Journal of Marine Science and Technology
https://doi.org/10.1007/s00773-026-01115-0

ORIGINAL ARTICLE

Proposing a hybrid ship maneuvering model through force/moment-level integration of MMG and FNN with output constraints

Kosuke Hatayama[1] · **Kouki Wakita**[1] · **Yasuo Ichinose**[1] · **Atsuo Maki**[1]

Received: 15 August 2025 / Accepted: 15 February 2026

Abstract

With recent advances in autonomous navigation technology, the importance of ship maneuvering models that can predict maneuvering motions with high accuracy and interpretability has increased. This study proposes a hybrid maneuvering model for predicting ship motions using the Maneuvering Modeling Group (MMG) model and the Feedforward Neural Network (FNN). In the proposed model, the modular structure of the MMG model is enhanced through the addition of a complementary FNN module. This approach improves the prediction accuracy while maintaining the interpretability of the model. Furthermore, to prevent the FNN output from depending too heavily on its training data, an output constraint and a mechanism for detecting Out-of-Distribution (OOD) based on k-Nearest Neighbors (kNN) distance are introduced. Numerical verification using multiple maneuvering test datasets demonstrates that the proposed model consistently exhibits higher prediction accuracy than both the MMG-only and the FNN-only models. Specifically, it attains an average reduction of 30–40% in Root-Mean-Squared Error (RMSE) compared to the MMG-only model. In particular, under highly nonlinear maneuvering conditions, such as Turning test and Random maneuver, the model shows up to a 70% RMSE reduction. Additionally, the model exhibits low accuracy degradation over long-term predictions, ensuring stable motion predictions. Based on these results, the proposed hybrid model successfully integrates the characteristics of physical models and data-driven models, providing a practical maneuvering model that offers high accuracy, lower data dependency, and interpretability. This model offers a valuable foundation for applications in autonomous navigation support systems.

Keywords Ship maneuvering · Maneuvering model · System identification · Hybrid model

1 Introduction

To realize autonomous ship navigation, a system-based mathematical model (hereinafter referred to as a maneuvering model) capable of accurately reproducing ship motions in real-world maritime environments is indispensable [1–3]. A maneuvering model can predict ship motions with limited computational resources, while high-fidelity models are essential for designing, tuning, and verifying control algorithms, as well as for developing autonomous navigation systems.

Maneuvering models are broadly categorized into physical models, data-driven models, and hybrid models. First, physical models provide a white-box approach with high interpretability, based on principles of fluid dynamics. Representative examples include the Maneuvering Modeling Group (MMG) model [4] and the Abkowitz model [5, 6], which remain the foundation of many research studies and practical applications. However, these models may not fully capture highly nonlinear maneuvering motions, such as those that occur during low-speed maneuvers.

Second, in contrast to physical models, data-driven models achieve superior prediction accuracy through a black-box approach, where machine learning is applied to operational data from real ships or model ships. Methods, such as Artificial Neural Networks (ANN) [7–9] and Support Vector Machines (SVM) [10], are commonly used. However, these models heavily depend on the quantity and quality of data, and lack a solid physical foundation, which leads to challenges in terms of interpretability [11].

Third, hybrid models integrate the physical knowledge of physical models with the nonlinear representational power of

✉ Kosuke Hatayama
hatayama_kousuke@naoe.eng.osaka-u.ac.jp

[1] Osaka University, 2-1 Yamadaoka, Suita, Osaka, Japan

Published online: 24 March 2026

data-driven models. By complementing the prediction errors of physical models with data-driven components, these models can achieve higher accuracy than physical models while requiring less data than purely data-driven models [12–17]. However, several challenges remain for hybrid models. As the model structure becomes more complex during integration, there is a risk that the internal computational processes may become black-boxed again [15, 18]. Furthermore, if the data quality or quantity is insufficient, the data-driven component may lose its complementary capability, potentially leading to a significant degradation in prediction accuracy. Therefore, ensuring compatibility between the physical and data-driven components and optimizing the balance of their contributions are crucial.

In this study, we propose a hybrid model that integrates the MMG model with an FNN in a novel way. The advantages of the proposed model compared to previously proposed hybrid models are as follows:

- To ensure high interpretability and physical consistency, a complementary FNN module is integrated at the force/moment level with output constraints. This design preserves the original MMG structure, enabling force decomposition and causal analysis while maximizing the use of physical insights.
- To avoid accuracy degradation, an Out-of-Distribution (OOD) detection mechanism based on k-Nearest Neighbor (kNN) distance is introduced. This prevents performance degradation by suppressing the FNN output whenever an input falls outside the training distribution.

Overall, the proposed approach achieves high prediction accuracy with reduced reliance on datasets, all while preserving the physical interpretability of the original MMG model. The effectiveness of this method in addressing the limitations of conventional hybrid models is validated through free-running scale model experiments.

2 Related work

This section classifies and summarizes existing methods for a maneuvering model into three categories: physical models, data-driven models, and hybrid models that integrate both approaches.

2.1 Physical models

Physical models are constructed based on fluid dynamic knowledge regarding ship motions, and their high interpretability has led to their widespread use in research and practical applications. Representative approaches include the MMG model [4, 19] and the Abkowitz model [5, 6].

The MMG model [4, 19] is a maneuvering model that separates and modularizes the forces and moments that act on the ship into components, such as the hull, propeller, and rudder. This model explicitly handles each contribution, allowing for the evaluation of individual effects. This approach also facilitates giving physical meaning to each parameter and force component, and it is advantageous for scalability.

The Abkowitz model [5, 6] is a linear maneuvering model based on dynamic parameters, which allows for a simple analysis of turning performance and steering responses. This model has been widely used in the past due to its merits.

While these physical models are excellent in terms of interpretability and scalability, they may not accurately represent highly nonlinear maneuvers, such as those at low speeds.

2.2 Data-driven models

Data-driven models predict ship motions based on motion data obtained directly from real ships or model ships, without relying on physical assumptions. Machine learning methods such as ANN and SVM are commonly used.

Data-driven models can learn flexibly nonlinear ship motions that may not be fully captured by physical models, demonstrating high prediction accuracy. Early studies by Yamato (1990) [20] and Hasegawa et al. (1993) [21] proposed steering models using ANN. Subsequently, Rajesh (2008) [7] and Zhang et al. (2013) [22] successfully replicated the nonlinear dynamic characteristics of maneuvering using FNN models. In recent years, attention has shifted to time-dependent models like Recurrent Neural Networks (RNN) and Long-Short Term Memory (LSTM). Wakita et al. (2021) [23] and Liu et al. (2024) [9] reported high accuracy in predicting low-speed maneuvers and trajectory using these models. Additionally, approaches using SVM [10] and ensemble Gaussian Process Regression (GPR) [11] have been proposed, all of which surpass physical models in terms of prediction accuracy.

However, data-driven models have several challenges. First, achieving high prediction accuracy requires large quantities of high-quality data. Second, they often show a large drop in accuracy with OOD inputs, making it hard to adapt to unknown situations. Third, the internal workings of these models are not transparent, which makes it difficult to understand the causal relationships behind their predictions and behaviors of the model.

2.3 Hybrid models

Hybrid models are gaining attention for combining the physical knowledges of physical models with the flexibility of data-driven models. Physical models, based on fluid dynamics, offer clear understanding but often show lower accuracy for nonlinear ship motions. In contrast, data-driven models can learn nonlinear maneuvering motions through training, offering high accuracy predictions. However, they face challenges such as data dependency and a lack of interpretability.

Against this background, numerous hybrid approaches have been proposed that integrate the strengths of both methods. Shi et al [12, 24] developed the "RM-SVM/RM-RF model", which integrates known physical models with SVM or Random Forest. It achieved higher accuracy than physical-only models while requiring less data than data-driven-only models would need. Skulstad et al. (2020, 2021) [13, 25] proposed the "Co-operative Hybrid Model" which corrects the predictions of physical models using LSTM or ANN, reducing the position prediction errors by an average of $4m$ for 30-second predictions.

Furthermore, Wang et al. (2022) [14] reported a 90% reduction in position error and improved stability by correcting prediction errors of physical models with ANN. However, the models proposed by Skulstad et al. (2020, 2021) [13, 25] and Wang et al. (2022) [14] face the challenge of unclear contributions of the physical and data-driven components due to the complexity of the integration. As a result, the individual contributions of each model to the overall behavior prediction remain unclear, leading to a loss of interpretability. Kanazawa et al. (2022) [15, 18] reported that in the construction of hybrid models, the balance between the contributions of physical and data-driven models is crucial, and a clear, simplified structure is effective for improving accuracy.

Other approaches, such as those by Li et al. (2024) [16] and Zhu et al. (2025) [26], used Automatic Identification System (AIS) data along with ANN or Multi-Objective Gaussian Process (MOGP) to achieve high prediction accuracy. However, the data dependency still presents a challenge. Additionally, Physics-Informed Neural Networks (PINNs), which integrate constraints based on physical formulas into the loss function, allow for learning with limited data while ensuring physical consistency [17, 27]. However, they still struggle to accurately replicate physical phenomena.

From these previous studies, it has been confirmed that hybrid models can achieve high accuracy predictions with less data than data-driven models by complementing the fluid dynamics knowledge of physical models. In contrast, the complexity of the integration structure leads to unclear contributions from both physical and data-driven models, resulting in reduced interpretability of the model. Issues such as significant prediction accuracy degradation for OOD inputs also remain.

3 Proposing model

This section describes the overall structure of the proposed hybrid model. The proposed hybrid maneuvering model operates in three degrees of freedom (surge, sway, and yaw). It consists of a physical modeling part (MMG model), a data-driven correction part (FNN), and an integration scheme that combines these two components. The model's inputs include the state variables of the ship, actuator states, and relative wind states. Using these inputs, the model predicts the its state variables for the next time step, with a 1 s time increment. The prediction flow within the model is shown in Fig. 1.

3.1 Coordinate system

Figure 2 shows the Earth-fixed coordinate system O-$x_0 y_0$ and the ship-fixed coordinate system O-xy used in this study. We define the ship's state variables using two vectors: the attitude vector $\boldsymbol{\eta} \equiv (x_0, y_0, \psi)^{\mathsf{T}} \in \mathbb{R}^3$, and the velocity vector $\boldsymbol{\nu} \equiv (u, v_{\mathrm{m}}, r)^{\mathsf{T}} \in \mathbb{R}^3$. Here, (x_0, y_0) represents the midship position in the Earth-fixed coordinate system, ψ is the heading angle, (u, v_{m}) are the surge speed and the sway speed of the ship, and r is the yaw rate. Additionally, the state variables of the ship are collectively defined as $\boldsymbol{x} \equiv (\boldsymbol{\eta}^{\mathsf{T}}, \boldsymbol{\nu}^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^6$. Note that the following relationship holds between the attitude vector $\boldsymbol{\eta}$ and the velocity vector $\boldsymbol{\nu}$:

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

where $\boldsymbol{R}(\boldsymbol{\eta}) \in \mathbb{R}^{3\times 3}$ is the rotation matrix from the O-xy to the O$_0$-$x_0 y_0$ coordinate system. The rotation matrix $\boldsymbol{R}(\boldsymbol{\eta})$ is given by

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

For actuator states, the angle of the rudder is denoted as δ_{r}, and the propeller rotation number is denoted as n_{p}. These are defined as $\boldsymbol{u} \equiv (\delta_{\mathrm{r}}, n_{\mathrm{p}})^{\mathsf{T}} \in \mathbb{R}^2$. The true wind speed and wind direction are denoted as U_{T} and γ_{T}, respectively, while the relative wind speed and direction as seen from the ship are denoted as U_{A} and γ_{A}, respectively. The relative wind state is defined as $\boldsymbol{\omega} \equiv (U_{\mathrm{A}}, \gamma_{\mathrm{A}})^{\mathsf{T}} \in \mathbb{R}^2$. In this study, we assume that the wind speed and direction are uniform in space (no spatial variation) and vary only with time.

