IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 20, 2019, accepted April 23, 2019, date of publication April 30, 2019, date of current version May 14, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2914120*

# White-Black-Box Hybrid Model Identification Based on RM-RF for Ship Maneuvering

**BIN MEI[1,2], LICHENG SUN[1], AND GUOYOU SHI[1,2]**
[1]Navigation College, Dalian Maritime University, Dalian 116026, China
[2]Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Dalian 116026, China

Corresponding author: Guoyou Shi (sgydmu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51579025, in part by the Liaoning National Natural Science Foundation under Grant 20170540090, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132018306.

**ABSTRACT** A new system identification scheme based on model reference and random forest (RM-RF), was used to model ship maneuvering. First, the scheme establishes the relationship between a ship and the RM using the similarity rule. Second, a suitable RM was selected from public ship maneuvering models to avoid the tuning process in the empirical Maneuvering Modeling Group (MMG), model. Third, RF creates a relationship to map accelerations between the ship and the RM. Finally, the study case was implemented with fewer free-running model test data. The results show the feasibility of the identification modeling scheme and validated the generalizability of the proposed approach.

**INDEX TERMS** Modeling, machine learning algorithms, hybrid intelligent systems, motion estimation, system identification.

## I. INTRODUCTION

With the increasing use of autonomous systems in the maritime sector, ship automation and intellectualization have advanced rapidly. A regulatory scoping exercise for the use of maritime autonomous surface ships (MASS) has been proposed with a target completion date of 2020 [1]. This regulatory scoping exercise is supported by the Maritime Safety Committee established by the International Maritime Organization (IMO) [2], [3]. Generally, autonomous navigation and collision avoidance systems are technological aspects of MASS [4]. According to the Maritime Unmanned Navigation through Intelligence in Networks (MUNIN) and machine-executable collision regulations for marine autonomous systems (MAXCMAS) projects, a ship maneuvering model is a crucial technology from the viewpoint of research on autonomous navigation and collision systems [5]–[7]. Furthermore, the modeling of ship maneuvering is an essential topic in the domains of controller design and maneuvering prediction [8], e.g. model predictive control and marine simulator.

A variety of different methods exist for modeling ship maneuvering. The reports in 2008 and 2017 International Towing Tank Conference (ITTC) provide broad overviews of many extant methods [9], [10]. Generally, ship maneuvering modeling methods include NO simulation, computational fluid dynamics (CFD)-based maneuvering simulation, and system-based simulation. The NO simulation method uses a statistical database of the trajectories of independent ship variables. The CFD method employs viscous CFD and potential flow theory. The system-based simulation methods consist of a mathematical structural model and parameters that can be ascertained using system identification, CFD, captive model test. To obtain various loading force and determine hydrodynamic coefficients, the planar motion mechanism (PMM) and the circular motion tests are repeated in the complicated and expensive facilities [11], [12]. Based on theoretical calculations, the CFD method can interpret physical phenomena from the flow field perspective, but requires considerable computer resources and artificial experience to mesh a grid and set turbulence model; what's more model test results are required for validation. According to an overview of the maneuvering model methods, although system identification is not as accurate as the captive test, free-running test, or CFD [9], it is advantageous in terms of its significantly lower cost and effort. Therefore, improving the accuracy of system identification based on the free-running test is crucial.

The associate editor coordinating the review of this manuscript and approving it for publication was Roberto Nardone.

System identification modeling is a modeling method involving regression analysis test data [14]. During sailing, ship motion status changes constantly, for instance, loading status including goods, oil, and water; waterway conditions such as ship draught and water depth; ship speed, including sea speed and port speed; and environmental disturbances, such as wind and waves [15], [16]. Note that online system identification algorithms for accurate modeling can capture the changes in ship motion status [17]. In control and prediction, it is expected that ship maneuvering data can be used for identification by installing sensors [18].

### A. SURVEY OF SHIP MANEUVERING IDENTIFICATION MODELING

Various system identification methods have been used for modeling ship maneuvering motion. Many important research achievements have been made in the field of identification modeling, the development routes are divided into white-box, grey-box, and black-box modeling.

Approaches to white-box identification include the extended Kalman filter (EKF), genetic algorithm (GA), linear forward neural network (NN), least squares–support vector machines (LS-SVMs), and maximum likelihood (ML). Abkowitz introduced the EKF to evaluate ship maneuvering analytically and proved the effectiveness of system identification [19]. However, in this method, the analysis of process and observation noise was complicated and required convoluted operations. Luo and Zou were the first to apply the LS-SVM method [20]. However, LS-SVMs are sensitive to sample size and cannot regress precisely. Kallstrom *et al*. employed the ML method to identify the linear state space equation coefficient and used a pseudo-random binary sequence (PRBS) signal to input the test rudder angle [21]. However, PRBS signals may not be suitable for ships with a large metacentric height, such as container carriers and ro–ro ships. Yoon and Rhee applied the EKF and improved the modified Bryson–Frazier smoother for data preprocessing, these used ridge regression to estimate model parameters, and improved the PRBS rudder angle signal. In addition, Yoon and Rhee proposed the D-type ship test scheme [22]. Luo and Zhang presented a two-layer forward NN to identify the parameters of a linear model, and the neural network was similar to nonlinear approximation [23]. Recently, Luo and Li reduced parameter drift by means of SVM sample reconstruction [24]. Zhu*et al*. optimized an SVM using the artificial bee colony algorithm to ascertain the parameters of a simple maneuvering model [25] but did not compare zigzag velocities. Sutulo and Guedes Soares presented an offline identification algorithm based on the learning metrics and optimized the learning process by using a GA [26]. However, the artificially measured noise value obtained was different from the actual environmental disturbances. Moreover, when following Sutulo and Guedes Soares approach, GA can only search for the parameters in an interval.

With regards to black-box identification, many artificial intelligence algorithms have been introduced. Haddara collected several sets of simulation data and effectively modeled the ship hydrodynamic polynomials by using a NN [27], but the process of training the network was time consuming. Wang and Han presented a fuzzy NN network [28], but from the performance viewpoint, the signal may be unsuitable for the steering gear in a real ship. Moreover, Wang's simulation scheme many not be implementable on actual equipment. Moreira and Soares proposed a maneuvering model based on a recursive neural network (RNN) [29], but the RNN needed a considerable number of training tests for turning circle and zigzag motion. Oskin presented an RNN create a response model that neglects ship speed predictions [30]. Wang *et al*. utilized SVM regression to fit the relationship between ship hydrodynamics and ship motion state [31]. Bai*et al*. used multi-innovation gradient iterative locally weighted learning as a black box to predict ship maneuvers [32], but the rudder angle and maneuvering scheme were severe from the viewpoint of ship handling, and the model is limited to deal with simulator only.

As for the grey-box model, Blanke and Knudsen employed Hessian matrix decomposition to show linear dependencies and the influence of variables [33]. In the case of a complex model, importance makes it easy to list out well-identifiable parameters. However, there are no examples and results in Blanke's reference. Wang*et al*. introduced a grey-model-based SVM [34], where third-order Taylor expansion was employed in place of the MMG or the Abkowitz model structure, and the SVM black box approximated the Taylor expansion. Wang's grey model ignored the hydrodynamic coefficients, and the computation time of the model was long.

Apart from white-box, grey-box, and black-box identification, other relevant models are available. Hess used full-scale trial data and constructed an RNN model using difference inputs and minimizing error of training datasets as the objective function [35]. However, its major drawbacks were high-order networks, for example, with a large number of iterations (100,000 order of magnitude), and complex validation. Furthermore, the RNN inputs consist of component force modules. The component force modules are detailed in the MMG model or the modular mathematical (MM) model. These modules require some bias altering and complicated debugging work to achieve the desired result. Soon afterwards Chiu presented reproduced results similar to those in Hess's paper [36].

### B. ANALYSIS OF REVIEWS

The following analysis and conclusions can be drawn from the cited publications and mentioned reviews:

1) The use of simulation data for system identification is not feasible. Because the data generated by means of simulation is ideal and simple, as explained in the following sentences. Firstly, when the simulation test data do not contain noise, data quality is better, which makes identification easy. In test data, by contrast, measurement errors

and environmental disturb exist. Secondly, the amplitude and frequency of the excitation signal, for instance, sinusoidal rudder angle signal, approximate pulse signal [28], PRBS signal [21], or D-type signal [22]. However, it is difficult to apply a sinusoidal rudder angle in a practical steering control system. Moreover, considering the dimensions of a basin, the PRBS and D-type signals are unsuitable for conducting tests. Moreover, in sea trials, the maximum allowable rudder angle or $+20°/-20°$ rudder angle is proposed in standards for ship maneuverability (IMO) [37]. It is not recommended to use large reverse rudder angles because of concerns about cargo shifting and steering gear conditions, especially when a ship is full loading. It can be concluded that in a free running model test, execution of the excitation signal is difficult. Therefore, although identification is easier when using simulation than when using test data, it is difficult to execute rudder angle signal in real conditions.

2) The ship is not a fully loaded system in sea trials. First, the ship loading condition is not full load but ballast. There is a certain difference between full load and ballast; therefore, it is necessary to make corrections from non-standard trial conditions [38], [39]. The test speed in the standards is a speed at further 90% of the ship speed corresponding to 85% of the maximum ship engine output [37]. However, because of safety-related concerns, shipping companies rarely allow testing of the zigzag test scheme with a large rudder angle under the full load condition in the normal speed domain and high power from the main engine. Therefore, the test data obtained from sea trials do not satisfy the requirement of exciting the ship maneuvering sufficiently.

3) Not all ship maneuvering models have strong generalizability [40]. Since ITTC 2008, the precision of system identification has been improved, but its generalizability remains poor. From the machine learning perspective [41], firstly, if zigzag data are set as training data, the higher precision predictions of turning test represents the better generalizability. When the training data and validation data are zigzag test, the prediction is not convincing. Secondly, when the ship is steering with a small angle, the ship motion characteristics can be represented using a linear model. However, ship maneuvering contains many nonlinear components. In the case that the linear and nonlinear components are identified simultaneously by using small rudder angle, the model indicates that the system identification method has strong generalizability. In this paper, zigzag and turning circle were used for identification and validation, respectively.

### C. WORK CONDUCTED IN THIS STUDY

#### 1) METHOD EMPLOYED IN THIS STUDY

We present a new concept called a white–black-box hybrid model that combines white box and black box. The proposed model improves the models presented by Faller [35] and Chiu *et al.* [36]. Furthermore, we enhanced the white–black-box hybrid model by introducing a few concepts and designs.

At present, machine learning that integrates knowledge-driven and data-driven modeling is a frontier research field [43]. Knowledge-driven modeling uses expert knowledge or experience formulas as modeling methods [44]. In ship motion research, Abkowitz model, MMG model and response model, can be regarded as knowledge-driven model. SVMs and NNs are data-driven modeling methods. However, it is complex to prune the parameters of SVM and NN algorithms [41]. By contrast, the random forest algorithm has the advantages of higher precision and easier pruning [41], [45].

