observations support the interpretation of the proposed method as an anchored and interpretable hybrid model. The learned weights are not merely free numerical fitting parameters; they indicate which existing ship models are locally used to reconstruct each component of the target response. In this sense, the prediction remains tied to the response characteristics of the dictionary vessels, while still allowing the relevant vessel contribution to vary with the operating state and with the motion component.

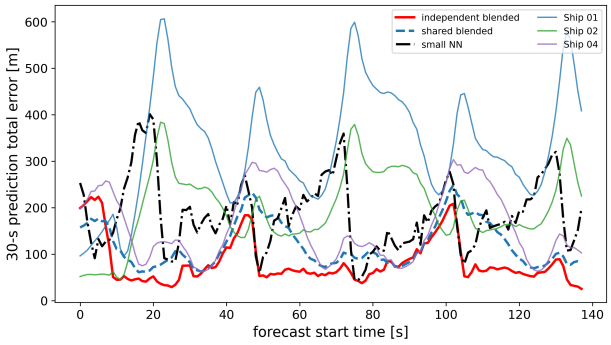


Fig. 10: A time history of the accumulated position prediction error over a 30-second prediction horizon at each time step.

Figure 10 presents the time series of the accumulated position prediction error over a 30-second prediction horizon at each time step. The error is defined as

$$\sum_{i=1}^{30} \sqrt{(N_i^{\text{true}} - N_i^{\text{pred}})^2 + (E_i^{\text{true}} - E_i^{\text{pred}})^2},$$

where N_i^{true} and E_i^{true} denote the ground-truth position components in the north and east directions at the i-th prediction step, respectively, and N_i^{pred} and E_i^{pred} denote the corresponding predicted position components.

The performance of each dictionary model varies substantially depending on the operating condition. The Ship 1 model generally exhibits poor agreement with the motion of the *R/V Gunnerus*. The Ship 2 model also shows relatively low overall accuracy, although it outperforms the Ship 4 model during certain periods. Among the individual dictionary models, the Ship 4 model achieves the best overall performance in most cases.

The proposed method consistently provides the most accurate predictions throughout almost the entire experiment by blending the available maneuvering responses in a state-dependent manner. This result indicates that no single dictionary ship is sufficient to represent the target vessel across all operating conditions and that the target dynamics can be represented more effectively through combinations of existing maneuvering behaviors.

Figure 10 also compares two blending strategies: a shared-weight formulation, in which a common blending vector is applied to surge, sway, and yaw motions, and the proposed component-wise formulation, in which independent blending weights are estimated for each motion component. The shared-weight formulation consistently produces larger prediction errors, with the degradation being particularly pronounced in sway and yaw motions.

This observation provides an important insight into the nature of ship representation. A shared-weight formulation implicitly assumes that the target vessel maintains the same relationship to the dictionary ships across all motion components. The experimental results suggest that this assumption is overly restrictive. In practice, a target vessel may exhibit surge characteristics similar to one dictionary ship while its sway and yaw responses are better represented by others. Consequently, allowing motion-dependent blending improves representation capability and leads to superior prediction performance.

It should be emphasized that the proposed component-wise blending does not ignore the coupling among surge, sway, and yaw motions. Each blending network receives the full vessel state vector as input, allowing the estimated weights to depend on the coupled vessel dynamics. The role of independent blending is therefore not to decouple the dynamics, but to provide sufficient flexibility to represent motion-dependent relationships between the target vessel and the dictionary ships.

In Fig. 10, the proposed blending method was also compared with the small direct neural network trained using the same target-ship data. The direct neural network achieved larger errors compared to the proposed blending. This result indicates that the improvement over the individual dictionary models is not merely an expected benefit of combining several constituent predictors. Instead, the acceleration responses generated by the validated dictionary models provide useful structural information for learning the target dynamics from the limited training data.

V. Discussion

The results of this study provide several insights into the representation of new ship dynamics through combinations of existing maneuvering models. Across both simulation and full-scale experiments, the proposed framework achieved accurate predictions when the target dynamics were sufficiently covered by the available dictionary models. In such cases, the target

vessel could be represented effectively through state-dependent blending of previously validated maneuvering behaviors, without requiring the construction of a dedicated ship-specific model.

An important implication of these findings is that accurate maneuvering prediction does not necessarily require identifying a new dynamic model from scratch. Instead, the results suggest that many target vessels can be represented within a response space generated by existing ships. This observation supports the central hypothesis of the present study: although individual ships possess unique characteristics, their maneuvering behaviors often share substantial common structure. When a suitable collection of validated models is available, a new ship may therefore be represented through combinations of existing responses rather than through complete model redevelopment.