3.2 Hybrid structure

In this study, corrections in terms of forces/moments are applied to the MMG model to complement the dynamics of ship motions that the MMG model cannot capture, while ensuring consistency with the maneuvering equation. Specifically, the proposed hybrid model predicts the ship's state variables at the next time step by incorporating the forces computed independently by the MMG model and the FNN into the right-hand side of the maneuvering equations, using

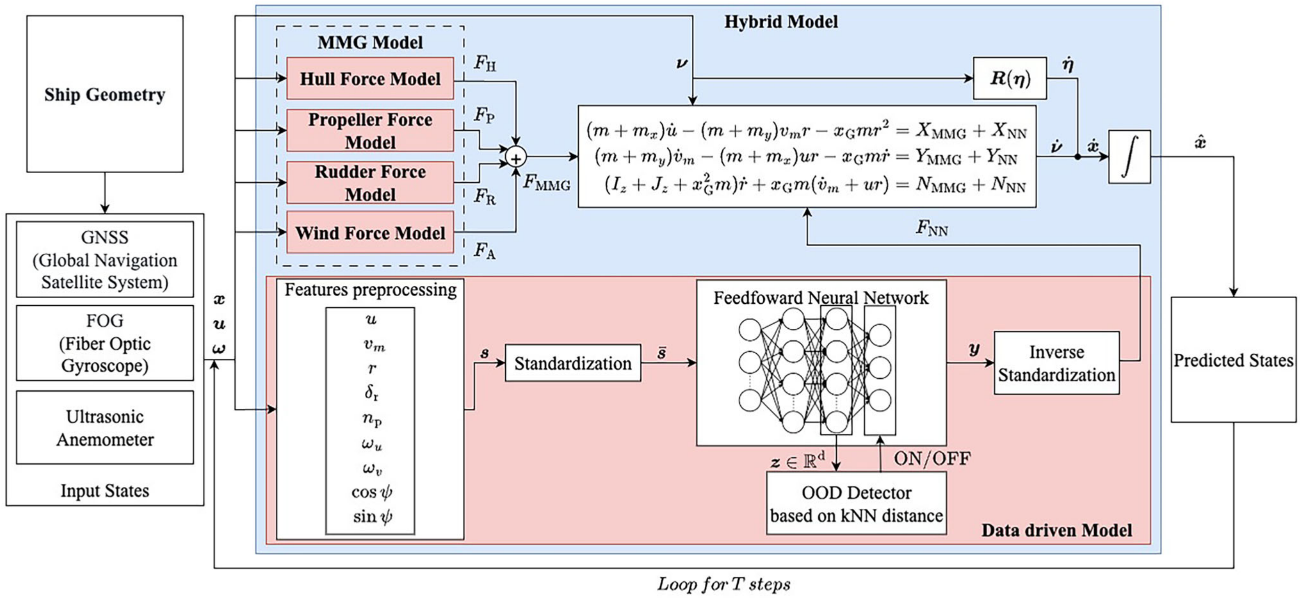

Fig. 1 Prediction flow of the proposed model: The state at a given step is used as input, and the fluid forces are calculated by the MMG model. The complementary module are added by the trained FNN. These results are subsequently incorporated into the right-hand side of the maneuvering equations, and the predicted state for the next step is output

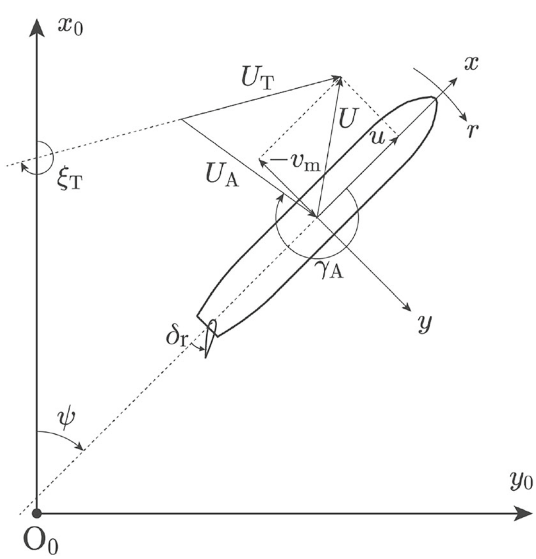

Fig. 2 Coordinate system

the current state variables as input. The external forces calculated by the MMG model and the FNN are defined as follows:

$$\begin{aligned} F_{\mathrm{MMG}} &\equiv (X_{\mathrm{MMG}}, Y_{\mathrm{MMG}}, N_{\mathrm{MMG}})^{\mathsf{T}} \in \mathbb{R}^3 \\ F_{\mathrm{NN}} &\equiv (X_{\mathrm{NN}}, Y_{\mathrm{NN}}, N_{\mathrm{NN}})^{\mathsf{T}} \in \mathbb{R}^3. \end{aligned} \tag{3}$$

In this case, the maneuvering equations are expressed as

$$\begin{cases} (m+m_x)\,\dot{u} - (m+m_y)v_m r - x_G m r^2 = X_{\mathrm{MMG}} + X_{\mathrm{NN}} \\ (m+m_y)\dot{v}_m + (m+m_x)ur + x_G m\dot{r} = Y_{\mathrm{MMG}} + Y_{\mathrm{NN}} \\ (I_z + J_z + x_G^2 m)\dot{r} + x_G m\,(\dot{v}_m + ur) = N_{\mathrm{MMG}} + N_{\mathrm{NN}}. \end{cases} \tag{4}$$

Table 1 Symbols and their meanings used in Eq. (4)

Symbol	Description
m	Mass of the ship
m_x	Added mass in the ship's surge direction
m_y	Added mass in the ship's sway direction
x_G	Position of the center of gravity from the ship's origin
I_z	Moment of inertia of the ship
J_z	Added moment of inertia

This integration method clarifies the distinct contributions of the hull, propeller, rudder, wind, and FNN corrections at the force and moment level, thereby facilitating analysis. Unlike conventional models that apply corrections directly to position [13, 14] or velocity [12, 24, 28], this approach offers a simple and highly compatible integration that preserves the traditional advantages of the MMG model. Furthermore, since the proposed method explicitly solves the equations of motion, it ensures kinematic continuity in velocity and position. This approach maintains consistency with the ship's mass and moment of inertia, thereby avoiding the physical inconsistencies often observed in data-driven models that directly correct state variables.

3.3 MMG model structure

In this study, the MMG model is adopted as the physical model. The MMG model mathematically represents ship maneuvering motions by modularizing the dynamic contributions of the hull, propeller, and rudder, taking into account their respective fluid forces [4, 29]. Since the physical contributions of each component are clearly separated in this model, it is easy to analyze and extend, making it widely used in many studies aimed at evaluating ship design and maneuvering performance.

The total force F_{MMG} is defined as the sum of fluid forces acting on each component

$$\begin{cases} X_{\mathrm{MMG}} = X_{\mathrm{H}} + X_{\mathrm{P}} + X_{\mathrm{R}} + X_{\mathrm{A}} \\ Y_{\mathrm{MMG}} = Y_{\mathrm{H}} + Y_{\mathrm{P}} + Y_{\mathrm{R}} + Y_{\mathrm{A}} \\ N_{\mathrm{MMG}} = N_{\mathrm{H}} + N_{\mathrm{P}} + N_{\mathrm{R}} + N_{\mathrm{A}}, \end{cases} \tag{5}$$

where the subscripts H, P, R, and A denote the hull, propeller, rudder, and wind (air), respectively. In this study, the unified models are utilized for the hull and wind forces to cover the entire maneuvering range. In contrast, the propeller and rudder forces are calculated using the varying models that switch equations depending on the operating conditions, such as forward/reverse rotation. The detailed mathematical formulation and specific equations for each component of the MMG model used in this study are provided in A.1. The model structure follows the standard MMG model with extensions for wind effects and propeller reverse rotation, based on the work of Miyauchi et al. [29].

3.4 FNN model structure

In this study, the hybrid model uses Artificial Neural Networks (ANN) [30], a type of machine learning, to realize a complementary FNN module. ANN is a mathematical model inspired by the human brain's neural system, known for its strong function-approximation ability. In this study, a fully connected FNN is used. The advantage of a FNN, compared to time-series models such as RNN and LSTM, is its simpler structure and superior computational efficiency. Additionally, considering the limited amount of training data available in this study, the FNN was selected as it is less prone to overfitting compared to complex time-series models.

3.4.1 Input parameters

The input vector $\boldsymbol{s}$ used for predicting the correction of the ship motion is defined as the following 9-dimensional vector:

$$\boldsymbol{s} \equiv \left(u,\ v_{\mathrm{m}},\ r,\ \delta_{\mathrm{r}},\ n_{\mathrm{p}},\ \omega_u,\ \omega_v,\ \cos\psi,\ \sin\psi\right)^{\mathsf{T}} \in \mathbb{R}^9. \tag{6}$$

The input vector includes velocity vectors, actuator states, attitude angles, and relative wind states. To avoid discontinuities from the 360° periodicity of angles [31], the cosine and sine of the attitude angle ψ and the relative wind state $\boldsymbol{\omega}_{\mathbf{A}} = (\omega_u, \omega_v) \in \mathbb{R}^2$ are transformed. ω_u and ω_v are the components x and y of the relative wind speed in the ship-fixed coordinate system, and are calculated as follows:

$$\begin{cases} \omega_u = U_{\mathrm{A}} \times \cos\gamma_{\mathrm{A}} \\ \omega_v = U_{\mathrm{A}} \times \sin\gamma_{\mathrm{A}}. \end{cases} \tag{7}$$

The input vectors involve variables with different scales; all variables are standardized to improve the stability and optimization performance of the training. All input variables are then standardized using the mean and standard deviation from the training dataset $\mathscr{D}^{(\mathrm{train})}$. For each component j of the input vector, the standardized value is defined as

$$\bar{s}_j = \frac{s_j - \mu_{s,j}^{(\mathrm{train})}}{\sigma_{s,j}^{(\mathrm{train})}}\ . \tag{8}$$

The standardized input vector $\bar{\boldsymbol{s}}$ is then used as the final input to the FNN

$$\bar{\boldsymbol{s}} = \left(\bar{u},\ \bar{v}_m,\ \bar{r},\ \bar{\delta}_r,\ \bar{n}_p,\ \bar{\omega}_u,\ \bar{\omega}_v,\ \overline{\cos\psi},\ \overline{\sin\psi}\right)^{\mathsf{T}} \in \mathbb{R}^9. \tag{9}$$

The FNN produces a 3-dimensional output vector $\boldsymbol{F}_{\mathrm{NN}} \in \mathbb{R}^3$ (representing predicted force/moment corrections). Each component $F_{\mathrm{NN},j}$ is then rescaled to physical units as

$$F_{\mathrm{NN},j} = \sigma_{F,j}^{(\mathrm{train})} \cdot y_j + \mu_{F,j}^{(\mathrm{train})}\ , \tag{10}$$

where $\boldsymbol{y} \in \mathbb{R}^3$ is the j-th output of the FNN's final layer (in standardized form), and $\mu_{F,j}^{(\mathrm{train})}$ and $\sigma_{F,j}^{(\mathrm{train})}$ are the training data mean and standard deviation for the j-th force/moment component, respectively. Therefore, the FNN model is expressed as the function $\boldsymbol{f}_{\mathrm{NN}}$ as shown in Eq. (11)

$$\boldsymbol{y} = \boldsymbol{f}_{\mathrm{NN}}\left(\bar{\boldsymbol{s}};\ \boldsymbol{\theta}\right), \tag{11}$$

where $\bar{\boldsymbol{s}}$ is the standardized input vector defined in the previous section, and $\boldsymbol{\theta}$ represents the parameters of the neural network.

3.4.2 Loss function

The optimization of model parameters during the training of the data-driven model is generally formulated as the minimization of the loss function. Given the dataset $\mathscr{D}$ during training, the loss function is evaluated and the model parameters $\boldsymbol{\theta}$ are adjusted to minimize it. The optimized parameters are denoted as $\boldsymbol{\theta}_{\mathrm{opt}}$, and the loss function is denoted as $\mathscr{L}(\boldsymbol{\theta};\mathscr{D})$. The optimization problem is then expressed as

$$\boldsymbol{\theta}_{\mathrm{opt}} = \operatorname{argmin} \mathscr{L}(\boldsymbol{\theta};\mathscr{D})\,. \tag{12}$$

By solving this optimization problem, we obtain $\boldsymbol{\theta}_{\mathrm{opt}}$, the set of parameters that best fit the given data. In this study, the Adam optimization method [32] is used.

The loss function for the proposed model is defined as

$$\mathscr{L}(\boldsymbol{\theta};\mathscr{D}) = \lambda_{\boldsymbol{\nu}}\mathscr{L}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\mathscr{D}) + \lambda_{\mathbf{F}}\mathscr{L}_{\mathbf{F}}(\boldsymbol{\theta};\mathscr{D}) + \lambda_{\mathrm{L2}}\|\boldsymbol{\theta}\|^2. \quad (13)$$

In Eq. (13), $\mathscr{L}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\mathscr{D})$ represents the state variable prediction error, $\mathscr{L}_{\mathbf{F}}(\boldsymbol{\theta};\mathscr{D})$ denotes the conditional regularization acting as a constraint for the FNN output, and $\|\boldsymbol{\theta}\|^2$ is the L2-regularization term to prevent overfitting. $\lambda_{\boldsymbol{\nu}}$, $\lambda_{\mathbf{F}}$, and λ_{L2} are scaling parameters that adjust the relative magnitude of each term. The first two terms, $\mathscr{L}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\mathscr{D})$ and $\mathscr{L}_{\mathbf{F}}(\boldsymbol{\theta};\mathscr{D})$, will be explained in detail later.

Loss of state variables: $\mathscr{L}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\mathscr{D})$

In model training, multi-step loss is used, where prediction errors for up to $T = 5$ steps ahead are included in the loss function. It has been reported that this method provides a smoothing effect against observation noise compared to single-step learning [33–35]. In theory, using an Root-Mean-Squared Error (RMSE) over T-step predictions shows that the error contribution from observation noise decays as $1/T$

$$\mathrm{MSE}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(\hat{v}_t - v_t)^2] \approx \mathrm{Bias}^2 + \mathrm{Var}_{\mathrm{model}} + \frac{\sigma_{\mathrm{obs}}^2}{T}. \quad (14)$$

However, if T is set too large, the loss function will overemphasize long-horizon averaging, potentially smoothing out the short-term errors that the FNN is supposed to learn to correct. This may reduce the FNN's correction capability. Moreover, with limited data, the state distribution at long-term steps may not be adequately covered by the training data, leading to instability in training. In this study, based on the assumption that the observation noise from velocity sensors follows a Gaussian distribution with a variance of $\sigma_{\mathrm{obs}} = (0.01, 0.01, 0.01°)$, we chose $T = 5$ in our training, as this value balanced noise suppression, data availability, and error smoothing. This choice allows the FNN to capture short-term prediction errors while maintaining stable training and prediction accuracy, even with limited data.