Furthermore, the MMG, Abkowitz models are treated as white-box models [14], whereas the random forest algorithm is a black-box model. Therefore, the method reference model-random forest(RM-RF) presented herein can be called a white–black-box hybrid model [14]. Figure 1 shows the conceptual evolution from knowledge-driven and data-driven modeling to the white–black-box hybrid model. The function of yellow blocks is first estimation and the blue ones means compensation of errors.

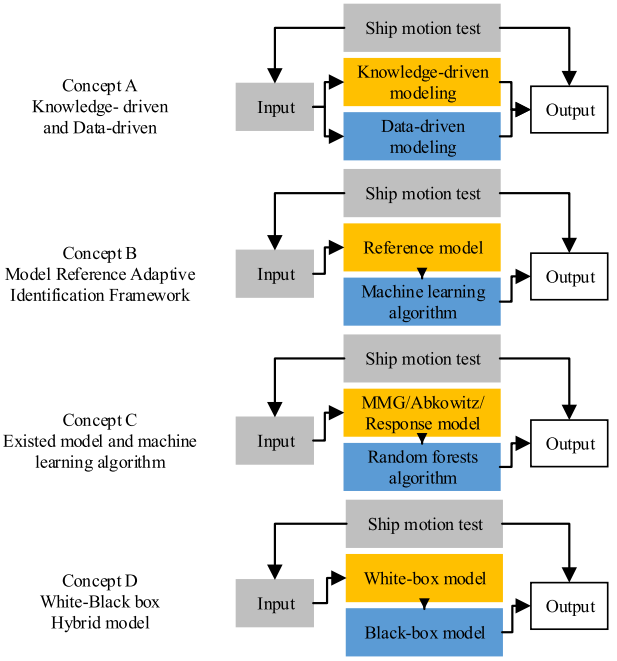


**FIGURE 1. Conceptual evolution of white-black-box hybrid model for ship motion system identification( Grey blocks mean inputs, yellows mean first estimation, blues mean compensation of errors ).**

#### 2) CONTRIBUTIONS OF THIS STUDY

The main work and contributions of this study can be summarized as follows:

(1) The selection method of a reference model was made. In robust adaptive control, the reference model must satisfy certain requirements to ensure that it is effective. We present a method of ship correlation analysis.

(2) State transformation employed normalization forms. In state transformation, we employed the similarity rule to transfer the target ship velocity to reference model velocities [47].

(3) The inputs of random forest are optimized. Because of the applicability and parameter tuning of the machine learning algorithm, the algorithms presented in other papers may not be suitable for the purpose of the present study. The input variables should be modified based on analysis [42]
(4) The innovations include a smaller (insufficient) excitation signal, and various validation cases different from the training test.
(5) The method based on the use of a reference model and a random forest algorithm creates a continuous-time system rather than a discrete-time system [31], [32], [48].

These presents RM method which is easy and stable, and RM-RF is more accurate. Litter data and small rudder is much more suitable for online modeling in real practice.

## II. DESCRIPTION AND PRELIMINARIES

Generally, a mathematical model is employed to describe ship maneuvers. However, such models contain a few assumptions related to hydrodynamic theory and methods. Assumption 1: Based on the linear condition, the ship motion and incoming wave are slight [49]. Therefore, the added mass is constant. Assumption 2: As surge, sway and rotate are important and different with pitch, roll and heave, hence a three-degree-of-freedom (3DOF) model is a specific horizontal plane model.

Consequently, as given by Eq. (1), the nonlinear equations of 3DOF ship maneuvering model [19] are expressed as Equation (1):

$$\begin{cases} (m - X_{\dot{i}})\,\dot{u} = f_1(u, v, r, \delta) \\ (m - Y_{\dot{v}})\,\dot{v} + (mx_G - Y_{\dot{r}})\,\dot{r} = f_2(u, v, r, \delta) \\ (mx_G - N_{\dot{i}})\,\dot{v} + (I_z - N_{\dot{r}})\,\dot{r} = f_3(u, v, r, \delta) \end{cases} \tag{1}$$

here $m$ denotes ship mass; $x_G$ denotes ship mass center of gravity; $u, v, r$ are velocity of surge, sway, and yaw motions, respectively; $I_z$ denotes moment of inertia; and $X_{\dot{u}}, Y_{\dot{v}}, Y_{\dot{r}}, N_{\dot{v}}, N_{\dot{r}}$ are constant numbers. The hydrodynamic force in the longitudinal or lateral direction and the generated moment in the lateral direction can be set as $f_1$, $f_2$ and $f_3$, respectively. Considering the relationship between kinematics and kinetics, Eq. (1) can be transformed to Eq. (2).

$$\begin{cases} \dot{u} = f_1 / (m - X_{\dot{u}}) \\ \dot{v} = \dfrac{(I_z - N_{\dot{r}}) \cdot f_2 - (mx_G - N_{\dot{v}}) \cdot f_3}{(m - Y_{\dot{v}})(I_z - N_{\dot{r}}) - (mx_G - N_{\dot{v}})(mx_G - Y_{\dot{r}})} \\ \dot{r} = \dfrac{(m - Y_{\dot{v}}) \cdot f_3 - (mx_G - N_{\dot{v}}) \cdot f_2}{(m - Y_{\dot{v}})(I_z - N_{\dot{r}}) - (mx_G - N_{\dot{v}})(mx_G - Y_{\dot{r}})} \end{cases} \tag{2}$$

where $m$, $I_z$, and $x_G$ are constant. Based on assumption 1, $X_{\dot{u}}$, $Y_{\dot{v}}$, $Y_{\dot{r}}$, $X_{\dot{r}}$, and $X_{\dot{v}}$ can be treated as constants and determined using potential theory, double-body model testing, or PMM testing [49]. Thus, on the right side of Eq. (2), the ship hydrodynamics $f_1$, $f_2$, and $f_3$ are related only to ship motion state $u$, $v$, $r$, $\delta$. In other words, the ship motion state can be used to ascertain the acceleration of the ship or $f_1$, $f_2$, and $f_3$, which are functions that depend on the variables of ship motion state $u$, $v$, $r$, and $\delta$. Finally, Eq.(2) changes to Eq. (3). expresses the relationship between ship kinematics and kinetics.

$$\begin{cases} \dot{u} = g_1\,(f_1(u, v, r, \delta)) = G_1(u, v, r, \delta) \\ \dot{v} = g_2\,(f_2(u, v, r, \delta), \quad f_3(u, v, r, \delta)) = G_2(u, v, r, \delta) \\ \dot{r} = g_3\,(f_2(u, v, r, \delta), \quad f_3(u, v, r, \delta)) = G_3(u, v, r, \delta) \\ \dot{\psi} = r \\ \dot{x} = u\cos\psi - v\sin\psi \\ \dot{y} = u\sin\psi + v\cos\psi \end{cases} \tag{3}$$

where $G_{i=1,2,3}\,(u, v, r, \delta)$ are functions depending on $u$, $v$, $r$, and $\delta$; $(x, y)$ is the trajectory of ship center of gravity on Earth; and $\psi$ is ship course.

## III. METHODOLOGY

RM-RF is consist of RM and RF. In this section, the work flow RM-RF in identification and prediction for ship maneuvering is introduced. And then the selection of RM and the similar rule are list behind.

### A. WORK FLOW OF RM-RF

In model reference adaptive control (MRAC), the reference model generates the desired trajectory, which is similar to the target trajectory [46]. By contrast, model reference adaptive system identification (MRASI) is a dual system of MRACs [17]. As an offline system, RM-RF is not an adaptive algorithm, which makes it different from MRASI.

In MRASI, the reference model is an approximate model of the target ship. The characteristics of the reference model are similar to those of the target model. As expressed by Eq. (3) and shown in Figure 2, the work flow consists of two part, identification progress and prediction progress. The identification progress is described as following steps. First, the target ship velocity is transferred to the reference model. Then transferred velocity and acceleration are the inputs and output of reference model.

Secondly, there is errors between reference model accelerations and target ship acceleration. So the random forest compensates the errors. The velocities and rudder angle consist of the inputs for random forest.

Generally, according to the reviews, ship motion is treated as a discrete-time system [34], [35]. In the case of a discrete-time system, difference transformation of Eq. (2) can be performed to predict the $k+1$ th $u(k+1)$, $v(k+1)$, $r(k+1)$, $\delta(k+1)$ state based on the $u(k)$, $v(k)$, $r(k)$,. Because of the nonlinearity of $f_1, f_2$, and $f_3$ in Eq. (2), the discrete operation increases the complexity of ship dynamics [50]. Furthermore, the sampling rate influences the equilibrium properties and asymptotic stability of the ship motion system [51], [52].

However, in the present study, ship maneuver prediction is a continuous-time system. RM-RF predicts the current $\dot{u}$, $\dot{v}$, $\dot{r}$ based on the current $u$, $v$, $r$, $\delta$. Therefore, RM-RF is not a discrete-time system, but a continuous-time system which different from these researches. [32], [34], [48].

### B. REFERENCE MODEL AND SELECTION METHOD

In terms of ship maneuvering, the MMG model and Abkowitz model are reference model. We collected the MMG model

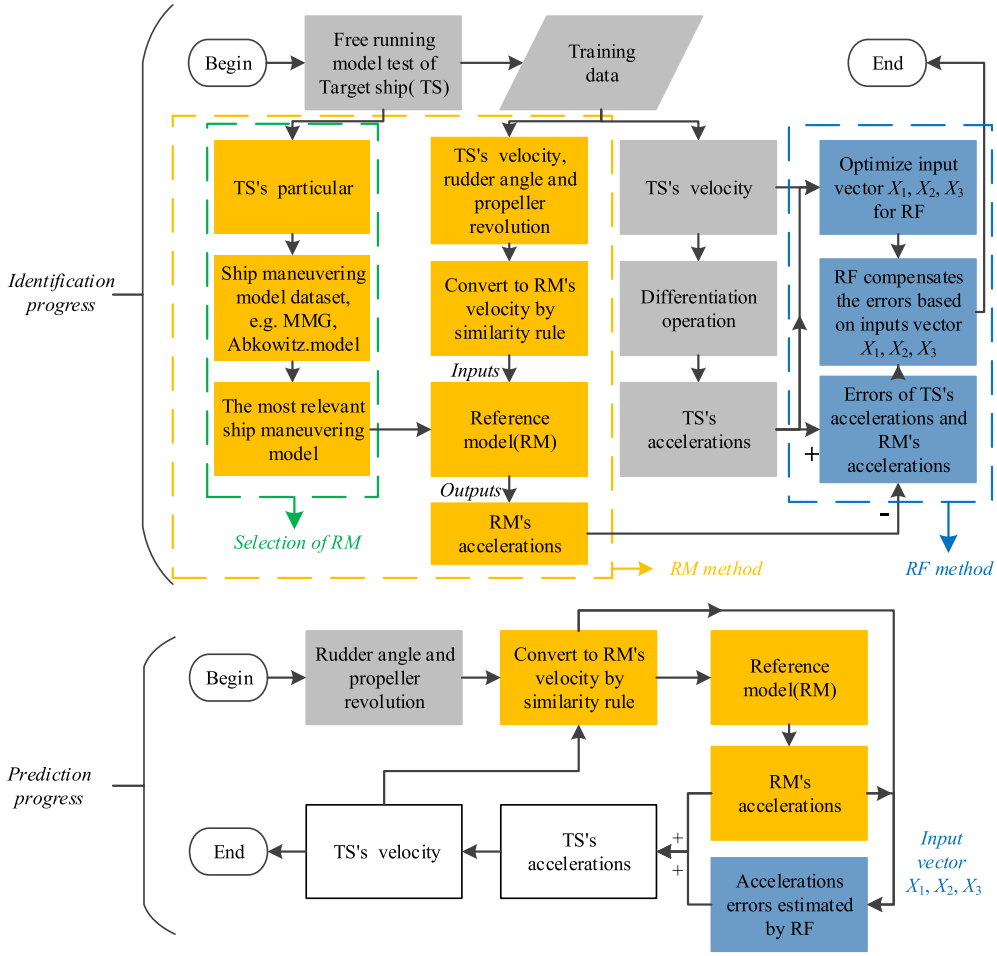


**FIGURE 2.** The identification and prediction progress of ship maneuvering based on RM-RF.

and Abkowitz model as RM [8], [54], [56]. In MRAC, as the same structure and similar order as target system, the reference model describes the performance of target system [46].