The simulation studies further reveal that prediction performance is governed primarily by response-space coverage rather than by the number of available models. Prediction accuracy improves rapidly when the target dynamics become sufficiently covered by the dictionary, but additional models provide only marginal benefits once adequate coverage has been achieved. This finding has practical significance because it suggests that the value of a dictionary lies not in its size but in its ability to span relevant maneuvering behaviors. Consequently, future efforts may be more effectively directed toward improving coverage of maneuvering characteristics rather than simply increasing the number of available models.

The behavior of the proposed framework outside the covered response space is also noteworthy. When the target dynamics are not represented by the available dictionary models, the prediction naturally converges toward the closest available maneuvering behavior rather than producing unrestricted extrapolations. Unlike conventional black-box learning approaches, the prediction therefore remains associated with previously validated ship responses. Although such behavior does not guarantee physical correctness outside the covered domain, it provides a predictable failure mode that may be advantageous in practical engineering applications where robustness and interpretability are important considerations.

Another interesting observation concerns the interpretation of blending weights. The weights are not merely numerical coefficients used to improve predictive performance. Instead, they provide information about how the target vessel relates to previously validated ships contained in the dictionary. In the full-scale experiment, different blending patterns emerged for surge, sway, and yaw motions, indicating that similarity between ships is not necessarily uniform across all maneuvering characteristics. A target vessel may exhibit longitudinal behavior resembling one ship while its lateral or yaw responses are better represented by others. This observation suggests that ship similarity may be motion-dependent rather than globally defined, providing additional justification for representing maneuvering dynamics through combinations of behaviors rather than through a single reference vessel.

The results also motivate a broader perspective on maneuvering-model development. Traditionally, a maneuvering model is regarded as the final product of a modeling effort associated with a particular ship. Under this viewpoint, the usefulness of a model is largely confined to the vessel for which it was developed. The present results suggest an alternative interpretation. Validated maneuvering models can also be viewed as reusable engineering assets that collectively define a maneuvering response space. From this perspective, the value of a model extends beyond direct prediction and includes its contribution to the representation of future vessels. This shifts the emphasis from constructing isolated ship models toward building and maintaining libraries of validated maneuvering behaviors.

Despite these promising findings, several limitations should be acknowledged. First, the full-scale validation was conducted using a single vessel under relatively calm environmental conditions. Additional studies involving diverse ship types, operating conditions, and environmental disturbances are required to establish the generality of the proposed framework. Second, the present study employs a simple filtering strategy based on principal particulars to construct the dictionary. While this approach proved sufficient to demonstrate the feasibility of the concept, it does not provide a principled method for determining which models should be included or excluded.

A related limitation concerns the notion of response-space coverage itself. Although the experiments demonstrate that coverage plays a central role in prediction performance, the present study evaluates coverage only indirectly through prediction results. A rigorous definition of maneuvering-response coverage and practical metrics for assessing whether a target vessel is adequately represented by a given dictionary remain open research questions. Addressing these issues may enable more systematic construction and management of maneuvering-model libraries in future applications.

Furthermore, the present study focuses on maneuvering dynamics under relatively benign environmental conditions and does not explicitly account for waves, wind, or current disturbances. Extending the proposed framework to incorporate environmental effects, uncertainty quantification, and more complex actuator configurations represents an important direction for future research. Nevertheless, the fundamental concept introduced in this study—the representation of new ship dynamics through combinations of validated maneuvering behaviors—remains applicable beyond the specific experimental setting considered here.

Future work should therefore focus on large-scale validation using diverse real-world datasets, principled methods for dictionary construction, quantitative measures of response-space coverage, and integration with environmental and operational uncertainties. Such developments would further strengthen the role of maneuvering-model libraries as reusable engineering assets for future maritime systems.

The comparison with the direct neural network also clarifies the intended position of the proposed framework. The objective is not to claim universal predictive superiority over unrestricted data-driven models. A direct neural network has greater freedom to generate arbitrary acceleration responses and may outperform the proposed method when sufficiently representative training data and model capacity are available.

However, this flexibility comes without explicit attribution to previously validated maneuvering models and without a structural restriction to their response space.

In the proposed framework, the neural network does not directly generate an acceleration response or an additive residual correction. Its role is restricted to determining how the responses of validated dictionary models should be combined. The resulting prediction is therefore anchored to identifiable source models and provides explicit model-level traceability. This characteristic distinguishes the proposed method from both purely black-box models and hybrid methods in which a learned correction or mapping directly modifies the physical-model output.

VI. Conclusion

This paper presented a dictionary-based framework for representing the maneuvering dynamics of a new ship through combinations of existing validated maneuvering models. Rather than constructing a dedicated model for the target vessel, the proposed approach expresses the target dynamics as a state-dependent blend of responses generated by a library of previously developed ship models.