The prediction error of state variables calculated as the squared error between the observed data $\boldsymbol{\nu}$ and the predicted values $\hat{\boldsymbol{\nu}}$ over 5 steps

$$\mathscr{L}_{\boldsymbol{\nu}} = \sum_{i=0}^{B}\left\{\frac{1}{T}\sum_{t=0}^{T} w_{\nu}^{\mathsf{T}}(\boldsymbol{\nu}_{i,t} - \hat{\boldsymbol{\nu}}_{i,t})^2\right\}. \quad (15)$$

Here, B represents the batch size, which is the number of data samples used for a single model update, and $w_{\boldsymbol{\nu}} \in \mathbb{R}^3$ is a parameter that adjusts the scale differences in the velocity vector.

Regularization term for FNN output: $\mathscr{L}_{\mathbf{F}}$

In this study, a conditional regularization term is introduced to the FNN output F_{NN}. Without any constraints, the FNN output could potentially dominate the prediction. Given that the training data available for the FNN in this study are not abundant, such dominance could result in a model with low interpretability and low validity that ignores the physical insights provided by the MMG model. Therefore, to ensure that the MMG model remains the primary driver of the prediction while the FNN acts as a supplementary corrector, we introduce a penalty as a conditional regularization term [36, 37].

Domain knowledge from previous studies on maneuvering performance supports this design. For instance, evaluations of standard MMG models and Computational Fluid Dynamics (CFD) by Yasukawa and Yoshimura [4] and Duan et al. [38] quantify prediction errors using the formula $100\% \times (1 - \mathrm{pred}/\mathrm{obs})$. These studies indicate that for steady turning tests, the errors in turning indices, such as Advance (AD) and Tactical Diameter (DT), are relatively small, typically falling within the range of 2–6%. In contrast, for Zigzag tests involving transient responses, the errors in Overshoot Angles (OSA) are larger, often observed to be approximately 20% (e.g., 19–20% in 20/20 Zigzag tests). These findings suggest that while the physical model is highly reliable for steady motions, its structural limitations in dynamic phases result in residuals averaging up to 20%. Consequently, we selected 20% as the threshold and defined the penalty term as follows:

$$\mathscr{L}_{\mathrm{F}} = \sum_{i=1}^{B}\sum_{t=1}^{T} F_{\mathrm{penalty}\ i,t} \quad (16)$$

$$\begin{aligned} &F_{\mathrm{penalty}\ i,t} = w_{\mathrm{F}}^{\mathsf{T}}(|F_{\mathrm{NN}\ i,t}| - 0.2|F_{\mathrm{MMG}\ i,t}|)^{+} \\ &\quad \text{where} \\ &|F| = (|F_X|, |F_Y|, |F_N|),\ x^{+} = \max(x, 0). \end{aligned} \quad (17)$$

Here, $w_{\mathrm{F}} \in \mathbb{R}^3$ is a parameter that scales for each component of F_{NN}. The FNN is permitted to learn corrections freely within the 20% error margin of the MMG model but is suppressed beyond this threshold. Crucially, this 20% threshold is not a hyperparameter tuned to maximize prediction accuracy, but a quantitative bound determined by the known structural limitations of the standard MMG model. Therefore, we defined this threshold as a physically valid range for correction. This ensures that the FNN compensates only for inherent model errors, rather than overfitting to noise or outliers.

3.4.3 OOD detection based on kNN distance

In this study, a kNN distance-based detection mechanism is introduced to prevent the FNN from making unreliable corrections on OOD inputs. If an OOD input is detected, the FNN output is set to zero, and only the output of MMG model is used [39].

Regarding the choice of the OOD detection method, comparisons were made with other approaches such as ensemble methods (e.g., MC Dropout)[40, 41], generative models (e.g., VAE)[42], and statistical distances (e.g., Mahalanobis distance)[43]. First, in terms of computational efficiency for real-time implementation, ensemble methods generally require multiple forward passes, significantly increasing the computational load. In contrast, the proposed kNN-based method utilizes the low-dimensional feature vector ($z_t \in \mathbb{R}^{64}$) extracted from the intermediate layer of the FNN. Distance calculations between a 64-dimensional vector and thousands of data points can be completed in the order of milliseconds on modern CPUs. Therefore, the computational load is negligible relative to the 1 Hz simulation cycle, making it highly suitable for real-time control loops. Second, regarding data characteristics, generative models risk overfitting when training data is limited. Furthermore, statistical methods like Mahalanobis distance assume a Gaussian distribution, which may not capture the complex manifold of ship maneuvering motions (e.g., Random maneuvers). The kNN approach is non-parametric and makes no assumptions about the data distribution, allowing for robust detection based on local density in the semantic feature space learned by the FNN. Therefore, considering the trade-off between computational cost and robustness with limited data, we selected the kNN distance method.

To increase the reliability of OOD detection, the kNN distance is calculated using the feature vector z_t extracted from the intermediate layers of the FNN [44]. This vector is a high-dimensional activation vector obtained just before the output layer when the input state $\boldsymbol{s}_t$ at time t is given to the FNN. It represents the learned semantic feature in the training process. We define $d_{\mathrm{kNN}}(z_t)$ as the average Euclidean distance between z_t and its k-nearest feature vectors from the training data

$$d_{\mathrm{kNN}}(z_t) = \frac{1}{k}\sum_{i=1}^{k}\left\| z_t - z_{\mathrm{NN}}^{(i)} \right\|_2 . \tag{18}$$

Based on this distance, the FNN output is switched ON/OFF according to an adaptive threshold δ_{kNN}. Let $\mathscr{D}_{\mathrm{train}} = \{d_1, d_2, \ldots, d_N\}$ be the set of kNN distances in the training dataset, and let $\alpha \in (0, 1)$ be a fixed hyperparameter representing the target coverage rate of the normal data. The threshold δ_{kNN} is determined as the α-quantile of the distribution of $\mathscr{D}_{\mathrm{train}}$. Formally, using the empirical cumulative distribution function (ECDF) $\hat{F}(x)$, the threshold is defined as

$$\delta_{\mathrm{kNN}} = \inf\{x \in \mathbb{R} : \hat{F}(x) \geq \alpha\}, \tag{19}$$

where

$$\hat{F}(x) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(d_i \leq x). \tag{20}$$

In this study, we set $\alpha = 0.9995$, meaning δ_{kNN} corresponds to the top 0.05% of the kNN distances in the training data. This high threshold was chosen, because the training data are sparse relative to the pattern diversity; valid data falling into gaps in the training distribution could be misclassified as OOD with a stricter threshold. By setting the threshold to cover 99.95% of the training data, we aim to detect only clear OOD inputs and avoid false positives

$$\boldsymbol{y}_t = \begin{cases} F_{\mathrm{MMG}} + F_{\mathrm{NN}} & \text{if } d_{\mathrm{kNN}}(z_t) \leq \delta_{\mathrm{kNN}} \\ F_{\mathrm{MMG}} + 0 & \text{if } d_{\mathrm{kNN}}(z_t) > \delta_{\mathrm{kNN}}. \end{cases} \tag{21}$$

This mechanism reduces the risk of the FNN making unreliable corrections on OOD-inputs-corrections that could otherwise degrade overall prediction accuracy. Although this hard switching logic inevitably introduces discontinuities in the acceleration vector, we determined that the risk of model destabilization caused by erroneous FNN predictions in OOD regions significantly outweighs the drawbacks of these transient discontinuities. Therefore, mitigating severe deterioration in predictive performance was given precedence over maintaining smoothness in the transition.

3.5 Simulation method

Solving Eq. (4) for $\dot{\boldsymbol{\nu}}$ yields the state derivative $\dot{\boldsymbol{x}}$. The entire model is formulated in the form of the following differential equation:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\omega}(t); \boldsymbol{\theta}\right). \tag{22}$$

Here, $\boldsymbol{f}$ is the model function that takes the given ship's states variables $\boldsymbol{x}$, actuator states $\boldsymbol{u}$, and relative wind states $\boldsymbol{\omega}$ as inputs and outputs $\dot{\boldsymbol{x}}$ based on Eq. (4). Also, $\boldsymbol{\theta}$ represents the model parameters that are trained by minimizing the loss function.

We then predict the state variables of the ship at the next time step by numerically integrating $\dot{\boldsymbol{x}}(t)$ forward using a simple Euler method. This process is applied repeatedly to simulate the trajectory sequentially. When time-series data for the ship's state variables $\boldsymbol{x}$, actuator states $\boldsymbol{u}$, and relative wind states $\boldsymbol{\omega}$ are provided, the entire dataset is defined as follows:

$$\mathscr{D} \equiv \left\{\boldsymbol{x}_n(t), \boldsymbol{u}_n(t), \boldsymbol{\omega}_n(t) \mid t \in [t_{0,n}, t_{1,n}]\right\}_{n=1,2,\ldots,N}. \quad (23)$$

Here, t represents the time step, and n represents the index of a time-series sample. Using this dataset, numerical simulation of the ship's state can be performed for any given time-series.

The predicted state variables $\hat{\boldsymbol{x}}_n(t;\boldsymbol{\theta})$ are defined as

$$\begin{aligned}\hat{\boldsymbol{x}}_n(t;\boldsymbol{\theta}) &= \boldsymbol{x}_n(t_{0,n}) \\ &\quad + \int_{t_{0,n}}^{t} \boldsymbol{f}\left(\hat{\boldsymbol{x}}_n(\tau;\boldsymbol{\theta}), \boldsymbol{u}_n(\tau), \boldsymbol{\omega}_n(\tau);\boldsymbol{\theta}\right) \mathrm{d}\tau \quad (24) \\ &\quad \text{for} \quad n = 1, 2, \ldots, N.\end{aligned}$$

Finally, the prediction accuracy is quantified by the prediction error, defined as the difference between the observed velocity components $\boldsymbol{\nu}$ and the simulated values $\hat{\boldsymbol{\nu}}$.

4 Validation by numerical experiments

This section aims to quantitatively validate the effectiveness of the proposed hybrid model by evaluating its motion prediction accuracy through comparative experiments with conventional models.

4.1 Measurement data

The dataset used for training and testing in this study was generated through free-running model tests conducted at the Inukai Pond facility of the University of Osaka. A 1/108.33 scale model of the VLCC *M.V. ESSO OSAKA* was selected as the subject ship. Although the full-scale vessel typically requires tug assistance for berthing, this ship was chosen due to the availability of detailed model parameters and an existing free-running model suitable for data acquisition. Each test was conducted at the center of the experimental pool, ignoring the presence of a quay and shallow water effects. The principal particulars of the model ship are listed in Table 2. The loading condition was set to correspond to the full-scale trial conditions.

The model ship is equipped with a high-precision onboard measurement system, which includes a Fiber Optic Gyroscope (FOG), three Global Navigation Satellite System (GNSS) receivers (Magellan Systems Japan MJ-3021-GM4-QZS-EVK), and two ultrasonic anemometers (Gill PGWS-100-3). These instruments were used to record the time-series of the ship's state vector $\mathbf{x}(t)$ and the true wind vector $\boldsymbol{\omega}_T(t)$.

Regarding the data processing, all measurements were originally performed at a sampling frequency of 10 Hz. The ship's position trajectory (x_0, y_0) was derived from the GNSS receivers and converted to the midship position. To ensure high positioning accuracy, the Centimeter Level Augmentation Service (CLAS) was utilized. Furthermore, the relative distance between the two GNSS receivers was monitored during the tests, and only data with a distance error of less than 5 cm were used to guarantee reliability. The surge and sway velocities were estimated by numerically differentiating the midship trajectory using a second-order central difference method (with a differential time of $\Delta t = 1.0$ s) and smoothing the results with a linear Kalman filter.

Table 2 Principal particulars of subject ship *Esso Osaka*

Item	Value
Length between perpendicular: L_{pp}(m)	3.0
Ship breadth: B(m)	0.489
Ship draft: d(m)	0.201
Diameter of propeller: D_{p}(m)	0.084
Area of rudder: A_{R} (m^2)	0.0106
Diameter of bow thruster: D_{BT} : (m)	0.050
Diameter of stern thruster: D_{ST} : (m)	0.050
Mass: m (kg)	244.6
Longitudinal center of gravity: x_{G}(m)	0.094
Transverse projected area: A_{T} (m^2)	0.135
Lateral projected area: A_{L} (m^2)	0.520
Block coefficient: C_{b}	0.830

The heading angle ψ was calculated based on the relative position vector between the fore and aft GNSS antennas. This approach was adopted to eliminate the influence of sensor drift observed in the FOG, which was approximately 5 degrees per 10 min. The yaw rate r was directly measured by the FOG and processed using a low-pass filter with a cutoff frequency of 0.2 Hz. Wind disturbances were modeled as a function of time. To mitigate spatial dependency, apparent wind velocities measured by the two onboard anemometers were converted into local true wind velocities and then averaged to determine the environmental parameters.