In MRAC, the reference model describe the characteristics, of target system a model with the same or similar order and structure [46].

Notably, Hayes [53] used the concept of a reference model and the parameters of the reference model converged adaptively. In the present work, the reference model parameters are not modified. Therefore, Hays' method is not the same as RM-RF.

We introduced a selection method of reference model from existing ship maneuvering models. According to [57], every ship has its own unique characteristics. The value $Cb$ denotes the ship block coefficient, $Lpp$ denotes length between perpendiculars, $B$ denotes ship beam, $T$ denotes draught, $V_0$ denotes designed ship speed, and $Ar$ denotes rudder area. The particulars vector $p = (Cb,\ Lpp/B,\ B/T,\ Lpp/V_0,\ Ar/(Lpp \cdot T))$ represents the characteristics of the ship particular. The particulars matrix consists of $p$. After normalization according to Eq.(4), the normalized particulars matrix $P'$ is obtained. Then, by using Eq.(5) for correlation analysis, the correlation coefficients (CCs) between the normalization particulars vectors $p'_1$ and $p'_i$ $(i = 2, 3, 4, 5)$ show the similarity between KVLCC2 and other existing models.

$$\begin{cases} P = Transpose(\ p_1,\ \ldots,\ p_i,\ \ldots,\ p_m\ ) \\ = \begin{pmatrix} Cb_1 & Cb_2 & \ldots & Cb_n \\ \dfrac{Lpp_1}{B_1} & \dfrac{Lpp_2}{B_2} & \ldots & \dfrac{Lpp_n}{B_n} \\ \dfrac{B_1}{T_1} & \dfrac{B_2}{T_2} & \ldots & \dfrac{B_n}{T_n} \\ \dfrac{Lpp_1}{V_{01}} & \dfrac{Lpp_2}{V_{02}} & \ldots & \dfrac{Lpp_n}{V_{0n}} \\ \dfrac{Ar_1}{Lpp_1T_1} & \dfrac{Ar_2}{Lpp_2T_2} & \ldots & \dfrac{Ar_n}{Lpp_nT_n} \end{pmatrix} \\ p_i = (p_{i1},\ \ldots,\ p_{ij},\ \ldots,\ p_{in}\ ) \\ m = 5;\ i = 1, 2, \ldots, m;\ j = 1, 2, \ldots n; \\ \|p_i\|_2 = (p_{i1}^2 + \ldots + p_{ij}^2 + \ldots + p_{in}^2)^{\frac{1}{2}} \\ p'_i = \dfrac{p_i}{\|p_i\|_2} \\ P' = Transpose(p'_1,\ p'_2,\ \ldots,\ p'_m\ ) \end{cases} \tag{4}$$

$$\begin{cases} P' = (q_1', q_2', \ldots, q_n') \\ CC = \dfrac{COV(q_1, q_i)}{\sigma(p_1)\sigma(p_i)} \\ = \dfrac{E(q_1 q_i) - E(q_1)E(q_i)}{\{E(q_1^2) - E^2(q_1)\}^{\frac{1}{2}}\{E(q_i^2) - E^2(q_i)\}^{\frac{1}{2}}} \\ = \dfrac{\sum q_1 q_i - \dfrac{\sum q_1 \sum q_i}{n}}{\{(\sum q_1^2 - \dfrac{(\sum q_1)^2}{n})(\sum q_1^2 - \dfrac{(\sum q_1)^2}{n})\}^{\frac{1}{2}}} \end{cases} \quad (5)$$

where $COV(\cdot)$ is covariance, $\sigma(\cdot)$ is variance, $E(\cdot)$ is standard deviation, $Transpose(\cdot)$ is the transpose of a matrix, and $CC$ is the correlation coefficient.

### C. SIMILARITY RULE

There is a difference between the target ship and the reference model in terms of ship design speed and propeller revolution. The difference can be compensated for by using the similarity rule from dimensional analysis or. [47], [58]. We introduce the normalization forms from SNAME [47], as in Eq.(6).

$$\begin{cases} u_R = u_T\ (V_{0R}/V_{0T}) \\ v_R = v_T\ (V_{0R}/V_{0T}) \\ r_R = r_T\ \{(V_{0R}/Lpp_R\}\ \{V_{0T}/Lpp_T)\}^{-1} \\ \delta_R = \delta_T \end{cases} \quad (6)$$

where $(\cdot)_{0(\cdot)}$ is ship speed when it is running straight in a steady speed, $(\cdot)_T$ is target ship state, and $(\cdot)_R$ is reference model state.

## IV. RANDOM FOREST AND ITS APPLICATION

The main purposes of the machine learning algorithm used herein are performing regression analysis on the sampled data and establishing a relationship between the reference model and the target ship. After training, the machine learning algorithm predicts the accelerations of the target ship based on the outputs of the reference model.

Regression analysis algorithms used by the machine learning community include SVMs, NNs, Gaussian process regression models, generalized linear models, and regression trees [41].Because of the nonlinear dynamic characteristics of ship maneuvering with uniform sampling, datasets covering different types of maneuvering tests are not independent identically distributed samples. Therefore the random forest algorithm [45] can be used to solve the previously described problem of different sampling ratios in different tests.

### A. DETAILS OF RANDOM FOREST

Random forest is an ensemble learning scheme that involves training regression trees. An important feature of the random forest is its use of out-of-bag samples. Another notable feature is that variable importance is constructed in the same way as that in gradient-boosted models [41].

The random forest algorithm is the most accurate algorithm in ensemble learning.

**TABLE 1. Basics of random forest.**

| Random forest algorithm for regression [41] |
|---|
| For $b = 1$ to $B$: |
| Draw a bootstrap sample of size $N$ from training data. |
| Grow a random-forest tree $Tree_b$ of bootstrapped data. |
| While not reaching the minimum node size $n_{min}$, select $m$ variables randomly from all variables. |
| Pick the best variable/split-point among the $m$ variables. |
| Split the node into two daughter nodes. End While |
| End For |
| Output the trees $Tree_b(X;\theta_b)$, $b \in \{1,2,...,B\}$. |
| Predict at a new point X: $f^B_{randomforest}(X) = \frac{1}{B}\sum_{b=1}^{B} Tree_b(X;\theta_b)$. |

### B. IDENTIFICATION AND PREDICTION BASED ON RANDOM FOREST

$Tree_b(X;\theta_b)$ denotes regression trees, $X \in \{X_1, X_2, X_3\}$; $X_1, X_2, X_3$ is the vector of variables and is composed of ship velocity, acceleration, and rudder angle. The specific $X_1, X_2, X_3$ will be ascertained and optimized in the next section; $\theta_b$ is the split of the $b$ th regression tree, and $f^B_{randomforest}(X)$ is a random forest consisting of regression trees. The ship accelerations $\dot{u}_T, \dot{v}_T, \dot{r}_T$ are functions $f^B_{randomforest}(X)$ of $X_1, X_2, X_3$ respectively.

The residual error of surge acceleration between the target ship and the reference model $\Delta\dot{u}_{TR}$ is defined in Eq.(7).

$$\Delta\dot{u}_{TR} = \dot{u}_T - \dot{u}_R \quad (7)$$

As given in Eq. (3), $G_{1T}$ and $G_{1R}$ are the surge accelerations of the target ship and the reference model, respectively. The residual error $\Delta\dot{u}_{TR}$ can be rewritten as Eq. (8)

$$\Delta\dot{u}_{TR} = G_{1T} - G_{1R} \quad (8)$$

The residual error $\Delta\dot{u}_{TR}$ can be approximated using a random forest $f^{B_1}_{randomforest}(X_1)$ depending on $X_1$. $B_1$ is the number of trees in the random forest $f^{B_1}_{randomforest}(X_1)$. $Tree_{b_1}(X_1;\theta_{b_1})$ is the $i$th tree depending on $X_1$, when the split is $b_i$.

$$\Delta\dot{u}_{TR} = f^{B_1}_{randonfforest}(X_1) = \frac{1}{B_1}\sum_{b_1=1}^{B_1} Tree_{b_1}(X_1;\theta_{b_1}) \quad (9)$$

As a generalized expression, ship hydrodynamics are a function of $u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R$. [59]. Thus, for the residual error $\Delta\dot{u}_{TR}$, the independent variables of the random forest $f^{B_1}_{randomforest}(X_1)$ are $u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R$ and $\theta_{b_1}$. Eq.(9) is then rewritten as Eq. (10)

$$\Delta\dot{u}_{TR} = \frac{1}{B_1}\sum_{h_1=1}^{B_1} Tree_{b_1}(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R; \theta_{b_1}) \quad (10)$$

In the same way as for surge acceleration, sway acceleration and yaw acceleration can be summarized by Eq. (11) and Eq.(12).

$$\begin{cases} \Delta\dot{v}_{TR} = \hat{v}_T - \dot{v}_R = G_{2T} - G_{2R} \\ = f^{B_2}_{randomforest}(X_2) = \dfrac{1}{B_2}\displaystyle\sum_{b_2=1}^{B_2} Tree_{b_2}\left(X_1; \theta_{b_2}\right) \\ = \dfrac{1}{B_2}\displaystyle\sum_{b_2=1}^{B_2} Tree_{b_2}\left(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R; \theta_{b_2}\right) \end{cases} \tag{11}$$

$$\begin{cases} \Delta\dot{r}_{TR} = \dot{r}_T - \dot{r}_R = G_{3T} - G_3R \\ = f^{B_3}_{randomforest}(X_3) = \dfrac{1}{B_3}\displaystyle\sum_{b_3=1}^{B_3} Tree_{b_3}\left(X_1; \theta_{b_3}\right) \\ = \dfrac{1}{B_3}\displaystyle\sum_{b_3=1}^{B_3} Tree_{b_3}\left(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \vec{r}_R; \theta_{b_3}\right) \end{cases} \tag{12}$$

Therefore, based on the training process of ship accelerations for surge, sway, and yaw, the prediction process is expressed by Eq.(10) and Eq.(12).

$$\begin{cases} \dot{u}_T = \dot{u}_R + \Delta\dot{u}_{TR} \\ = \dot{u}_R + \dfrac{1}{B_1}\displaystyle\sum_{b_i=1}^{B_1} Tree_i\left(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R; \theta_{b_1}\right) \\ \dot{v}_T = \dot{v}_R + \Delta\dot{v}_{TR} \\ = \dot{v}_R + \dfrac{1}{B_2}\displaystyle\sum_{b_i=1}^{B_2} Tree_i\left(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R; \theta_{b_2}\right) \\ \dot{T}_T = \dot{r}_R + \Delta\dot{r}_{TR} \\ = \dot{r}_R + \dfrac{1}{B_2}\displaystyle\sum_{h=1}^{B_3} Tree_i\left(u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R; \theta_{b_3}\right) \end{cases} \tag{13}$$

The input vectors $X_1$, $X_2$, and $X_3$ for the random forests in Eq.(13) are $u_R, v_R, r_R, \delta_R, \dot{u}_R, \dot{v}_R, \dot{r}_R$; this translates into a complicated and multidimensional scenario. Thus, simplified expressions are derived, as described in the following paragraph.