Simulation studies and full-scale experiments demonstrated that the proposed framework can accurately represent the maneuvering behavior of a target vessel when the relevant dynamics are sufficiently covered by the available dictionary. The results further showed that prediction performance depends primarily on response-space coverage rather than on the number of available models. When adequate coverage exists, the target dynamics can be reconstructed effectively through blending of existing responses. When coverage is insufficient, the framework naturally converges toward the closest available maneuvering behavior contained in the dictionary.

The results also highlight a broader implication for maneuvering-model development. Traditionally, maneuvering models are treated as ship-specific products whose usefulness is largely confined to the vessels for which they were developed. In contrast, the proposed framework demonstrates that validated maneuvering models can also serve as reusable engineering assets that contribute to the representation of future ships. From this perspective, the value of a maneuvering model extends beyond direct prediction and includes its role within a broader library of accumulated maneuvering knowledge.

The main contribution of this study is therefore not only a new prediction framework, but also a new viewpoint on maneuvering-model utilization. By treating existing ship models as reusable components and representing new vessels through their combinations, the proposed approach provides a practical mechanism for leveraging accumulated modeling assets while reducing the effort required to model future ships.

Future work should focus on systematic dictionary construction, quantitative measures of response-space coverage, validation across diverse ship types and operating conditions, and extension to environmental disturbances and uncertainty-aware prediction. Such developments may further establish maneuvering-model libraries as reusable knowledge infrastructures for future maritime systems.

References

[1] P. Zhang, Y. Ma, H. Chen, C. Liu, and Z. Wu, "The development and challenges of intelligent ship," in *The Proceedings of 2024 International Conference on Artificial Intelligence and Autonomous Transportation*, J. Liu, Y. Wang, B. Wu, Z. Jiang, and Y. Xiao, Eds. Singapore: Springer Nature Singapore, 2025, pp. 148–154.

[2] I. Im, D. Shin, and J. Jeong, "Components for smart autonomous ship architecture based on intelligent information technology," *Procedia Computer Science*, vol. 134, pp. 91–98, 2018, the 15th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2018) / The 13th International Conference on Future Networks and Communications (FNC-2018) / Affiliated Workshops. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918311116

[3] A. Bärlund, J. Linder, H. Feyzmahdavian, M. Lundh, and K. Tervo, "Nonlinear mpc for combined motion control and thrust allocation of ships," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 14 698–14 703, 2020, 21st IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896320324472

[4] J. Yin, H. Ren, and Y. Zhou, "The whole ship simulation training platform based on virtual reality," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 2, pp. 207–215, 2021.

[5] H. Yasukawa and Y. Yoshimura, "Introduction of MMG standard method for ship maneuvering predictions," *Journal of Marine Science and Technology*, vol. 20, no. 1, pp. 37–52, Mar. 2015.

[6] Y. Yoshimura, M. Nakamura, T. Taniguchi, and H. Yasukawa, "Empirical formulas of hydrodynamic parameters for predicting ship maneuvering based on the mmg-model," *Ocean Engineering*, vol. 337, p. 121831, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801825015379

[7] G. Rajesh and S. Bhattacharyya, "System identification for nonlinear maneuvering of large tankers using artificial neural network," *Applied Ocean Research*, vol. 30, no. 4, pp. 256–263, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141118708000667

[8] W. Luo and Z. Zhang, "Modeling of ship maneuvering motion using neural networks," *Journal of Marine Science and Application*, vol. 15, no. 4, pp. 426–432, Dec. 2016.

[9] Y. Jiang, X.-R. Hou, X.-G. Wang, Z.-H. Wang, Z.-L. Yang, and Z.-J. Zou, "Identification modeling and prediction of ship maneuvering motion based on LSTM deep neural network," *Journal of Marine Science and Technology*, vol. 27, no. 1, pp. 125–137, Mar. 2022.

[10] L. Hao, Y. Han, C. Shi, and Z. Pan, "Recurrent neural networks for nonparametric modeling of ship maneuvering motion," *International Journal of Naval Architecture and Ocean Engineering*, vol. 14, p. 100436, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2092678222000024

[11] Y. Xue, Y. Liu, G. Xue, and G. Chen, "Identification and prediction of ship maneuvering motion based on a gaussian process with uncertainty propagation," *Journal of Marine Science and Engineering*, vol. 9, no. 8, 2021. [Online]. Available: https://www.mdpi.com/2077-1312/9/8/804

[12] R. Skulstad, G. Li, T. I. Fossen, B. Vik, and H. Zhang, "A hybrid approach to motion prediction for ship docking—integration of a neural network model into the ship dynamic model," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[13] T. Wang, G. Li, L. I. Hatledal, R. Skulstad, V. Æsøy, and H. Zhang, "Incorporating approximate dynamics into data-driven calibrator: A representative model for ship maneuvering prediction," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1781–1789, 2022.