The maneuvering data collected include the following:

- Berthing maneuver: These were included to cover complex maneuvering motion data for berthing and unberthing in the low-speed range, which are difficult to obtain in standard turning and zigzag tests.
- Turning test and zigzag test: These were selected as the baseline data for verifying model accuracy, because they are standard maneuvers widely used to evaluate a ship's turning ability, course-keeping ability, and yaw-checking performance
- Random maneuver: These inputs were introduced to efficiently enhance the model's generalization performance

Table 3 Control conditions for the rudder angle and the propeller rotation number in each maneuvering test

Test type	Rudder angle [°]	Propeller rotation number [rps]
Turning test	±20, ±35	8, 10
Zigzag test	±15, ±20, ±30	10, 12, 15, 17
Random maneuver	[−35, 35]	[−10, 10]

by ensuring data diversity. Specifically, they encompass diverse motion states independent of specific patterns—such as astern motion and propeller reversal—which are not fully covered by standard Turning and Zigzag tests. As demonstrated in the previous studies [45, 46], learning from such randomized inputs is an effective approach to achieving robust estimation even with a limited amount of data.

Table 3 summarizes the rudder angle and propeller rotation number settings for each maneuvering test. We obtained all data at 10 Hz and downsampled to 1 Hz for the experiment. The detailed measurement procedures and equipment configuration follow those described in [23, 29].

4.2 Datasets and comparison models

In this paper, we refer to a Turning test with a rudder deflection of $\delta°$ as the "Turning $\delta°$ Test", and a Zigzag test with a rudder deflection of $\delta°$/ and a heading change of $\psi°$ as the "Zigzag $\delta°/\psi°$ Test". The datasets used in this study are classified into two types based on their characteristics, as shown in Table 4:

- $\mathscr{D}_1$: Composed of scenarios with relatively limited control conditions, intended for model accuracy validation within a narrow distribution range.
- $\mathscr{D}_2$: Includes all data from $\mathscr{D}_1$ and additionally incorporates random maneuver data (including scenarios with reverse motion and propeller reversal) to expose the model to a wider, more diverse set of motion patterns and thus validate the maneuvering model's accuracy over a broader domain.

We applied the multi-step sampling method to augment the training data, by sliding a 5-s window through each time-series in 2-s increments. This approach increases the diversity of training samples extracted from the limited dataset.

For comparison of performance, the following three types of models were used:

- **MMG-only model**: a traditional physics-based MMG model. Forces and moments are calculated from the contributions of the hull, rudder, propeller, and wind using theoretical formulas.
- **NN-only model**: a fully data-driven model consisting solely of an FNN. Forces and moments are approximated by the FNN.
- **Hybrid model**: the proposed MMG+FNN hybrid. It combines the MMG model's output $\boldsymbol{F}_{\mathrm{MMG}}$ with an additional FNN-generated force/moment term $\boldsymbol{F}_{\mathrm{NN}}$, thereby capturing complex nonlinear motions.

Table 4 Time durations [s] of each test type in the datasets

	Berthing Maneuver	Turning Test	Zigzag Test	Random Maneuver
$\mathscr{D}_1$	1042.4	5269.6	914.1	—
$\mathscr{D}_2$	1042.4	5269.6	914.1	5281.0

Table 5 Maneuvering tests and their time durations used for test data

Test type	Duration [s]
Berthing Maneuver	127.7
Turning 20° Test	311.2
Turning 35° Test	519.7
Zigzag 20°/20° Test	140.2
Zigzag 30°/30° Test	121.1
Random Maneuver	671.6
Total	1891.5

For each model, two training datasets ($\mathscr{D}_1$ and $\mathscr{D}_2$) were used. Prediction accuracy was evaluated using the RMSE of the predicted velocity vector $\hat{\boldsymbol{v}}$. Six types of maneuvering tests were used as test scenarios to evaluate model performance (Table 5).

The prediction period was set to 1 s per step and predictions were made for medium-term (30-s) and long-term (100-s) horizons. The medium-term prediction is suitable for accuracy comparison with minimal cumulative error, while the long-term prediction highlights the model stability. A comprehensive evaluation was performed using both periods [15, 23].

4.3 Model parameters

The details of the FNN model used to construct $\boldsymbol{f}_{\mathrm{NN}}$ for simulation experiments are shown in Table 6. The datasets $\mathscr{D}_1$ and $\mathscr{D}_2$ were split into training and validation sets using a 0.7 ratio (70% for training and 30% for validation). The numbers of training samples were $N_{\mathscr{D}_1} = 2330$ and $N_{\mathscr{D}_2} = 4338$. The total number of parameters in the FNN is

$$P = (9 \cdot 64 + 64) + (64 \cdot 64 + 64) + (64 \cdot 3 + 3) = 4995.$$

Table 6 Configuration of the FNN model used in this study

Layer	Units	Activation function
Input	9	Tanh
Hidden 1	64	Tanh
Hidden 2	64	Tanh
Output	3	Linear

(25)

Accordingly, the parameter-to-sample ratios are $P/N_{\mathscr{D}_1} \approx 2.14$ and $P/N_{\mathscr{D}_2} \approx 1.15$. Since it has been pointed out that, for deep models, the relationship between the number of parameters and the generalization error is not necessarily monotonic[47], we did not assess the appropriateness of the model size based solely on these ratios. Instead, overfitting was mitigated by regularization and early stopping based on the validation data[48].

Other parameters, such as scaling parameters ($\lambda_{\boldsymbol{v}}$, $\lambda_{\mathbf{F}}$, $\lambda_{\mathbf{L2}}$ etc.), were tuned iteratively to suitable values for each model.

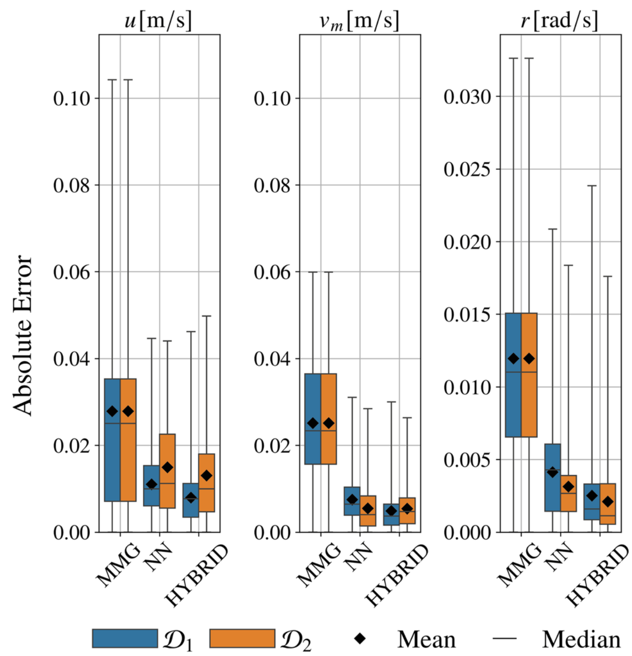

Fig. 3 Absolute error in the rurning 20° test (30-s prediction)

4.4 Simulation results

Tables 7 and 8 show the RMSE for u (Surge), v (Sway), and r (Yaw) components for 30-s and 100-s predictions in each test, categorized by model (MMG, NN, Hybrid) and dataset ($\mathscr{D}_1$, $\mathscr{D}_2$). We use these RMSE values as quantitative metrics to evaluate each model's prediction accuracy.

Additionally, Figs. 3 to 5 present box plots of the 30-s prediction errors for the Turning 20° test, Zigzag 30°/30° test, and Random maneuver. These figures visually show the interquartile range (IQR), mean, and median of the absolute error (AE) for the velocity vector to compare the error distributions among the different models (see Fig. 4).

Table 7 RMSE [m/s, rad/s] for u, v, and r predictions at 30 s for each test (by model and dataset)

Test Name	Model	$\mathscr{D}_1$			$\mathscr{D}_2$		
		u	v_m	r	u	v_m	r
Berthing maneuver	MMG	0.0103	0.0132	0.0115	0.0103	0.0132	0.0115
	NN	0.0076	0.0063	0.0071	0.0114	0.0062	0.0062
	Hybrid	0.0069	0.0071	0.0052	0.0062	0.0076	0.0048
Turning 20° test	MMG	0.0189	0.0227	0.0105	0.0189	0.0227	0.0105
	NN	0.0086	0.0076	0.0048	0.0107	0.0074	0.0044
	Hybrid	0.0087	0.0069	0.0041	0.0094	0.0068	0.0035
Turning 35° test	MMG	0.0220	0.0173	0.0089	0.0220	0.0173	0.0089
	NN	0.0093	0.0081	0.0052	0.0075	0.0090	0.0036
	Hybrid	0.0101	0.0084	0.0044	0.0085	0.0081	0.0022
Zigzag 20°/20° test	MMG	0.0262	0.0363	0.0126	0.0262	0.0363	0.0126
	NN	0.0247	0.0199	0.0106	0.0245	0.0217	0.0124
	Hybrid	0.0199	0.0134	0.0057	0.0237	0.0120	0.0042
Zigzag 30°/30° test	MMG	0.0354	0.0529	0.0241	0.0354	0.0529	0.0241
	NN	0.0336	0.0404	0.0287	0.0264	0.0499	0.0339
	Hybrid	0.0239	0.0262	0.0153	0.0315	0.0282	0.0132
Random maneuver	MMG	0.0210	0.0120	0.0075	0.0210	0.0120	0.0075
	NN	0.0340	0.0182	0.0086	0.0164	0.0062	0.0055
	Hybrid	0.0199	0.0109	0.0063	0.0149	0.0071	0.0034

Table 8 RMSE [m/s, rad/s] for u, v, r predictions at 100 s for each test (by model and dataset)

Test Name	Model	$\mathscr{D}_1$			$\mathscr{D}_2$		
		u	v_m	r	u	v_m	r
Berthing maneuver	MMG	0.0237	0.0173	0.0161	0.0237	0.0173	0.0161
	NN	0.0207	0.0062	0.0080	0.0331	0.0074	0.0066
	Hybrid	0.0170	0.0095	0.0082	0.0222	0.0060	0.0061
Turning 20° test	MMG	0.0372	0.0294	0.0138	0.0372	0.0294	0.0138
	NN	0.0132	0.0091	0.0051	0.0192	0.0075	0.0040
	Hybrid	0.0099	0.0067	0.0039	0.0169	0.0068	0.0032
Turning 35° test	MMG	0.0455	0.0224	0.0098	0.0455	0.0224	0.0098
	NN	0.0101	0.0096	0.0065	0.0087	0.0126	0.0052
	Hybrid	0.0173	0.0122	0.0055	0.0107	0.0109	0.0032
Zigzag 20°/20° test	MMG	0.0645	0.0398	0.0144	0.0645	0.0398	0.0144
	NN	0.0170	0.0239	0.0114	0.0286	0.0270	0.0140
	Hybrid	0.0233	0.0163	0.0093	0.0145	0.0115	0.0056
Zigzag 30°/30° test	MMG	0.0273	0.0520	0.0217	0.0273	0.0520	0.0217
	NN	0.0659	0.0355	0.0254	0.0682	0.0478	0.0323
	Hybrid	0.0283	0.0260	0.0160	0.0347	0.0288	0.0144
Random maneuver	MMG	0.0299	0.0154	0.0085	0.0299	0.0154	0.0085
	NN	0.0373	0.0273	0.0082	0.0138	0.0079	0.0070
	Hybrid	0.0230	0.0148	0.0086	0.0126	0.0086	0.0041

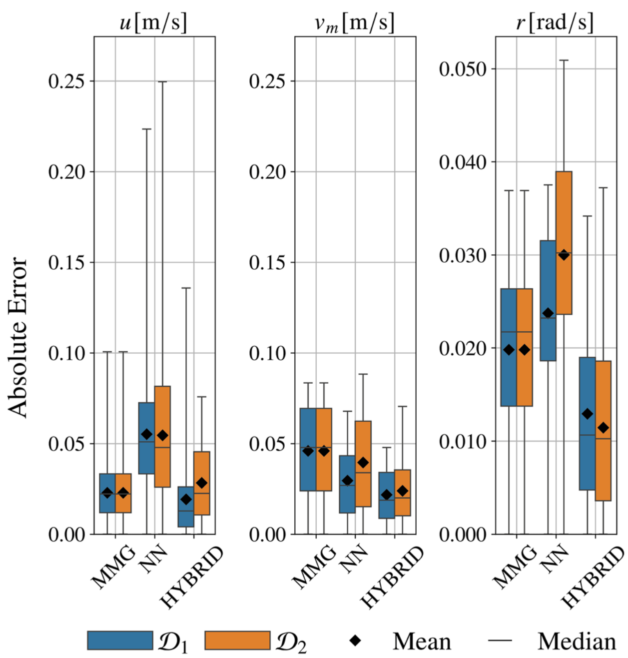

Fig. 4 Absolute error in the zigzag 30°/30° test (30-s prediction)

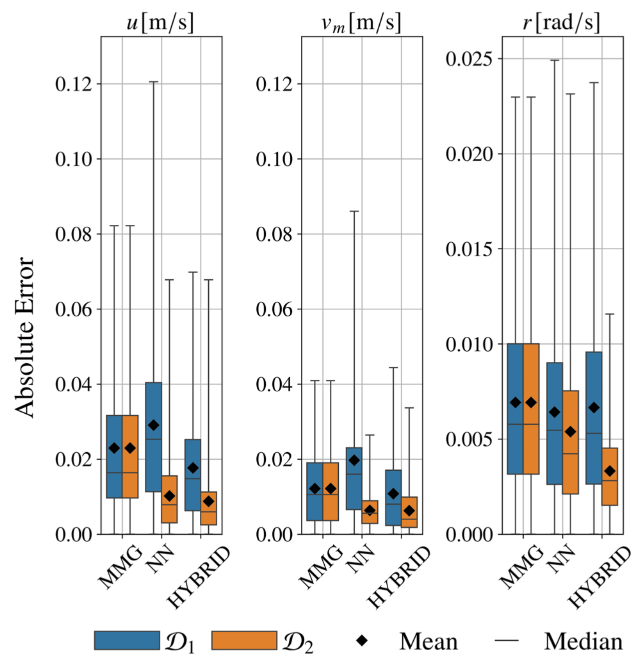

Fig. 5 Absolute error in the random maneuver (30-s prediction)

Figures 7 and 9 show the 100-s prediction results for each model (MMG, NN, Hybrid), trained using $\mathscr{D}_2$. The results correspond to the Berthing maneuver (0–100 s) and Random maneuver (200–300 s), and compare the predicted attitude vector $\hat{\boldsymbol{\eta}}$ and velocity vector $\hat{\boldsymbol{\nu}}$ with the observed data. Additionally, Figs. 6 and 8 show the corresponding actuator states and the relative wind states. Both time intervals involve irregular and unsteady maneuvers, revealing the characteristics of each model and the supportive role of the FNN. These figures are used to observe and evaluate how well each model can predict the characteristics of short-term ship motion under unsteady actuator states.