Referring to the Abkowitz model [59], [60], surge motion includes straight-ahead resistance and rudder force components. Assuming no significant interaction between the viscosity and inertia properties of fluid, terms such as $X_{\dot{u}\dot{u}}$, $X_{\dot{v}\dot{v}}$, and $X_{\dot{r}\dot{r}}$ are treated as zero. Surge acceleration is decoupled from sway and yaw motions. Therefore, the optimized input vector $X_1$ is $[u_R, v_R, r_R, \delta_R]^{\mathrm{T}}$.

Furthermore, because of the symmetry about the $xz$-plane, $Y(u) = N(u) = x0$, and the derivatives $Y_u$, $Y_{uu}$, $Y_{uuu}$, $Y_{\dot{u}}$, $N_u$, $N_{uu}$, $N_{uuu}$, and $N_{\dot{u}}$ are all zero. By contrast, the acceleration derivatives $Y_{\dot{v}}$, $Y_{\dot{r}}$, $N_{\dot{v}}$ and $N_{\dot{r}}$ are not zero in the Taylor expansion of lateral force $Y$ and yaw moment $N$. Thus, the sway and yaw motions are coupled. Moreover, the drift angle $\beta_R = \arctan(-\frac{v_R}{u_R})$ is obtained via numerical integration of ship yaw velocity and the index of relationship between the longitudinal and lateral force. Therefore, $\beta_R$ contains the effects of $u_R$, and the optimized input vectors $X_2$ and $X_3$ are $[v_R, r_R, \dot{v}_R, \dot{r}_R, \beta_R, \delta_R]^{\mathrm{T}}$.

According to the previous analysis of input vectors of random forests, the training and prediction process can be summarized by Eq.(14) and Eq.(15). optimized training:

$$\begin{cases} \Delta\dot{u}_{TR} = \dfrac{1}{B_1}\displaystyle\sum_{b_i=1}^{B_1} Tree_i\left(u_R, v_R, r_R, \delta_R; \theta_{b_1}\right) \\ \Delta\dot{v}_{TR} = \dfrac{1}{B_2}\displaystyle\sum_{b_i=1}^{B_2} Tree_i\left(v_R, \dot{v}_R, r_R, \dot{r}_R, \beta_R, \delta_R; \theta_{b_2}\right) \\ \Delta\dot{r}_{TR} = \dfrac{1}{B}\displaystyle\sum_{b_i=1}^{B_3} Tree_i\left(v_R, \dot{v}_R, r_R, \dot{r}_R, \beta_R, \delta_R; \theta_{b_3}\right) \end{cases} \tag{14}$$

optimized prediction:

$$\begin{cases} \dot{u}_T = \dot{u}_R + \Delta\dot{u}_{TR} \\ = \dot{u}_R + \frac{1}{B_1}\displaystyle\sum_{b_i=1}^{B_1} Tree_i\left(u_R, v_R, r_R, \delta_R; \theta_{b_1}\right) \\ \dot{v}_T = \dot{v}_R + \Delta\dot{v}_{TR} \\ \dot{v}_T = \dot{v}_R + \Delta\dot{v}_{TR} \\ = \dot{v}_R + \frac{1}{B_2}\displaystyle\sum_{b_i=1}^{B_2} Tree_i\left(v_R, \dot{v}_R, r_R, \dot{r}_R, \beta_R, \delta_R; \theta_{b_2}\right) \\ \dot{r}_T = \dot{r}_R + \Delta\dot{r}_{TR} \\ = \dot{r}_R + \frac{1}{B_3}\displaystyle\sum_{b_i=1}^{B_3} Tree_i\left(v_R, \dot{v}_R, r_R, \dot{r}_R, \beta_R, \delta_R; \theta_{b_3}\right) \end{cases} \tag{15}$$

## V. PREPARATION IN STUDY CASES

In this section, illustrative examples are presented to verify the performance of the proposed identification scheme. For this purpose, data from ship model tests were collected. In the following subsections, we describe the ship model particulars, model test data, data process, application of the reference model, and training and evaluation of the RM-RF.

**TABLE 2. Particulars of KVLCC2.**

| | Full-scale | model |
|---|---|---|
| Scale | 1 | 45.714 |
| $Lpp$ (m) | 320.0 | 7.0 |
| $B$ (m) | 58.0 | 1.1688 |
| $D$ (m) | 30.0 | 0.6563 |
| Draught(m) | 20.8 | 0.4550 |
| $Cb$ | 0.8098 | 0.8098 |
| Max. rudder rate (°) | 2.34 | 15.8 |
| Service speed (m/s) | 7.974 | 1.179 |

### A. TARGET SHIP PARTICULARS

In the ship motion modeling community, KVLCC2 is a benchmark ship for various methods. KVLCC2 is a very large crude carrier (VLCC) ship, and its particulars are listed in TABLE 2.

**TABLE 3. Test conditions.**

| Training data | | | | |
|---|---|---|---|---|
| Test type | Rudder angle | Heading angle | Port/ starboard | Sample points |
| zigzag 1 | 10° | 10° | starboard | 520 |
| zigzag 2 | 10° | 10° | port | 565 |
| zigzag 3 | 20° | 20° | starboard | 606 |
| zigzag 4 | 20° | 20° | port | 684 |
| Validation data | | | | |
| Test type | Rudder angle | Heading angle(°) | Port/ starboard | Sample number |
| turning 1 | 35° | * | starboard | 1610 |
| zigzag 5 | 20° | 10° | port | 1278 |

The samples were obtained from a free-running model test with a sampling period of 1s. The test type and other details of training and validation data are presented in TABLE 3. The tests were completed in a tank or basin, and the maximum rudder angle in the zigzag tests was not more than 20°. Therefore, we used fewer small-angle test data to train the RM-RF.

### B. DATA PROCESSING

Measured data contain noise. Thus, we employed a spline to smooth the data, that is, $u_T$, $v_T$, and $r_T$. The second derivatives of the spline $u_T$, $v_T$, and $r_T$ are the accelerations $\dot{u}_T$, $\dot{v}_T$, and $\dot{r}_T$. As in the example, in the $+20°/-20°$ zigzag test, the spline smoothed ship lateral velocity and calculated lateral acceleration.

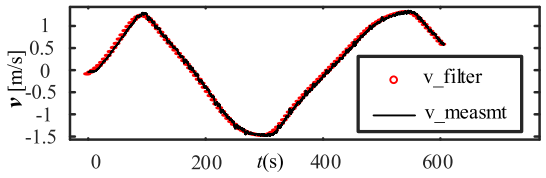


**FIGURE 3. Lateral velocity from filter.**

FIGURE **3** shows the effective smoothing and calculation of sway motion. FIGURE **4** shows the numerical results for accelerations. The accelerations obtained in this way are of the same order as those obtained using Yasukawa's MMG model in conjunction with an oblique towing test and a circular motion test [61].

The difference between accelerations indicates that smooth acceleration is not suitable for the regression. Thus, further approximation is needed, such as the RM-RF, as described section 4.b.

### C. SELECTION OF REFERENCE MODEL

Existing models are stable, such as the Abkowitz models "Mariner" [56] and "Tanker" [8] and MMG models "SR108" [54] and "PCC" [54], and it is easy to acquire research results using them. The target ship analyzed in this paper is a VLCC named KVLCC2. We selected a ship that was the closest to the ships used in the existing models.

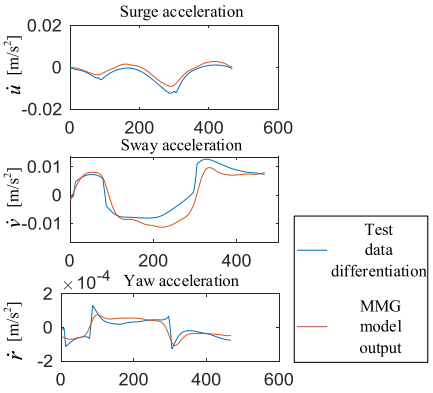


**FIGURE 4. Comparison of accelerations between numerical differentiation of experience data from MARIN and Yasukawa's MMG model from NMRI [61].**

**TABLE 4. Particulars matrix of KVLCC2 and other existing models.**

| Name | KVLCC2 | Mariner | Tanker | SR108 | PCC |
|---|---|---|---|---|---|
| Model | * | Abkowitz | Abkowitz | MMG | MMG |
| Type | Crude Oil | Cargo | Crude Oil | Container | RORO |
| $Cb$ | 0.898 | 0.590 | 0.83 | 0.562 | 0.547 |
| $Lpp/B$ | 5.517 | 6.952 | 6.389 | 6.890 | 5.590 |
| $B/T$ | 2.788 | 3.1 | 2.584 | 2.988 | 3.926 |
| $Lpp/V_0$ | 40.13 | 20.853 | 37.035 | 14.085 | ≈18 |
| $\frac{Ar}{Lpp*T}$ | 1/48.69 | 1/83.05 | ≈1/61 | 1/45.8 | 1/39.83 |

The particulars matrix is summarized in TABLE 4. After normalization, TABLE 4 is transformed to TABLE 5. Then, correlation analysis was performed, and the CCs obtained between KVLCC2 and other four existing models are listed in TABLE 6. The maximum coefficient shows that Tanker is the most similar to KVLCC2; therefore, it was selected as the reference model. The symbol "≈" indicates that the particulars are not recorded in papers [8], [54], [57] and were estimated by referring to the appendix of a book [62].

### D. OUTPUTS OF REFERENCE MODEL AND TARGET SHIP

In this paper, Tanker was selected as the reference model. The particulars of "Tanker" are summarized in TABLE 7.