[14] B. Mei, L. Sun, and G. Shi, "White-black-box hybrid model identification based on rm-rf for ship maneuvering," *IEEE Access*, vol. 7, pp. 57 691–57 705, 2019.

[15] H. Yasukawa, R. Sakuno, and Y. Yoshimura, "Practical maneuvering simulation method of ships considering the roll-coupling effect," *Journal of Marine Science and Technology*, vol. 24, no. 4, pp. 1280–1296, Dec. 2019.

[16] Y. YOSHIMURA, "Mathematical model for the manoeuvring ship motion in shallow water," *Journal of the Kansai Society of Naval Architects, Japan*, vol. 200, pp. 41–51, 1986.

[17] M. Alexandersson, W. Mao, and J. W. Ringsberg, "System identification of vessel manoeuvring models," *Ocean Engineering*, vol. 266, p. 112940, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801822022235

[18] T. Mitsuyuki, K. Kuribayashi, R. F. S. Fernandez, H. Shimozawa, R. Kakuta, R. Niki, and R. Matsushita, "MMG 3DOF model identification with uncertainty of observation and hydrodynamic maneuvering coefficients using MCMC method," *Journal of Marine Science and Technology*, vol. 29, no. 3, pp. 668–682, Sep. 2024.

[19] G. An, G. Xiang, X. Xiang, and C. G. Soares, "Physics informed neural networks based identification modelling of ship maneuvering motion and associated optimal excitation design," *Engineering Applications of Computational Fluid Mechanics*, vol. 19, no. 1, p. 2566860, 2025. [Online]. Available: https://doi.org/10.1080/19942060.2025.2566860

[20] X. Zhou, L. Zou, and Z.-J. Zou, "A unified physics-data-driven modeling method for predicting time-series ship maneuvering motions in regular waves," *Ocean Engineering*, vol. 341, p. 122747, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801825024308

[21] K. Tian, Y.-Q. Wen, M. Zhu, T. Lei, S.-Y. Wang, and Y.-M. Huang, "Hybrid modeling and online prediction of 4-dof ship maneuvering motion in disturbed waters," *Ocean Engineering*, vol. 341, p. 122566, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801825022498

[22] K. Hatayama, K. Wakita, Y. Ichinose, and A. Maki, "Proposing a hybrid ship maneuvering model through force/moment-level integration of MMG and FNN with output constraints," *Journal of Marine Science and Technology*, Mar. 2026.

[23] M. Kanazawa, T. Wang, R. Skulstad, G. Li, and H. Zhang, "Knowledge and data in cooperative modeling: Case studies on ship trajectory prediction," *Ocean Engineering*, vol. 266, p. 112998, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801822022818

[24] V. Kanatas, N. Passalis, T. Wang, B. Sanguino, H. Zhang, G. Li, and A. Tefas, "Anchor-based adaptation for modeling vessel maneuvering dynamics," *Ocean Engineering*, vol. 341, p. 122722, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801825024059

[25] N. Shi, F. Lai, R. Al Kontar, and M. Chowdhury, "Fed-ensemble: Ensemble models in federated learning for improved generalization and uncertainty quantification," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 2792–2803, 2024.

[26] M. Garouani, A. Barhrhouj, and O. Teste, "Xstacking : An effective and inherently explainable framework for stacked ensemble learning," *Information Fusion*, vol. 124, p. 103358, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253525004312

[27] B. Zhang, L. Tang, and M. Roemer, "Probabilistic planning and risk evaluation based on ensemble weather forecasting," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 556–566, 2018.

[28] P. Dion, J.-L. Martel, and R. Arsenault, "Hydrological ensemble forecasting using a multi-model framework," *Journal of Hydrology*, vol. 600, p. 126537, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169421005849

[29] J. De Caigny, J. F. Camino, and J. Swevers, "Interpolation-based modeling of mimo lpv systems," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 1, pp. 46–63, 2011.

[30] F. Gatti and D. Clouteau, "Towards blending physics-based numerical simulations and seismic databases using generative adversarial network," *Computer Methods in Applied Mechanics and Engineering*, vol. 372, p. 113421, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004578252030606X

[31] T. Bikmukhametov and J. Jäschke, "Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models," *Computers & Chemical Engineering*, vol. 138, p. 106834, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135419313675

[32] T. Wang, R. Skulstad, G. Li, and H. Zhang, "An interactive multiple-model approach for accurate and interpretable trajectory prediction in autonomous docking," *IEEE Transactions on Automation Science and Engineering*, vol. 23, pp. 8166–8177, 2026.

[33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[34] I. T. T. C. (ITTC), "Ittc recommended procedures and guidelines: Full scale manoeuvring trials," International Towing Tank Conference (ITTC), Tech. Rep. 7.5-04-02-01, 2017, iTTC Quality System Manual. [Online]. Available: https://www.ittc.info/media/8179/75-04-02-01.pdf