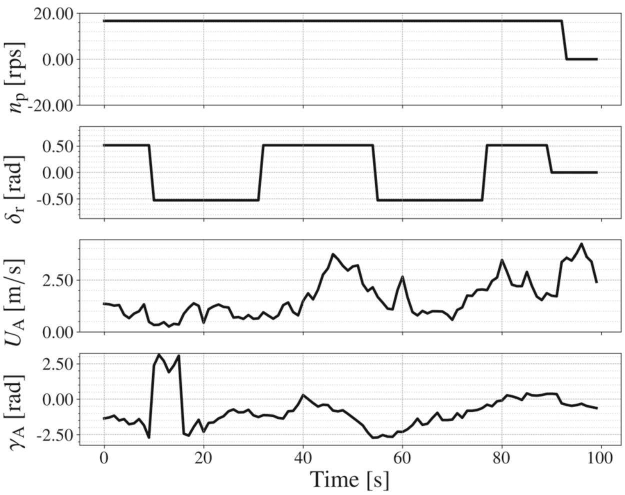

Fig. 6 Actuator state and relative wind state in the Zigzag 30°/30° test (trained with $\mathscr{D}_2$: 0–100 s)

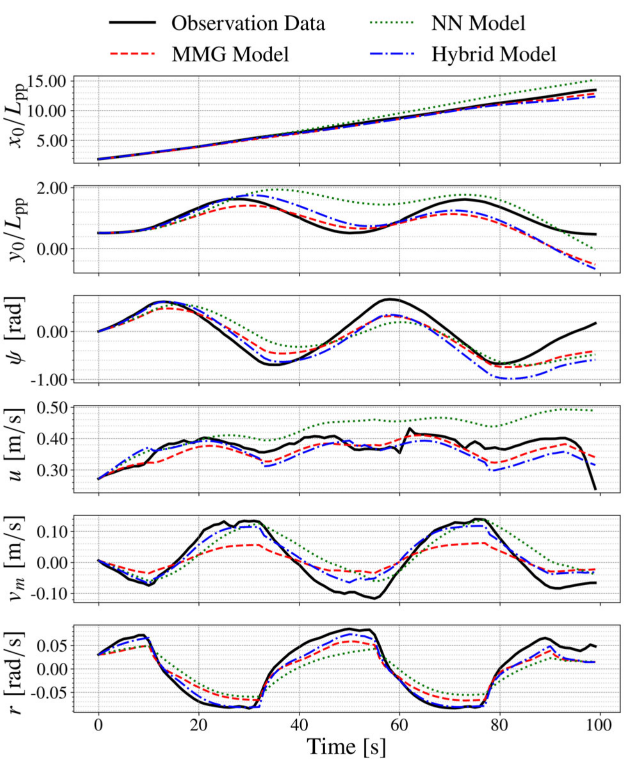

Fig. 7 Time variation of the attitude vector and velocity vector in the Zigzag 30°/30° test (trained with $\mathscr{D}_2$: 0–100 s)

Figures 10 to 12 show the comparison results between the predicted trajectories and the observed trajectories for each model (MMG, NN, Hybrid) trained by $\mathscr{D}_2$. The evaluation includes the Turning 20° test, the Zigzag 30°/30° test, and the Random maneuver. These tests are used to visualize the spatial prediction performance for each motion and to compare the differences between models.

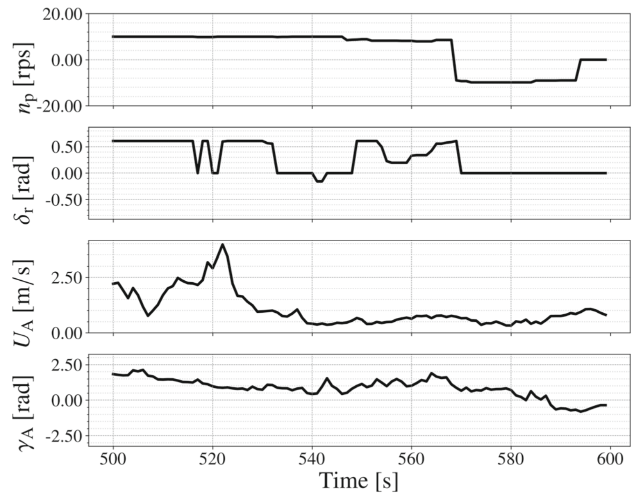

Fig. 8 Actuator state and relative wind state in the Random maneuver (trained with $\mathscr{D}_2$: 500–600 s)

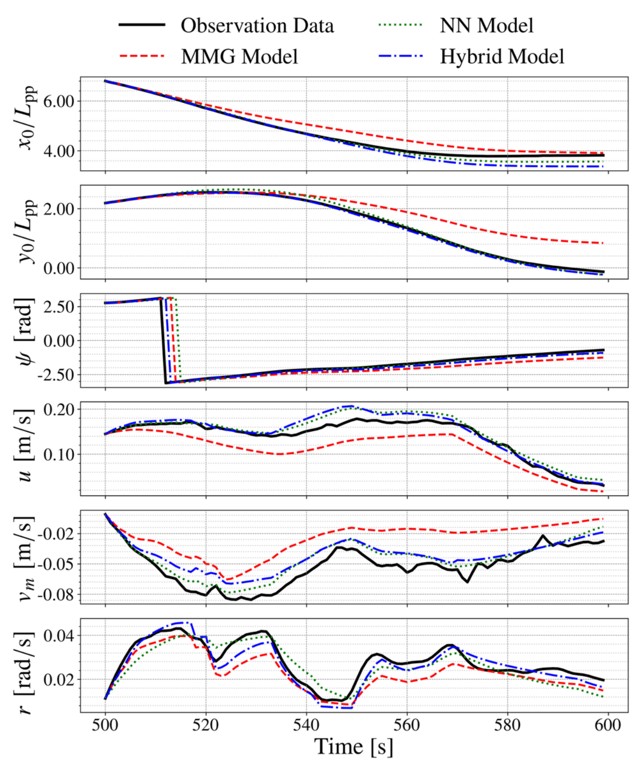

Fig. 9 Time variation of the attitude vector and velocity vector in the Random maneuver (trained with $\mathscr{D}_2$: 500–600 s)

Figures 13 to 16 show the results of OOD detection based on kNN distance in the Hybrid model for the 30-s predictions. The horizontal axis represents the time steps (seconds), and the vertical axis represents the kNN distance ($k = 10$). The blue line indicates the kNN distance at each time step, the green dashed line shows the predefined detection threshold, and the red shaded area represents the periods identified as OOD input. Figures 13 and 14 show the behavior when

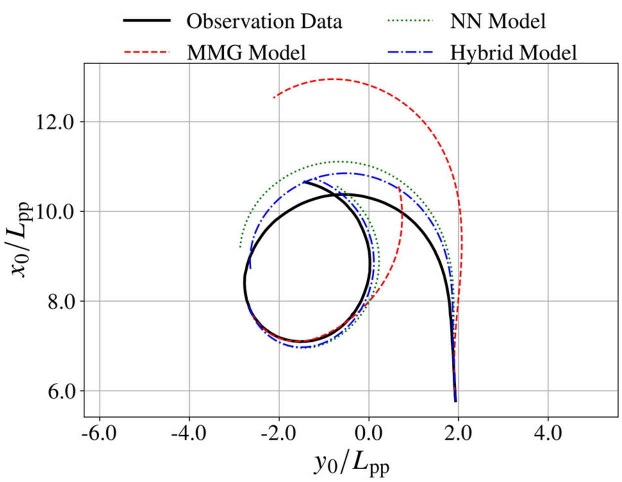

Fig. 10 Trajectory prediction results for each model in the Turning 20° test (trained with $\mathscr{D}_2$)

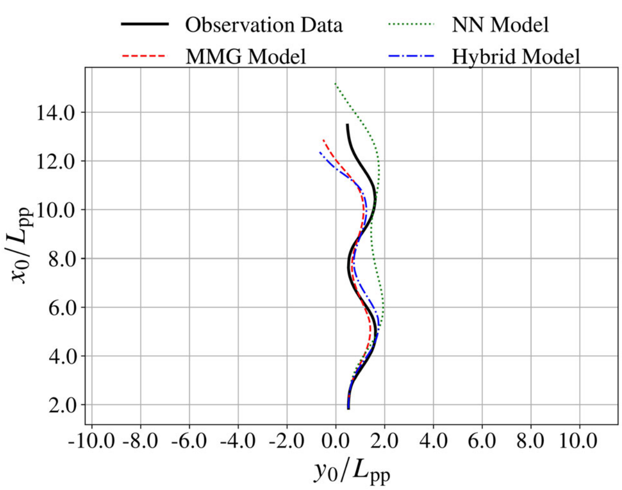

Fig. 11 Trajectory prediction results for each model in the Zigzag 30°/30° test (trained with $\mathscr{D}_2$)

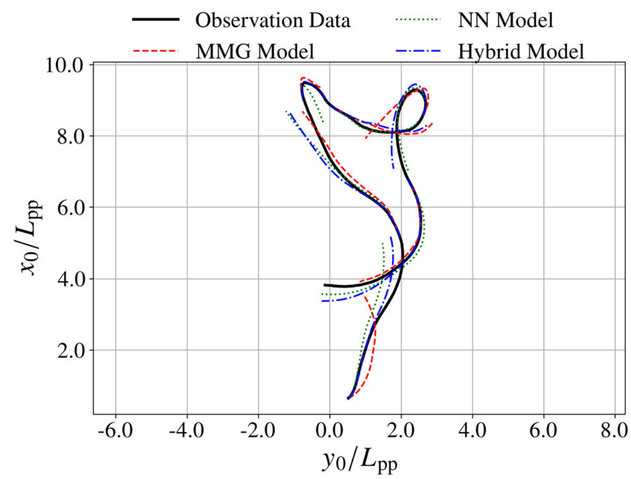

Fig. 12 Trajectory prediction results for each model in the Random maneuver (trained with $\mathscr{D}_2$)

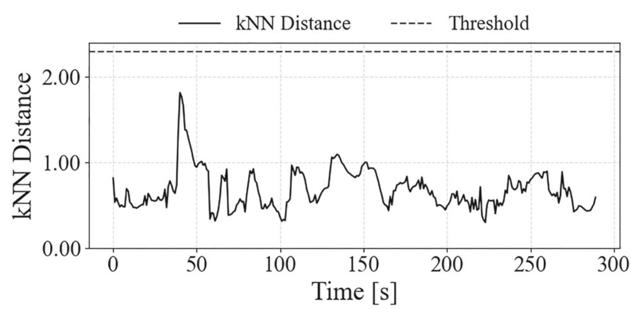

Fig. 13 kNN distance and OOD detection for the Turning 20° test (trained with $\mathscr{D}_1$)

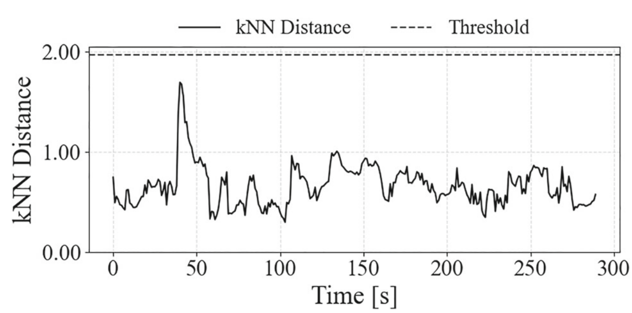

Fig. 14 kNN distance and OOD detection for the Turning 20° test (trained with $\mathscr{D}_2$)

Turning 20° test data are used, showing output results for models trained with $\mathscr{D}_1$ (which includes many Turning tests but excludes Random maneuver data) and $\mathscr{D}_2$ (an extended dataset including Random maneuver). In contrast, Figs. 15 and 16 show the behavior when Random maneuver data are used, where $\mathscr{D}_1$ represents the untrained state, and $\mathscr{D}_2$ represents the trained state. In each figure, you can compare how the inclusion of training data affects the kNN distance and the OOD detection behavior. These results visually confirm the effectiveness of the OOD detection mechanism using the kNN distance to distinguish between data inside and outside the input state distribution. They also allow for a detailed evaluation of the detection system's behavior and overall performance.