The training data comprised rudder angle, velocity and acceleration of surge, and sway and yaw motions. With the state transition using normalization forms, the target ship state $u_T$, $v_T$, and $r_T$ is converted to the reference model state $u_R$, $v_R$ and $r_R$. The reference model state $u_R$, $v_R$, and $r_R$ is the input to the reference model, and the accelerations $\dot{u}_R$, $\dot{v}_R$, and $\dot{r}_R$ are its outputs. As can be inferred from TABLE 7, the reference model is Tanker. As expressed by 0 the accelerations of the reference model are an approximation of the accelerations of the target ship.

**TABLE 5.** Normalized particulars matrix of KVLCC2 and other existing models.

| Name | KVLCC2 | Mariner | Tanker | SR108 | PCC |
|---|---|---|---|---|---|
| Model | * | Abkowitz | Abkowitz | MMG | MMG |
| Type | Crude Oil | Cargo | Crude Oil | Container | RORO |
| $(Cb)'$ | 0.573 | 0.376 | 0.529 | 0.358 | 0.349 |
| $(Lpp/B)'$ | 0.392 | 0.494 | 0.454 | 0.490 | 0.397 |
| $(B/T)'$ | 0.401 | 0.446 | 0.371 | 0.430 | 0.564 |
| $(Lpp/V_0)'$ | 0.639 | 0.332 | 0.590 | 0.224 | 0.287 |
| $(\frac{Ar}{Lpp*T})'$ | 0.466 | 0.273 | 0.372 | 0.495 | 0.570 |

**TABLE 6.** Correlation coefficients between KVLCC2 and other existing models.

| | | | | |
|---|---|---|---|---|
| *CC* | -0.576 | 0.851 | -0.887 | -0.723 |

**TABLE 7.** Particulars of tanker.

| | |
|---|---|
| *Lpp* (*m*) | 304.8 |
| *B* (*m*) | 47.17 |
| *T* (*m*) | 18.46 |
| *Cb* | 0.83 |
| Nominal propeller (*rpm*) | 2.34 |
| Service speed(*m/s*) | 8.23 |

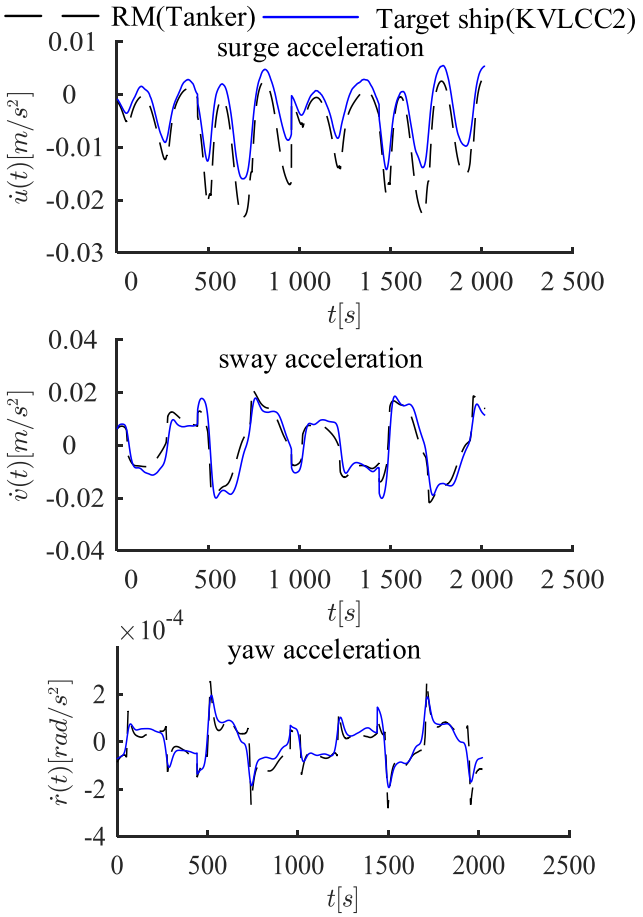


**FIGURE 5.** Kinetic variables of target ship approximated by outputs of reference model.

As shown in FIGURE 5, the correlation coefficients of these accelerations were 0.9731, 0.9438, and 0.9106. Therefore, Tanker is a good approximation of KVLCC2. Linear independence between the acceleration of 'Tanker' and KVLCC2 is readily apparent.

## VI. TRAINING AND VALIDATION IN STUDY CASES

To illustrate the modeling performance of the proposed RM-RF method, a training case consisting of $10°/10°$, $-10°/-10°$, $20°/20°$, $-20°/-20°$ zigzag tests were considered. Validation cases consisting of 35° turning circle and $20°/10°$ zigzag test are presented.

### A. TRAINING CASE

Based on Eq.(14), the training process was implemented with the training tests $10°/10°$ zigzag test, $-10°/-10°$ zigzag, $20°/20°$ zigzag3, and $-20°/-20°$ zigzag4, as summarized in TABLE 3.

After the training was completed, the simulation was performed using Eq.(15). Then, in contract with two results including the EFD method from NMRI by Yasukawa and Yoshimura [61] and the CFD method from HMRI by Sung and Park [63], a comparison of heading angles in training data zigzag $20°/20°$ is shown in Figure 6. The first and second overshoot angles are listed in TABLE 8. These comparisons show that the RM-RF performs well and yields accurate predictions for the training data.

As for the function of RM, Figure 6 shows that RM gives the base estimation of target ship. Based on RM, the prediction of RM-RF is improved for RF compensates the errors between RM and target ship.

Furthermore, the time histories of velocities and accelerations of surge, sway, and yaw in Norrbin's normalized forms are shown in

FIGURE 6 [58]. Good correspondence can be observed between the results of MARIN's free-running test and those obtained using RM-RF. The simulation results illustrate that the RM-RF can converge to the training data. To quantify the precision of the proposed method, various evaluation indices, including mean square error (MSE), normalized root-mean-square error (NRMSE), and CCs, were employed. The precision of the heading angle was evaluated using the data in TABLE 8.

The predictions obtained in the training test cannot be accurate with high precision. As mentioned in the introduction, overfitting of the training data leads to more effective fitting but lower accuracy from the viewpoint of validation. Thus, the precision of predictions for the $20°/20°$ zigzag

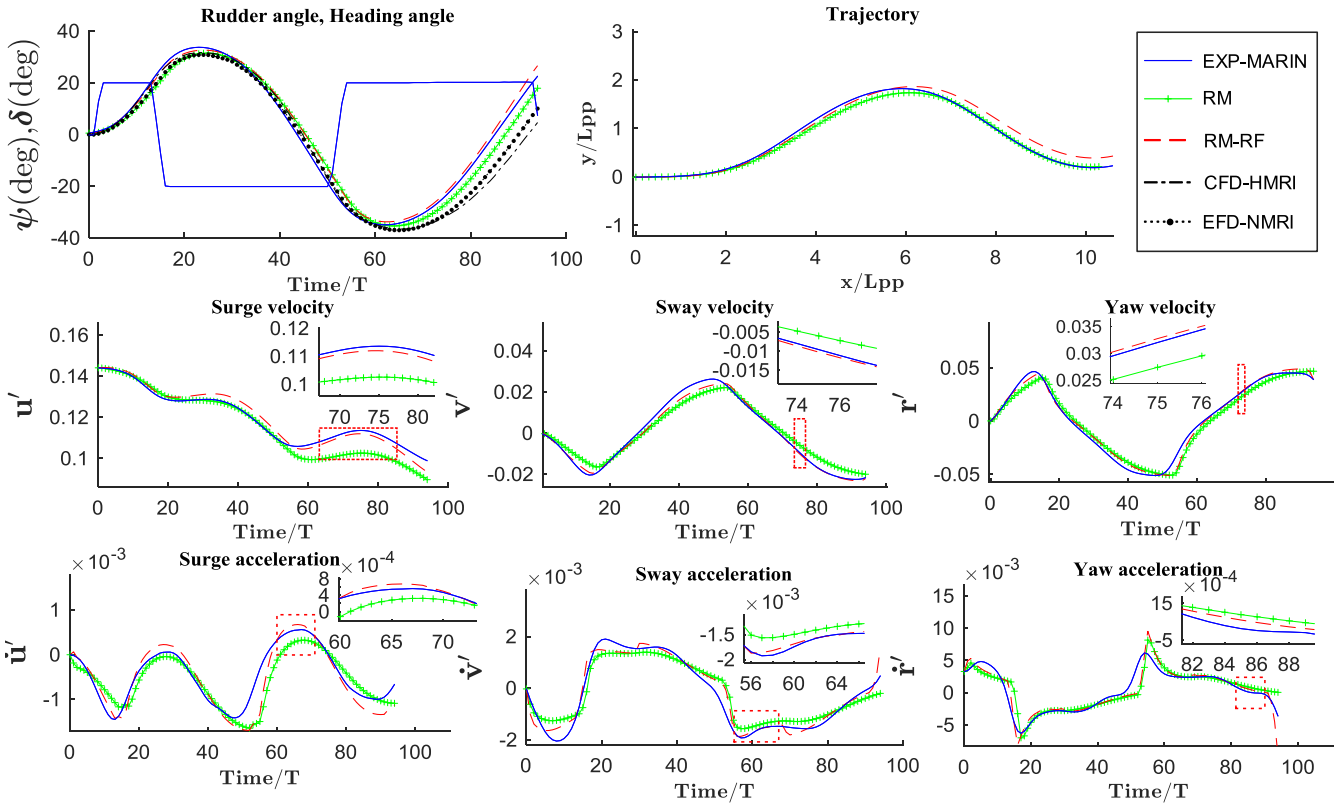


**FIGURE 6.** Prediction and comparison of 20°/20° zigzag test [61], [63].

**TABLE 8.** Comparison of overshoot angles between different methods for 20°/20° zigzag test.

| Methods | EXP -MARIN | RM-RF | EFD -NMRI | CFD -HMRI |
|---|---|---|---|---|
| 1st OSA | 13.8 | 12.7 | 10.9 | 11.7 |
| 2nd OSA | 14.9 | 13.7 | 17.0 | 16.5 |

**TABLE 9.** MSE, NRMSE, and CCs of heading angle between free-running test from MARIN and other methods.

| | Methods | RM-RF | EFD-NMRI | CFD-HMRI |
|---|---|---|---|---|
| $\psi$ (°) | MSE | 5.9948 | 31.6271 | 50.8925 |
| | NRMSE | 0.0357 | 0.0819 | 0.1039 |
| | CC | 0.9967 | 0.9808 | 0.9655 |

training test is acceptable and necessary, which represents and implies the good generalizability of the validation test.

### B. VALIDATION CASE 1

To demonstrate the effectiveness of the proposed RM-RF method, 35° turning circle validation case from TABLE 3 was simulated. Based on Eq.(15) of RM-RF, prediction of the 35° turning circle was considered for validation. The CFD simulation results obtained by Shanghai Jiao tong University (CFD-SJT) [64], the EFD method, and the CFD method were gathered [61], [63]. The results obtained using these methods and RM-RF were compared with the results of the free-running test from MARIN. Velocities of surge, sway, and yaw were involved only in CFD-SJT [64]. Thus, the velocities obtained in the present study were compared with the results of CFD-SJT. The trajectory obtained herein contrasts CFD-SJT [64], EFD-NMRI [61], and CFD-HMRI [63].

The prediction results corresponding to the 35° turning circle obtained using these methods are shown in Figure 7. Furthermore, the velocities of surge, sway, and yaw obtained using CFD-SJT and RM-RF are compared in Figure 7. The accelerations obtained using RM-RF were compared with those obtained using MARIN. These comparisons show that although fewer training data with a small rudder angle were used, the RM-RF converged in validation case 1. In other words, the simulation results were not affected by the use of few training or sample data. The effectiveness of RM-RF is thus demonstrated.

As for RM's function, the appreciate of RM is obvious. And based on RM, RFs decrease the error between RM and EXP-MARIN.

The turning circle geometry is compared in TABLE 10. The prediction precision of advance obtained with RM-RF is outstanding. The tactical diameter obtained using RM-RF is acceptable and medium.

The prediction accuracy indicates the generalizability of the proposed method. As in the training test, evaluation indices, including MSE, NRMSE, and CC, were used to

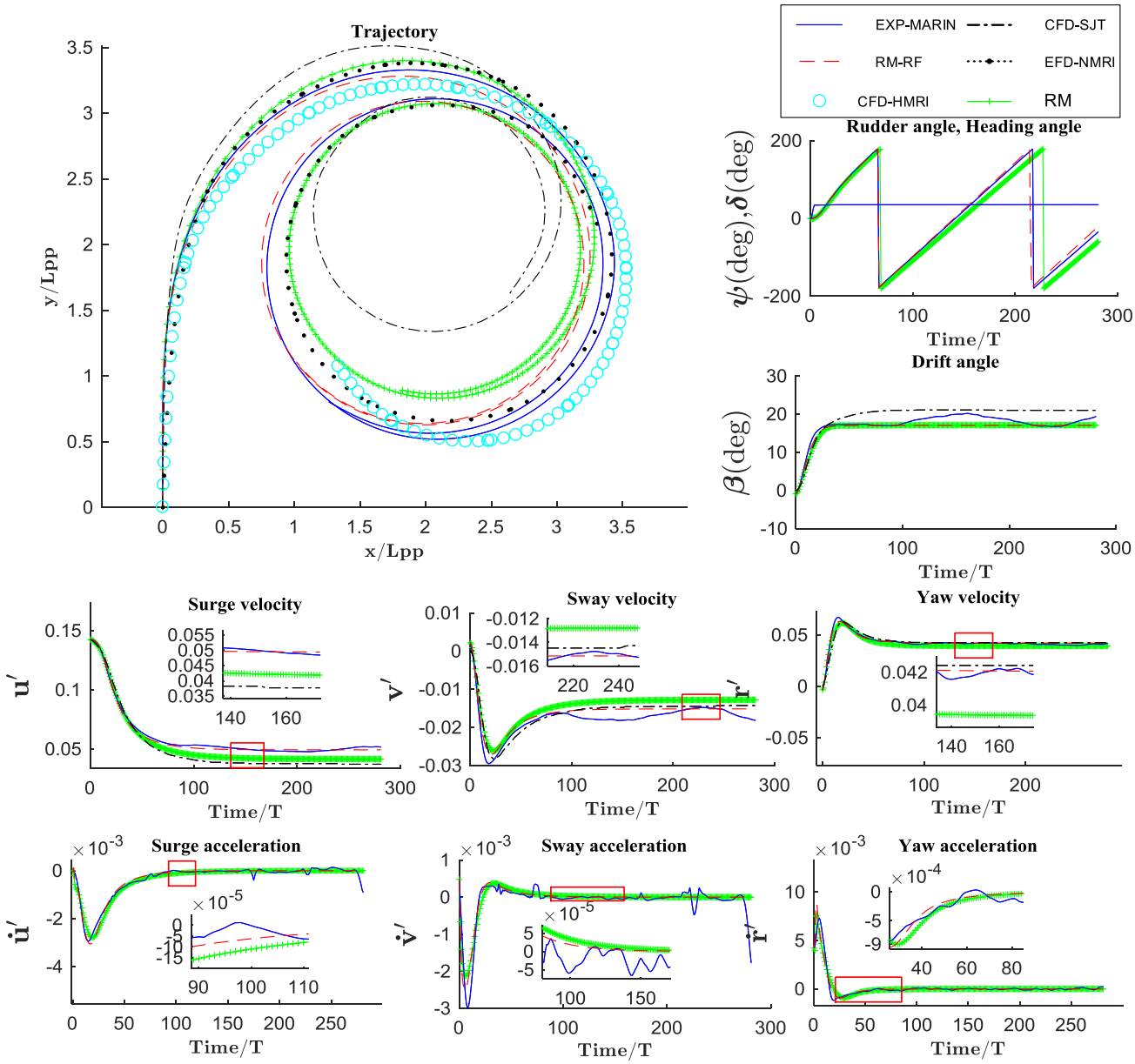


**FIGURE 7.** Prediction and comparison of 35° turning test and compared with free running test and other methods [61], [63], [64].

**TABLE 10.** Comparison of advance and tactical diameter among different methods for 35° turning circle.

| Methods | EXP -MARIN | RM -RF | CFD -SJT | EFD -NMRI | CFD -HMRI |
|---|---|---|---|---|---|
| $A_D^{'}$ | 3.25 | 3.21 | 3.63 | 3.31 | 3.12 |
| $D_T^{'}$ | 3.34 | 3.21 | 2.87 | 3.36 | 3.4 |

qualify the precision between the free-running test from MARIN and the other methods, as shown in TABLE 11. RM-RF had higher precision than CFD-SJT. This means RM-RF has good generalizability as a modeling method.

### C. VALIDATION CASE 2

In addition to validation of the 35° turning circle, a 20°/10° zigzag test was simulated using the RM-RF as another validation test. In ship motion, the ship state is depicted using velocities. In the case of the 35° turning circle, the range of ship states is more extensive than in the 10°/10° zigzag and 20°/20° zigzag, and range of turning circle motion covers the two zigzag. Therefore, the results of validation case 1 verify the extrapolation ability of the RM-RF. Furthermore, validation case 2 tested the interpolation ability of RM-RF.

For validation case 2, predictions of heading, trajectory, and velocities and accelerations of surge, sway, and yaw were obtained and are shown and compared with the SVM from

**TABLE 11.** MSE, NRMSE, and CCs of heading angle between free-running test from MARIN and other methods.

| | Qualify methods | RM-RF | CFD-SJT |
|---|---|---|---|
| $u$ (m/s) | MSE | 2.5729e-06 | 1.1261e-04 |
| | NRMSE | 0.0170 | 0.1122 |
| | CC | 0.9980 | 0.9926 |
| $v$ (m/s) | MSE | 2.9854e-06 | 5.2496e-06 |
| | NRMSE | 0.0546 | 0.0723 |
| | CC | 0.9685 | 0.9154 |
| $r$ (rad/s) | MSE | 1.6502e-06 | 7.9759e-06 |
| | NRMSE | 0.0182 | 0.0400 |
| | CC | 0.9900 | 0.9323 |
| $\beta$ (rad) | MSE | 6.5764e-04 | 0.0022 |
| | NRMSE | 0.0697 | 0.1289 |
| | CC | 0.9515 | 0.9315 |

**TABLE 12.** MSE, NRMSE, and CC of 20°/10° zigzag test between free-running test from MARIN and RM-RF.

| | MSE | NRMSE | CC |
|---|---|---|---|
| $u$ (m/s) | 3.3683e-05 | 0.1579 | 0.9831 |
| $v$ (m/s) | 1.1621e-05 | 0.0831 | 0.9625 |
| $r$ (rad/s) | 5.2275e-05 | 0.0739 | 0.9670 |
| $\dot{u}$ (m/s$^2$) | 1.0337e-07 | 0.1760 | 0.8464 |
| $\dot{v}$ (m/s$^2$) | 2.0456e-07 | 0.1218 | 0.9470 |
| $\dot{r}$ (rad/s$^2$) | 3.5256e-06 | 0.1357 | 0.8601 |
| $\psi$ (°) | 16.4577 | 0.0845 | 0.9732 |
| $x$ (m) | 0.0772 | 0.0252 | 1.0000 |
| $y$ (m) | 0.0118 | 0.1040 | 0.9500 |

reference [24] in Figure 8. It is obvious that the prediction of sway velocities by RM-SVM is more accurate than SVM. As for the evaluation indices showing the precision of RM-RF, MSE, NRMSE, and CC, are listed in TABLE 12. The prediction shows the good interpolation ability of RM-RF. the evaluation indices represent accurate modeling progress achieved using the proposed RM-RF method.

## VII. DISCUSSION

The white-black-box hybrid model of RM-RF was presented for ship maneuvering. And in the study cases, the RM-RF was trained and validated. However, some points about remark, stability of model and further research should be discussed.

### A. REMARK OF METHOD SIGNIFICANT

As required in standards for ship maneuverability, full-scale trial should be conducted in calm environment. However, wave and wind disturbs effecting ship maneuvering and the calm water data used in study cases, RM-RF has its own significant. Ship maneuvering in waves consist of wave and calm water induced motion. Therefore, ship maneuvering in calm water contributes to ship navigate on the sea.

On the other hand, as analysis in the introduction, there still remains some problems of ship maneuvering identification modeling in calm water. This paper aimed at improving model precision at small rudder angle and less training data.

Furthermore, the simulation of the 35° turning circle test was conducted using fewer and smaller rudder angle test data. Therefore, the validation case shows the generalizability of the model, indicated by the low training data demand of the RM-RF. Moreover, the prediction of acceleration items is extremely precise, and it could be used for decision-making in ship motion control.

However, the following points should be considered. (1) As reproduction of the EFD-NMRI method is not consistent with the paper [61], then there is no comparison of velocities between EFD-NMRI and RM-RF. (2) Because a 3DOF model was used in this paper, the effect of ship rolling on ship maneuvering motion was not considered. (3) The 35° turning circle test represents only one test result, and uncertainty analysis of the test results was not performed. Thus, the results include a certain error with the advance and transfer of the circle. (4) Although the ship model scale used was 45.714, the scale effect in other methods, including CFD-SJT and EXP-NMRI, was not considered or analyzed. (5) As compared with maneuvering basin, the accelerations in sea trials used in RM-RF are obtained after considerably greater amounts of effort.

### B. STABILITY OF MODEL

As a white-black-box hybrid model, RM-RF model consists of two nonlinear systems. As expressed in reference [65], the stability of simple hybrid system is either undecidable or computationally. For AI based modeling, the generalization ability replaces stability analysis by validation. Validation cases for RM-RF which different from training case prove the stability to some extent. Furthermore, RM is stable and linear correlation with target ship. Then RM-RF can be replaced by the RM when the out of work. Although now the RM-RF cannot be proved to be stable, but the hybrid system is a new model to ship maneuvering.