Figures 17 and 19 show the time variations of the fluid forces/moments in the Hybrid model, trained using $\mathscr{D}_2$, for the predictions of 100 s of the Turning 20° test and Zigzag 30°/30° test. These figures show the time-series of the components F_{MMG} from the MMG model and the complementary FNN module F_{NN}. Subscripts H, P, R, A, NN represent Hull, Propeller, Rudder, Air, and FNN, respectively. Figures 18 and 20 separately display the components F_{MMG}, F_{NN}, and their combined term $F_{\mathrm{MMG}} + F_{\mathrm{NN}}$. These figures allow for visualizing the contributions of each module and their correction relationships. They also illustrate how the FNN complements the MMG model and how the interaction between the physical and data-driven model evolves over time.

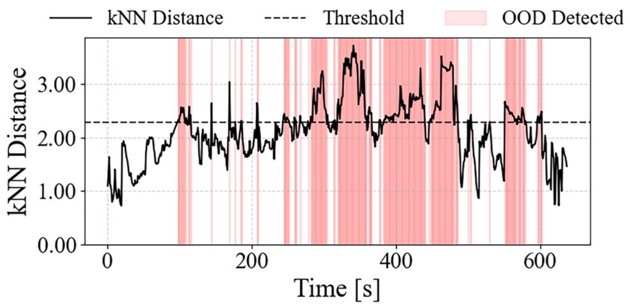

Fig. 15 kNN distance and OOD detection for the Random maneuver (trained with $\mathscr{D}_1$)

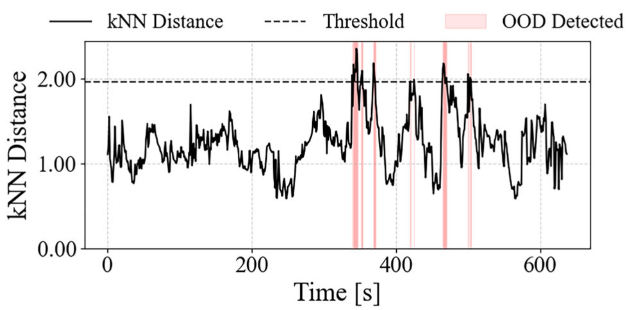

Fig. 16 kNN distance and OOD detection for the Random maneuver (trained with $\mathscr{D}_2$)

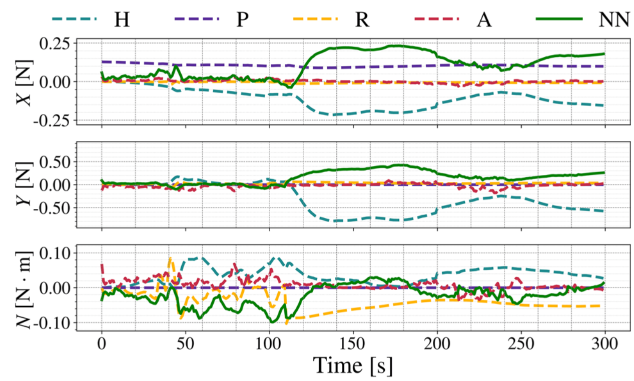

Fig. 17 External forces (by module) in the Turning 20° test (trained with $\mathscr{D}_2$)

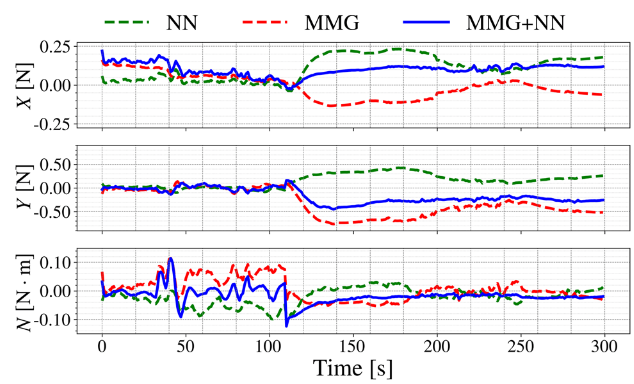

Fig. 18 External forces (by internal model) in the Turning 20° test (trained with $\mathscr{D}_2$)

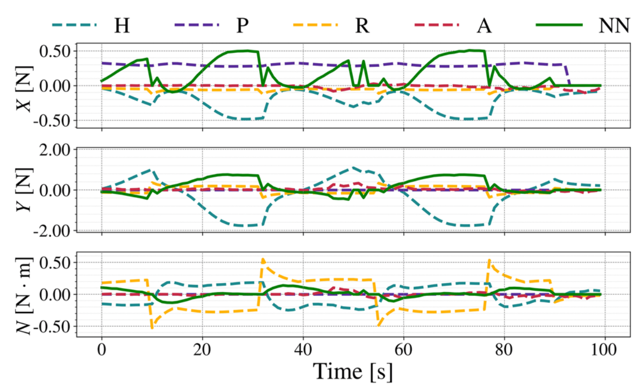

Fig. 19 External forces (by module) in the Zigzag 30°/30° test (trained with $\mathscr{D}_2$)

Figure 21 shows a scatter plot of the relationship between the absolute value of F_{NN} and velocity prediction errors in the Hybrid model trained with $\mathscr{D}_2$. This demonstrates the effectiveness of the FNN as a corrective component within the model.

5 Discussion

Based on the results of numerical experiments, this section discusses the advantages of the proposed MMG+FNN hybrid model over the MMG-only model and the NN-only model. Specifically, we compare these models in terms of prediction accuracy, data dependency, and interpretability.

As shown in Tables 7 and 8, the proposed model achieves an average improvement of approximately 30–50% in prediction accuracy compared to the MMG-only model. The 30-s prediction results show a clear improvement in terms of accuracy. This is especially evident in the Turning 20° test (trained with $\mathscr{D}_2$), where the RMSE values for the u, v_m, and r components are reduced by 50.3%, 70.0%, and 66.7%, respectively. Similarly, a significant reduction in errors was observed in the Turning 35° test, confirming that the FNN complement effectively corrects the MMG model. Furthermore, from the error distribution shown in Figs. 3 to 5, compared to the NN-only model, the quartiles and mean are smaller in many cases. This suggests that the Hybrid model, which is based on the MMG model, can produce more stable predictions with less variation in error.

The 100-s prediction results (from Table 8) also confirm that the Hybrid model shows improved prediction accuracy in many test conditions compared to the MMG-only model. Specifically, in tests involving significant actuator operations such as the Turning 20° and Zigzag 30°/30° the Hybrid model outperforms the NN-only model. This indicates that the Hybrid model remains effective across a broad opera-

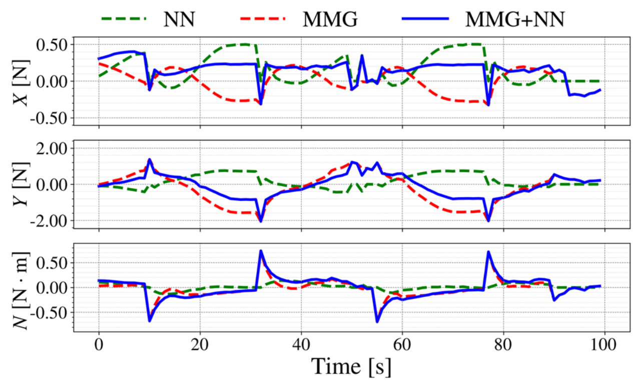

Fig. 20 External forces (by internal model) in the Zigzag 30°/30° test (trained with $\mathscr{D}_2$)

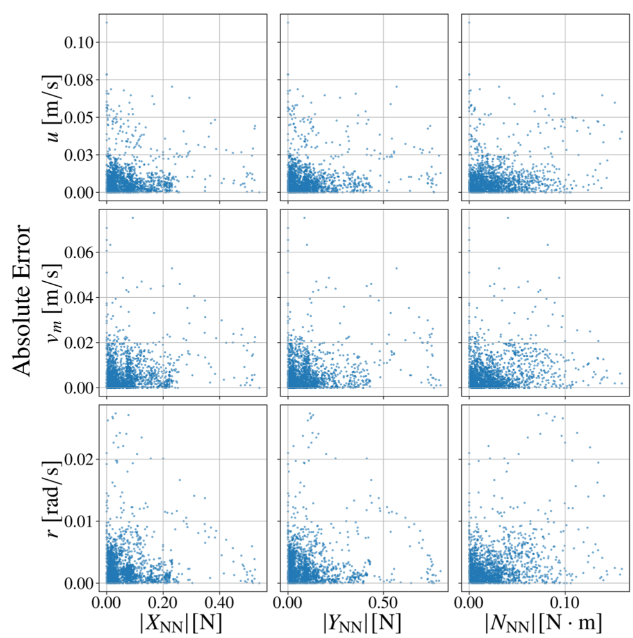

Fig. 21 Relationship between $|F_{\mathrm{NN}}|$ and absolute prediction errors for the velocity vector (trained with $\mathscr{D}_2$)

tional range. Furthermore, Figs. 6 to 9 confirm that, even under complex and unsteady actuator operations, the Hybrid model retains features similar to those of the MMG-only model while locally correcting predictions using the FNN. This demonstrates the integrated strength of the proposed model, in which the MMG model provides a physical foundation and the FNN flexibly applies corrections according to the situation.

In Figs. 10 to 12, it is clear that the Hybrid model generally follows the behavior of the MMG-only model, making stable predictions without significant deviation. Especially in the Zigzag test, the NN-only model shows significant deviation due to insufficient training, whereas the hybrid model maintains a trajectory close to the actual motion through reasonable predictions based on MMG. These figures visually demonstrate the effective integration of data-independent physical knowledge from the MMG model and appropriate corrections by the FNN in the hybrid model. They also show that this integration maintains high stability even in long-term predictions (see Fig. 11).

In the proposed model, the OOD detection based on kNN distance introduces a mechanism to switch to the MMG model in OOD inputs. As a result, as shown in Table 7, for the Random maneuver (trained with $\mathscr{D}_1$), while the NN-only model shows declined accuracy, the Hybrid model maintains good performance. This is because the FNN's ON/OFF system, based on OOD detection, appropriately switches to the only MMG based prediction, as shown in Fig. 15. Additionally, when Random maneuver are included in $\mathscr{D}_2$, as seen in Fig. 16, most of the data are judged as being within the in distribution, and FNN corrections are utilized. This allows for effective correction with minimal data and avoids excessive data dependency compared to the NN-only model.

In Figs. 17 and 19, the external force in the x direction is found to be strongly negatively correlated with the fluid force X_{H} during turnings, with corrections being applied accordingly. Similarly, in the y direction, large fluid forces Y_{H} occur when the rudder angle is switched, and the FNN outputs Y_{NN} to mitigate their impact. For yaw moments, the corrections are smaller during turnings, and Fig. 17 shows that N_{NN} helps reduce the variation in moments. Focusing on the internal models, as shown in Figs. 18 and 20, the FNN generally follows the forces/moments F_{MMG} of the MMG model and contributes to accuracy improvement through local corrections. In the x-direction, there are exceptions where X_{NN} notably exceeds X_{MMG}. These results suggest that while the MMG model cannot fully predict nonlinear maneuvers such as sharp turnings, the FNN learns these features and confidently applies corrections. This correction ultimately leads to a reduction in RMSE, supporting the effectiveness of the proposed model. Moreover, the structure of the Hybrid model is highly interpretable, which makes it straightforward to analyze each internal module. This flexibility allows for easy modification and extension of individual components, enabling clear evaluation of changes in behavior when a component is modified.

As shown in Fig. 21, the scatter plot of the relationship between the absolute values of F_{NN} and prediction errors visually demonstrates that the correction terms effectively reduce errors by preventing unnecessary growth of F_{NN}, particularly in the lower half of the plot.

Summarizing the results: the proposed hybrid model significantly outperforms the MMG-only model in prediction accuracy, and it is more stable (especially in long-term predictions) than the NN-only model. At the same time, it reduces the dependency on training data and maintains high interpretability, especially in comparison to previously reported hybrid methods. Furthermore, the hybrid model's computation time is lower than that of sequence-dependent architectures like RNNs, and the kNN-based OOD detection is computationally lightweight. These factors make our approach suitable for real-time applications.

As a future prospect, the adoption of time-series models (e.g., RNN, LSTM, or Transformer) instead of the FNN could improve performance by capturing temporal dependencies. Additionally, introducing water depth and the distance to the quay as inputs would help compensate for unmodeled dynamics in the MMG framework. Nevertheless, these extensions necessitate a substantially larger dataset than is currently available, making their immediate implementation difficult.

6 Conclusion

In this study, we proposed a hybrid maneuvering model that integrates the MMG model with a force/moment-level FNN module, incorporating output constraints. The goal was to achieve a maneuvering model with high prediction accuracy including low data dependence and maintaining the interpretability of the MMG model.

We specifically designed the hybrid model to maintain the MMG model's modular structure, through the addition of complementary FNN module to account for those force/moment contributions that the MMG formulation cannot fully capture. This design keeps the model highly interpretable, as each module's role is clear and it is easy to understand the contribution of each module to the overall prediction. In addition, to prevent the FNN output from becoming dominant and correcting the motion by ignoring the MMG model's output, a conditional regularization term based on the MMG output is introduced during the FNN learning process. This reflects the design philosophy that the MMG model serves as the main predictor, and corrections are preferentially performed only within the range of the predicted error. Furthermore, the significant deterioration of predictability for out-of-distribution data, often seen in conventional models, is suppressed by OOD detection using kNN distance.

Numerical experiments confirmed that the hybrid model improves predictive accuracy substantially: in middle-term (30-s) predictions, it reduced RMSE by roughly 30–50% relative to the MMG-only model, and in highly nonlinear scenarios, it achieved error reductions up to about 70%. The hybrid model also maintained stable performance over long-term predictions (100 s), whereas the NN-only model's accuracy degraded in such cases. Particularly, in tests involving strong nonlinear maneuvering motions, such as the Turning test and Random maneuver, an RMSE reduction of up to 70% was achieved. Overall, the proposed approach successfully leverages the strengths of both the physical model and data-driven model, resulting in a maneuvering model that offers improved accuracy, reduced data dependency, and maintained high interpretability.

Based on these findings, this study presents an effective approach for the design of maneuvering models that integrate the advantages of physical models and data-driven models. In the future, the model is expected to be applied to real-time autonomous navigation support systems, providing a technical foundation for its practical implementation.

A Appendix

A.1 Structure of the MMG model

In this study, the MMG model is adopted as the physical model. The MMG model mathematically represents ship maneuvering motions by modularizing dynamic elements, such as the hull, propeller, and rudder, considering the fluid forces acting on each [4]. Since the physical contributions of each component are clearly separated, this model is easy to analyze and extend, and has been utilized in many studies aimed at evaluating ship design and maneuvering performance. In this study, the model is constructed based on the MMG model used by Miyauchi et al. (2021) [29], which incorporates wind conditions and propeller reverse rotation [49] into the standard MMG model [4].

A.2 Fluid forces and moments acting on hull

The fluid force terms acting on the hull ($X_{\mathrm{H}}, Y_{\mathrm{H}}, N_{\mathrm{H}}$) are calculated based on Yoshimura's estimation model [50] as follows:

$$
\begin{aligned}
X_{\mathrm{H}} &= \left(\frac{\rho}{2}\right) L_{\mathrm{pp}} d \Big[\{X'_{0(\mathrm{F})} + (X'_{0(\mathrm{A})} - X'_{0(\mathrm{F})})(\beta/\pi)\} u U \\
&\quad + X'_{vr} L_{\mathrm{pp}} \cdot v_m r\Big] \\
Y_{\mathrm{H}} &= \left(\frac{\rho}{2}\right) L_{\mathrm{pp}} d \\
&\quad \Big[Y'_v v_m |u| + Y'_r L_{\mathrm{pp}} \cdot r u \\
&\quad - \left(\frac{C_{\mathrm{D}}}{L_{\mathrm{pp}}}\right) \int_{-L_{\mathrm{pp}}/2}^{L_{\mathrm{pp}}/2} |v_m + C_{rY} r x| (v_m + C_{rY} r x)\, dx\Big] \\
N_{\mathrm{H}} &= \left(\frac{\rho}{2}\right) L_{\mathrm{pp}}^2 d \\
&\quad \Big[N'_v v_m u + N'_r L_{\mathrm{pp}} \cdot r|u| \\
&\quad - \left(\frac{C_{\mathrm{D}}}{L_{\mathrm{pp}}^2}\right) \int_{-L_{\mathrm{pp}}/2}^{L_{\mathrm{pp}}/2} |v_m + C_{rN} r x| (v_m + C_{rN} r x)\, x\, dx\Big],
\end{aligned}
\tag{26}
$$

where ρ is the water density, L_{pp} is the length between perpendiculars, d is the draft, $X'_{0(\mathrm{F})}$ and $X'_{0(\mathrm{A})}$ are the resistance coefficients during forward and backward motion, C_{D} is the cross-flow drag coefficient, C_{rY} and C_{rN} are correction coefficients for force and yaw moment, and $X'_{0(\mathrm{F})}, Y'_v, Y'_r, N'_v, N'_r$ represent dimensionless hydrodynamic coefficients. In the following descriptions, symbols with a prime mark indicate dimensionless values.

A.3 Fluid forces and moments due to propeller

Since the standard MMG model [4] assumes only forward motion ($u > 0, n_\mathrm{p} > 0$), additional sub-models corresponding to the propeller operating conditions are used to calculate propeller forces. These are classified into the following four quadrants: 1st quadrant ($u \geq 0, n_\mathrm{p} > 0$), 2nd quadrant ($u < 0, n_\mathrm{p} > 0$), 3rd quadrant ($u \geq 0, n_\mathrm{p} < 0$), and 4th quadrant ($u < 0, n_\mathrm{p} < 0$).

In the 1st and 2nd quadrants, the propeller thrust X_P is calculated by the standard MMG model as follows:

$$X_\mathrm{P} = \rho n_\mathrm{p}^2 D_\mathrm{p}^4 (1 - t_\mathrm{p}) K_\mathrm{T}. \tag{27}$$

Here, the thrust coefficient K_T is expressed as a polynomial of the advance coefficient $J_\mathrm{p} = (1 - w_\mathrm{p})u/(n_\mathrm{p} D_\mathrm{p})$. The effective propeller wake fraction w_p is calculated as follows [51]:

$$1 - w_\mathrm{p} = 1 - w_{\mathrm{p}0} + \tau |v'_m + x'_\mathrm{p} r'| + C'_\mathrm{p}(v'_m + x'_\mathrm{p} r')^2, \tag{28}$$

where $w_{\mathrm{p}0}$ is the wake fraction at $v_m = r = 0$, and $\tau, C'_\mathrm{p}, x'_\mathrm{p}$ are coefficients determined based on experiments. The thrust deduction factor t_p and wake fraction w_p vary depending on the propeller operating conditions, but in this study, they are modeled simply as follows [51, 52]:

$$t_\mathrm{p} = 0 \quad (\text{for } n_\mathrm{p} < 0) \tag{29}$$
$$w_\mathrm{p} = 0 \quad (\text{for } u < 0). \tag{30}$$

The propeller-induced lateral force Y_P and yaw moment N_P in the 1st and 2nd quadrants ($n_\mathrm{p} \geq 0$) are usually ignored in the standard MMG model, but in this model, they are calculated using the following polynomials based on training ship experiments [53]:

$$Y_\mathrm{P} = \begin{cases} 0 & (u \geq 0) \\ \frac{1}{2}\rho L_\mathrm{pp}^2 d (n_\mathrm{p} P)^2 (A_6 J_\mathrm{s}^2 + A_7 J_\mathrm{s} + A_8) & (u < 0) \end{cases} \tag{31}$$
$$N_\mathrm{P} = \begin{cases} 0 & (u \geq 0) \\ \frac{1}{2}\rho L_\mathrm{pp}^2 d (n_\mathrm{p} P)^2 (B_6 J_\mathrm{s}^2 + B_7 J_\mathrm{s} + B_8) & (u < 0), \end{cases} \tag{32}$$

where P is the propeller pitch, $J_\mathrm{s} = u/(n_\mathrm{p} D_\mathrm{p})$, and $A_6 \sim A_8$ and $B_6 \sim B_8$ are polynomial coefficients.

For propeller reverse rotation conditions (3rd and 4th quadrants), polynomial expressions based on CMT [54] are used

$$X_\mathrm{P} = \rho n_\mathrm{p}^2 D_\mathrm{p}^4 \begin{cases} C_6 + C_7 J_\mathrm{s} & (J_\mathrm{s} \geq C_{10}) \\ C_3 & (J_\mathrm{s} < C_{10}) \end{cases} \tag{33}$$

$$Y_\mathrm{P} = \frac{1}{2}\rho L_\mathrm{pp} d (n_\mathrm{p} D_\mathrm{p})^2 \begin{cases} A_1 + A_2 J_\mathrm{s} & (-0.35 \leq J_\mathrm{s} \leq -0.06) \\ A_3 + A_4 J_\mathrm{s} & (J_\mathrm{s} < -0.35) \\ A_5 & (-0.06 < J_\mathrm{s}) \end{cases} \tag{34}$$

$$N_\mathrm{P} = \frac{1}{2}\rho L_\mathrm{pp}^2 d (n_\mathrm{p} D_\mathrm{p})^2 \begin{cases} B_1 + B_2 J_\mathrm{s} & (-0.35 \leq J_\mathrm{s} \leq -0.06) \\ B_3 + B_4 J_\mathrm{s} & (J_\mathrm{s} < -0.35) \\ B_5 & (-0.06 < J_\mathrm{s}), \end{cases} \tag{35}$$

where $A_1 \sim A_5, B_1 \sim B_5, C_3, C_6, C_7, C_{10}$ are polynomial coefficients.

A.4 Fluid forces and moments due to rudder

In the standard MMG model [4], the forces and moments generated by the rudder are expressed as follows:

$$X_\mathrm{R} = -(1 - t_\mathrm{R}) F_\mathrm{N} \sin\delta \tag{36}$$
$$Y_\mathrm{R} = -(1 + a_\mathrm{H}) F_\mathrm{N} \cos\delta \tag{37}$$
$$N_\mathrm{R} = -(x_\mathrm{R} + a_\mathrm{H} x_\mathrm{H}) F_\mathrm{N} \cos\delta, \tag{38}$$

where t_R, a_H, and x_H are interaction coefficients of the rudder on the hull and propeller. Also, x_R is the x-coordinate of the rudder position. The rudder normal force F_N is defined by the following equation:

$$F_\mathrm{N} = \frac{1}{2}\rho A_\mathrm{R} U_\mathrm{R}^2 f_\alpha \sin\alpha_\mathrm{R}. \tag{39}$$

In the equation, the rudder inflow velocity U_R and the effective rudder inflow angle α_R are expressed using the longitudinal and lateral inflow velocities $u_\mathrm{R}, v_\mathrm{R}$ as follows:

$$U_\mathrm{R} = \sqrt{u_\mathrm{R}^2 + v_\mathrm{R}^2} \tag{40}$$
$$\alpha_\mathrm{R} = \delta - \mathrm{atan2}\left(\frac{v_\mathrm{R}}{u_\mathrm{R}}\right). \tag{41}$$

The normal force gradient is calculated based on Fujii's formula [55]

$$f_\alpha = \frac{6.13\lambda}{2.25 + \lambda}. \tag{42}$$

In the model by Miyauchi et al. [29], to apply to berthing maneuvers, the standard MMG model was extended by introducing $\mathrm{atan2}(y, x)$ which returns values in the range $(-\pi, \pi]$. The lateral inflow velocity v_R is expressed using the flow rectification coefficient γ and the experimental constant l_R as follows:

$$v_R = \begin{cases} -\gamma_p(v_m + l_R r) & (v_m + x_R r \geq 0) \\ -\gamma_N(v_m + l_R r) & (v_m + x_R r < 0). \end{cases} \tag{43}$$

The longitudinal inflow velocity u_R is strongly influenced by the ship's motion direction and the propeller-induced flow. During forward rotation ($n_p \geq 0$), a correction model for the low-speed range [50] is used

$$u_R = \epsilon \sqrt{\eta \left\{ u_p + \frac{k_x}{\epsilon} \left(\sqrt{u_p^2 + \frac{8K_T(n_p D_p)^2}{\pi}} - u_p \right) \right\}^2 + (1-\eta)u_P^2}, \tag{44}$$

where $u_P = (1 - w_P)u$, $\eta = D_p/H_R$ (H_R is the rudder height), ϵ is the wake ratio, and k_x is an empirical coefficient.

For reverse rotation ($n_p < 0$) and forward motion ($u \geq 0$), Kitagawa's model [56] is applied

$$u_R = \mathrm{sgn}(u_{Rsq}) \cdot \sqrt{|u_{Rsq}|}, \tag{45}$$

where

$$u_{Rsq} = \eta \cdot \mathrm{sgn}(u_{RPR1}) \cdot u_{RPR1}^2 + (1-\eta)\,\mathrm{sgn}(u_{RPR2}) \cdot u_{RPR2}^2 + C_{PR} \cdot u \tag{46}$$

$$u_{RPR1} = u\epsilon(1 - w_p) + n_p D_p k_{xPR} \sqrt{8|K_T|/\pi} \tag{47}$$

$$u_{RPR2} = u\epsilon(1 - w_p), \tag{48}$$

k_{xPR} is the velocity increase coefficient, and C_{PR} is the correction coefficient during propeller reverse rotation.

For reverse rotation ($n_p < 0$) and backward motion ($u < 0$), it is assumed that the inflow velocity is equal to the ship's motion ($u_R = u$) [51].

A.5 Fluid forces and moments due to wind

Regarding the external forces induced by wind turbulence, Fujiwara's regression formula [57] is used to estimate the wind pressure coefficients. The wind forces and moment are expressed as follows:

$$X_A = \frac{1}{2}\rho_A U_A^2 A_T \cdot C_X \tag{49}$$

$$Y_A = \frac{1}{2}\rho_A U_A^2 A_L \cdot C_Y \tag{50}$$

$$N_A = \frac{1}{2}\rho_A U_A^2 A_L L_{OA} \cdot C_N. \tag{51}$$

The wind pressure coefficients C_X, C_Y, C_N in the equations are as follows:

$$C_X = X_0 + X_1 \cos(2\pi - \gamma_A) + X_3 \cos 3(2\pi - \gamma_A) + X_5 \cos 5(2\pi - \gamma_A) \tag{52}$$

$$C_Y = Y_1 \sin(2\pi - \gamma_A) + Y_3 \sin 3(2\pi - \gamma_A) + Y_5 \sin 5(2\pi - \gamma_A) \tag{53}$$

$$C_N = N_1 \sin(2\pi - \gamma_A) + N_2 \sin 2(2\pi - \gamma_A) + N_3 \sin 3(2\pi - \gamma_A), \tag{54}$$

where ρ_A is the air density, U_A is the relative wind speed, γ_A is the relative wind direction angle, A_T is the transverse projected area, A_L is the lateral projected area, and L_{OA} is the length over all. The coefficients X_i, Y_i, N_i are wind pressure coefficients derived by regression formulas [57] using the ship's geometric parameters as explanatory variables, based on wind tunnel experiment data using numerous scale models.