### C. FURTHER RESEARCH OF SHIP MANEUVERING IN WAVES

As a hot topic, ship maneuvering in waves is also under consideration. The author is attempt to address the problem in following work. Limited to the pages of the paper, we only described the maneuvering in calm water and make the following statement. Firstly, based on the decouple hypothesis of seakeeping and maneuvering, with precise estimation of waves and wind effect, the calm water model will predict ship maneuvering in waves accurately. Therefore, calm water model is functional. Secondly, based on inviscid flow and ideal liquid, RM approximate other ship maneuvering by amplification.

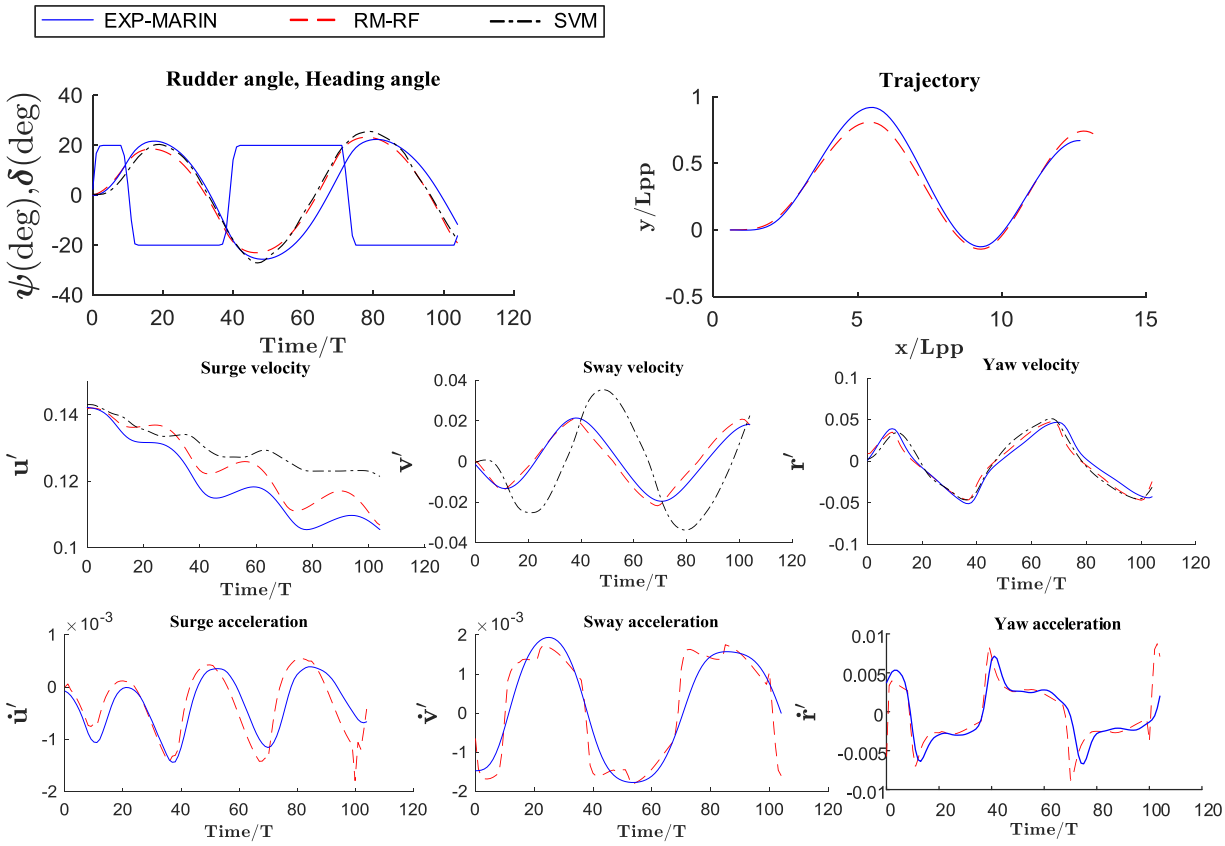


**FIGURE 8.** Prediction of 20°/10° zigzag test and compared with free running test and other methods14 [24].

However, there still some difficulty for validation as the lack of ship maneuvering in waves test. Finally, the ship maneuvering in calm water benefits for the maneuvering in waves, and the application of RM-RF for maneuvering in waves still needs further work.

## VIII. CONCLUSION

In this paper, the RM-RF method based on an improved model structure and a random forest is presented for ship maneuvering model identification. The advantages of the proposed method include less work of the reference model relative to that required by existing models and the small scale of training data. The vital skills relevant to the proposed method are reference model, similarity rules, and random forests. We obtained two important results in this study. First, compared with CFD and experimental methods, the prediction accuracy of the RM-RF is high. As one of identification methods for ship maneuvering, the accuracy of the RM-RF has improved since ITTC 2008. Secondly, compared with other identification methods, the RM-RF uses smaller amounts of data and requires tests with smaller rudder angles for generating identification data, which indicates the better generalizability of the method. Furthermore, zigzag test data were used for training, and turning circle test data were used for validation.

The RM-RF is a hybrid method. There are no specific coefficients, as in the MMG or Abkowitz models, and no descriptions of flow field, as in CFD. However, the input vector for the SVM is not ideal, and it requires considerably more data about the mechanism principle. Considering the complex sea environment, physical understanding and perception of this environment are beneficial to a system identification method for ship maneuvering. Based on our analysis, the system identification and mechanism method serve as complementary methods.

## ACKNOWLEDGMENT

Authors are grateful to the Maritime Research Institute Netherlands (MARIN) for providing the experimental data.

## REFERENCES

[1] *Work Programme-Maritime Autonomous Surface Ships Proposal for a Regulatory Scoping Exercise. MSC 98th Session*, IMO, London, U.K., Jun. 2017.

[2] *Report of the Maritime Safety Committee on Its 98th Session. MSC 98th Session*, IMO, London, U.K., Jun. 2017.

[3] *Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS) Report of the Working Group. MSC 99th Session*, IMO, London, U.K., Aug. 2018.

[4] *Any Other Business—The IMO Regulatory Framework and Its Application to Marine Autonomous Systems Submitted by the U.K., IAIN, IMarEST. MSC 95th Session*, IMO, London, U.K., Aug. 2018.

[5] Ø. J. Rødseth, Å. Tjora, and P. Baltzersen. (2014). MUNIN Deliverable D4.5: Architecture Specification. Hamburg, Germany. [Online]. Available: http://www.unmanned-ship.org
[6] K. Benedict, C. Krueger, G. Milbradt, and M. Schaub, "Simulation-augmented maneuvering system to support autonomous ships from the shore in different loading conditions," presented at the Auto. Ship Technol. Symp., 2016.
[7] J. M. Varas , "MAXCMAS project: Autonomous COLREGs compliant ship navigation," in *Proc. 16th Conf. Comput. Appl. Inf. Technol. Maritime Industries (COMPIT)*, Cardiff, U.K., 2017, pp. 454–464.
[8] T. I. Fossen, *Handbook of Marine Craft Hydrodynamics and Motion Control*. Noida, India: Wiley, 2011, pp. 109–132.
[9] *The Maneuvering Committee, Final Report and Recommendations to the 25th ITTC*, ITTC, Fukuoka, Japan, Sep. 2008.
[10] *The Maneuvering Committee, Final Report and Recommendations to the 28th ITTC*, ITTC, Wuxi, China, Sep. 2017.
[11] S. Bhushan *et al.*, "Verification and validation of CFD for surface combatant 5415 for straight ahead and 20 degree static drift conditions," in *Proc. 6th World Maritime Technol. Conf.*, Providence, RI, USA, 2015, pp. 124–149.
[12] R. E. D. Bishop and A. G. Parkinson, "On the planar motion mechanism used in ship model testing," *Philos. Trans Roy. Soc. A, Math., Phys., Eng. Sci.*, vol. 266, no. 1171, pp. 35–61, 1970. doi: 10.1098/rsta.1970.0002.
[13] F. Stern *et al.*, "Experience from SIMMAN 2008—The first workshop on verification and validation of ship maneuvering simulation methods," *J. Ship Res.*, vol. 55, no. 2, pp. 135–147, Jun. 2011.
[14] L. Ljung, "Perspectives on system identification," *Annu. Rev. Control*, vol. 34, no. 1, pp. 1–12, Apr. 2008. doi: 10.1016/j.arcontrol.2009.12.001.
[15] K.-G. Oh, K. Hasegawa, "Low speed ship manoeuvrability: Mathematical model and its simulation," in *Proc. ASME 32nd OMAE*, Nantes, France, 2013, pp. 11487–11487. doi: 10.1115/OMAE2013-11489.
[16] W. Van Hoydonck, G. Delefortrie, K. Eloot, P. Peeters, and F. Mostaert, "SIMMAN 2014: Shallow-water CFD computations. Version 3.0," Flanders Hydraulics Res., Antwerp, Belgium, Tech. Rep. WL Rapporten 00_066, Aug. 2015.
[17] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Autom. Control*, vol. AC-21, no. 2, pp. 194–202, Apr. 1976. doi: 10.1109/TAC.1978.1101673.
[18] S. Takase *et al.*, "Accuracy and Application for measurement of ship motion by kinematic GPS," *J. Jpn. Inst. Navigat.*, vol. 97, pp. 15–22, Sep. 1997. doi: 10.9749/jin.97.15.
[19] M. A. Abkowitz, "Measurement of hydrodynamic characteristics from ship maneuvering trials by system identification," *Trans. Soc. Nav. Archit. Mar. Eng.*, vol. 88, no. 10, pp. 283–318, 1980.
[20] W.-L. Luo and Z.-J. Zou, "Identification of response models of ship maneuvering motion using support vector machines," *J. Ship Res.*, vol. 11, no. 6, pp. 832–838, Jun. 2007. doi: 10.4028/www.scientific.net/AMR.211-212.106.
[21] K. J. Åström and C. Källström, "Identification and modelling of ship dynamics," Dept. Autom. Control, Lund Univ., Lund, Sweden, Tech. Rep. 7202, Mar. 1976.
[22] H. K. Yoon and K. P. Rhee, "Identification of hydrodynamic coefficients in ship maneuvering equations of motion by Estimation-Before-Modeling technique," *Ocean Eng.*, vol. 30, no. 18, pp. 2379–2404, Dec. 2003. doi: 10.1016/S0029-8018(03)00106-9.
[23] W. Luo and Z. Zhang, "Modeling of ship maneuvering motion using neural networks," *J. Marine Sci. Appl.*, vol. 15, no. 4, pp. 426–432, Dec. 2016. doi: 10.1007/s11804-016-1380-8.
[24] W. Luo and X. Li, "Measures to diminish the parameter drift in the modeling of ship manoeuvring using system identification," *Appl. Ocean Res.*, vol. 67, pp. 9–20, Sep. 2017. doi: 10.1016/j.apor.2017.06.008.
[25] M. Zhu, A. Hahn, Y.-Q. Wen, and A. Bolles, "Identification-based simplified model of large container ships using support vector machines and artificial bee colony algorithm," *Appl. Ocean Res.*, vol. 68, pp. 249–261, Oct. 2017. doi: 10.1016/j.apor.2017.09.006.
[26] S. Sutulo and C. G. Soares, "An algorithm for offline identification of ship manoeuvring mathematical models from free-running tests," *Ocean Eng.*, vol. 79, no. 4, pp. 10–25 Mar. 201. doi: 10.1016/j.oceaneng.2014.01.007.
[27] M. R. Haddara and Y. Wang, "Parametric identification of manoeuvring models for ships," *Int. Shipbuilding Prog.*, vol. 46, no. 445, pp. 5–27 Mar. 1999. doi: 10.3233/ISP-1999-4644501.
[28] N. Wang, M. J. Er, and M. Han, "Large tanker motion model identification using generalized ellipsoidal basis function-based fuzzy neural networks," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2732–2743, Dec. 2015. doi: 10.1109/TCYB.2014.2382679.
[29] L. Moreira and C. G. Soares, "Dynamic model of manoeuvrability using recursive neural networks," *Ocean Eng.*, vol. 30, no. 13, pp. 1669–1697, Sep. 2003. doi: 10.1016/S0029-8018(02)00147-6.
[30] D. A. Oskin, A. A. Dyda, and V. E. Markin, "Neural network identification of marine ship dynamics," in *Proc. 9th IFAC Conf. Control Appl. Mar. Syst.*, Osaka, Japan, vol. 46, Sep. 2013, pp. 191–196. doi: 10.3182/20130918-4-JP-3022.00018.
[31] X. G. Wang, Z. J. Zou, and L. Cheng, "Blind prediction of ship maneuvering motion in 4 degrees of freedom based on support vector machines," *J. Ship Mech.*, vol. 18, no. 9, pp. 1013–1023, Sep. 2014. doi: 10.3969/j.issn.1007-7294.2014.09.001.
[32] W.-W. Bai, J.-S. Ren, and T.-S. Li, "Multi-innovation gradient iterative locally weighted learning identification for a nonlinear ship maneuvering system," *China Ocean Eng.*, vol. 32, no. 3, pp. 288–300, Jun. 2018. doi: 10.1007/s13344-018-0030-0.
[33] M. Blanke and M. Knudsen, "Efficient parameterization for grey-box model identification of complex physical systems," in *Proc. 14th IFAC Symp. Syst. Identificat.*, Newcastle, NSW, Australia, Mar. 2006, vol. 39, no. 1, pp. 338–343. doi: 10.3182/20060329-3-AU-2901.00049.
[34] X.-G. Wang *et al.*, "System identification modeling of ship manoeuvring motion in 4 degrees of freedom based on support vector machines," *China Ocean Eng.*, vol. 29, no. 4, pp. 519–534, Jun. 2015. doi: 10.1007/s13344-015-0036-9.
[35] D. W. Faller, "Simulation of ship maneuvers using recursive neural networks," in *Proc. 23th Symp. Nav. Hydrodyn.* Val-de-Reuil, France: Academies, vol. 23, Sep. 2001, pp. 223–242.
[36] F.-C. Chiu, T.-L. Chang, J. Go, S.-K. Chou, and W.-C. Chen, "A recursive neural networks model for ship maneuverability prediction," in *Proc. Oceans MTS/IEEE Techno-Ocean*, Kobe, Japan, vol. 4, Nov. 2004, pp. 1211–1218. doi: 10.1109/OCEANS.2004.1405752.
[37] *Standards for Ship Maneuverability*, Resolution MSC.l37(76), IMO, London, U.K., 2002.
[38] K. Kijima, "Some studies on the prediction for ship maneuverability," in *Proc. MARSIM*, Tokyo, Japan, vol. 26, Aug. 2003, pp. 26–35.
[39] X.-J. Liu, G.-L. Huang, and D.-H. Deng, "Research of interaction force coefficients based on model test in loading condition," *J. Shanghai Jiaotong Univ., Sci.*, vol. 15, no. 2, pp. 168–171, Apr. 2010. doi: 10.1007/s12204-010-8033-x.
[40] M. Viviani *et al.*, "Alternative methods for the identification of hydrodynamic coefficients from standard maneuvers," in *Proc. MARSIM*, Tokyo, Japan, vol. 26, Aug. 2003, pp. 5–14.
[41] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Cham, Switzerland: Springer, 2009, pp. 219–223. doi: 10.1007/978-0-387-84858-7.
[42] B. Mei, L. C. Sun, and G. Y. Shi, "An approach based on model reference and random forest for ship maneuvering identification modeling," *J. Dalian Maritime Univ.*, vol. 44, no. 2, pp. 15–21, Apr. 2018. doi: 10.16411/j.cnki.issn1006-7736.2018.02.003.
[43] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, "Data-driven vs. model-driven: Fast face sketch synthesis," *Neurocomputing*, vol. 257, pp. 214–221, Sep. 2017.
[44] M. Kwiatkowska and A. S. Atkins, "Integrating knowledge-driven and data-driven approaches for the derivation of clinical prediction rules," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, Los Angeles, CA, USA, Dec. 2005, pp. 171–176. doi: 10.1109/ICMLA.2005.41.
[45] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[46] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996, pp. 313–330.
[47] *Nomenclature for Treating the Motion of a Submerged Body Through a Fluid*, Tech. Res. Bull., Soc. Nav. Archit. Mar. Eng., New York, NY, USA, 1950.
[48] E. K. Larsson and T. Söderström, "Identification of continuous-time AR processes from unevenly sampled data," *Automatica*, vol. 38, no. 4, pp. 709–718. Apr. 2002. doi: 10.1016/S0005-1098(01)00244-8.
[49] J. Newman, *Marine Hydrodynamics*. Cambridge, MA, USA: MIT Press, 1977, pp. 343–346.
[50] X. J. Zhu, H. H. Shao, and Z. J. Zhang, "The effect of behavior discrete analysis on the dynamic of nonlinear control systems," *Inf. Control*, vol. 25, no. 2, pp. 115–120, Apr. 1996. doi: 10.13976/j.cnki.xk.1996.02.010.
[51] Z. Zhang, D. U. An, H. Kim, and K. ToChong, "Comparative study of Matrix exponential and Taylor series discretization methods for nonlinear ODEs," *Simul. Model. Pract. Theory*, vol. 17, no. 2, pp. 471–484, Feb. 2009. doi: 10.1016/j.simpat.2008.10.003.

[52] N. Kazantzis and C. Kravaris, "Time-discretization of nonlinear control systems via Taylor methods," *Comput. Chem. Eng.*, vol. 23, no. 6, pp. 763–784, Jun. 1999. doi: 10.1016/S0098-1354(99)00007-1.
[53] M. N. Hayes, "Parametric identification of nonlinear stochastic systems applied to ocean vehicle dynamics," Ph.D. dissertation, Dept. Ocean. Eng, Massachusetts Inst. Technol., Cambridge, MA, USA, 1971.
[54] Y. Yoshimura and Y. Masumoto, "Hydrodynamic force database with medium high speed merchant ships including fishing vessels and investigation into a manoeuvring prediction method," *J. Jpn. Soc. Nav. Archit. Ocean Eng.*, vol. 14, no. 2, pp. 63–73, Dec. 2011. doi: 10.2534/jjasnaoe.14.63.
[55] Y. Yoshimura and K. Nomoto, "Modeling of manoeuvring behaviour of ships with a propeller idling, boosting and reversing," *J. Soc. Nav. Archit. Jpn.*, vol. 1978, no. 144, pp. 57–69, Nov. 1978. doi: 10.2534/jjasnaoe1968.1978.144_57.
[56] M. S. Chislett and J. Strom-Tejsen, "Planar motion mechanism tests and full-scale steering and manoeuvring predictions for a Mariner class vessel," *Int. Shipbuilding Prog.*, vol. 12, no. 129, pp. 201–224, May 1965.
[57] *Revision of the Interim Standards for Ship Maneuverability Ship Maneuvering Data Submitted by the Republic of Korea, Sub-Committee on Ship Design and Equipment, 44th Session*, IMO, London, U.K., Nov. 2001.
[58] N. H. Norrbin, "Theory and observations on the use of a mathematical model for ship manoeuvring in deep and confined waters," SSPA, Stockholm, Sweden, Tech. Rep. 68, 1970.
[59] M. A Abkowitz, *Lectures on Ship Hydrodynamics: Steering and Manoeuvrability*. Lyngby, Denmark: Hydro-and Aerodynamics Laboratoay, 1964.
[60] E. V. Lewis , *Principles of Naval Architecture*, 2nd ed. New York, NY, USA: Society of Naval Architects, 1988, p. 291.
[61] H. Yasukawa and Y. Yoshimura, "Introduction of MMG standard method for ship maneuvering predictions," *J. Mar. Sci. Technol.*, vol. 20, no. 1, pp. 37–52, Nov. 2015. doi: 10.1007/s00773-014-0293-y.
[62] B. G. Hong, *Technology and Principle of Ship Maneuvering*. Dalian, China: Dalian Maritime University Press, 2007, pp. 121–143.
[63] Y. J. Sung and S. H. Park, "Prediction of ship manoeuvring performance based on virtual captive model tests," *J. Soc. Nav. Archit. Korea*, vol. 52, no. 5, pp. 407–417, 2015. doi: 10.3744/SNAK.2015.52.5.407.
[64] H. Liu, N. Ma, and X. Gu, "Maneuvering prediction of a VLCC model based on CFD simulation for PMM tests by using a circulating water channel," in *Proc. Int. Conf. Ocean, Offshore Arctic Eng.*, St. John's, NF, Canada, 2015, pp. 41545–41545. doi: 10.1115/OMAE2015-41548.
[65] V. D. Blondel and J. N. Tsitsiklis, "Complexity of stability and controllability of elementary hybrid systems," *Automatica*, vol. 35, no. 3, pp. 479–489, Mar. 1999. doi: 10.1016/S0005-1098(98)00175-7.

**BIN MEI** received the B.Eng. degree in nautical science and the M.Eng. degree in maritime science and technology from Dalian Maritime University (DMU), Dalian, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in traffic information engineering and control. His research interests include the development of ship maneuvering modeling based on free running model test or sea trials using system identification techniques, fundamental study of traffic flow on the sea for channel design, and ship collision avoidance.

**LICHENG SUN** received the B.Eng., M.Eng., and Ph.D. degrees from Dalian Marine College (DMC), Dalian, China, in 1982, 1988 and 2000, respectively. From 1982 to 1994, he was an Assistant, a Lecturer, and an Associate Professor with the Navigation Department, DMC, where he is currently a Professor. His research interests include ship maneuvering and ship collision avoidance.

**GUOYOU SHI** received the B.Eng., M.Eng., and Ph.D. degrees from Dalian Maritime University (DMU), Dalian, China, in 1993, 1996, and 2006, respectively. He was an Assistant, a Lecturer, and an Associate Professor with the Navigation College, DMU, where he is currently a Professor. His research interests include ship maneuvering, intelligent ship collision avoidance, and technologies for maritime autonomous surface ship.