A.6 MMG model parameters

Table 9 shows the parameters of the MMG model used in this study. These parameters are adopted from the EFD (Experimental Fluid Dynamics) model parameters used by Miyauchi et al. (2021) [29] (Table 8 in their paper).

Table 9 MMG model parameters

Parameter	Value	Parameter	Value	Parameter	Value
$m_x/(0.5\rho L_{pp}^2 d)$	0.0467	A_3	−0.00493	a_H	0.393
$m_y/(0.5\rho L_{pp}^2 d)$	0.2579	A_4	−0.00587	x_H/L_{pp}	−0.45
$(I_{zz}+J_{zz})/(0.5\rho L_{pp}^4 d)$	0.0286	A_5	−0.000558	l_R/L_{pp}	−1.08
X'_{vr}	0.4535	A_6	0.00508	k_x	0.288
Y'_v	−0.3728	A_7	0.0017	k_{xPR}	0.144
Y'_r	0.116	A_8	0.000916	ε	1.42
N'_v	−0.1458	B_1	−0.00317	C_{PR}	−0.176
N'_r	−0.04849	B_2	−0.000558	γ_N	0.4406
C_D	0.9665	B_3	0.00916	γ_p	0.3506
C_{rY}	2.217	B_4	0.00233	X_0	−0.010466205
C_{rN}	0.8097	B_5	0.000225	X_1	−0.821991956
$X'_{0(A)}$	−0.03189	B_6	−0.00216	X_3	0.060273613
t_p	0.22	B_7	$6.27E-05$	X_5	0.044558221
w_{p0}	0.614	B_8	$6.27E-05$	Y_1	1.058172083
τ	0.871	C_3	−0.251	Y_3	0.002321463
C'_p	−0.359	C_6	−0.175	Y_5	−0.015293452
x'_p	0.517	C_7	0.33	N_1	−0.030727798
A_1	$-7.90E-05$	C_{10}	−0.233	N_2	0.10411637
A_2	0.00799	t_R	0.19	N_3	−0.000880301

Acknowledgements This work was supported by The Shipbuilders' Association of Japan (SAJ), The Organization for Promoting Research and Development in Shipbuilding (REDAS), under Grant No. REDAS25-5(9). The authors would like to express our sincere gratitude to the organization for their support.

Funding Open Access funding provided by The University of Osaka

Data Availability The datasets used and/or analyzed during the current study are available in the GitHub repository at https://github.com/NAOE-5thLab/FRT-DS-ESSO.

References

1. NationalMaritimeResearchInstitute(NMRI), Development of autonomous navigation technology - meguri2040 unmanned navigation demonstration project, NMRI News (119), 2 (2021)
2. Shimizu E (2021) Recent trends and challenges for the practical application of autonomous ships. ClassNK Technical Research Institute Report 8:1
3. Abudu R, Bridgelall R (2024) Autonomous ships: a thematic review. World 5(2):347
4. Hiroshi Y, Yuichi Y (2014) Introduction of mmg standard method for ship maneuvering predictions. J Mar Sci Technol 20(1):37
5. Abkowitz M (1964) Lectures on ship hydrodynamics: Steering and manoeuvrability
6. Abkowitz M (1980) Measurement of hydrodynamic characteristics from ship maneuvering trials by system identification. Trans Soc Naval Architects Marine Eng 88:283
7. Rajesh G, Bhattacharyya S (2008) System identification for nonlinear maneuvering of large tankers using artificial neural networks. Appl Ocean Res 30(3):144
8. Zhang X, Zou Z (2013) Neural network identification of marine ship dynamics. J Mar Sci Technol 18:1
9. Liu Y, Wang H, Zhang X (2024) Research on lstm-based maneuvering motion prediction for usvs. J Marine Sci Eng 12(9):1661
10. Zhang Y, Soares C (2019) Real-time parameter estimation of a nonlinear vessel steering model using a support vector machine. J Offshore Mech Arct Eng 141(6):061606
11. Jiang L, Shang X, Qi X, Ouyang Z, Zhang Z (2025) Variance-tuned ensemble gaussian process regression for learning and prediction of ship maneuvering motion. Ocean Eng 285:116234
12. Mei B, Sun L, Shi G, Liu X (2019) Ship maneuvering prediction using grey box framework via adaptive rm-svm with minor rudder. Polish Maritime Res 26(3):115
13. Skulstad R, Li G, Fossen T, Vik B, Zhang H (2020) A hybrid approach to motion prediction for ship docking-integration of a neural network model into the ship dynamic model. IEEE Trans Instrum Meas 70:1
14. Wang T, Li G, Hatledal L, Skulstad R, Æsøy V, Zhang H (2022) Incorporating approximate dynamics into data-driven calibrator: A

representative model for ship maneuvering prediction. IEEE Trans Industr Inf 18(3):1781
15. Kanazawa M, Wang T, Skulstad R, Li G, Zhang H (2022) Knowledge and data in cooperative modeling: Case studies on ship trajectory prediction. Ocean Eng 266:112998
16. Li S, Xiang J, Xia G (2024) Enhancing ship maneuvering models through hybrid integration of data-driven velocity calibrator for improved performance, OCEANS 2024 - Singapore pp. 1–6
17. Gomi S, Mitsuyuki T (2025) A study in identifying maneuvering models using a physics-informed neural network with the mmg model. Proc Jpn Soc Naval Architects Ocean Eng 40:353
18. Kanazawa M, Skulstad R, Wang T, Li G, Hatledal L, Zhang H (2022) A physics-data co-operative ship dynamic model for a docking operation. IEEE Sens J 22(11):11173
19. Kuniji K, Hiroyoshi H, Yasuhisa H, Eijiro F (1984) On a mathematical model of maneuvering motions of ships in low speeds. J Soc Naval Architects Japan 1984(155):132
20. Yamato H (1990) Automatic berthing by the neural controller. in Proceedings of the Ninth Ship Control Systems Symposium, pp. 3183–3201
21. Hasegawa K, Kitera K (1993) Automatic berthing control system using network and knowledge-base. J Kansai Soc Naval Architects Jpn 220:135
22. Zhang X, Zou Z (2013) Black-box modeling of ship manoeuvring motion based on feed-forward neural network with chebyshev orthogonal basis function. J Mar Sci Technol 18:1
23. Wakita K, Maki A, Umeda N, Miyauchi Y, Shimoji T, Rachman D, Akimoto Y (2022) On neural network identification for low-speed ship maneuvering model. J Mar Sci Technol 27:772
24. Mei B, Sun L, Shi G (2019) White-black-box hybrid model identification based on rm-rf for ship maneuvering. IEEE Access 7:57691
25. Skulstad R, Li G, Fossen T, Wang T, Zhang H (2021) A co-operative hybrid model for ship motion prediction. MIC Model Identif Control 42:17
26. Zhu M, Han P, Wang C, Skulstad R, Zhang H, Li G (2025) An ais data-driven hybrid approach to ship trajectory prediction. IEEE Trans Syst Man Cybern Syst 55(1):96
27. Xu P, Han C, Cheng H, Cheng C, Ge T (2022) A physics-informed neural network for the prediction of unmanned surface vehicle dynamics. J Marine Sci Eng 10(2):148
28. Nielsen R, Papageorgiou D, Nalpantidis L, Jensen B, Blanke M (2022) Machine learning enhancement of manoeuvring prediction for ship digital twin using full-scale recordings. Ocean Eng 257:111579
29. Miyauchi Y, Maki A, Umeda N, Rachman D, Akimoto Y (2022) System parameter exploration of ship maneuvering model for automatic docking/berthing using cma-es. J Marine Sci Technol 27:1065
30. Rumelhart D, Hinton G, Williams R (1986) Learning internal representations by error propagation. Parallel Distributed Processing Explorations in the Microstructure of Cognition 1:318
31. Wang Z, Cheng J, Xu L, Hao L, Peng Y (2024) Hybrid physics-ml modeling for marine vehicle maneuvering motions in the presence of environmental disturbances, arXiv preprint arXiv:2411.13908. Accessed 6 Mar 2026
32. Kingma D, Ba J (2014) Adam: A method for stochastic optimization, Arxiv:1412.6980 Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015
33. Narendra K, Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. IEEE Trans Neu Netw 1(1):4
34. Benechehab A, Paolo G, Thomas A, Filippone M, Kégl B (2023) Multi-timestep models for model-based reinforcement learning. In: Proceedings of the 40th International Conference on Machine Learning, PMLR 202:1963–1983
35. Gibson J, Vlahov B, Fan D, Spieler P, Pastor D, Agha-mohammadi A, Theodorou EA (2023) A multi-step dynamics modeling framework for autonomous driving in multiple environments. arXiv preprint arXiv:2305.02241. Accessed 6 March 2026
36. Yin Y, Le Guen V, Dona J, de Bézenac E, Ayed I, Thome N, Gallinari P (2021) Augmenting physical models with deep networks for complex dynamics forecasting. Journal of Statistical Mechanics: Theory and Experiment 2021(12):124012. https://doi.org/10.1088/1742-5468/ac3ae5. Accessed 6 March 2026
37. Takeishi N, Kalousis A (2023) Deep grey-box modeling with adaptive data-driven models toward trustworthy estimation of theory-driven models. in Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 206, vol. 206, pp. 4089–4100
38. Duan W, Cao G, Huang L, Duan S (2019) Ship maneuvering prediction based on CFD method. in Conference proceedings, the Japan Society of Naval Architects and Ocean Engineers, vol. 28 (The Japan Society of Naval Architects and Ocean Engineers), vol. 28
39. Sun Y, Guo C, Li Y (2022) Out-of-distribution detection with deep nearest neighbors. in Proceedings of the 39th International Conference on Machine Learning (ICML), vol. 162 , pp. 20887–20900
40. Gal Y, Ghahramani Z (2016) Dropout as a bayesian approximation: Representing model uncertainty in deep learning. https://arxiv.org/abs/1506.02142. Accessed 6 Mar 2026
41. Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. https://arxiv.org/abs/1612.01474. Accessed 6 Mar 2026
42. An J, Cho S (2015) Variational autoencoder based anomaly detection using reconstruction probability. Technical Report, Seoul National University
43. Lee K, Lee K, Lee H, Shin J (2018) A simple unified framework for detecting out-of-distribution samples and adversarial attacks. https://arxiv.org/abs/1807.03888. Accessed 6 Mar 2026
44. Gianluca G, Marc M (2025) Leveraging intermediate representations for better out-of-distribution detection, Proceedings of the 28th Computer Vision Winter Workshop CVWW pp. 53–61
45. Nonaka K, Mori M, Matsumoto N (1972) Estimating Maneuvering Coefficients from Free-Running Trials. Tech. rep., 20th National Maritime Research Institute Report
46. Yoon HK, Rhee KP (2003) Identification of hydrodynamic coefficients in ship maneuvering equations of motion by estimation-before-modeling technique. Ocean Eng 30(18):2379
47. Belkin M, Hsu D, Xu J (2020) Two models of double descent for weak features. SIAM J Math Data Sci 2(4):1167–1180
48. Prechelt L (2012) Early stopping – but when? Springer, Berlin Heidelberg, pp 53–67
49. Hachii T (2004) The prediction of manoeuvring motion on ships with low speed using standard mmg model. Master's thesis, Osaka University (in Japanese)
50. Yoshimura AIY, Nakao I (2009) Model for ocean and harbour manoeuvring, MARSIM2009 pp. 116–233
51. Kobayashi H, Blok J, Barr R, Kim J, Nowick J (2002) The specialist committee on esso osaka final report and recommendations to the 23rd ittc, Proc. of 23rd International Towing Tank Conference II, 581
52. Yasukawa H, Kose K (2003) Simulation of stopping maneuver of a tanker in wind and waves. Trans West-Jpn Soc Naval Architects 106:57
53. Ueno M, Nimura T, Miyazaki H, Fujiwara T, Nonaka K, Yabuki H (2001) Model experiment and sea trial for investigating manoeuvrability of a training ship. J Soc Naval Architects Jpn 2001:71
54. Hasegawa K, Fukutomi T (1994) On harbour manoeuvring and neural control system for berthing with tug operation, Proc. of 3rd International Conference Manoeuvring and Control of Marine Craft (MCMC'94) pp. 197–210

55. Fujii H, Tuda T (1961) Experimental researches on rudder performance. (2). j zosen kiokai 1961(110):31, J Jpn Inst Navig 91, 263
56. M.H. Kitagawa Y, Tsukada Y (2015) A study on mathematical models of propeller and rudder under maneuvering with propeller reverse rotation, Conf Proc Jpn Soc Naval Architect Ocean Eng 117(20)–120
57. N.T. Fujiwara T, Ueno M (1998) Estimation of wind forces and moments acting on ships. J Soc Naval Architect Jpn 183: 77–90.https://doi.org/10.2534/jjasnaoe1968.1998.77